Health information systems

From local to pervasive medical data

Christine Verdier

Lyon Research Center in Images and Intelligent Information Systems (LIRIS)
UMR 5205 CNRS, Claude Bernard University Lyon 1, Lumière Lyon 2 University
Ecole Centrale of Lyon

INSA of Lyon, Bat. Blaise Pascal, 7 av. Jean Capelle, F-69621 Villeurbanne cedex
christine.verdier@liris.cnrs.fr

ABSTRACT. *This paper concerns the evolution of medical information systems. The evolution is due to two main factors: the evolution of the society (a patient more concerned with his health and more informed) and the evolution of Internet-based technology. We see a double consequence: medical informatics must be pushed to medical professionals rather than the opposite that prevailed until the nineties and health information systems (HIS) must move from structuring specifications to visualisation and interface ones. The legacy health information systems gather numerous data concerning patients with different structure, data type, name, etc. To rebuild health information systems is not possible considering time-consuming, conceptual and financial problems. But to rebuild a lifelong medical record is also difficult and with no semantics guarantee. We propose in this paper to present what is hoped about health information system design, what was really done and what we consider as the future of this domain from legacy systems.*

RÉSUMÉ. *Nous proposons dans ce papier d'étudier l'évolution des systèmes d'information de santé. Cette évolution est relative à deux facteurs : l'évolution de la société (le patient est de plus en plus concerné par sa santé et plus informé) et l'évolution de la technologie, notamment internet. Nous y voyons une double conséquence : l'informatique médicale doit être adaptée au professionnel de santé contrairement à ce qui a prévalu jusque dans les années 90 : adapter le professionnel de santé à l'informatique. La conception des systèmes d'information de santé doit se modifier : de la structuration des données vers le développement d'interfaces de visualisation et d'interrogation des données. Les systèmes d'information de santé existants concentrent un nombre considérable de données de santé ayant des structures différentes, des types de données différents, des noms différents, etc. Reconstruire les systèmes existants n'est pas possible au regard du temps prohibitif, de problèmes conceptuels et financiers. Mais reconstruire l'histoire santé d'un patient est également difficile et sémantiquement hasardeux. Nous présentons dans cet article ce qui est souhaité en matière de conception de système d'information de santé, ce qui a été réalisé et ce qu'il nous semble devoir être fait pour le futur à partir des systèmes existants.*

KEYWORDS: *medical information system, electronic medical record, adaptive interface, pervasive data.*

MOTS-CLÉS : *système d'information médical, dossier patient informatisé, interface adaptative, données pervasives.*

1. Introduction

At the beginning of the early research programs concerning medical information systems (beginning of the 90th), the world of health was not really got used to computer science applied to medical data. The early applications only concerned management data, bed management, scheduling and so on. Two main problems can be analyzed. First of all, the medical data are fickle, non-structured. Paper-based medical records, medical summaries, notes, comments and other documents are not easy media to be computerized. The second reason is that health professionals have not considered, at least at the beginning of the computerization movement, it was a priority. To moderate our language, we can say that the approach lead by the computer scientist was not the better one: the difficulty to translate non-structured medical document into structured ones, attributes with fixed forms and repetitive access was certainly discouraging. Let us notice that, before XML, relational databases were the only way to capture efficiently data.

First medical information systems were essentially based on the user interface quality. This quality was interpreted as coloured, overloaded screens containing free text fields to capture diseases, prescription. This type of interfaces is explained by the difficulty to represent medicine through data model. The first designers were not really competent in data modeling or not really felt concern about data representation. They project onto the interface their own view of the medicine.

The second evolution was the prerogative of the computer scientist. The success of the relational model had a considerable impact on medical information system design. We, and many research teams, used the relational model to design the medical record. Structured data and field were obtained to represent every item of medicine. Prescription screens were built with a field for each item (posology, drug name, drug indications, price, etc.). Diseases screens were described with the problem-oriented medical record structure. These HIS were totally structured and offer two real advantages: to make the data capture (pulldown menu for diseases and drugs, check box, etc.) easy for the GP and to help for statistical studies and epidemiology. The major disadvantage of these all-structured HIS were the use-slowness because all fields have to be filled in.

The today representation of medical record is oriented to XML structure because of its unstructured data representation. This language is well-adapted to medical practice because doctors use many documents (operational report, clinical summary of stay, drug prescription, etc.). A second advantage can be found: XML is an interchange data format useful for communicating data through web applications.

