A Proposed Data Driven Architecture for Cardiology Network Application

Horea Adrian Greblă
Faculty of Mathematics and Computer Science, “Babeş-Bolyai” University of Cluj-Napoca
Mihail Kogalniceanu, 1, Cluj-Napoca, 400084, Romania,
horea@cs.ubbcluj.ro

Călin Ovidiu Cenan
Technical University of Cluj-Napoca
Memorandumului, 28, Cluj-Napoca, 400114, Romania
calin.cenan@cs.utcluj.ro

Abstract
The paper presents a framework for a distributed medical system meant to bring a modern approach and enhance the quality of medical services offered to patients with chronic cardio-vascular diseases, via the latest IT&C technologies. The proposed system provides online interaction between the main actors of a medical system: patients, doctors, medical entities (e.g. hospitals, clinics) and medical authorities (e.g. social services). With the aid of widely accepted medical standards, the system supplies storage for medical records and offers services for data integration between different kinds of healthcare applications and entities. The proposed ontological approach allows for the interchange of medical knowledge and best practices with conceptually organized patients’ records. The proposed solution allows computer assisted diagnoses and multi-criteria medical data analysis, with the possible extent of building a data warehouse for complex medical data mining.

Keywords: data driven architecture, cardiology, multi-agent system

1. Introduction
We live in modern times where connectivity becomes necessity but the health information system is still archaic, stored in a disparate manner and not shareable by clinicians. Shared health records for patients are increasingly needed, preferably in an electronic form so they can provide fast, comprehensive and supervised healthcare. In order to support interoperability of health information, such a system that manages the electronic patient’s health record needs clinical content to be standardized according to some ontology. The standards allow communication of complex health information between systems and the opening of these vertical domains and organizational silos, supplying accurate and semantically computable health information flow.

The modern solution to delivering medical care is possible only in conjunction with a coordinated approach – medical services providers need to collaborate and work in real time (especially in cardiology), which is not suitable when operated by traditional means and methods, such as faxes, appointments and phone calls. For that, an interoperable and integrated patient record in which all healthcare providers can participate, within a framework providing governance, authorization and security measures, is mandatory.

2. Health Care System
The modern approach for the concept of health care should involve access for professionals from various medical fields as well as modern medical technologies such as medication, imagistics and surgical techniques. In a modern medical care system all these services should be focused on the “consumer”, a patient who receives medical attention, supervised care or treatment. This sector is one of the most dynamic of life and the activities
are conducted on many levels, such as primary medical care, secondary, tertiary or even quaternary care. Usually, there are some other actors involved in this health care process as stipulated by governing laws. These actors are usually the Health Ministry or another national authority together with a national social insurance system, insurance companies that sell health related products, and many others. Joining the European Union consequently resulted in boundary extension so the perspective is to involve multinational authorities in healthcare management in a unified manner.

Generally, the term of “health care” is related to prevention, treatment, and management of illness together with the preservation of mental and physical well being through the services offered by medical professionals. It can be said that it involves all the products and services designed to promote health. It can be regarded as a domain in fault intolerant, in which the errors can have severe consequences, a domain which deals with large volumes of not always very well structured data. Our approach proposes the automation of all primary care (seen as family medicine) and secondary care (policlinics, laboratories) sectors.

By “primary care” is usually designated the activity of a health care provider who interacts with patients as a first point of consultation. Generally, physicians that offer primary care services are based in the community, as opposed to a clinic or a hospital. The secondary care services are supplied by specialists who do not generally have the first contact with patients. Some examples are cardiologists, urologists and dermatologists. Usually a second care physician will be appointed only with patients who have already seen a primary care provider. As the EU laws state and it becomes more and more a reality, the primary care provider can be part of another country’s health care system, so the secondary care system needs access to information filled by the primary care and this information has to be presented in a manner easy to understand.

