CardioML: Integrating Personal Cardiac Information for Ubiquous Diagnosis and Analysis

Luis Coelho[1], Ricardo Queirós[1],

[1] ESEIG - IPP, Rua D. Sancho I, 981,
4480-876 Vila do Conde, Portugal
{lcoelho, rqueiros}@eu.ipp.pt

Abstract. The latest medical diagnosis devices enable the performance of e-diagnosis making the access to these services easier, faster and available in remote areas. However this imposes new communications and data interchange challenges. In this paper a new XML based format for storing cardiac signals and related information is presented. The proposed structure encompasses data acquisition devices, patient information, data description, pathological diagnosis and waveform annotation. When compared with similar purpose formats several advantages arise. Besides the full integrated data model it may also be noted the available geographical references for e-diagnosis, the multi stream data description, the ability to handle several simultaneous devices, the possibility of independent waveform annotation and a HL7 compliant structure for common contents. These features represent an enhanced integration with existent systems and an improved flexibility for cardiac data representation.

Keywords: XML, e-diagnosis, cardiopathology.

1 Introduction

In the last few years information technologies and electronic devices acquired an increasingly important role on clinical practice and medical services. Most activities, from patient's clinical history management to assisted diagnosis, have benefited with the latest technological developments. Cardiology, a highly requested area in modern societies, is not an exception. The electrocardiogram (EKG), the primary diagnostic instrument for cardiopathology detection, can now be obtained with portable and accurate equipments which can provide reports in the classical graph paper or in a digital format. Phonocardiograms (PCG) and others are also important complements for accurate diagnosis. However the structure and definition of a file format for integrating these informations still finds no agreements and will be the subject of discussion in this paper.

In an effort to improve organization efficiency and eliminate paper as a support for medical exams many healthcare institutions adopted medical image examination and management environments with Picture Archiving and Communication (PAC) system. In PAC systems paper based electrocardiograms are digitalized and stored in an image format which is inappropriate for a waveform type signal. The main disadvantages are the inability to directly perform signal processing operations, the difficulty in waveform segmentation/annotation, the absence of complementary diagnosis information and, not less important, the higher database storage requirements. In some case image processing algorithms are used to extract waveform information but the process in highly inaccurate. In 2000, the National Electrical Manufactures Association (NEMA) extended their Digital Imaging and Communications in Medicine (DICOM) format [1] in order to include ECG waveform information, the DICOM-ECG. This format uses two possible file types, Standard Communications Protocol ECG (SCP-ECG) in binary format and Extensible Markup Language (XML-ECG), that are encoded in a new file containing patient information and other examination related data. Additionally the International Standards Organization (ISO) released his own SCP-ECG [2] while the American Food and Drug Administration (FDA) announced the FDA-XML format [3], different from the first. The FDA-XML format is based on the results of the openECG project [4], an European Commission sponsored project for enhancing the interoperability between medical devices and information systems that gathered the interest of several organizations and companies of the medical area. However there is still no consensus between equipment manufacturers which continue to make their own interpretations of the available file formats and provide proprietary data encoding. In addition to the information storage mode, binary or text, the main differences between these formats are related with the quantity and type of content that complements the waveform and how it is structured. Some authors [5,6] have proposed new extensions to the existing data descriptions structure but they still present issues like lack of flexibility, insufficient number of parameters and/or non-compliance with current standards.

In this paper we propose a new file/data structure designed to bring together the best features of each of the above mentioned formats and simultaneously provide an extended and wide range of possibilities in order to make it suitable for most applications. The proposed format uses an XML structure and encoding in order to provide flexibility, portability and an enhanced integration with other existent architectures and technologies such as Health Level Seven (HL7) (Health Level Seven International) [7] recommendations (which evolved, in his last version 3, also to an XML based document format).

The rest of this paper is organized as follows. In the next section we describe the range of applications for which the proposed file format can be used and a possible usage architecture. In section 3 we thoroughly describe the proposed EKG format giving special details of the included features. Finally we provide the main conclusions and some envisioned work.

2 System Architecture

In this section we describe an architecture to acquire cardiac related data and how to display it in multiple targets for later interpretation of specialized medical personnel. We formalize this data proposing a new XML language, called CardioML, an XML based electrocardiograph language for ubiquous diagnosis and analysis.

