Data Models' Synchronization in MUDRlite EHR

Nagy, Miroslav
2006

Data Models' Synchronization in MUDRlite EHR

Post-Graduate Student:
MGR. MIROSLAV NAGY

Department of Medical Informatics
Instutite of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2
182 07 Praha 8

Czech Republic

nagy@euromise.cz

Supervisor:
RNDR. ANTONÍN ŘÍHA, CSC.

Department of Medical Informatics
Instutite of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2
182 07 Praha 8

Czech Republic

riha@euromise.cz

Field of Study:
Biomedical Informatics

The work was supported by the project number 1ET200300413 of the Academy of Sciences of the Czech Republic.

Abstract

Contemporary electronic health record (EHR) applications incorporate comprehensive medical knowledge that is being updated very fast. Even EHRs that are built considering best practices vary slightly in their data models (DM) over time. There are various software tools supporting "DM synchronizations" on the market. Within this paper we describe special requirements that are to be set on these tools to be successfully applicable for synchronization of DMs in healthcare domain. As the available tools do not meet the requirements well we also present a new open source DM synchronization tool Schemagic that is particularly suitable for these synchronizations. Schemagic was developed as a supportive tool within the project Information Technologies for Shared Health Care. Several partners from academy as well as from industry joined for collaboration on this project in order to cooperate on new approaches to support sharing of medical data and knowledge among heterogeneous information systems. The Schemagic synchronization tool has been successfully tested while formalizing the clinical contents of EHRs in dental medicine. A rapid evolution of the underlying DM that started with a few clinical concepts and evolved stepwise into hundreds of attributes and relations has confirmed that Schemagic considerably multiplies the power and scalability of EHR systems and significantly simplifies further research and development.

1. Introduction

Variation in the use of clinical resources, outcomes, costs, access to health care, clinical content of electronic health records (EHR), and quality of provided health care are well recognized, ever present aspects. It is a phenomenon that affects all sectors of the health care delivery process that is important to clinicians, administrators, and patients [1]. New generation EHRs should be based on an architecture flexible enough to last several generations and allowing for improvements, but without starting from scratch [2].

An EHR for integrated care is defined as a repository of information regarding the health status of a subject of care in computer processable form, stored and transmitted securely, and accessible by multiple authorized users [3]. It has a standardized or commonly agreed logical information model which is independent of EHR systems. Its primary purpose is the support of continuing, efficient and quality integrated health care and it contains information which is retrospective, concurrent and prospective.

The clinical content (i.e., collected concepts) of an EHR for integrated care evolves over time. There remains a wide void between the basic definition and the ideal existence of a standardized EHR. The reality is that we lack a common data model (DM), a common set of data elements, common vocabulary, and a common set of scenarios [4]; however, there is some hope in the CEN prEN 13606 becoming an EU standard in the near future. Still, even the EHRs built considering best practices vary slightly in their

DMs over time, which is an apparent consequence of the variation mentioned at the beginning of this paper. Moreover, contemporary EHR systems incorporate comprehensive medical knowledge that is being updated regularly and that also might force the DMs to slightly evolve. A seamless support of information flow for increasingly distributed healthcare processes requires integrating of heterogeneous IT systems into a comprehensive distributed information system. Different standards contribute to ease this integration [5]; however, a kind of synchronization of the underlying DMs is often enforced.

Within the project Information Technologies for Shared Health Care several partners from academy as well as from industry joined in order to cooperate on new approaches to the EHR design with the main goal of solving issues connected with sharing medical data among heterogeneous hospital information systems (HIS) and EHR applications. Together, they are preparing solutions to model various EHRs and HISs using worldwide-used standards, based mainly on the HL7 version 3 specification [6]. The MUDRLite EHR system [7] is among pilot applications chosen to test new approaches. It provides interfaces to include user-defined modules and components. These interfaces enable to develop and integrate special components to share data among other systems based on a defined communication standard. MUDRLite architecture is based on two tiers. The first one is a relational database and the second one is a MUDRLite User Interface. The database schema corresponds to particular needs and varies therefore in different medical environments. An important part of the project was to incorporate and structuralize clinical information in dental medicine using the Dental Cross component [8], [9].

Evolution of the incorporated dental medicine knowledge obviously forced the evolution of the underlying DM over time. This paper discusses issues we faced and requirements we analyzed while synchronizing DMs in order to support continuous health care. Approaches we came up with can not only be applied to EHRs in dental medicine but to information systems in healthcare in general terms.

