Towards interoperability of heterogeneous health databases: application to a tumor samples bank.

Michel Joubert\textsuperscript{a,b}, Jean-Charles Dufour\textsuperscript{a,b}, Laurent Falco\textsuperscript{b}, Sylvain Aymard\textsuperscript{a}, Marius Fieschi\textsuperscript{a,b}

\textsuperscript{a} LERTIM, Faculté de Médecine, Université de la Méditerranée, Marseille, France
\textsuperscript{b} SIM, Hôpital de la Timone-Adultes, Assistance Publique – Hôpitaux de Marseille, France

Abstract

Objectives: to define principles and methods that allow heterogeneous database in the health sector to be interoperable. Material and method: to design a component-based middleware able to provide flexible and efficient means of communication between end-users and databases, and that exploits the standard nomenclatures of the health sector. Results: according to these principles, to implement a prototype of a tumor samples bank in the University hospitals of Marseille, France. Discussion: to discuss the benefits that the approach brings and the progress in prototyping.

Keywords:
Medical Informatics ; Information Systems ; Computer Communication Networks ; Databanks ; Tumors.

Introduction

During the last twenty years, medical informatics designers implemented and tested a lot of hospital information systems all over the world. It appears that, to this day, many questions raised during the 80’s have not been answered [1]. Hospital information systems have difficulties in evolving. Their future depends either on their capability to make their architecture evolve [2], or in providing new functionalities such as access to knowledge, decision support tools, and clinical guidelines ease of use [3]. Encouraging experiments have been carried out in this respect by IAIMS (Integrated Advanced Information Management Systems) [4-7]. Today, the focus is on patient’s security, decision support and evidence-based medicine. Thus, the notion of health information system should enlarge the restrictive notion of hospital information system [8]. At the present time, the departmental systems developed during the 80’s must dialog inside information systems. The IHE (Integrating the Healthcare Enterprise) approach started in the USA several years ago, and extended in Europe, with the intent to facilitate the communication of medical imaging data [9]. Based on the same idea which consists of making systems interoperable by means of structured information exchanges, the DICOM (Digital Imaging and Communication in Medicine) standard, in the medical imaging field again, stands out as a norm [10], while HL7 (Health Level Seven) evolves towards a universal medical data interchange format [11].

Today, heterogeneity is a reality. From a technical point of view, it is primarily caused by [12]: network technologies, devices, and operating systems; middleware solutions and communication paradigms; programming languages; services and interface technologies; domains and architectures; data and document formats. From an informational point of view, as soon as applications exchange data and documents, problems appear due to differences in coding rules. Therefore, shared coding rules and transcoding capabilities are of great importance to make systems interoperable.

Coding and transcoding processes need to use standard nomenclatures and thesauri. The NLM’s (U.S. National Library of Medicine) project for UMLS (Unified Medical Language System) [13] aims at unifying them into a coherent system that allows means, among other features, to translate a given concept denoted into a nomenclature into its representation in another one. Besides coding, structure of exchanged data is very important for the process of computing systems interoperability. XML (eXtended Markup Language) [14], a file format emerging from the Internet technologies, provides for this need since data conveyed in formatted files have a predefined structure made of tags which meaning can be shared to allow a semantic dimension for this language. A way to proceed is to agree regarding an ontology and to tag data according to it. Computer scientists are currently working towards a semantic Web [15] for which the medical applications are appearing [16].

