Best Practices in the Design and Development of Health Care Information Systems

Marius Muji, Radu Ciupa, Peter Olah, Vladimir Bacarea, Tudor Calinici, Marius Marusteri

Abstract — The objective of this paper is to present design patterns and development methodologies for Health Care Information Systems. First, there will be presented general purpose issues concerning the design of Database Information Systems, focusing on the differences and similarities between Health Care and Business Information Systems. Different development methodologies are compared in terms of system quality and skills required for the development team. Second, specialized data modeling solutions for the medical field are described. Two specific requirements for the data model are identified and treated: scalability in design and semantic consistency of the historical data.

Keywords: health care information systems, database design patterns.

1. INTRODUCTION

Information Systems were first designed for business purposes. For this reason, systems architecture, development methodologies and people profile are generally biased by business information requirements. Although the design principles and the development methodologies are similar, Health Care Information Systems have some specific requirements, which lead to particular approaches for system design and development.

One of the most important constraints of the medical data is the need to permanently adjust the items of data to be stored in the database. Some data architecture patterns are presented, which enable the information systems to support those adjustments without any change of the data model or software application, providing scalability in design to the systems. Another issue related to medical data is the possible existence of different terms for the same concept. This is usually the case when legacy data are imported, or when medical teams from different specializations are using the same system. Design solutions are offered to enable the specialist to “merge” two different terms or to keep track of their correlation, assuring thus the semantic consistency of the database.

2. HEALTH CARE INFORMATION SYSTEMS VS. BUSINESS INFORMATION SYSTEMS

With regard of the system design and development methodology, Health Care Information Systems share the same techniques and tools with Business Information Systems. Despite the appearance that IT technology evolves at an exponential annual rate, the industry is in fact at an early stage, which determines a lack of standards for development methodologies. The methodology used by each software company, if there is one, depends more on the project managers’ profile, than on some industry’s best practices. There are two opposite profiles for IT people involved in this kind of projects: the programmer, and the data architect.

When the project management is done by programmers – and, unfortunately, this is the most common situation – their process-oriented approach leads to a development methodology which focuses on output and processing logic [1]. The system’s design in this case is based upon what the system is supposed to do, not on the ideal organization of data. Data are considered secondary to the applications, it just having to match the specifications established in each application. Moreover, each application is considered separately. One of the major problems reached by this approach is the existence of several files of data, each of them locked within different applications and programs. Many of the files in these different applications may contain the same data elements, the change of which would require changes in each application.

When the project team has a data-oriented approach, the focus is on the ideal organization of data, rather on where and how they are used [1]. This determines application independence – the separation of the data, and their definition, from the applications that use these data. In this case, a central database is created, all the applications accessing the same data.

Even if the data-oriented approach is becoming largely accepted in IT industry, the systems development techniques often reveal the process orientation of the people which use them. Thus, in many cases, the central database is used just for data storage, most of the business rules of the organization being captured in the system at the application level, instead of the database...
level. Contrary, data orientation requires that the business rules have to be modelled, if possible, by the database logical design, or by database triggers and stored procedures, or by application programs, in this priority order. Usually, all the business rules contained by the system can be captured at the database level, the accessing applications serving in this case just as the user interface with the system. This approach should be followed as best practice for database information systems development methodologies.

Besides the common aspects regarding development methodologies, Health Care Information Systems raise some specific problems related to the information requirements of the domain:

a. textual data, instead of numerical data - health care data is mostly textual data, gathering descriptive information, by contrast with business data, dominated by numbers;

b. the need for the integration of text, through a common vocabulary;

c. the importance of the historical data.

A general assessment of health care information systems would first reveal the difference on content and frequency between health care transactions and business transactions. In business, we usually have to deal with simple transactions which associate partners, business units, products, raw materials, places and numbers, with similar pattern and great frequency. In health care the information is more complex, the transactions are fairly unique, and they contain rather textual data than numbers.

The greatest problem with textual data is the integration of terms and concepts. Usually, health care information systems have to deal with data from different sources - hospitals, outpatient clinics, doctor’s offices, emergency rooms, and so forth. In addition, the doctors that feed the database represent different disciplines – pediatrics, cardiology, epidemiology, orthopedics, gynecology, and so forth. Each of these disciplines have their own terminology. Furthermore, the information going into the database is written by different levels of people – physicians, nurses, technicians, accountants and so forth. By ingesting all this diverse textual information, it is discovered that what is happening is a recreation of the Tower of Babel. Everyone is speaking a different language and no one really understands what anyone else is saying. What is needed is a common vocabulary [2].

There is another major difference between data in the institutionalized world and data in the healthcare world, especially the research portion of the healthcare world. That difference is the time value of information. In a business environment, there is a short amount of time in which the data is useful: once the business event linked with a transaction passed, the information contained lose its relevance, and is usually classified in a few years. But in the healthcare environment data has a long life indeed. For the purpose of studying a disease, medical records that are 50 years old may be extremely valuable, even where many of the people whose records are represented are departed [2].

