# XML Based Mediation for Automating the Storage of SCP-ECG Data into Relational Databases

H Jumaa[1], J Fayn[2], P Rubel[1]

[1]Université de Lyon, Université Lyon1, INSA-Lyon, MTIC, EA4171, F69621, France
[2]Université de Lyon, Université Lyon1, INSERM, MTIC, EA4171, F69677, France

**Abstract**

*The aim of this paper is to present a generic mediator for automating the storage and retrieval of an SCP-ECG message into and from a relational database. The architecture of the mediator is based on the XML technology and its related languages and derivatives (XML Schema, XSLT…), which provide powerful tools for sharing, converting and exchanging information. The adopted methodology consists in the design and in the interconnection of meta-models of the source and target data representations using the XML language. The SCP-ECG format is easily represented in XML, whereas the database meta-model must be appropriately created according to a tree structure by reverse engineering. Then, our method includes a rule base allowing a coherent and secure mapping between the source and the target data for ensuring the database integrity.*

## 1. Introduction

Electrocardiography (ECG) is one of the most important non-invasive diagnostic tools for early recognition of coronary heart disease. An important issue for improving the quality of decision-making in the field of quantitative electrocardiography is to have a rapid access to the patient's medical history and previous ECGs. Hence, ECG databases have been required in Healthcare to manage ECG data for patient care as well as for medical research to discover new medical knowledge using data mining techniques. Numerous ECG data management systems have been designed and developed by healthcare organisations and manufacturers through the last few decades [1]. In addition, the Structured Query Language (SQL) allows to efficiently store, query and retrieve a huge volume of data. Thus, relational databases remain a well-experimented, dominant technology for structured data storage and retrieval, especially for the management of ECG data. A reference core database model is the OEDIPE [2] model which has been proposed within the framework of an European project for the storage of serial ECG data and related metadata. The model, composed of 50 tables and 200 attributes, has been designed to support a modular implementation in various sub-models according to different scenarios of use of the ECG data.

On the other hand, ECG patient data may be recorded by different acquisition devices and stored in the SCP-ECG format [3]. The SCP-ECG European Standard Communication Protocol, which has recently been approved has an ISO standard (ISO/DIS 11073-91064), is the best way for the exchange of ECG data. However, SCP-ECG being an open protocol, it has numerous implementations since most of the sections and of the data fields are not mandatory. As a consequence, the extraction of the protocol fields from an SCP-ECG file and their insertion into a database require a costly development of specific interfaces for each pair of source and target representations. Therefore, appropriate means should be developed for facilitating the process of data extraction and storage from any SCP-ECG file coming from any manufactured device into any relational database of health records.

The aim of this paper is to address this objective. We thus propose a software architecture solution enabling the development of generic components, and which easily results in the setup of specific interfaces adapted to any source or target SCP-ECG data. The challenge is to automate as much as possible the storage and retrieval of any SCP-ECG protocol message into/from a relational database transparently for the user. We propose hereafter an XML based mediation approach for data integration between SCP-ECG files and relational databases. Extensible Markup Language (XML) [4] is a simple, very flexible and standard meta-language which is recommended by the World Wide Web Consortium (W3C) as a mechanism for structuring data for large-scale electronic publishing. XML plays an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. Indeed, XML technology and its related languages and derivatives (XML Schema, XSLT…) now provide powerful tools for sharing, converting and exchanging information via networked

computers, while saving the physical structure of the data. The role of our mediator is to store any SCP-ECG file into a relational database as well as to retrieve the results of any query over the database and to create the corresponding SCP-ECG data sections.

## 2. Mediator architecture design

The adopted methodology consists in the design and in the interconnection of meta-models of the source and target data representations using the XML language (Fig. 1). The XML representation of the SCP-ECG format is easily designed and specified using XML Schema. The database meta-model can be first created by reverse engineering, and then appropriately refined according to a tree structure related to the hierarchy of the tables being updated or queried to ensure the database integrity. Moreover, our method includes a rule base for the support of coherent and secure ECG databases updates. Finally, the design of the interconnection between the source and target data representations is specified using XSLT. All the concepts of the exchanged data are specified and represented in XML, i.e. a standard, general structure with a clear data format.