2. Synchronization of Data Models in Healthcare-oriented Information Systems

The rapid progress in database systems research over the past couple of decades has resulted in the evolution of diverse database environments. Consequently, developing a declarative approach to schema integration in the context of heterogeneous database systems was major goal of various research activities [10]. Lately, relational databases with object-oriented extensions became a standard in the data management field. As Entity-Relationship (ER) modelling approach became a widely used methodology of designing database models and schemas [11] various issues connected with the synchronization of DMs and maintenance of data integrity developed to an extensively investigated issue [12].

Resulting from the ongoing research and answering to the presented demand various software tools supporting DM comparisons and synchronizations have been produced, e.g., [13], [14], [15], or [16] to name but a few. All the tools are generally intuitive to visually compare and synchronize two databases. They scan both databases and provide the combined schema tree view, with all differences marked. However, even that these tools are fast and easy to operate we have realized that they do not meet all requirements that are to be implemented while synchronizing of DMs in information systems in medicine. Basic functionality of a DM synchronizing tool consists of being capable to read metadata of relational database schemas, comparing two database schemas in order to find changes that are to be solved during the synchronization process, and generation of the synchronizing SQL script. However, further analysis shows that a successful tool synchronizing DMs in healthcare domain should not only provide the basic functionality but also it needs to fulfil other requirements stated as follows:

- *Universality.* As information systems in healthcare domain are increasingly distributed over various heterogeneous IT systems the tool should be universal enough to be run on various operating system platforms (e.g., MS Windows, various Linux based systems, Mac OS) and capable of co-operating with many database system, e.g., Oracle, MS SQL Server, DB2, PostgreSQL, MySQL.
- *No data loss.* The tool must not affect (delete or change) any data stored in the database that is being synchronized. This is especially important in the healthcare domain where patients' data are very

sensitive and the importance of this requirement is well recognized. Even though this requirement seems to be obvious, it is often not the case. A simple data loss may be caused by rounding errors while converting between incompatible data types or by truncating large text fields. Less obvious (and thus more dangerous) errors may be introduced while losing of some functional relations and dependencies that exist in the original database and cannot be held in the new model. Best practices of database normalization and DMs design shall always be followed [17].

- *Extended security*. Patient identification is one of the highest and most controversial priorities for the implementation of HISs [18]. There is a broadly shared goal to better understand the long-term health status of patients when addressing their immediate needs, to study the effectiveness of different patterns of care, to investigate the long-term outcomes of proposed interventions through clinical research studies, and to optimize the system of healthcare delivery; however, a misuse of these sensitive information shall always be avoided. Within a HIS it may be often the case that the clinical data are separated from the administrative ones so that database administrators cannot get the full-context information they do not need. A successful tool used to synchronize DMs in healthcare information systems shall support these scenarios.

- *Extensibility*. The tool should be extensible in different ways in order allow the user (developer) adding of new functionality as this may be of significant importance in the continuously evolving medical domain, e.g., the user should be able to extend the set of supported database objects. Synchronizing capabilities ought to be improved or extended in a convenient way, which means, for example, without the necessity of developing pieces of program code (e.g., new Java classes). However, it also should be possible to add new features by adding plug- ins, which is lately a common way of extending contemporary software.

- *Off-line usage*. The schema synchronization capability should be available also in an off-line mode without a direct connection to a particular database at the time of synchronization. This feature is generally required in 24x7x365 systems providing year-round performance with critical consequences when the system goes down. Certainly, HISs are the case. The offline usage supports synchronizing the developers' and customers' schemas before the application deployment.

- *Automation support*. As the synchronization process in distributed environments, e.g., complex healthcare information systems, is usually not a simple one, a command line interface becomes useful. It helps to manipulate schemas of various servers and to synchronize their models by previously well tested scripts.

- *Standards-based design*. In order to ease the end-user usage as well as potential extensibility by third parties the application of standards is always essential.

- *Extensive reporting*. A successful synchronizing tool should be capable to generate a well structured report about the synchronization performed.

- *DM documentation*. The ability of keeping updated documentation about the current DM is essential. This usually does not take much additional effort as the tool needs to learn the DM details to perform the synchronization anyway.

- *National languages support.* Correct handling of database objects named or in any other way linked with terms using language specific characters is an inherent feature that can be essential in localized information systems. Even though this seems obvious it is often not the case.

