Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Technology 5 (2012) 885 – 894

Procedia Technology

CENTERIS 2012 - Conference on ENTERprise Information Systems / HCIST 2012 - International Conference on Health and Social Care Information Systems and Technologies

Design and Implementation of a Data Warehouse for Benchmarking in Clinical Rehabilitation

João Completo[a,b,1], Rui Santos Cruz[a], Luisa Coheur[a], Manuel Delgado[b]

[a]*Instituto Superior Técnico, Av. Prof. Doutor Aníbal Cavaco Silva, 2744-016 Porto Salvo, Portugal*
[b]*IASIST Portugal, Campo Grande 56 6-A, 1700-093 Lisboa, Portugal*

Abstract

Clinical Benchmarking provides comparative analysis among healthcare institutions in order to improve their quality and efficiency. This paper describes the development and implementation of a Data Warehouse, following Kimball's Business Dimensional Lifecycle framework, to support a dynamic Decision Support System for Benchmarking in Clinical Rehabilitation. For the Data Warehouse development, the identification of the most important concepts as well as the data modeling process are the most challenging and important tasks, requiring deep knowledge on the subjects in matter. Consequently, the developed Data Warehouse respects the specific characteristics related with Clinical Rehabilitation, addressed through a data modeling process using different Data Marts for each of the major business requirements. The evaluation of the solution, using real clinical records, demonstrated a clear orientation and suitability to the subject in matter as well as the capability to provide accurate current and historical analysis.

Keywords: Clinical Benchmarking; Data Warehouse; Decision Support Systems; Rehabilitation.

[1] Corresponding author. Tel.: +351 96 531 42 55; fax: +351 21 340 41 89.
E-mail address: joaocompleto@gmail.com.

doi:10.1016/j.protcy.2012.09.098

1. Introduction

In public healthcare, clinical data evolved tremendously with the implementation of instruments for financing purposes. The urge to estimate patients' needs and their associated costs to the National Health Service (NHS) institutions led to invest on management Healthcare Information Systems (HISs) where relevant clinical data could be stored and accessed more easily. These systems often integrate with patients' administrative data, in order to allow more comprehensive and complete clinical records.

Countries with NHS have the most available and standardized clinical data, allowing comparisons for Clinical Benchmarking. Standardization of clinical data, in this context, means the ability to share the same informational concepts (What is an episode? What is a procedure? etc.) and their representation (specific diagnosis codes, specific rehabilitation scales, etc.).

Clinical Benchmarking systems are specific Decision Support Systems (DSSs) that allow the analysis and comparison of clinical data, providing objective information to healthcare institutions, in order to improve clinical quality and efficiency.

IASIST Portugal[2], a healthcare management company, has developed a pioneering Benchmarking system for Clinical Rehabilitation named Clinical Rehabilitation Profile (Perfil de Reabilitação Clínica - PRC). This system takes clinical data as input and produces standard clinical benchmarking reports. Although the system provides useful analytical reports with several indicators for benchmarking, it is static in time, meaning that the information draws on data previously sent to IASIST from a specific period. The reports produced for the client are predefined, not allowing historical or trend analysis and the information is also static, being difficult to change the produced standard indicators.

As far as known, no other system or tool addresses this increasingly important area in a dynamic fashion. As such, this paper describes the design and development of a Data Warehouse to support a dynamic DSS for Benchmarking in Clinical Rehabilitation. For that purpose, Section 2 starts by giving an overview of Management Healthcare Information Systems in Portugal and of specific Clinical Benchmarking Systems, Section 3 describes the main concepts related with Benchmarking in Clinical Rehabilitation, Section 4 presents the Data Warehouse development process, followed by the corresponding Design and Implementation in Section 5. To finalize, Section 6 presents the Evaluation methodology followed and the results, and Section 7 concludes the paper and proposes future work.

2. Healthcare Information Systems and Clinical Benchmarking

This section describes some of the most important clinical management Healthcare Information Systems in Portuguese NHS hospitals and the state-of-the-art regarding Clinical Benchmarking Systems.

