e-Health – For Continuity of Care
C. Lovis et al. (Eds.)
© 2014 European Federation for Medical Informatics and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License.
doi:10.3233/978-1-61499-432-9-278

A conceptual Framework to Design a Dimensional Model Based on the HL7 Clinical Document Architecture

Fabrizio PECORARO[a,1], Daniela LUZI[a] and Fabrizio L. RICCI[b]
[a]*Institute for Research on Population and Social Policies, National Research Council*
[b]*Institute of Molecular Bioimaging and Physiology, National Research Council, Italy*

Abstract. This paper proposes a conceptual framework to design a dimensional model based on the HL7 Clinical Document Architecture (CDA) standard. The adoption of this framework can represent a possible solution to facilitate the integration of heterogeneous information systems in a clinical data warehouse. This can simplify the Extract, Transform and Load (ETL) procedures that are considered the most time-consuming and expensive part of the data warehouse development process. The paper describes the main activities to be carried out to design the dimensional model outlining the main advantages in the application of the proposed framework. The feasibility of our approach is also demonstrated providing a case study to define clinical indicators for quality assessment.

Keywords. Data Warehouse, Extract, Transform and Load, Dimensional Model, HL7, Clinical Document Architecture, Electronic Health Record.

Introduction

Enterprise data warehouses are increasingly adopted in the healthcare domain to improve decision-making, business processes as well as to discover new treatments and improve patient care [1]. Their development makes it necessary to integrate data provided by heterogeneous information systems (e.g. Laboratory Information Systems, Electronic Health Records, Radiology Information Systems), developed for different specialties, by different organizations and for different purposes [2]. In data warehouse implementation and maintenance process this integration is called Extract, Transform and Load (ETL), that is considered the most time-consuming and expensive activity (around 70% of the resources used) [3]. This implies the adoption of transformation procedures to convert data from source systems in a common data model to be loaded in the target system (i.e. data warehouse and/or data marts).

In the healthcare domain Health Level 7 (HL7) plays an important role to improve interoperability providing a set of standards for the exchange, integration, sharing and retrieval of healthcare information. These standards are based on the Reference Information Model (RIM) [4], which represents a conceptual model covering both clinical and administrative aspects. One of the most widely adopted implementation of the RIM is the Clinical Document Architecture (CDA) [5] that specifies the structure and semantics of a document (e.g. discharge summary, continuity of care document).

[1] **Fabrizio Pecoraro**, IRPPS-CNR, Via Palestro, 32, 00185 Rome, Italy (f.pecoraro@irpps.cnr.it).

The aim of this paper is to propose a conceptual framework to map the dimensional model primitives with the HL7 CDA classes in order to facilitate the definition and implementation of ETL tools developed in a data warehouse architecture. This study is part of the Smart Health 2.0 national project that aims to develop a regional healthcare infrastructure based on HL7 standards. It also intends to explore the use of Electronic Health Record (EHR) for secondary purposes in a clinical governance framework to assess the quality of care from the structural, organizational, financial and professional points of view [6]. The framework proposed in this paper facilitates the identification of clinical indicators for quality assessment to be calculated from data contained in CDA documents stored in a regional EHR.

1. Conceptual framework based on the dimensional model developing life cycle

The approach proposed in this paper is described considering the 4-step process described by Kimball and Ross in [7]. The feasibility of our approach is demonstrated providing a case study based on the Continuity of Care Document (CCD) [8] to define clinical outcome indicators for quality assessment, such as percentage of patients with a vital sign parameter within a specific threshold (e.g. glycated hemoglobin under 7%).

1.1. Choose the business process and declare the grain

The determination of a business process can be based on the HL7 RIM Act class that represents "an intentional action that can be either executed, ordered, planned and must be documented". For this reason it is important to identify which Act is a feasible representation of the event taking into account also the level of granularity to be represented. In this perspective, the CDA is mainly described by three Acts: 1) ClinicalDocument that provides an high level description of the document; 2) Section that aggregates in a single "narrative block" core patient-specific data based on common clinical conventions (e.g. laboratory test results, medications, problems, procures); 3) clinicalStatement a choice structure that represents the content of a specific action (e.g. observation, procedure, substance administration).

In particular, in the CCD implementation the class Section through the attribute code is a standardized identifier to classify clinical business processes coded using the LOINC (Logical Observation Identifiers Names and Codes) nomenclature, such as functional status, family and social history, medical equipment and encounters. In this paper the attention is posed on the vital sign section that contains individual's clinical findings, such as blood pressure, heart rate, body mass index and glycated hemoglobin.

