Clinical Document Data Warehouse (CD2W)

Josep Vilalta Marzo [a], Diego Kaminker [b], Josep M. Picas Vidal [c], M. Lluisa Bernard Antoranz [d], Cristina Siles [c], Rafael Rosa Prat [a]

[a] Vico Open Modeling S.L., [b] Kern Information Technology SRL, [c] Hospital de la Santa Creu i Sant Pau , [d] Institut Català de la Salut

Abstract

This paper shows the development of the Clinical Document Data Warehouse project (CD2W) and its implementation using CDA R2 documents (Clinical Document Architecture Release 2).

The main objectives of this project was to prototype a web based portal allowing access to a clinical document repository and a data warehouse with clinical information about patients from several medical organizations in Barcelona, Spain (a major hospital and 4 primary care centers), giving access to clinical patient information for primary care and leveraging the same standardized information to populate the data warehouse (secondary use).

The project was developed during the first half of 2009 under the general direction of the Hospital de la Santa Creu i Sant Pau. (HSCSP)

Due to the prototype nature of the project, the scope was limited to patients with Congestive Heart Disease (CHD) who consented the use of their information for clinical research to the HSCSP and with the current vocabularies used by the providers (ICD9 and other local terminologies).

The data warehouse was developed using HL7 RIM basic concepts.

The project mission was to improve patient care with the use of global standards, open technology, low exploitation cost and ease of use. CDA R2 documents.

During this project, we used the SCRUM agile methodology allowing scalable, progressive and incremental software development process.

Introduction – Business Case

Solving the questions arising when a patient shows several clinical issues is one of the greatest challenges for healthcare providers.

Fast clinical decisions making at the healthcare location generates the need to put all relevant and up-to-date clinical information to support the process.

Usually, we don't have efficient filtering flagging the critical issues where to focus our attention. We encounter scenarios with data overload demanding a big effort to synthesize useful information, or sparse data demanding the use of imagination to connect and reach any conclusion.

Another problem is that usually our available clinical information source is the sole organization supporting the healthcare provider. Other information generated by other channels where the patient was attended is usually brought by the patient in several paper formats. Consolidation of relevant and up to date information from distinct healthcare organizations from different authorities is a pending issue waiting for an agreement from the healthcare authorities and harmonization of different data base schemes.

Nowadays, there are several on-going projects with the shared goal of consolidation of clinical information, both at the Spain level or at a autonomic community level.

These projects also share a great level of complexity and costs to achieve their goals. This small project CD2W aspires to help in achieving this consolidation of clinical information with focus on easing the task for the healthcare providers, organizations administering huge data bases with several difficulties to integrate and no reference information model available.

Materials and Methods

Hospital de la Santa Creu i Sant Pau is a high complexity hospital which dates back six centuries, making it the oldest hospital in Spain. Healthcare is centered on Barcelona but extends to the rest of Catalonia. The center plays a prominent role in Spain and is internationally renowned.

The hospital has distinguished itself in the healthcare provided in many fields, making it a reference centre in several specialties. The center attends over 34,000 admissions each year and more than 150,000 emergencies. Approximately 300,000 people are visited at the ambulatory services annually and the Day Hospital attends over 60,000 users. There are 71 day hospital beds, 634 hospitalization beds and 19 surgical rooms.

Teaching and training programmes at the Hospital de la Santa Creu i Sant Pau cover many levels, comprising the UAB (Universitat Autonoma de Barcelona) Faculty of Medicine Teaching Unit, the University School of Nursing, participation in the State Residency Programmes to train specialists, masters and doctorate course, continuing education, etc. In the field of research, Hospital de la Santa Creu i Sant Pau is one of the most prominent centers in Spain, as can be appreciated from the volume of papers published and their input factor, the number and quality of projects which receive funding and the grants awarded. The Hospital de la Santa Creu i Sant Pau is governed by the Patronat de la Fundació de Gestió Sanitària (FGSHSCSP), a board with representatives from the Regional Catalan Government (Generalitat de Catalunya), the City Hall of Barcelona and the Archbishopric of Barcelona.