As mentioned above there are various tools supporting DMs synchronization. However, these tools can mainly serve as good inspirations but they cannot be directly used to synchronize DMs of EHRs and HISs as they do not meet all the requirements well. This was mainly the reason why we have developed a new synchronizing tool in frame of our research aimed to support continuous shared health care.

3. Schemagic – Data Models Synchronization Tool

Schemagic is a tool developed for DM synchronization with the focus on healthcare environment. It was developed in frame of the applied research within the project *Information Technologies for Shared Health Care* number 1ET200300413 of the Academy of Sciences of the Czech Republic. Further description of Schemagic's functionally, features, and usability can be found in [19]. It is based on current state of the art in this field and on ideas gained while studying various synchronization tools with the main goal to meet all the requirements stated especially for DMs synchronization in healthcare domain. The tool is open source and can be used by anyone for free. Its source codes are based on design patterns [20] in order to be easily modifiable and extendable by independent developers in the future.

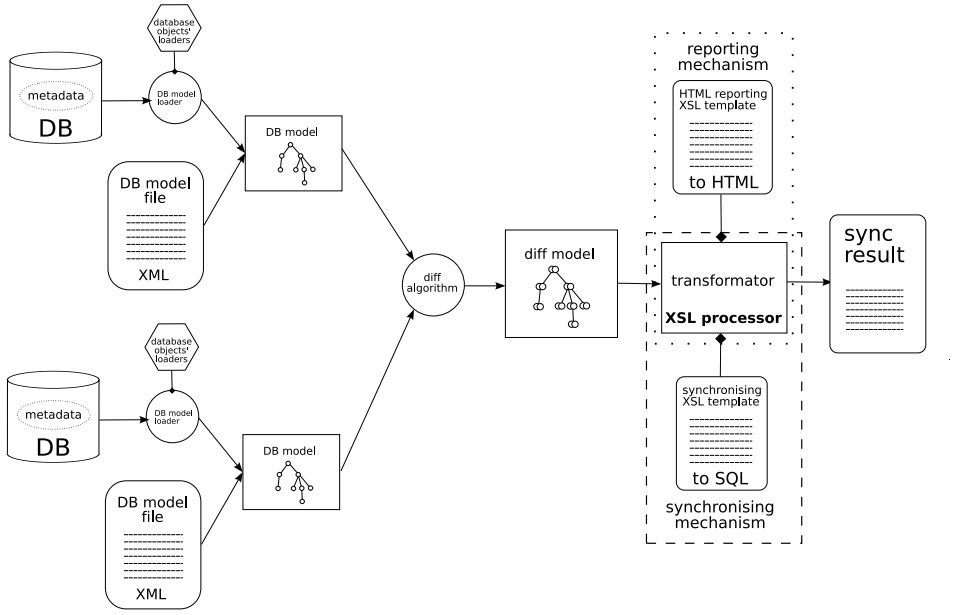

Figure 1: Schemagic's functionality model

The synchronization process in a simplified form is described in Figure 1. During the synchronization source schemas' meta-data are obtained either from a database or from an XML file describing the DM. The next step of the synchronization process is the execution of a specially designed *diff* algorithm that compares both the target and the source schemas and finds differences. The diff algorithm produces a *diff model* that holds all divergences. Final step of the process consists of creation of the synchronization SQL script via XSL transformation [21] mechanism. The script can be run anytime to perform the synchronization itself. The Schemagic tool ready for download as well as its documentation can be found at WWW pages [22].

4. Results

The Schemagic synchronization tool has been successfully tested while applying the MUDRLite EHR into the dental medicine domain [8], [9]. The fact that the detailed structuralization of clinical content of EHRs in dental medicine was a significant part of our research project caused a rapid evolution of the underlying DM that started with a few clinical concepts and evolved stepwise into hundreds of attributes and relations. The Czech national patent application number PV 2005-229 contains some results of this particular research.

Owing mostly to universality of the tool there were no problems with compatibilities of various DMs and their distributed parts that were evolving during the formalization of the clinical EHR content in dental medicine. The capability to generate the DM documentation has made our research and development very convenient. Moreover, we have successfully applied Schemagic while formalizing the clinical content of EHR systems in neurosurgery.