Regardless of current heterogeneity of data, in order to build information systems, one must carry on with systems able to exploit the various pre-existing databases. The aim of the project CoMeDIAS (Computerized Medical Data through Internet Access Services) is to propose a platform built on the means offered by current software tools, based on terminology standards, and to provide healthcare professionals with a homogeneous way to access data, information and knowledge which are naturally heterogeneous. Through the development of the tumor samples bank of the University hospitals of Marseille, we show in this article how to implement a system according to the principles of CoMeDIAS. The implementation is component-based with the intent to be modular and to make sure that developed components will be reusable [17].
Objectives of the project CoMeDIAS

This project is the continuation of the work we carried out in order to make it easier for healthcare professionals to access documentary resources and knowledge bases which they need in their daily practice [18-20]. The work we did in the project ARIANE is based on the knowledge sources of the UMLS. These resources have been developed with the intent to federate the most regularly used nomenclatures, and many others, in the biomedical field. We exploited them for establishing gateways to heterogeneous documentary databases and knowledge resources. The fundamental results of this semantic integration produced by the project ARIANE have been reused in the framework of the European project WRAPIN1 which aim is to propose a gateway for certified health information over the Internet. Again in this case, the semantic contribution furnished by the UMLS knowledge sources is convincing since it allows to improve the performance of the tool used to index the documents over the Internet, and also the capability of the associated information retrieval robot.

The two above mentioned projects were dedicated to documentary resources. The aim of the project CoMeDIAS is to include factual databases into the sphere of activity while exploiting the assets of the two previous projects: ARIANE with regard to the method, and WRAPIN regarding the software development technique. Its aim is not to make various resources communicate by means of messages exchanges, but to produce a platform for applications dedicated to gather information and data in various databases and to wrap them into a unique user interface.

Interoperability of heterogeneous databases in applications is operated in CoMeDIAS by a middleware made of components interfaced with existing communication protocols by the way of connectors able to physically communicate with databases according to the appropriate formats. Data may concern a patient: clinical or surgical data, medical imaging, etc. It may also concern a medical problem: clinical guidelines, operating or medical data, medical imaging, etc. It may also concern a medical problem: clinical guidelines, operating or medical protocols, drugs prescriptions, etc. CoMeDIAS must not be seen as an application, but as a set of services thanks to which it is possible to build user interfaces that allow users to “intelligently” navigate inside databases and easily gather useful information in databases inside or outside their institution. This kind of navigation is possible when using standard nomenclatures for coding information as well as for information retrieving.

Implementation of the platform CoMeDIAS

Principles of implementation

Component-based programming allows a flexible and efficient mechanism for software development. Following this approach, a Web service [21] is made of components which cooperate with the achievement of a requested action by some other software modules, such as a query to a database. Interactions with a Web service are operated by the means of XML files exchanged by the way of Internet protocols. This global view on Web service is schematized by the diagram of Figure 1, where they appear as mediators between users and databases.

Figure 1 - The three layers (client, mediation, and data) involved in the software architecture of CoMeDIAS

Web services exchange XML files which contain tagged text including the carried queries. A query is a character string that will be sent to a database management system, an information retrieval robot, or the API (Application Programmatic Interface) of a given server. A query is built by components which operate transformations on an initial query in order to formulate it in the query language of a selected information resource. For instance, a component may replace a diagnostic code by the term or keyword used to refer to the related disease in a factual database or for the purpose of a documentary server. Another component may replace, by means of the UMLS knowledge sources, a code coming from a nomenclature with its designation in another one. Knowing the syntax to use for querying a given resource, another component may contribute to formulate an initial query according to the expected syntax. Etc. Conversely, when results are returned by a queried resource, a series of transformations are operated on these results. Such series of transformations are presented in Figure 2, which shows the treatment of PubMed results. The data schema on the left side, gets wrapped in a standard format, which schema is shown on the right side. This example shows how a set of authors (each of them is represented in the PubMed schema by a last name, a forename, and initials) is transformed into a list. It is the same kind of transformation in the case of key-words. While a title and an abstract are not modified and directly transferred to the related user interface element. These components are used to build up a Web service. The set of these Web services constitutes the CoMeDIAS middleware.

The user interface and Web services communicate by the way of an Internet protocol. With the intent to communicate with a database, a Web service must use the appropriate communication means, that is the appropriate protocol. At this level, connectors are coming up: a connector is able to physically transmit a query to a data resource and to receive results according to appropriate protocols and formats. The place and function of such connectors are illustrated by the Figure 3.