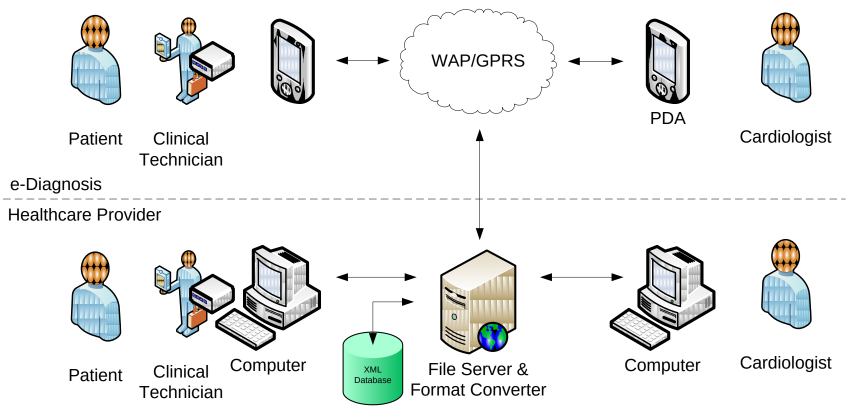

Fig. 1. Overall system's architecture.

2.1 Architecture

Latest cardiac signal monitors are highly portable and with good power autonomy which allows the execution of exams in remote areas either for routine exams on villages which are far from medical centres or in emergency situations where fast action is paramount. However this data requires interpretation from specialized medical personnel which is not always available. In this case the acquired data can be transmitted to a server that can forward it to a personal digital assistant (PDA) or mobile phone where the physician's professional can interpret and suggest an adequate procedure. The overall system architecture is illustrated in Fig.1.

The flow execution sequence for this system is summarized in 5 steps:

1. The Clinical Technician makes the examination to the Patient using one or several equipments (with or without web access);
2. The cardio-physiological signal and related data is passed from the equipment who made the examination to the server; When the equipment does not have this functionality the acquired data is sent to an equipment (computer, PDA or smartphone) with web access (local or remote) who will then forward it to the web server.
3. The acquired data is then converted to the language CardioML (this format conversion will depend on the source file type and data description structure). A first-in first-out document queue is created for managing the conversion process. Server side conversion allows an always updated source file format database and a faster conversion time than client side conversions;
4. The server keeps all the XML files in a XML Database. (Since one of the motivations for the development of this new format was system integration and data portability the database should be supported be a native XML structure);

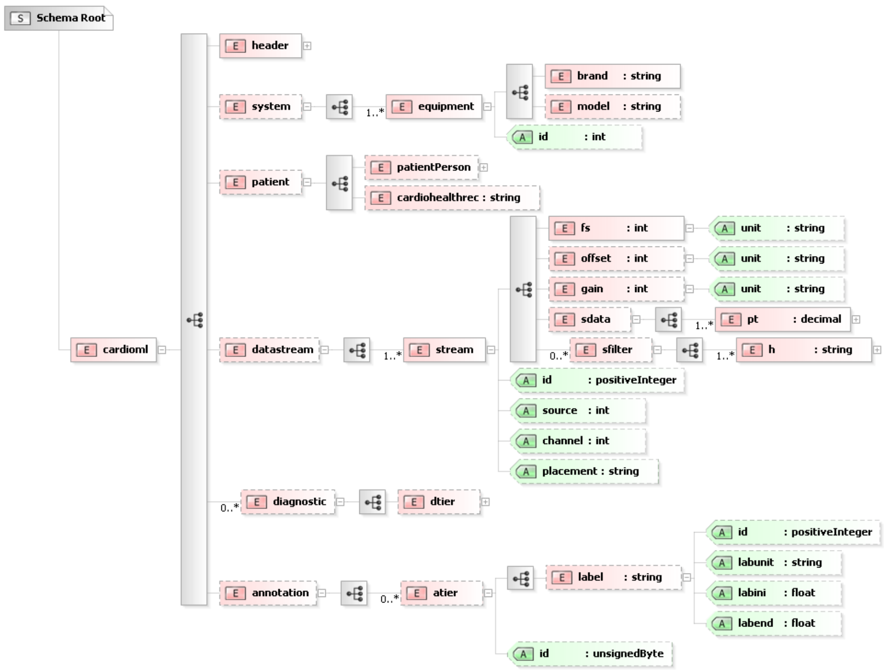

Fig. 2. Partial view of the CardioML schema.

5. The cardioML encoded data can be viewed or annotated using a web based server-side application or downloaded and analysed on the client side. For devices with low computational resources the server can transform the XML file in a specific graphic format (HTML, SVG, Flash, etc.) for viewing purposes only;

2.2 The CardioML language

As we have seen in the previous subsection, the acquired data can be transmitted to servers which can then forward it to a mobile phone where the medical professional can interpret and suggest an adequate procedure. This data transmission over several equipments demands a portable file format and a common description language.

Figure 2 exposes the CardioML schema language. It consists of an element root `cardioml` composed by six elements: `Header`, `System`, `Patient`, `Datastream`, `Diagnostic` and `Annotation`. Elements such as `patientPersonType` or `diagnosticType` are not fully expanded in figure 2 due to space constraints but their content will be thoroughly covered in the next paragraphs.