After a primary care physician concludes that a patient is suspect of some disease and needs a supervised medical investigation in a secondary system, a suite of medical tests and procedures are performed to diagnose disease or evaluate disease processes. A major challenge for computer science applied in medicine is to build a clinical management system that manages electronic health records and medical practice management information. The most important feature of such a system is to deal with information stored in many places where the patient may have had some medical treatment: national health authority, social insurance system, primary or secondary care. The best approach to solve this problem is a distributed one, as in [1]. One may say that the use of a personal storage system, as a personal memory card for each patient can overcome this problem if all actors record their data there, but it will not solve the problem of data representation as they can use different ways to store similar or related information using different systems. It is obvious that an approach that represents entities as an ontology, together with ideas and events, along with their properties and relations, as a form of knowledge representation about the medical world [2], can be the answer to the problem.

Even if the medical system is not always up to date with the technology, a computer system is generally used to register medical information such as medical records, laboratory results. Accredited by the American National Standards Institute (ANSI), Health Level Seven (HL7) is an organization whose mission is to develop standards for unambiguous transmission of clinical and administrative health care information between computers. HL7 works to impose standards for clinical terminology used in each patient’s electronic health record, to impose order and uniformity in health information [3]. The standards that are imposed by HL7 for medical terminology can be used as the base of our system, and some tools can be used to easily transform and integrate data in a unified manner. Our approach investigates the possibility of building such a system for a cardiology network, and deals with some of the problems that might occur.
3. Distributed systems and data integration

Some of the main problems addressed above can be overcome by using a software system that is divided into multiple components, and data, together with some of the tasks and knowledge, is distributed among these components. In a distributed system we can identify two dimensions of distribution: distributed data and distributed processing. As identified above, in a medical care system data can reside on different computers used by various medical professionals, so our approach is based on distributed databases.

A distributed database system [20] is a database system which is fragmented and/or replicated on the same configurations of hardware and software, running on different geographical sites. In distributed database theory it is useful to make distinctions between horizontal and vertical fragments. Distribution of data should be transparent to the user.

In general we shall distinguish between at least two types of distributed database systems: a homogeneous distributed database system or, like in our case, a heterogeneous database system. A heterogeneous distributed database system is one in which the hardware and software configuration is diverse. One site might be running ORACLE under UNIX operating system and yet another site could use MS SQL Server under Windows.

Because of the variety of the data sources we have to offer an increased autonomy in our system. Generally, autonomy refers to the level of control distribution (without considering data) throughout the system. The capacity of a Database Management System (DBMS) to run independently is influenced by various factors such as information exchange or sharing between its components, the ability to independently execute transactions, and not least the capability to modify an individual component without compromising or affecting the entire system.

Medical standards for computational models are important for design of terminologies, which form the basis for data used to populate clinical databases. In our approach the ontology aims to model some general medicine knowledge (primary health care), some specialized medicine (secondary health care), and the health care networks and their actors [4]. The ontology will represent the conceptual vocabulary common to all the actors of the investigated health care network.

A possible approach to the general design can be a data warehouse system. Generically speaking, a data warehouse system is built from components that have the role to extract, transform, and load data from distributed sources and consolidate them into a single schema where queries can be performed. From the architecture perspective, this offers a tightly coupled approach provided the data reside together in a single repository when queries are performed. The problems that may occur are related to the freshness of data. Difficulties also arise in constructing data warehouses when one has only a query interface to a data source and no access to the full data.

Data integration has favored loosening the coupling between data. This may involve providing a uniform query interface over a mediated schema (see figure 1), thus transforming a query into specialized queries over the original databases. One can also term this process "view-based query-answering" because each of the data sources functions as a view over the (nonexistent) mediated schema.
Some of the current work in data-integration research regards the problem of semantic integration. This problem does not address the structuring of the architecture of the integration, but how to solve conflicts caused by the semantics of heterogeneous data sources. For example, if two health care physicians, or networks, merge their data stored in the database, some of the concepts and terms can have different meanings or representations. A common strategy for the resolution of such problems involves the use of ontologies which explicitly define schema terms and thus help to solve semantic conflicts. This approach represents ontology-based data integration.