The proposed mediator architecture and its data flow are depicted in figure 1. The protocol and the database models are wrapped by a converter into their XML representation. The converter α transforms ECG data from an SCP-ECG format into an XML representation and vice versa, by using the protocol meta-model. The converter β has two roles. First it is an SQL generator for each table's row to be updated in the database, from an XML-infoset. SQL queries are created according to the database meta-model, and by applying the rules specified in the Rule-base related to each updated table of the ECG Relational Database. The second role of converter β is to be an XML publisher of the result of each database query into an XML-infoset that is compliant with the XML schema of the database meta-model.

The schemas' mapping between the source and target formats is an essential building block for interoperability management of data exchange. This mapping is achieved by the main component of the mediator, that we call the Meta-Model Manager. This component includes four sub-components. The first sub-component is the protocol meta-model, which is specified using XML Schema to describe the layout of the SCP-ECG protocol. The second one is the database meta-model, which is specified using also XML Schema to describe the structure of the relational database and its physical model. The third sub-component is the mapping meta-model, which is specified using XSLT to describe the interconnection between the protocol fields in the protocol meta-model and the database columns in the database meta-model. The last sub-component is the rule-base, which contains a set of preprocessing rules that will be used to verify the possible pre-existence of the data to be inserted in each table of the database, when generating the SQL script of database update with new SCP-ECG data.

In the following sections we present in detail the Meta-Model Manager component and we will explain the two basic functions of the mediator.

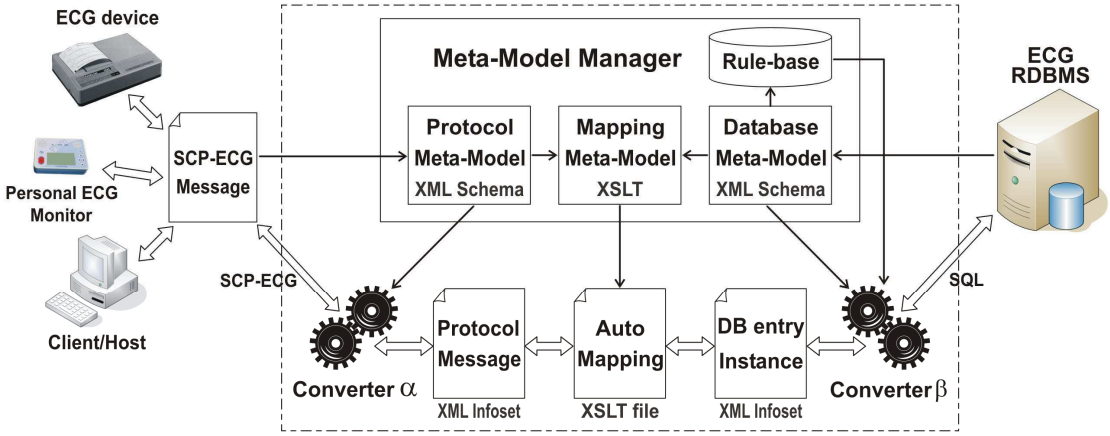


Figure 1. The XML-based mediator architecture

### 2.1. Meta-model manager

The Meta-Model Manager consists of four sub-components (Figure 1).

#### 2.1.1. Protocol meta-model

This sub-component is the XML representation of the SCP-ECG protocol. The protocol meta-model has two roles: (1) first, it is used in the specification of the mapping meta-model with the database meta-model; (2)

secondly, it is used by the converter α to ensure the correct representation of ECG data from a SCP-ECG file into the XML protocol message, compliant with its schema.