2.1. Healthcare Information Systems in Portugal

In Portugal, the creation of a NHS in 1979, the Serviço Nacional de Saúde (SNS), rushed the deployment of clinical management Healthcare Information Systems across every public hospital institution. Many of the these early systems, and their successors, were still in use at time of writing. These Healthcare Information Systems made current Clinical Benchmarking possible, not just in Portugal but also in other countries with

[2] http://www.iasist.pt

NHS. Figure 1 shows a timeline[3] of the most important clinical management systems installed in Portuguese NHS hospitals, which are further described[4].

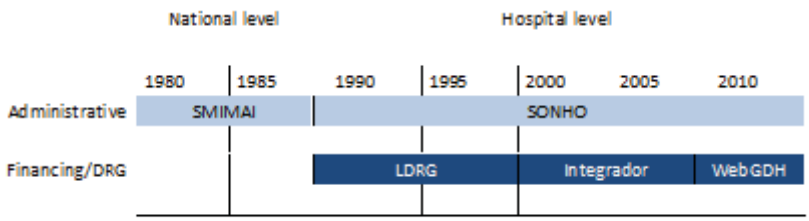

Fig. 1. Healthcare Information Systems in National Health Service

SMIMAI. Sistema Mínimo de Informação Médico-Administrativa no Internamento (SMIMAI) was the first system designed to store administrative and clinical data from hospitals, which was then used to produce hospital statistics.

SONHO. Sistema Integrado de Informação Hospitalar (SONHO) is a patient management system, still used in most Portuguese hospitals, which keeps the master patient index, including administrative and clinical data, functioning as a fundamental support to patient referencing among healthcare institutions.

LDRG. The LDRG is a relational dBASE[5] system created to collect the clinical data from inpatient episodes in acute hospitals and cluster those episodes into Diagnostic Related Groups (DRGs), a patient classification system (further described in this paper).

Integrador. Integrador is the successor of the LDRG. It was developed to collect and manage data related to inpatient episodes.

WebGDH. WebGDH[6] is a system designed to collect and edit records from inpatient and ambulatory episodes and group them into DRGs.

2.2. Clinical Benchmarking Systems

Clinical Benchmarking Systems are a special collection of systems that allow not only a descriptive information analysis, but also a comparative one, highlighting areas of excellence or areas where there is still potential for improvement.

Clinical Management Profile (Perfil de Direcção Clínica - PDC). The Clinical Management Profile (PDA)[7] is the first clinical benchmarking system developed by IASIST with the objective to provide specific indicators to evaluate quality and efficiency in acute hospitals. The system can identify areas of improvement when comparing the hospital with a "comparison standard" composed of other similar hospitals, and also, with its own performance in a homologous period.

---

[3] Based in a figure from http://portalcodgdh.min-saude.pt/index.php/SONHO
[4] Extensive descriptions of SMIMAI, SONHO, LDRG, and Integrador can be found in http://portalcodgdh.min-saude.pt
[5] http://en.wikipedia.org/wiki/DBase
[6] http://www.timestamp.pt/portal/page/portal/TIMESTAMP/PRODUTOS/WEBGDH_v10.pdf
[7] http://www.iasist.pt/en/content/cmp-pda-care-management-profile

IAmetrics. IAmetrics[8] is the first on-line clinical benchmarking system. It was developed by IASIST as the successor of the PDA. In IAmetrics, hospitals have at their disposal information updated as often as they are able to encode the clinical data, and, unlike the PDA where the periods of analysis are fixed, in IAmetrics users can change the analysis period whenever they want to.

Clinical Rehabilitation Profile (Perfil de Reabilitação Clínica - PRC). The PRC[9] is a benchmarking tool created by IASIST in order to periodically produce structured and contextualized reports for Clinical Rehabilitation. Similarly to the PDA and IAmetrics, the PRC provides both positioning and evolutive analysis, comparing the Institution, in a fixed period, to a standard of comparison or to an homologous period.

3. Benchmarking in Clinical Rehabilitation

Rehabilitation reality is quite different from the reality of acute hospitals as the health conditions of patients in rehabilitation require more attention and help from human staff and the inpatient periods are much longer, in order to guarantee that the admitted patient is likely to have major health condition improvements. But health condition improvements require measurement systems allowing those conditions to be monitored.