1.2. Identify the dimensions

In our approach dimensions are determined on the basis of the Zachman framework [9] that provides a systematic information representation starting from the following questions related with the investigated event: who (persons), what (the fact), when (the time), where (the place), why (the reason) and how (the manner). Each Participation included in the CDA model represents a suitable candidate element to derive a dimension table, given that it "captures the functions of subjects and objects involved in a specific process, identifying who performed it (i.e. performer), for whom it was done (i.e. recordTarget), where it was done (i.e. location), etc.". For instance, the Act

Observation that represents an action in the CDA, can be associated with the following Participations: a) performer to identify the "person who actually and principally carries out or will carry out the action"; b) consumable to determine the "substance that is taken up or consumed as part of the substance administration"; c) subject to define the "principle target that the service acts on". For each applicable Participation a dimension table can be determined collapsing the hierarchy represented by the 4-ple <$Entity_{player}$, $Entity_{scoper}$, Role, Participation> in a single denormalized table. This process that introduces data redundancy with a less complex queries is particularly suitable for one-to-many relationships between the fact and the relevant dimension. For instance the Act ClinicalDocument is related with only one Participation recordTarget. Thus, the hierarchy can be mapped in a single dimension table composed by relevant attributes of both Role (i.e. PatientRole) and related Entities (i.e. Patient and Organization), as shown in figure 1. However, some requirements need to be represented with a many-to-many relationship in the dimensional model to be modeled using a bridge table as proposed in [10]. For instance, when different practitioners deliver care to an individual over different distinct time intervals.

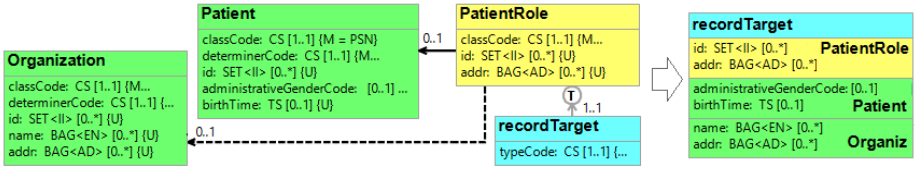

Figure 1. recordTarget dimension defined through the denormalization of the RIM 4-ple hierarchy.

Table 1. Example of dimensions derived from the HL7 RIM stereotypes and CDA features

Question	RIM stereotype	CDA feature	Example
Who	Participation	recordTarget	The patient involved in the event
		performer	Physician/Practitioner that carried out the event
		responsibleParty	Participant with legal responsibility
Where	Participation	location	Healthcare facility where the event occurred
What	Participation	consumable	Material consumed during the event
		product	Material dispensed during the event
	Act	Act.code	Type of event performed
	ActRelationship	referenceRange	Range of values for the measured parameter
When	Act	Act.effectiveTime	Time when the event has been performed
How	Act	Act.methodCode	Technique used to establish the event
Why	ActRelationship	entryRelationship	A principal cause of an event

Table 1 highlights some examples of possible mapping between HL7 stereotypes and dimensional model primitives. The attribute *typecode* of the stereotype Participation can be also used to define who/whom is involved in the business process, where the event has been carried out (i.e. location) and what are the equipment involved to perform the event (i.e. substance consumed or product dispensed). Other information describing the event can be captured through the Act that describes the event under investigation and its related Act described by the stereotype ActRelationship. Moreover, there are different attributes of the stereotype Act and its specialization in the CDA model that can describe the business process, such as time when the event took place (i.e. *effectiveTime*) and type of the event observed (i.e. *code*). Considering the ActRelationship class, in the CDA model there are two main specializations: 1) entryRelationship that links two clinicalStatements of the same Section to identify, for instance, a principal cause of an event (e.g. an allergy reaction

to penicillin that causes hives or rashes); 2) referenceRange that shows the thresholds of a specific parameter (e.g. blood pressure). The mapping of the ActRelationship class in a dimension can require the adoption of a bridge table to be associated with the fact.

The dimensions detected to describe the vital sign business process based on the CCD are described as follows and depicted in figure 2: recordTarget and performer using the denormalization approach described previously including the Entity Organization that identifies the location where the service has been performed. Moreover, type of clinical test (vitalSignObservation), the technique used to determine it (method) and the time when it has been performed (Time_Date) are defined using, respectively, the attribute *code*, *methodCode* and *effectiveTime* of the Act Observation.

1.3. Identify the Fact

A fact table contains qualitative and quantitative measures to represent the performance of the business process. In the RIM these information are collected in specific attributes of the stereotype Act representing "measurement of healthcare business processes". Data to be collected strictly depend on the event under investigation. For instance, the vital sign section of the CCD includes two attributes: 1) *value* to capture a physical quantity (i.e. weight) and its unit of measure (i.e. kg); 2) *interpretationCode* to provide a qualitative interpretation of the measure (e.g. normal, low). Both attributes have been included in the fact table of the dimensional model shown in figure 2.

1.4. Modelling the HL7 RIM Data Types

Most of the attributes of the CDA are modelled using complex data types that need to be transformed in a database data type. A possible solution to represent of a complex data type is to store each property in a single column of the relevant table. For instance a Physical Quantity (PQ) can be mapped into two distinct attributes: value and unit. However, different attributes of the RIM can assume multiple values. For instance, the Act Observation captures a set of *interpretationCode* (e.g. the value is decreased, is below lower alert threshold, is moderately susceptible, etc.). These data types can be modelled either creating a separate table to store each multiple valued attribute or revising the model to have only a single value for each attribute.