The general design of the system followed and tried to solve a number of requirements, considered important for a modern medical system:

- the proposed framework has to be a distributed solution, with autonomous applications in medical domain, adapted to different medical entities
- applications exchange medical information through standard interfaces/protocols
- a centralized application (portal) must ensure long term storage and general purpose medical services
- patients may access medical services (including interaction with their doctors) through the Internet, using common browsers
- the system must facilitate remote monitoring of patients, with the use of mobile medical devices.

Figure 2 presents a general schematic overview of the medical informational system, underlying the main roles in the system together with their interactions. From an architectural point of view one can identify the following main components:

- Host systems, named local (medical data) production systems
- General Practitioner system
- Analysis Laboratory system
- Hospital system
- Client system (interface to remote devices)
- Local portal (regional, national) for medical assistance, and long term data storage

Host systems consist of database servers connected through access points with fixed/mobile medical devices. The portal will supply the functionality necessary to collect information from the local interconnected systems into a centralized repository. Medical services will be provided at portal level using specialized web-services that will allow access to data stored in the repository.
Because it is a pretty “old” discipline, medicine has a long tradition in structuring knowledge, such as disease taxonomies, anatomical terms, medical procedures, in a wide variety of forms: medical terminology dictionaries, thesauri, classification systems. Therefore, there exist several terminological or ontological resources that model parts of the medical domain: controlled medical vocabularies such as the Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT), GALEN [8], and [9], MENELAS [10], and [11], ONIONS library [12], or the highly complex UMLS [13] used to allow a standard, accurate exchange of data content between systems and providers.

We have focused, in the current paper, upon SNOMED (Systematized Nomenclature of Medicine). SNOMED is a standardized medical vocabulary that has been accepted internationally [14]. Intended to completely and logically interrelate groupings of defined medical terms, SNOMED is a formalized, information-packed set of more than 300,000 coded medical terms. SNOMED defines codes for a wide spectrum of clinical concepts: Diseases and Findings; Procedures; Biological Functions; Body Structures; Living Organisms; Physical Agents, Activities, and Forces; Substances; Specimens; Occupations; Social Contexts; and Modifier Concepts.

UMLS, the Unified Medical Language System is an umbrella system which covers many medical thesauri and classifications. From a conceptual perspective, the UMLS can be divided into a Semantic Network (SN) which forms the upper ontology and consists of semantic types linked by semantic relations and a Meta-thesaurus which contains concepts assigned to one or more types. These concepts are linked by the semantic relations also supplied by the UMLS SN. The whole of this UMLS form a huge semantic network. Its semantics is shallow and entirely intuitive, which is due to the fact that their usage was primarily intended for humans in order to support health-related knowledge. Given the size, the evolutionary diversity and inherent heterogeneity of the UMLS, there is no surprise that the lack of a formal semantic foundation leads to inconsistencies and circular definitions [15].

The proposed ontology will support applications both individually but, more importantly, within an environment of heterogeneous interoperating clinical information systems. Some researchers assimilate a database conceptual schema to ontology. But it is important to represent the ontology explicitly in knowledge representation formalism and in a form understandable by a human user. We built then a medical ontology encoded in an internal format of a database but without a representation that is imperative to be readable by a human user (as the system will extract and make it readable in the presentation layer).
4. The Cardiology Database

Because of the advances in technology and application development we can state that a complex and dynamic system such as a clinical medical care system developed using traditional development methods (encodes all information in a proprietary manner inside the application) may be out-of-date at its launch! The inherent difficulties in enlarging the definitions of knowledge structures that are stored into the application and database inevitably leads to a gradual deterioration of the integrity and performance of the system.

In an information exchange between two or more traditional proprietary system architectures there are some mappings and assumptions to be performed. Because of that the integrity and exact meaning of the information that is passed from a system and used in another may be lost or corrupted. That can be a major factor that creates threats to patient safety. Even if such a system is built as a “state of the art” system with great functionalities and outstanding user interface, it might not help that much if it can not share data collected with other systems to provide a historical and complete overview of the patient’s record.

Some of the new commonly used clinical applications are usually able to exchange point-to-point messages using a format previously known and agreed and based on a standard such as the Health Level 7 (HL7). The software may also incorporate a terminology to assist in capturing and storing common clinical concepts. However, there is a common misconception that if we simply have a messaging standard paired with a terminology, then we have “achieved interoperability”.