The goals for this project were:

1. Define a scheme to integrate information from disjoint information platforms.

2. Implement a process for periodical data and document exchange with minimal workload implications for the primary care centers.

3. Generate a clinical data store enabling the coexistence of clinical documents and relational data in a longitudinal patient healthcare record..

4. Create a simple user interface to ease fast queries to the relevant clinical information about a patient..

5. Restrict access to clinical information to professionals authorized by the participating organizations (HSP and ICS)

6. Use open technology for design, development and implementation of the data warehouse, minimizing exploitation costs and use global interoperability standards to enable universal access.

7. Evaluate the impact of normalizing data from different organizations and code systems..

8.Evaluate useability and added value chain of CD2W to the physicians for healthcare decisions

9. Evaluate results of this prototype to study the possible extension of the access to the patients or to other professionals.

10 .Evaluate the results for a new project with a broader scope.

Implementation, Methodology and Tools

The project has four main design aspects:
1. Clinical Datawarehouse Design
2. Standard Document Design
3. Datawarehouse Population Process Design
4. User Interface Design
5. Technological implementation.
Lets review them:

1. Clinical Datawarehouse Design

The model for CD2W was derived from the RIM base classes (role,entity,act,participation) [3] following a three step methodology:

a- Development of a conceptual framework, to be discussed with the CD2W stakeholders, including which were the measures or facts, and how should the information be classified (dimensions).

Figure 1 – CD2W conceptual model

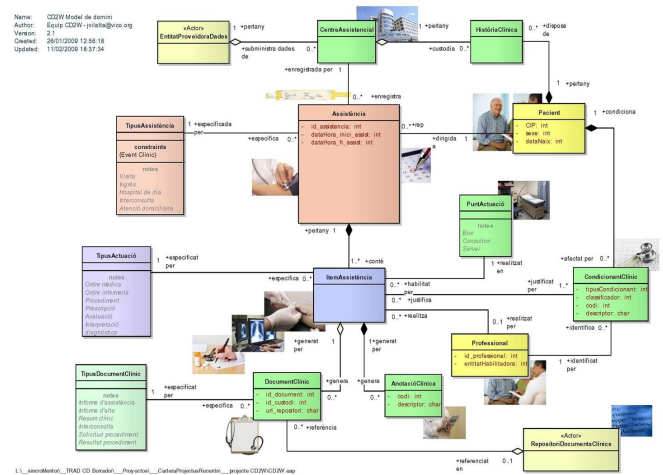b. Then, on a more technical level, a domain analysis model was generated.

Figure 2 – CD2W Domain Analysis

And finally the derivation of a datawarehouse model to store facts, dimensions and supporting standard clinical documents.
RIM-wise, facts are acts, and main dimensions are entities and roles and their attributes.

2. Standard Document Design.

The documents were the 'lingua franca' between disparate systems used by the participating organizations (primary care centers, Hospital her and discharge system) [1].

We designed two different clinical document templates, one for the evolution note from the primary care centers and one for the discharge note from the Hospital discharge system.

The templates shared the same information at the header level, but differed in their section contents.
Since this data warehouse was intended for secondary use, patient and physician information was de-identified [2] (names, identifications and addresses were removed or replaced).

The information generated by the local provider applications was transformed using an XSL to clinical documents and this standard documents were processed and stored into the C2DW.

We tested our mapping, process and query interface with 16125 clinical documents from the hospital and primary care centers.

In order to guide the development team a table was built with the main required elements from each primary care center and its location inside of the standard clinical document.