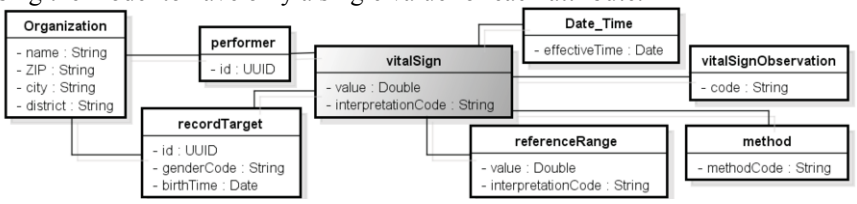

Figure 2. Dimensional model with complex data types normalized

Starting from the specification defined in the previous paragraph attributes with complex data types have been normalized and a dimensional model is proposed in figure 2. In particular: 1) in the vitalSign and referenceRange tables the unit of measure of the attribute value has been removed and only the measurement reported using a double data type (*value*); 2) *address* of the sub-dimension organization has been exploded in three distinct attributes: ZIP, city and district; 3) coded attribute (*interpretationCode*, *methodCode*, *code*, *administrativeGenderCode*) are stored as a

string attribute based on the relevant vocabulary. Finally, *birthTime* and *effectiveTime* are mapped in a Date column.

2. Discussion

In this paper we demonstrated the feasibility of a conceptual framework to define the primitives of a dimensional model based on the HL7 CDA entities with the purpose of developing a clinical data warehouse supporting clinical quality assessment. The adoption of this approach can lead to different advantages both at the design and execution phases of the data warehouse implementation process. At the design level, a dimensional model based on HL7 CDA can simplify the definition of the business process to be analysed as well as the dimensional model primitives that describe it. This makes data transformation a simplified straightforward process thus reducing the resources to be invested to implement ETL tools and procedures. From the implementation point of view, information collected in HL7 CDA messages can be loaded in the target data warehouse without implementing tools to transform source data models to the destination dimensional model. This makes the integration of heterogeneous systems more robust limiting the ambiguity in the semantics of messages and avoiding the proliferation of dialects. Moreover, as highlighted in our previous paper focused on the database model [11], this approach can also facilitate end-users in interpreting and defining query in the data warehouse. From the architectural point of view this approach enhances the extensibility of the data warehouse architecture facilitating the integration of other systems in the future given that: 1) rules and parsing procedures are standardized in a common framework and 2) clinical information are coded using international nomenclatures and dictionaries, such as LOINC. This approach is going to be tested within the Smart Health 2.0 project.

References

[1] T.R. Sahama, P.R. Croll, A data warehouse architecture for clinical data warehousing, *Fifth Australasian symposium on ACSW frontiers* **68** (2007).
[2] N. Wickramasinghe, J.L. Schaffer, Creating knowledge-driven healthcare processes with the intelligence continuum, *Int J Electron Healthc*, **2(2)** (2006), 164-174.
[3] R. Kimball, J. Caserta, *The Data Warehouse ETL Toolkit*, Wiley, Indianapolis, USA, 2006.
[4] G. Schadow, C.N. Mead, D. M. Walker, The HL7 Reference Information Model under scrutiny, *Stud Health Tech Informat*, **124** (2006), 151-156.
[5] R.H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F.M. Behlen, P.V. Biron, A. Shabo, HL7 Clinical Document Architecture, Release 2, *J Am Med Inform Assoc* **13** (2006), 30-39.
[6] F. Pecoraro, D. Luzi, F.L. Ricci, Secondary uses of EHR systems: a feasibility study. *IEEE E-Health and Bioengineering Conference (EHB)* (2013).
[7] R. Kimball, M. Ross, The Data Warehouse Toolkit Second Edition: The Complete Guide to Dimensional Modelling, Wiley Computing Publishing, 2002.
[8] O. Lupse, M. Vida, L. Stoicu-Tivadar, V. Stoicu-Tivadar, Using HL7 CDA and CCD standards to improve communication between healthcare information systems, *IEEE 9th International Symposium on Intelligent Systems and Informatics (SISY)* (2011), 453-457.
[9] W.H. Inmon, J.A. Zachman, J.G. Geiger, Data Stores, Data Warehousing, and the Zachman Framework: Managing Enterprise Knowledge, McGraw-Hill, New York, NY, 1997.
[10] T.J. Eggebraaten, J.W. Tenner, J.C. Dubbels, A health-care data model based on the HL7 Reference Information Model. *IBM Journal of Research and Development* **46** (2006), 5-18.
[11] F. Pecoraro, M. Contenti, D. Luzi, Clinical Investigation on Medical Device: A data model based on HL7 Reference Information Model, *IADIS MCCSIS* 2010.