One can state that for two clinical systems to be able to cooperate and exchange health data unambiguously, such that clinicians can read it and the computer can process it, more than message passing is required. In a modern system that can interoperate at knowledge level it is a must to incorporate a standard clinical model that can be used to share patients’ health records and also to provide complementary functionality such as clinical decision support and workflow/care planning. Moreover, an efficient and extensible system can exist when this clinical content structure is agreed on ontology at a local, regional, national, or international level. The broader the level of clinical content model agreement, the broader the potential for semantic health information exchange.

The main ontology implemented in our project is inspired by the Heart Failure ontology (http://www.heartfaid.org/) a knowledge base that should assist in management of heart failure patients. The current version of the constructed heart failure ontology is presented in [16] and it is available at http://lis.irb.hr/heartfaid/ontology/. The platform can intelligently assist users in a set of very different tasks ranging from monitoring patients in their home environment to decision support in specialized hospitals.

The design of the heart failure ontology has started from the terms defined by the European Society of Cardiology in the “Guidelines for diagnosis and treatment of the chronic heart failure” available at http://www.escardio.org. The ontology consists of classes, properties or slots, relationships between classes and instances.

In order to test the prototype along with the concepts presented above, we based the central part of our architecture on a MS SQL Server database, used to implement our proposed medical ontology. The database stores all the ontology elements, properly translated in our database. The database includes ontology of medical procedures, medical guidelines and clinical activities similar to that in [17]. So we have tables in our database to store the medical knowledge from the ontology like concepts, instances of concepts and concept hierarchy. We also have tables to store synonyms and UMLS synonyms relationships between different concepts and instances in our database. The slots from ontology are a little bit difficult to store in the database so we need many tables to store these properties: tables to store slot’s characteristics like name, tables to store possible associations, and tables to store the actual values of these slot. According to this we have for example tables to store the fact
that “Indicated” is the name of a slot connecting instances and not a slot connecting an instance with a proper value, either a character string or a number. We also store, in the database, the fact that this slot is a valid property only for instances of concepts “Patient characteristics”, “Testing”, “CHF risks”, “Classification” and “Treatment”. In the end, we can store the fact that for “Felodipine” (which is an instance of “Calcium antagonist” which is an instance of the super class “Heart failure medication group”), this slot is related to instances like “Diastolic hypertension”, “Unstable angina” and “Systolic hypertension” which are instances of concept “Patient characteristics” and thus in accordance with the previous statement. The instances “Diastolic hypertension” and “Systolic hypertension” of the concept “Hypertension” are “Patient characteristics” via sub classes “Diagnosis”, “Cardiovascular system related”, “Artery and blood disorder” and “Blood pressure disorder”. The other instance “Unstable angina” is an instance of the concept “Syndromes and effects” which has as super class “Diagnosis”, which has as super class the concept “Patient characteristics”.

We also have, in the database, tables to store information about the medical personnel, medical specialties, patients and findings. All the findings are divided in many episodes, one episode from the medical life of a patient grouping all the records and findings related to one single problem or medical status at a given time.

Besides plans and patients the medical ontology stored in the database for Heart Failure is mainly organized in the following concepts:

- Heart Failure concept
- Cardiac Risks
  - Blood, Homodynamic, Clinical, Demographical and historical, ...
- Classification
  - Clinical symptomatic classification, New York Heart Association symptoms
- Terms
  - Synonyms, UMLS synonyms
- Patient characteristics
- Demographic characteristics
- Diagnosis
  - Cardiovascular system related, Related to other organs, Syndromes and effects
- Prognosis
- Signs and symptoms
- Medical tests
- Treatment
  - Devices and surgery, Medical procedures, Medication, Recommendations

After the design process, we had to validate the ontology by using both an automatic checking through automated processing performed by the translation program and a human validation by doctors through visualization and navigation in the ontology. There are also some tests which enabled us to detect various errors in the database: relational induced cycles for the specialization relation, redundancies.