The protocol meta-model, specified using the XML Schema language, describes the physical structure of the protocol fields and their related metadata.

The XML Schema Definition Language (XSD) [5], published as a W3C Recommendation in May 2001, has been defined to formulate the constraints over the structure and the data types of a class of XML documents. XSD offers powerful and rich data modeling capabilities which are especially relevant within the framework of our application field with respect to the complexity and to the open structure of the SCP-ECG protocol. Thanks to the concepts of XML Schema, an infinite number of combinations of the XML messages may be built and validated by the same schema. Furthermore, the possibility of potential reuse of schemas may reduce the work and time of meta-model design. For instance adding a manufacturer specific section only requires the insertion of this new section by means of *include* or *import* commands.

Although defining and creating schemas using XSD language is quite complicated, fortunately many XML editors are now available with graphical functions, which make easier the design task.

### 2.1.2. Database meta-model

The database meta-model is the XML representation of the relational database. It has two roles: (1) first, it is used for specifying the mapping meta-model with the protocol meta-model; (2) secondly, it ensures the validation of any instance of database entry with respect to its schema.

The database meta-model, specified using the XSD language, describes the database structure and the information related to the database tables and columns.

XSD offers all the capabilities for capturing and modelling the physical structure of a relational database with respect to the hierarchy of the tables according to the relations and functional dependencies between the tables.

### 2.1.3. Mapping meta-model

The role of this sub-component is to perform the interconnection between the protocol and the database meta-models. It is specified by using the XSLT language.

Extensible Stylesheet Language Transformations (XSLT) [6], published as a W3C Recommendation in November 1999, have been defined for transforming XML documents into other XML documents. In fact, XSLT is an evolving standard and the W3C has recently released a working draft for XSLT2.0. The XSLT syntax allows style sheets composition for dynamically transforming XML documents by associating every element or attribute of the source document with an element or attribute of the target document by a rule template.

In this sub-component two instances of stylesheets are specified: the protocol to database stylesheet and the database to protocol stylesheet.

### 2.1.4. Rule-base

To ensure the coherence of the database while inserting new ECG data, we introduce this sub-component to specify a set of rules to verify the possible pre-existence of information in the database to guarantee the coherence and the integrity of the database during the update.

## 2.2. Mediator functions

Two main tasks are distinguished in the mediator, the protocol to database function and the reverse function, which are depicted in figure 2.

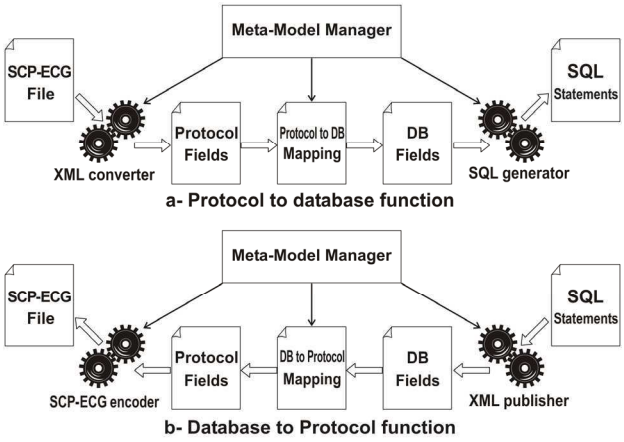


Figure 2. The two main functions of the mediator

### 2.2.1. Protocol to database function

Through this task, as shown in figure 2-a, first, a new SPC-ECG file is received from an ECG device or a Host. Then, the converter reads the file, extracts the protocol fields and stores them into an XML file, compliant with

the protocol meta-model in the Meta-Model Manager. Then, the protocol to database mapping stylesheet is applied to the XML file by the XSLT processor. This operation results in a new XML file with the corresponding database fields. Finally, the SQL statements are created by the SQL generator module, for the update of the database, taking into account the rules stored in the rule-base.