---

Implementation of a tumor samples bank

The University hospitals in Marseille (AP-HM), France, are equipped with old-style computer systems which link hospitals into a unique and huge network. Computer systems are relatively independent.

For the most part they are: a patient management system (identity, stay number, entry and exit dates, identification of visited units, etc.); a system for managing clinical and surgical data; a laboratory system, including pathology; a blood products management system. These three systems are independent from each other, but they are connected through the patient management system. Moreover, some medical units maintain complementary and independent databases for care and research. This is the case of a specific database which manages the biological resources (tumor samples) that the biology unit stores in its freezers.

French Ministry of Health and Ministry of Research propose to standardize the existing or “under construction” tumor samples banks all over France. The aim of the harmonization of tumor samples banks is to contribute to cancer research, as it was decided in the UK by the NCRI (National Cancer Research Institute) initiative [22], for instance. The head of the AP-HM and clinicians of this institution who are concerned with diagnosis and treatment of cancer, decided to build a tumor samples bank with the intent to give access to different databases with a unique tool, namely the patient management system, the related clinical and surgical databases, the pathology database, and the tumor samples database management system. A preliminary prototype has been implemented according to the above principles. The various systems and databases continue to be exploited as before, the prototype is only a tool to query them simultaneously with the intent to retrieve and display to users relevant data within the framework of the tumor samples management and the related clinical activity. The current release of the prototype includes data from organs oncology, that is neuro-oncology, gastroenterology, etc.

Figure 4 shows the principles introduced in Figure 1 instantiated in the case of the tumor samples bank of AP-HM: standard nomenclatures are CIM-10 and SNOMED; databases are clinical and surgical databases, pathology data, and tumor samples management; to which is joined the patients management system; the user interface is a standard Internet browser. The transformations operated in order to translate queries and connections with the proprietary systems have been realized for the occasion. Appropriate connectors have been developed in order to operate the
expected mediation between the Web services and the various data resources.

**Discussion**

Independent databases and sub-systems have existed in hospitals for numerous years. Their access is managed by the technology used when they were implemented. The first implementation we realized has shown how the principles of CoMeDIAS may be effectively implemented without previously challenging neither databases, nor existing systems. Moreover, displayed data are always up-to-date since they come from the source applications. In this way, this approach allows to quickly have at one’s disposal an expected service and to have time to let a hospital information system evolve coherently. This approach is also interesting in various integration projects because it proposes a complementary capability to data warehousing which imposes often long and costly treatments and duplication of data [23]. Without intending to substitute for data warehousing for which the objectives are different, CoMeDIAS proposes a logical integration in order to query directly the source databases, and no necessary physical construction of data warehouses. Transformations operated in flight by Web services allow a representation of data different from their representation in the source databases.

Integration and interoperability are an affair of architecture [12]. From this point of view, the solution we propose is a middleware built on a set of Web services which cooperate with existing systems. Following this method, the objectives of CoMeDIAS encourage the use of an ontology and standard nomenclatures because the Web services are conceived to exchange data and documents. Sharing coding rules is of utmost importance at this level. Moreover, this should allow to move towards the promising direction of a semantic medical Web.

For clinical research reasons and needs of experiments on samples, the final tumor samples bank system of AP-HM will allow conditional research of samples having common properties. The system is designed to retrieve, for instance, all the samples concerned with oligodendroglioma and related to patients treated by chemotherapy. The UMLS-based methods and techniques we developed previously for information retrieval find here again an area of application.

**Acknowledgements**

The authors thank the clinicians who brought their competencies in this first building phase of the tumor samples bank of the AP-HM, as well as the “Direction de l’Informatique et des Réseaux” of AP-HM for its support and its active participation in the project. We thank the French Ministry of Health for its support.

**References**


**Address for correspondence**

Michel Joubert
LERTIM, Faculté de Médecine
27, boulevard Jean Moulin
13005 Marseille, France
E-mail: Michel.Joubert@ap-hm.fr
http://cybertim.timone.univ-mrs.fr/