5. Proposed Software System for the Cardiology Network

The task of matching health care processes to patient needs has led further to an increased interest in health care information systems that are easily adjustable to changes in resource availability and in organizational structures. Let us note that the proposed software system is not at all an expert system. It only allows people from the health care network to visualize the reasoning and the decision-making process. The whole reasoning is carried out by the members of the health care network and the inferences based on the ontology enable the system to filter the choices offered to the user. The role of the ontology graph can be compared to a guide of best practices.
The architecture is divided in two main parts:

- policy guides - how information is protected
- technical guides - how information is exchanged

For the first part we need architecture for privacy in a networked health information environment, and for the second part we need standards at least for medication and for medical laboratory tests. We build our approach based on the Food and Drug Administration Orange Book (http://www.fda.gov/cder/orange/obreadme.htm) and on the Logical Observation Identifiers Names and Codes (LOINC) standards (http://www.regenstrief.org/medinformatics/loinc/). In USA, more than 90% of pharmacy transactions are computerized and increasingly available through national clearinghouses, consistent with the FDA Orange Book of approved drug products and as many as half of all medical laboratory tests results available electronically use LOINC standards.

We introduce the concept of episodes in the medical life of a patient and together with the ontology stored in the database we will try to design and implement a software system for creating and editing the patient’s medical record. The concepts describing this patient record are thus useful when reasoning for diagnosis or for choice of the medical treatment. In the context of such a tool aimed at supporting a health care network, it seems interesting to extend the ontology by concepts enabling to describe this health care network and its actors. We thus built concept hierarchies on health care networks, on health centers and on patient’s record.

In a health care network, our software system aims to be a collaborative work supporting tool, allowing the real time update and history of therapeutic decisions as an electronic board where each one can note information readable by the other members of the team. Starting from the patient’s health problems, the members of the team will formulate diagnostic hypotheses and proposals for a treatment. Via our software system, the health care team will connect the various elements of the patient record useful for the discussion, and thus will converge in an asynchronous way towards the definition of new health problems and of new therapeutic actions. The formulation of diagnostic hypotheses is a priori reserved to the medical actors, whereas the discussion on the treatment could sometimes imply non-medical professionals.

A typical pattern of interaction by an application recording patient’s health and the ontology encoded in the database is to connect, perform a series of requests, and then to disconnect. An important feature of the database will be that of mapping between different external representations (languages and coding schemes) to provide an invisible, automatic coercion mechanism. This mechanism performs the mapping to ontology concept entities from codes on input, and allows the application caller to specify a series of required output formats which are then produced from the underlying concept entity.

The dependencies between the various diagnostic and therapeutic hypotheses can be represented through a graph using the concepts defined in the ontology. The doctor will reason by linking the health problems to the symptoms, the clinical signs and the observations in order to propose health care procedures. The software system can thus rely on the SOAP model (Subjective, Objective, Assessment, Plan) used by the medical community [18]. In this model the S nodes describe current symptoms and clinical signs of the patient, the O nodes describe medical tests, analyses or observations of the physician, the A nodes correspond to the diseases or health problems of the patient, and the P nodes correspond to the procedures or action plans set up in order to solve the health problems.

This SOAP model is used in the medical community to structure a patient record. When a patient consults his/her doctor for new symptoms, the physician will create a new instance of an episode of the medical life for that patient. The system will initialize a SOAP graph with all currently open pathologies and prescriptions for this patient.
Considering all these we propose to use in our SOAP software system (subjective, objective, assessment, plan) formalism to store and visualize the medical record. Thus we have extended the ontology with relations useful for representing SOAP graphs (has_for_cause, has_for_consequence, confirmed_by, treated_by). Finally, using the ontology the system can propose a list of possible concepts to help the user to build the SOAP graphs:

- the $S$ nodes will correspond to instances of selected concepts among the sub-concepts of Symptom
- the $O$ nodes will be chosen among sub-concepts of Laboratory-Test
- the $A$ nodes among those of Pathology or Psychological-Problem
- the $P$ nodes among those of Treatment or Diagnostic-Therapeutically-Gesture

The arcs between the nodes will correspond to relations among concepts:

- Symptom - has_for_cause - Pathology
- Pathology - has_for_consequence - Symptom
- Pathology - confirmed_by - Laboratory-Test
- Pathology - treated_by - Treatment
- Symptom - treated_by - Treatment

The first task of the database encoded ontology, to support a user interface in helping clinicians enter the information, will be to examine the expression, convert it to a sanctioned expression if possible, and then see whether a corresponding concept entity exists. The database will answer questions such as:

- Is this a legal expression, and what is its simplest form (e.g. with any redundancies removed)?
- If it is legal, how is it classified, what more general concepts subsume it? What more specialized concept entities does it subsume?
- What is known about this concept entity from the ontology?
- What are the external expressions for this concept entity in a particular external system? What is the preferred term for this concept entity in that system?

An important aspect of such a medical software system is the security of data contained in the patient’s health record. The model for authorization and access control in such a distributed health information systems has to deal with policy description and negotiation including policy agreements, authentication, certification, and directory services. In our approach for managing privileges and access control in the shared health care environment we use an extended and refined schema considering medical workflow as presented in [19].

In order to assure maximum interoperability, the portal was designed based on SOAP architecture using Pentaho, allowing the heterogeneous systems’ integration.

Pentaho Data Integration represents a powerful tool that delivers means of extraction, transformation and data load using an innovative, metadata-driven approach [20]. Because it offers an intuitive, graphical, drag and drop design environment, and a highly scalable, standards-based architecture, Pentaho Data Integration represents a reliable and cost effective choice for organizations over traditional, proprietary ETL or data integration tools.
Pentaho Data Integration has implemented a metadata-driven approach where you only specify the data you want integrated, but you do not specify the way you want it done. One of the most important advantages of Pentaho is that one can create complex transformations and jobs in a graphical, drag-and-drop environment without having to create proprietary custom code that will work only with some proprietary application.

Common use cases for Pentaho Data Integration include [20]:

- Data warehouse population
- Information enrichment by integrating data from various sources
- Data migration between applications
- Imports of data into databases from text-files, Excel spreadsheets, relational systems and more
- Exports of data to other databases or text files
- Data cleansing by applying complex conditions in data transformations
- Exploration of data in existing databases (tables, views, etc.)

Because it provides a flexible architecture, the system offers the possibility to use various application platforms as sources for the integration. The system is based on translation functions (adapters) that provide interoperability with minimum implications for the interacting systems.

The system was designed to work both as a service provider and a service consumer, as it can both supply and request medical data from various systems. The portal services are implemented using a service-oriented architecture by web services stored in repositories.

The applications that are based on a SOA environment can be configured to work both as services exchanging data in point to point protocol and as a communication channel between two entities that call a service provided by an entity that has the role of mediator or broker. The letter can be seen as intermediary level that represents an “enterprise service bus” (ESB) between service producers and consumers, offering messaging exchange services.
5. Conclusions

The main purpose of the research was, starting from [14], to supply means for the creation and use of an electronic health record which can be used in a distributed environment by interoperable clinical systems. The introduction of an ontology based on medical terms used to populate an HL7 standardized message or document using standardized XML syntax will allow medical information to be exchanged between systems connected to the Web portal. Due to the use of the ontology, the system can be translated to other medical domains, but it is especially suitable to cardiology because of its distributed nature of record keeping and treatment.

The standardization based on a logically interrelated and searchable ontology ensures that the information related to a patient’s health record that the sender thought was sent is the same (from the point of view of medical significance and completeness) with the information the receiver received and interpreted.

These standards, in conjunction with the data integration platform, will provide the health care network a greater access to important patient information, even between different platforms.

An experimental implementation of the proposed methodology in a small health care network can emphasize the usefulness of the architecture. For the purpose of the experimental implementation a set of appropriate data must be developed for the representation of a patient’s health related data, concerning the drugs delivery, the examinations orders and results.

For the identification of the sites, a general model of the health care network divided into primary and secondary is used. Different user roles are also identified and an initial fragmentation of the data sets has already been obtained, based on the different user needs and sensitivities of the data. The allocation of the fragments has been based on the best fit and all beneficial sites approaches.

6. References