The element `system` is used to store all the information related with the equipments that were used to perform data acquisition. One novel feature of cardioML is that its structure allows the registration of several devices on an unique file. This is an important feature since the diagnosis of several cardiopahotologies can be enhanced by simultaneously acquiring and analysing several cardiac related signals (for example electrocardiography and phonocardiography) that require independent specific devices. Child elements are not widely described on figure 2 but informations such as next calibration date, accuracy, stability or environmental operating conditions can be stored and are necessary to assess the confidence of the measurements.

Header element contains file related information such as date of creation and owner. For e-diagnosis purposes a set of novel tags are included for registering the geographical information where the data acquisition took place. A document status element was also included for enabling the integration of cardioML files in workflow management systems. Patient information, in the `patient` element, is stored in two distinct child elements, one for the patient's personal information and other for relevant clinical history. CardioML only stores cardiac related information but it allows possible observations or references to extended health records. Both sub-elements are based on the HL7 representation since this format encompasses patient information, most pathologies and medical acts and its usage is quickly being adopted on most healthcare information systems. Cardiac signals are stored as independent data streams that can be provided by any of the existent channel in the given equipments. Each stream is linked to his related source by their unique `id` numbers while the related acquisition conditions are described by elements gain, offset and fundamental frequency (`fs`), all with a related measurement unit. Each acquired signal data is organized in the sub-element `sdata` that is composed by a sequence of `pt` elements. These have an optional attribute that allows the time indexation of samples. Additionally, the behaviour of the several data acquisition chain components (active or passive electrodes, connectors, filters, amplifiers, etc.) can be described by the inclusion of its frequency response in `sfilter` elements. When compared with existent formats [5], the described schema for cardiac signal description provides a more flexible structure, with the possibility of managing several data streams with distinct sources, all independently characterized, it has new important features like the delineation of the data acquisition chain. After signal visualization and analysis the physician can perform a diagnosis that can be stored on the `diagnostic` element. Several diagnostics are allowed, always dated and signed by an author, and, for pathology description, the HL7 structure should be used. Finally, an annotation element is included for enabling the inclusion of waveform metadata. Each annotation tier, composed by several `label` elements, can be related with a specific data stream, using an id reference, or can have a global scope. Besides their intrinsic description, the labels are associated with a time frame defined by their `labini` and `labend` attributes. The separation between annotation and waveform elements is a novel feature that presents advantages such as multi-authoring and local or global scopes.

3 Application

A software infrastructure that supports data interchange using the cardioML format was implemented already implemented. Some screenshots of the client application running on a smartphone are presented in figure 3.

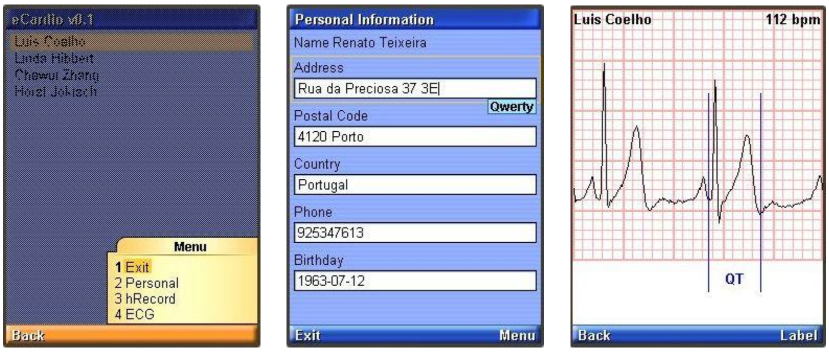

Fig. 3. Screenshots of a mobile application making use of the cardioML structure.

4 Conclusions

In this paper a new XML based format for storing cardiac signals and related information was presented. The evolution of management and archiving techniques for health records in this scope was thoroughly covered and the main challenges identified. An application architecture for e-diagnosis services was shown which creates additional motivations for the development of the CardioML language. CardioML was then presented giving special attention to the main elements and the advantages they provide when compared with distinct formats recently proposed by other authors. The presentation is complemented with the related schema and an application example. The proposed format can provides an enhanced integration with existent systems and an extended flexibility for cardiac data structuring.

References

1. DICOM Homepage, http://medical.nema.org/
2. 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
3. Brown B, Kohls M, Stockbridge N: FDA-XML data format design specification, revision C. www.openecg.net, 1–27, 2003
4. OpenECG Portal, http://www.openecg.net/
5. Xudong, L., Huilong, D., Zheng, H.: XML-ECG: An XML-Based ECG Presentation for Data Exchanging. In: Proceedings of the 1st International Conference on Bioinformatics and Biomedical Engineering ICBBE 2007, pp. 1141-1144, China, (2007)
6. Badilini, F., Isola, L.: Freeware ECG Viewer for the XML FDA Format. In: 2nd Open OpenECG Workshop, Germany, Berlin (2004)
7. Health Level Seven International, http://www.hl7.org/