The main concepts in Benchmarking for Clinical Rehabilitation are the Functional Independence Scales, addressing "what to measure and how to measure?" and the Patient Classification Systems, addressing "how to compare?" the data.

Clinical Benchmarking mechanisms rely on this two elements, allowing to compare episodes classified within the same clinical groups and to calculate the expected results for specific clinical indicators. For example, it is possible to calculate the expected mortality rate, based on the average mortality rate observed in the comparison group, for a specific pathology, in a certain time period, and compare it with the mortality rate observed in a hospital, for the same pathology, in the same period. Naturally, this is only possible with a considerable amount of clinical records, evaluated with the same scales and classified with the same classification systems.

3.1. Functional Independence Scales

The use of evaluation scales allows to objectively evaluate the capacity of an individual, to measure the results of the rehabilitation process, and to provide a common language for the different members of the rehabilitation healthcare community.

Functional Independence Measure (FIM). FIM is a standardized measurement of the type and level of assistance a person needs to perform daily life activities [2, section III]. This is a disability scale and not an impairment scale, hence, it quantifies what the patients truly can and cannot do, regardless of their health condition.

3.2. Patient Classification Systems

Patient Classification Systems or Case-Mix systems [5] aim to estimate the real needs of each patient using ones functional and clinical conditions, allowing an adequate allocation of resources to provide health care with quality and efficiency, classifying patients into groups with similar care needs and, consequently, similar consumption of resources.

---

[8] http://www.iasist.pt/en/contenido/iametrics-1
[9] http://www.iasist.pt/en/contenido/prc-perfil-de-reabilitacao-clinica-0

Diagnostic Related Groups (DRGs). The DRGs is a classification system for patients admitted in acute hospitals, which groups patients into clinical groups that are similar in terms of resources consumption. It allows defining the set of goods and services each patient receives, based on their health conditions.

Case-Mix Groups (CMGs). CMGs are categories developed for classifying rehabilitation inpatient episodes. Patients within the same CMG group have similar clinical characteristics (e.g., diagnosis, procedures) and are expected to consume similar amounts of resources.

4. The Development Process of the Data Warehouse

There are two key aspects concerning the design of a Data Warehouse (DW) to support a dynamic DSS for Clinical Benchmarking. The first is that it consists of the application of a known methodology, typically used in sales and trading businesses scenarios, but to a different problem, i.e., to Healthcare in Rehabilitation. Secondly, unlike common DWs designed to serve the business requirements of a company and its employees, the objective in this case is to serve the needs of clients from a healthcare management company, in the form of a DSS for Clinical Benchmarking.

Technically, the data in a DW is organized to fulfill the needs of business analysis, as a decision support instrument, rather than transactional operations. Particularly, when containing a logical subset of data concerning a specific topic or a specific business function, is named Data Mart.

The basic elements of a DW [4] are rather simple, involving the extraction of data from On-line Transaction Processing (OLTP) systems into a Data Staging Area, where it is cleaned and standardized in order to be loaded to a Data Presentation Area where the data is dimensionally modeled and Data Marts are created. To facilitate complex analysis and visualization, which are the main characteristics of On-line Analytical Processing (OLAP) systems, the data in a DW is typically modeled multidimensionally [1], commonly represented with a Star Schema or a hierarchical Snowflake Schema.

Taking these fundamental aspects into consideration, the development methodology for the DW project closely followed Kimball's Business Dimensional Lifecycle framework [4]. The diagram presented in Figure 2 shows the sequence of high-level tasks proposed by that framework.

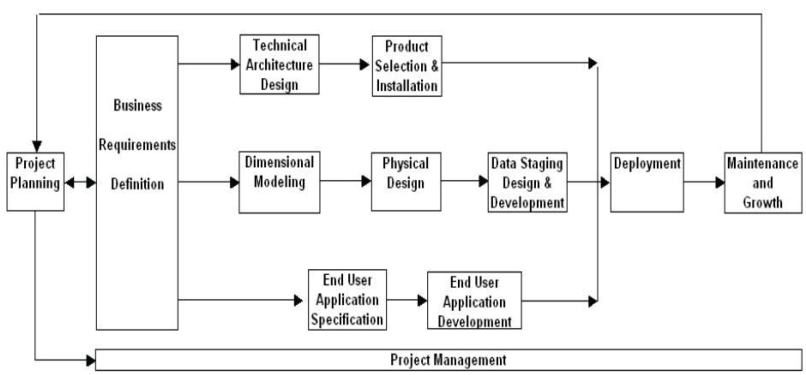

Fig. 2. Business Dimensional Lifecycle

Project Planning and Management. The activities of Project Planning and Project Management consist of defining, planning and managing all the other project tasks.