### 2.2.2. Database to protocol function

During this task, as shown in figure 2-b, an SQL query result is converted into an XML file by the XML publisher module, which gives in output the XML fields from the database fields. Then, the database to protocol mapping stylesheet is applied to this XML file. This transformation is performed by the XSLT processor which results in a new XML file with the corresponding protocol fields. Finally, the converter creates an SCP-ECG file with the protocol fields from the last XML file.

## 3. Discussion and conclusion

We have proposed an XML based mediator as a generic solution for the exchange of ECG data between any SCP-ECG compliant device and any relational database storing ECG data, open to the management of the evolution of data representation and to the dynamicity of their schemas.

Converting ECG data into XML has already been investigated in the literature. Many efforts have been made for a standard XML representation of ECG data such as XML-ECG [7] and ecgML [8], which vary in complexity and completeness. These approaches have the same purpose of providing ECG data files in a generic standard format for their exchange over the network. However, we preferred to design an XML representation which is more alike the general structure of the SCP-ECG protocol, and which is thus compliant with all the SCP-ECG instances. Indeed, our SCP-ECG XML representation using the XML Schema better reflects the flexibility and the openness of the SCP-ECG protocol. This is all the more relevant since SCP-ECG has now become an ISO standard.

The integration between XML data and Relational Databases has been an active topic of research in the last decade [9]. Two main objectives were considered: publishing relational databases into XML documents and storing XML documents into relational databases. However, to our knowledge no solution has yet been proposed for the storage of XML data into existing relational databases. In addition, the base of rules we have introduced for a coherent management of the well-ordered successive updates to be performed in the database, yields for data integrity.

The choice of XML as the common representation language in the mediation is motivated by all the existing tools and standards of the XML family. In particular, the mapping design and composition may be performed by using some commercial tool which allows, through a graphical interface, to specify the relationships between two XML schemas, and to finally automatically produce the mapping output in XSLT.

XML provides a level of interoperability for heterogeneous IT systems thanks to the separation of data content from their definition or presentation. It would be interesting to extend our mediator to take into account other standard formats such as DICOM, in order to assess the genericity of the architecture design.

## References


[1] Zhou S, Guillemette G, Antinoro R, Fulton F. New approaches in Philips ECG database management system design. Computers in Cardiology 2003, 2003:267-270.

[2] Fayn J, Rubel P, Willems J, Conti L, Reniers R. Design and implementation strategies of a core database model for the storage and retrieval of serial ECG data. Computers in Cardiology 1994, 1994:189-192.

[3] Health informatics-Standard communication protocol-Computer-assisted electrocardiography ICS: 35.240.80 IT applications in health care technology; reference number EN 1064:2005+A1:2007 and ISO/DIS 11073-91064.

[4] W3C: Extensible Markup Language (XML). http://www.w3.org/XML/.

[5] W3C: XML Schema Definition Language (XSD). http://www.w3.org/XML/Schema.

[6] W3C: Extensible Stylesheet Language Transformations (XSLT). http://www.w3.org/TR/xslt.

[7] Lu X, Duan H, Zheng H. XML-ECG: An XML-Based ECG Presentation for Data Exchanging. Proceeding of the 1st International Conference on Bioinformatics and Biomedical Engineering (ICBBE 2007), Wuhan, China, 2007:1141-1144.

[8] Wang H, Jung B, Azuaje F, Black N. ecgML: Tools and Technologies for Multimedia ECG Presentation. Proc. of XML Europe Conference, London, UK, 2003.

[9] Krishnamurthy R, Kaushik R, Naughton J. XML-to-SQL Query Translation Literature: The State of the Art and Open Problems. Database and XML Technologies, 2003:1-8.



Address for correspondence

Hossam Jumaa
MTIC - EA4171 Hôpital Cardiologique de Lyon
28 avenue Doyen Lépine - 69677 BRON France
hossam.jumaa@insa-lyon.fr