3. SCALABILITY IN DESIGN

The concept of scalability in design can be explained as the capability of an information system to extent its designed structure and/or functionality, without changes in the database structure or in the system’s application modules. This property of the information system is generally very important, because of the requirement changes in time, determined by the natural changes of the ‘business’ modeled by the system. The benefits of the systems with this property consist in significant maintenance costs savings.

In healthcare information systems, considering the complexity of the domain and its continuous grow, scalability in design is no more an order winner, but it’s becoming an order qualifier. The systems which require the intervention of a specialized team (analysts, database designers, application developers/programmers) for every change in data requirements would be simply unusable in a health care environment.

In a database which gather standard characteristics about some entities (e.g. standardized medical terms associated with every patient tracked by the system), the best way to ensure scalability is the creation of an open dictionary for the standard terms used in the system and some association tables to link them to the entities (patients) described, as shown below (see figure 1).

If a patient can have associated more than one option for each term in the dictionary, this association would be kept in a table with a similar structure as Patient_Data1, where the primary key is compound by both fields: Patient_ID and Option_ID. If no more than one option is allowed for each term to be associated with a patient, a structure like the one described by the table Patient_Data2 should be used instead. The primary key compound by Patient_ID and Term_ID avoid multiple associations between patients and the set of terms defined in the dictionary.

About the table Patient_Data2 must be said that, although it complies with 3rd normal form, it contains a functional dependency, Option_ID -> Term_Id, which violates the Boyce-Codd normal form. In order to preserve the consistency of the database, a trigger should be defined on update and insert operations for the Patient_Data2 table.
4. SEMANTIC CONSISTENCY OF THE HISTORICAL DATA

In business environments, transactional data have a relatively short life (usually less than five years), historical data summarizing values related to business units, geographical locations, or periods of time. In health care environments, data usually represents concepts and terms associated with people, so it cannot be summarized. Moreover, data with all the original details are needed for a long period of time, both for patients’ tracking and for statistical studies. This is why health care information systems are more sensitive to data evolution.

The most challenging problem related to database consistency preservation is the semantic consistency of the data. When data were imported from legacy systems, which is often the case, or they were added in the system by different people with different specialties or approaches, there is a great chance for the same concept to bear different names in the system.

There are two different ways to gather medical data in a database: writing free textual observations about each patient, or using a standard vocabulary predefined in a system dictionary.

Considering the first case, although personal data of every patient can be tracked without information loss, for research purposes free textual data is hardly useful. There are, however, some rules for textual data integration, like those proposed by Bill Inmon [2]:

- removal of stop words such as “a”, “an”, “and”, “the”, “which”, “what”, and so forth
- word reduction to the Latin (or other) common stem – move, moved, moving, mover all recognized as coming from the stem “mov”
- synonym resolution by replacement – there are at least 20 ways to say a bone is broken. Recognition of these terms all being synonyms for the same thing is one of most important first steps in the integration of text,
- synonym resolution by concatenation. In some cases synonyms are replaced; in other cases synonyms are recognized and are concatenated with the verbiage for the common term,
- homograph resolution by replacement. The term “ha” means heart attack to a cardiologist and “hepatitis A” to an endocrinologist. These homographs need to be recognized and resolved if the text is to make sense.
- alternate word spelling resolution. Names and procedures all can have alternate spellings and acronyms. The system must recognize the terms and prepare a resolution,
- negativity exclusion. When “no” or “not” precedes a term, the term needs to be excluded from indexing or further processing,
- external categorization of words. Healthcare text must be exposed to and analyzed versus different external categories of data to determine its relevancy,
- treatment of both words and phrases. It is not sufficient to look just at words. Phrases must be able to be processed as well,
- creation of internal themes form the text that is being read,
- and other activities.

Following the above rules, the resulting data will gain some accuracy, but not enough to support statistical studies. The most important condition for a medical database in order to support research studies is the existence of a standard vocabulary. All terms that will be used later for statistical analysis have to be previously defined in the system’s data dictionary.

When data from legacy systems have to be integrated in the actual system, there is often a problem to match the dictionaries of all systems. The best way to do this is to

Figure 1. An open structure which confers scalability in design to the system
Once the synonyms table is set up, the system should be able to provide the users with the possibility to use transparently all historical data, regardless of their origin. However, the completion of this table has to be done manually, term by term, by specialized personnel.

5. Conclusions

Similar with any other specialized information systems, health care information systems should have a data-driven architecture. The only ‘business’ requirements that absolutely, unconditionally must be satisfied by the information system are data requirements [3]. However, data constraints specific for health care environment require special treatment in health care applications. The continuous evolution of the medical science determines continuous changes of the information needs, which require design flexibility for the system’s data model. While business data occurs in repetitive numerical transactions with a simple pattern, healthcare data means textual transactions, with lower frequency, but more complicated patterns. Historical value of data is also greater in health care than in business.

In order to integrate health care data from different institutions, different medical disciplines, or subsequent generations of information systems, a common vocabulary is needed. Moreover, it is of great importance that this common vocabulary – which is in fact a dictionary with standard medical terms – is kept up to date by a medical team, so that all the data inputs would find their right match in the system.

6. References

[3] Lewis B., Data-oriented application engineering: an idea whose time has returned, TDAN.com January 2007