Business Requirements. Business requirements determine which data must be available in the DW, how it is organized and how often it is updated.

Development Tracks. After defining the Business Requirements, the project is in conditions to initiate the three development tracks:

- **Technology Track** is composed of two main high-level tasks, the Technical Design Architecture, where the technological environment is defined, and the Product Selection and Installation.
- **Data Track** comprises the Dimensional Modeling, the Physical Design, and the Data Staging Design and Development tasks.
- **Application Track** is composed by two high-level tasks, the End-User Application Specification, which defines the standard end-user applications, like calculation formulas and templates, and the End-User Application Development, which involves the development of the specified reports.

Deployment and **Maintenance and Growth** tasks come after the development phase and, therefore, are not addressed in this paper.

5. Design and Implementation

Kimball addresses an organization's receptiveness to move forward with a DW project through the Readiness Litmus Test [4, Chapter 3] which comprises five factors: Strong Business Management Sponsor; Compelling Business Motivation; Information Systems (IS)/Business Partnership; Current Analytic Culture; Feasibility.

When applied to the context of the DW project described in this paper, it was possible to verify that there was a high readiness level for almost every question in each of those factors.

5.1. Business Requirements

Using Kimball's interviewing method a common nomenclature was established, origination a glossary with the most relevant concepts expressed graphically in the Concept Map presented in Figure 3.

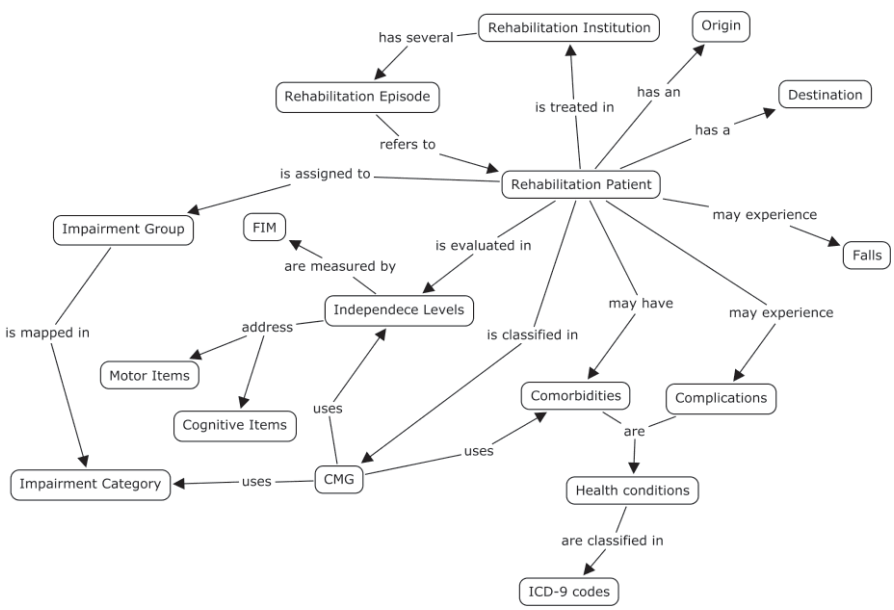

Fig. 3. Concept Map for Clinical Rehabilitation

Additionally, five major business processes were identified: Characterization of the Casuistic, Length of Stay (LOS) Management Evaluation, Functional Improvement, Efficiency and Quality Evaluation. According to these major five business requirements selected, Table 1 describes the DW Bus Architecture Matrix. The table presents the five major Business Processes versus the set of candidate conformed Dimensions.

