

Biomedical Information Infrastructure based on HL7 v3 Templates

A. Shabo

Abstract— This paper describes the Biomedical Information Infrastructure (BII) with its core of a general-purpose health information warehouse which can be deployed to serve various projects and solutions. The BII employs a methodology of developing a project-specific data model based on HL7 v3 standard specifications and their templates.

Index Terms— HL7 v3 templates, semantic warehousing.

I. INTRODUCTION

The methodology of developing project-specific data models for deploying semantic warehouse solution starts by selecting v3 standards that are relevant to the project needs in terms of the data involved in the integration phase. Then, each standard is constrained to precisely meet the requirements set forth by the project stakeholders. Finally, the constrained standards are interconnected and being put together into a coherent data model for the implementation at stake. The persistency layer is based on the v3 RIM and therefore v3 domain instances are ‘RIMified’ to simplify their persistence and retrieval in an XML-based database, i.e., each XML element is replaced with its respective RIM class name.

II. THE BIOMEDICAL INFORMATION INFRASTRUCTURE

The Biomedical Information Infrastructure (BII) is being developed as part of a European Commission funded project called Hypergenes (for more details, see <http://www.hypergenes.eu/>). The project goal is to develop a disease model for essential hypertension. Data is aggregated from about ten historical cohorts of ~12k hypertensive and normotensive subjects from all over Europe. Genome-wide scan is performed on DNA specimen of those subjects to find the genetic variations associated with hypertension. Then, clinical and environmental data of those subjects are taken into account when developing the disease model. When using the above-mentioned methodology, it is evident that the Hypergenes project needs standards for describing medical health histories of subjects along with their family health history and genetic data generated during the project.

Manuscript received February 28, 2010. The research leading to these results has received funding from the European Community's Seventh Framework Program FP7/2007-2013 under grant agreement n° 201550.

A. Shabo is with the IBM Research Lab, Haifa, Israel (phone: +972-4-8296358; fax: +972-4-8296116; e-mail: shabo@il.ibm.com).

Therefore, the v3 standards selected for Hypergenes BII data model are the CDA, Pedigree and Genetic Variation. CDA is the most generic standard out of those three specifications; therefore it is the standards that got constrained resulting in a CDA template for hypertension data. From the CDA instance there are references to the subject's Pedigree as well as to several instances of Genetic Variation instance. Figure 1 depicts the roadmap from data entry of proprietary data through RIM-based warehousing to data marts representing a user schema of interest. A main RIM-based specification is the CDA that serves as an organizer to other types of data such as genomic, imaging and fully structured family history data.

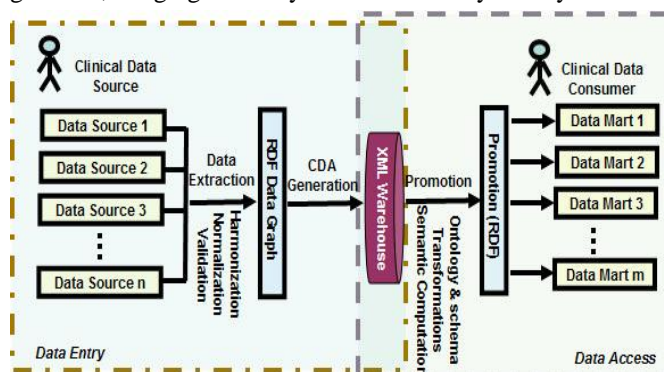


Fig. 1. End-to-end data flow from integration through warehousing to data marts.

III. CONCLUSION

The presentation will describe in detail the standards-based data model development methodology and its use in the Hypergenes project, including in-depth look in to the CDA template and attempts made to use the OHT CDA editor to formally represent the template and automatically create CDA instances compliant with the template.

ACKNOWLEDGMENT

A. Shabo thanks all the Hypergenes consortium members who contributed to the development of the BII: Anna Burla, Ariel Farkash, Boaz Carmeli, Carmel Kent, Roi Adadi, Yevgenia Tsimerman, Yonatan Maman.

A. Shabo is a co-editor of the CDA R2, CCD, Pedigree and Genetic Variation specifications. He is also a founder, co-chair and modeling facilitator of the HL7 Clinical Genomics Work Group.