Figure 4 – Transformation Table

Elemento de CDA	Descripción del valor	Generación o Extra extraer de datos originales
SECCION: MOTIVO		
/ClinicalDocument/component/structuredBody/component/section/title	Título de Sección	
/ClinicalDocument/component/structuredBody/component/section/text	Motivo de la visita	texto_motivo
SECCION: EXPLORACION		
/ClinicalDocument/component/structuredBody/component/section/title	Título de Sección	
/ClinicalDocument/component/structuredBody/component/section/text	Motivo de la visita	texto_exploracion
SECCION: EVALUACION		
/ClinicalDocument/component/structuredBody/component/section/title	Título de Sección	
/ClinicalDocument/component/structuredBody/component/section/text	Motivo de la visita	texto_evaluacion
SECCION: PLAN		
/ClinicalDocument/component/structuredBody/component/section/title	Título de Sección	
/ClinicalDocument/component/structuredBody/component/section/text	Motivo de la visita	texto_plan
SECCION: ACTIVIDAD		
/ClinicalDocument/component/structuredBody/component/section/title	Título de Sección	
/ClinicalDocument/component/structuredBody/component/section/text/table/tbody/tr[n]/td	Codigo Servicio Act	codi_servei_activitat
/ClinicalDocument/component/structuredBody/component/section/text/table/tbody/tr[n]/td[2]	Descripción Activid	desc_servei_activitat
/ClinicalDocument/component/structuredBody/component/section/text/table/tbody/tr[n]/td[3]	Fecha Actividad	data_item_assistencia
/ClinicalDocument/component/structuredBody/component/section/text/table/tbody/tr[n]/td[4]	Situación	desc_codi_situacio
SECCION: SITUACION		
/ClinicalDocument/component/structuredBody/component/section/title	Título de Sección	
/ClinicalDocument/component/structuredBody/component/section/text/table/tbody/tr[n]/td	Codigo de Situacio	codi_tipus_actuacio
/ClinicalDocument/component/structuredBody/component/section/text/table/tbody/tr[n]/td[2]	Descripcion	desc_tipus_actuacio
SECCION: CONDICIONANTES		
/ClinicalDocument/component/structuredBody/component/section/title	Título de Sección	
	Tipo de Condiciona	tipus_condicionant
/ClinicalDocument/component/structuredBody/component/section/text/table/tbody/tr[n]/td	Codigo de Condicio	codi_condicionant
/ClinicalDocument/component/structuredBody/component/section/text/table/tbody/tr[n]/td[2]	Descripcion	desc_condicionant
SECCION: ANOTACIONES CLINICAS		
/ClinicalDocument/component/structuredBody/component/section/title	Título de Sección	
/ClinicalDocument/component/structuredBody/component/section/text/table/tbody/tr[n]/td	Descripción	tipus_anotacio_clinica
/ClinicalDocument/component/structuredBody/component/section/text/table/tbody/tr[n]/td[2]	Valor	desc_anotacio_clinica

3. Data warehouse population process

The process to populate the data ware house included several steps.

[At the primary care center]
1. Select encounters for patients suffering of CHD with signed consents.
2. Create a basic, shallow XML for each primary care center encounter

[At the data processing center]
3. Create a CDA conformant document using the defined mapping for the center, for each instance
4. Populate the CD2W database with the information from each document header, and the document itself.

This process was triggered periodically by the primary care centers and HSP.

4. User interface design

The user interface design was use-case based.
Identified use cases were:

A. App Administration

Application setup and parametrization
- Define healthcare agent
- Define healthcare agent type
- Define healthcare agent role
- Define app parameters
- Define service catalog
- Define service

B. Data Generation Process

Use Cases related to the DW population

Login
- User validation
- Batch load
- Stat Processing
- Document Processing
- Update audit log

C. Queries

Data access and retrieval from the users
- Retrieve app parameters
- Query control panel (figure 3)
- Query Patient Monitor (figure 4)
- Query Encounter Monitor
- Query Conditions Monitor
- Query Healthcare Centers
- Query Healthcare Professionals
- Query Healthcare Services
- Query Stats (figure 5,6)
- Query audit logs
- Browse CDA R2 document (figure 7)

The following figures illustrate the main user interface forms for CD2W:

Figure 3 – CD2W Control Panel

Figure 4 – CD2W Patient Query

Figure 5 – Services for one patient

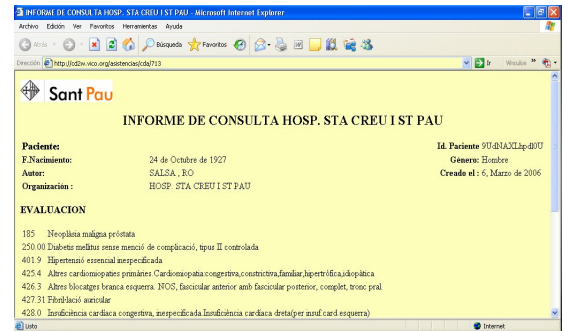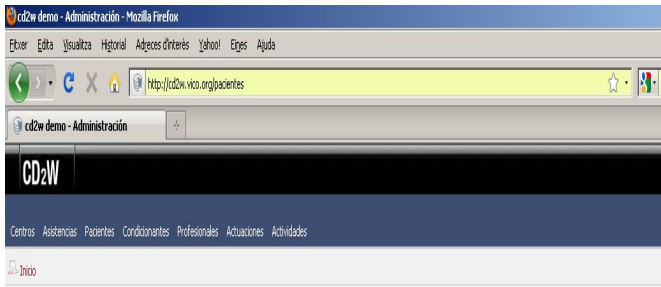
Figure 6- Temporal series for a patient

Figure 7- CDA R2 Document for an encounter

5. Technological Implementation

The system was implemented using an open source Apache application service, a MySQL database, PHP development environment and the amCharts data analysis and graphing component.

Evaluation/Assessment/Lessons Learned

The HL7 Reference Information Model was very useful to aid modeling our specific domain and generate a scheme to integrate all participating applications into an information model.

The process for exchange could be established, but we needed to educate the participating centers on the use of CDA R2 and the rationale for asking each piece of information. Nevertheless, we needed to bridge their data model to CDA R2 by providing them with a customized XSL for each center. Coexistence of documents with the relational information needed to explore the embedded information was possible, although we need to test with more volume.

The user interface was enough for our pilot users from three centers to achieve their data exploration and verification needs. The use of open technologies and standards was a key factor to minimize development time. Our normalizing efforts were not finished, we end up using ICD-9 and internal ICS vocabularies.

Future Plans

Extension of this tool to be accesible to patients and other professionals will be studied, but current policies make very difficult to gain access to patient data, even for this approved project. We also want to explore using native XML open source databases. A project with a greater scope will be studied but the whole approach and the generated model are suitable for other domains.
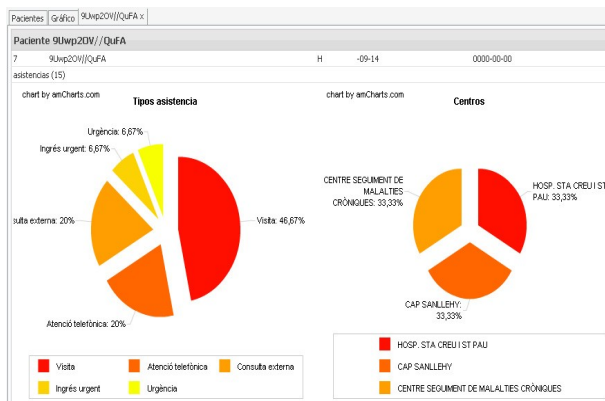
Acknowledgements

Bibliography

[1] HL7 Clinical Document Architecture, Release 2
Robert H Dolin et al
JAMIA 2006;13:30-39 doi:10.1197/jamia.M1888

[2] Ley Orgánica 15/1999, de 13 de diciembre, de Protección de
Datos de Carácter Personal
(BOE núm. 298, de 14-12-1999, pp. 43088-43099)

[3] The Design of the HL7 RIM-based Sharing
Components for Clinical Information Systems
Wei-Yi Yang, Li-Hui Lee, Hsiao-Li Gien, Hsing-Yi Chu, Yi-Ting
Chou, and Der-Ming Liou
World Academy of Science, Engineering and Technology 53 2009

Contact

Josep Vilalta Marzo
Francesc Layret 24, Badalona, Barcelona, Spain
Email: jvilalta@vico.org