These processes were modeled into four Data Marts (presented in Table 2) that include conformed Fact Tables containing the key business measures. The grain chosen for the Elementary Fact Tables in each Data Mart corresponds to one Rehabilitation Episode for each Fact Table row.

Table 1. DW Bus Architecture Matrix

Business Process	Time	Patient	Origin	Destination	Case-Mix Group	Rehab. Group	Rehab. Category	Institution	Comorbidities	Complications
Characterization of the Casuistic	X	X	X	X	X	X	X	X	X	
Length of Stay Management Evaluation	X	X			X	X	X	X		
Functional Improvement Evaluation	X	X			X	X	X	X		
Functional Efficiency Evaluation	X	X			X	X	X	X		
Quality Evaluation	X	X		X	X	X	X	X		X

The Dimensional Modeling considered not just the conformed dimensions presented in Table 1 but also AgeClass and AnalysisPeriod. Table 3 shows the Patient Dimension attribute detail. Each dimension will use a surrogate key.

Table 2. Mapping Business Processes into Data Marts

Data Mart	Business Processes
Data Mart #1	Characterization of the Casuistic
Data Mart #2	Length of Stay Management Evaluation
Data Mart #3	Functional Improvement and Efficiency Evaluation
Data Mart #4	Quality Evaluation

Table 3. Patient Dimension attribute detail

Attribute	Description	Cardinality	Sample
Patient File	Represents the patient file number.	N	123456789
Birth Date	Represents the patient birth date.	Max. N	15-08-1990
Gender	Represents the patient sex.	2	M

The defined Data Marts address specific business requirements, based on elementary facts. These facts also provide the basis for the implementation of other models with other Aggregated and Derived facts. Figure 4 shows an extract of the Elementary Fact table for Data Mart #3.

Fig. 4. Extract of Data Mart #3: Improvement-Efficiency Elementary Fact Table

5.2. Physical Design

Data Marts #1 and #4 include dimensions that correspond to Many-to-Many relationships. These relationships relate Episodes to a set of Comorbidities (Data Mart #1) or Episodes to a set of Complications (Data Mart #4). To address this issue, a "bridge" table is included between the Fact Table and the Comorbidities/Complications Table, to which each episode will have a specific surrogate key. The "bridge" table will then contain a set of records for each episode in which one or more Comorbidities/Complications are verified.

Sometimes the basic measures provided by the Elementary Fact Tables do not provide the most useful data for a particular task. Aggregated Facts consist on summarizing facts from the Elementary Fact Table and Derived Facts are facts calculated using other Aggregated or Elementary Facts. Figure 5 shows the Aggregated Fact Table for Data Mart #3, including useful aggregated and derived facts for the calculation of some benchmarking indicators for Improvement Evaluation.

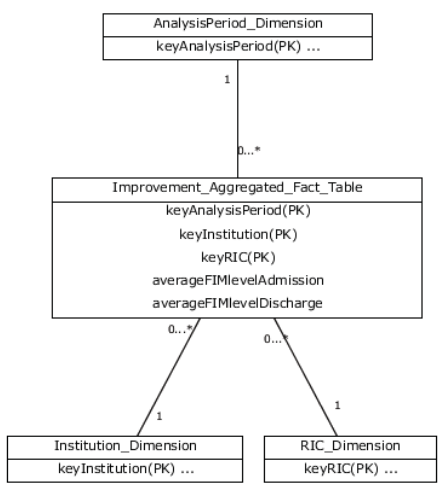

Fig. 5. Data Mart #3: Improvement Aggregated Fact Table

The new aggregated Fact table presented in Figure 5 is a simple example of how to address the indicators related to the FIM levels. The grain of the Fact Table is one record for each Rehabilitation Impairment Category (RIC) group, by Analysis Period and by Institution.

6. Evaluation

Following the path of other DSSs in Healthcare, as is the case of the Lymphoma Diagnosis and Treatment DSS [6], which holds on to Inmon's Data Warehouse definition [3], the evaluation methodology for this project was also based on Inmon's DW characteristics, "subject-oriented, integrated, time-variant and non-volatile", in two distinct perspectives, one "Formal" and the other "Operational".