5. Conclusion

There are various tools supporting DMs synchronization available; however, these tools are not particularly suitable for synchronizing of DMs in healthcare domain. We propose a new open source software tool Schemagic that may be used for these synchronizations. Owing to its extensibility there is no problem with support of additional representations of EHR DMs. This tool has been successfully tested while synchronizing various DMs developed within the project Information Technologies for Shared Health Care. Using Schemagic synchronizing tool multiplies the power and scalability of EHR systems and significantly simplifies further research and development. Moreover, the open source based character and the openness of the tool enables additional extensions and modifications of the tool's features according to particular users' needs.

References

[1] T.J. Hannan, "Variation in Health Care – the Roles of the Electronic Medical Record", *Int J Med Inform*, 54 (2), pp. 127–36, 1999.

[2] H. Linden, J. Talmon, H. Tange, G. Boers, A. Hasman, "An architecture for a virtual electronic health record", *Studies in Health Technology and Informatics*, 90: pp. 220–5, 2002.

[3] International Organization for Standardization: ISO/TR 20514:2005, "Health informatics – Electronic health record – Definition, scope and context"

[4] G.C. Liu, J.G. Cooper, K.M. Schoefflerb, W.E. Hammond, "Standards for the Electronic Health Record Emerging from Health Care's Tower of Babel", *Proc AMIA Symp*, pp. 388–92, 2001.

[5] R. Lenz, M. Beyer, K.A. Kuhn, "Semantic Integration in Healthcare Networks", *In: Connecting Medical Informatics and Bio-Informatics: Proceedings of MIE2005 – The XIXth International Congress of the European Federation for Medical Informatics*, vol. 116, pp. 385–90, 2005.

[6] Health Level Seven, http://www.hl7.org (last accessed July 9, 2006).

[7] J. Spidlen, P. Hanzlicek, J. Zvarova, "MUDRLite – Health Record Tailored to Your Particular Needs", *Transformation of Healthcare with Information Technologies*, pp. 202–209, 2004.

[8] J. Spidlen, M. Pies, Z. Teuberova, M. Nagy, P. Hanzlicek, J. Zvarova, T. Dostalova, "MUDRLite – an Electronic Health Record Applied to Dentistry by the Usage of a Dental-Cross Component", *EMBEC'05 Conference*, Prague, 2005.

[9] J. Spidlen, M. Pies, Z. Teuberova, M. Nagy, P. Hanzlicek, J. Zvarova, T. Dostalova, "MUDRLite Components Usage for Sharing EHR Data in Dental Medicine", *In: Sharing Health Information*, Merkantilie, Prague, pp. 83–87, 2005.

[10] L. Lakshmanan, F. Sadri, I. Subramanian, "On the Logical Foundations of Schema Integration and Evolution in Heterogeneous Database Systems", *Proceedings of 2nd Int. Conf. on Deductive and Object-Oriented Databases*, Phoenix, pp. 81–100, 1993.

[11] S. Spaccapietra, S.T. March, Y. Kambayashi, editors, "Conceptual Modeling – ER 2002", *21st International Conference on Conceptual Modeling Tampere*, Heidelberg: Springer, 2002.

[12] C. Batini, M. Lenzerini, S.B. Navathe, "A comparative analysis of methodologies for database schema integration", *ACM Computing Surveys (CSUR)*, vol. 18 n.4, pp. 323–364, 1986.

[13] AdeptSQLTools, "The Adept SQL Diff tool", http://www.adeptsql.com (last accessed Jan 27, 2006).

[14] Spectral Core, "Spectral Core Sync Database", http://www.spectralcore.com (last accessed Jan 27, 2006).

[15] Best SoftTool, "SQLDBCompare", http://bestsofttool.com (last accessed Jan 27, 2006).

[16] Red Gate Software, "SQL Compare", http://www.red-gate.com (last accessed Jan 27, 2006).

[17] P. Rob, C. Coronel, "Database systems, design implementation and management", *Wadsworth Publishing Company*, Belmont, California, 1993.

[18] Board of Directors of the American Medical Informatics Association, "Standards for medical identifiers, codes, and messages needed to create an efficient computer-stored medical record", *The Journal of the American Medical Informatics Association*, 1, pp. 1–7, 1994.

[19] M. Nagy, "Synchronisation of Relational Schemas", *Master Thesis at Faculty of Mathematics and Physics*, Charles University in Prague, 2005.

[20] E. Freeman, "Head First Design Patterns", O'Reilly, 2004.

[21] D. Tidwell, "XSLT", O'Reilly, 2001.

[22] EuroMISE Centre, "Schemagic project", http://schemagic.sourceforge.net (last accessed May 20, 2006).