IASIST Portugal provided the necessary conditions on its main servers and the environment used for the evaluation was composed of a MySQL Database Management System and its administration software package.

The Dimension Tables were loaded with the proper values for the Institutions, Rehabilitation Impairment Categories (RICs), Rehabilitation Impairment Groups (RIGs), Case-Mix Groups (CMGs), Origins, Destinations, and the additional Age Classes and Analysis Periods.

The Elementary Fact Tables were loaded with real clinical data corresponding to three years of Rehabilitation hospitalization episodes, for three different Rehabilitation Institutions, in which the Clinical Rehabilitation Profile (PRC) was already installed and being used. From these institutions, two were Portuguese, specifically, the "Centro de Medicina de Reabilitação do Alcoitão" and the "Centro de Medicina de Reabilitação do Sul", and the other Spanish, specifically, the "Institut Guttmann".

A set of simple tests, as exemplified in Listing 1, were ran in order to assess the conformity of the Data Warehouse architecture and the simplicity of the querying involved.

```
SELECT SUM(numberOfFalls) , SUM(rehabilitationProcessInterruption)
FROM Quality_Elementary_Fact_Table f ,
INNER JOIN AnalysisPeriod_Dimension p ON f.keyAnalysisPeriod = p.keyAnalysisPeriod
INNER JOIN Institution_Dimension i ON f.keyInstitution = i.keyInstitution
INNER JOIN RIC_Dimension r ON f.keyRIC = r.keyRIC
WHERE p.analysisYear = '2012' AND i.institutionName = "Institut Guttmann"
AND r.ricName = "Orthopedic";
```

List. 1. Example of a test query to the Elementary Quality Fact Table

Data Warehouse performance tests were not considered relevant in this phase, according to the evaluation criteria proposed on the Lymphoma Diagnosis and Treatment Decision Support System [6].

This qualitative evaluation, supported by real clinical records, showed a complete orientation to the subject in matter, with integrated data and multiple time-tracking information for each clinical episode, providing both current and historical or trend analysis.

7. Conclusions and Future Work

In Clinical Rehabilitation, the use of standard functional measurement scales and a proper patient classification system allows to perform Clinical Benchmarking, being the analytical requirements of Clinical Benchmarking systems well addressed by Data Warehouse architectures and On-line Analytical Processing (OLAP) mechanisms.

Kimball's Business Dimensional Lifecycle has proven to be a detailed but also a very straight-forward framework. The identification of the important concepts and the modeling of the subject into the proper Data Marts were the two most challenging but also important tasks in the Data Warehouse development. The definition of the Data Marts according to the main business processes, and the use of fact aggregations, allowed to easily address the main indicators in the business requirements, being these indicators easily extensible as well.

The evaluation showed a careful orientation to the subject in matter, with integrated data from clinical records with multiple time-tracking information for each episode, providing accurate analysis, both current and historical.

Topics to be addressed in future work are Data Staging, the definition of a DW Metadata structure and a User Interface based on controlled Natural Language.

Acknowledgements

The work presented in this paper has been performed at IASIST Portugal, a health management company, which provided all the support for the development, implementation and testing of the DW.

References

[1] Chaudhuri, S., Dayal, U.: An Overview of Data Warehousing and OLAP Technology. SIGMOD Record 26(1), p. 65–74 (1997).
[2] CMS: The Impatient Rehabilitation Facility-Patient Assessment Instrument. Centers for Medicare & Medicaid Services (Apr 2004).
[3] Inmon, W.H.: Building the Data Warehouse. Wiley, Indianapolis, 4th ed. (2005).
[4] Kimball, R.: The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses. Wiley, New York (1998).
[5] Stineman, M.G., Granger, C.V.: A modular case-mix classification system for medical rehabilitation illustrated. Health Care Financing Review 19(1), p. 87–103 (1997).
[6] Wah, T.Y., Sim, O.S.: Evaluating a Data Warehouse for Lymphoma Diagnosis and Treatment Decision Support System. In: Proceedings of the 2010 Fourth UKSim European Symposium on Computer Modeling and Simulation. p. 57–62. EMS, IEEE Computer Society, Washington, DC, USA (Nov 2010).