We present in the first paragraph the definition of a health information system and particularly what are the constraints for a good HIS. We continue with some reasons for relative HIS failures. In paragraph 3, we will speak about related works concerning HIS design and interoperability problems. We will go on with what we consider as the future in medical computing: interfaces and grids.

2. Health information systems

2.1. *Definitions*

The field of information systems deals with systems for delivering information and communications services in an organization and the activities and management of the information systems function in planning, designing, developing, implementing, and operating the systems and providing services. These systems capture, store, process, and communicate data, information and knowledge. The systems combine both technical components and human operators and users (Davis). That definition enounced by Gordon Davis expresses the real environment of information systems, composed of data types, technical aspects and wide-spread of organisational effects. To make the users and operators efficiently communicate through the information system (IS), three factors have to be considered: the data to send, the time to convey, and the person to reach. **Theses three factors can be synthesize into a simple sentence: to give the good data to the good person at the good time**. This simple idea is opposite to the complexity of designing and implementing an efficient (economic sense) information system that is able to connect people (problems of systems heterogeneity), to deliver the data or information necessary for decision support at the good time (not to late, not to early). This triple constraint represents the foundation of a good information system. It can be adapted to every domain of IS and is naturally true for health information systems (HIS).

A HIS represents a class of IS. It includes the management IS that takes into account the administrative part of the health domain and the medical information system that covers medical, medico-economic and environmental aspects. That second part will also deal with every connection data between the two categories (Verdier *et al*., 1994). The medical information system (MIS) contains the part of information which has a direct bearing on medical action and the automated procedure for assisting users in decision making (Flory *et al*., 1997).

2.2. *Description*

An information system constitutes a coordinated whole of procedures, software, materials and human resources making it possible to store, transform and share data. Each element of this physical composition is useful to manage the organization as well as possible to optimize the decision-making. The IS organization can be represented by the following diagram:

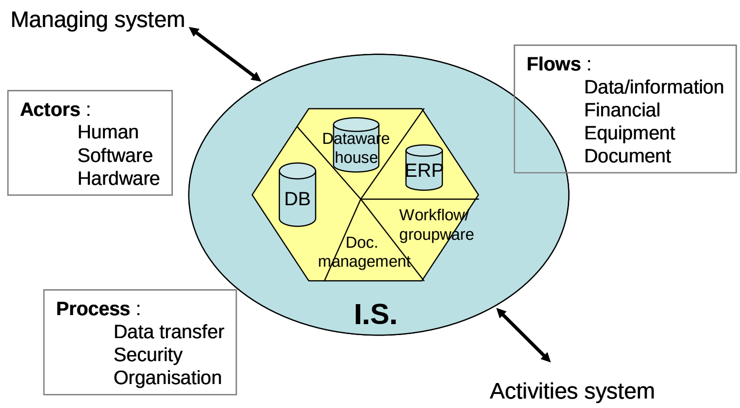

Figure 1. *Information system components*

It does not seem useful to detail the components of IS of which description is in all the good books in the field. Simply let us recall in some lines these components in order to establish a parallel with the IS medical ones. (I)Actors: They are of three different types; according to the terminology UML which widens the concept of actor (compared to the Merise method) (Booch *et al.*, 1998). Actors represent the human beings, hardware and software which interact with the system. (II)Procedures: they are diversified with the new complex aspects of IS; to the traditional organisational procedures, we can add security settings (field very significant since the introduction of Internet into the companies) and the procedures of data transfer. (III)Flows: the systemic representation of the organization underlies that it has a well-defined purpose and border. That also supposes that one can define inter-systems and intra-systems interactions. The representation of flows thus constitutes an important stage of the modeling of the information system. (IV)Materials: The materials represent the physical part of an IS. They are of different nature according to the user, the place of information collection, its availability. We have little interest to describe the existing materials. Let us note however that the awaited pervasivity of IS directs the material choices towards the use of mobile tools (cellular telephone, portable computer, Wifi network, etc).

2.3. *Conceptual specificities*

Medical information systems were built gradually without very constraining ministerial directives and official indications on medical information codification, structuring of patient’s file or possibility of ambulatory medicine computerization. In the absence of any regulation, medical information management was free for decision makers, doctors and academics. If the result is overall interesting (innovative ideas, adaptation to the users always closer to their daily job), the other side of the coin is particularly heavy: systems developed are heterogeneous (as well at the conceptual level as structural), awaited goals are different, numerous medical (or non medical) partners with very dissimilar formation levels. That led to a plethora of information systems (which besides are often reduced to a database) non communicating which contain medical files entirely burst in the various places of care and which makes absolutely impossible to reconstitute patient's medical history. A vertical structuring of the information system (by service) was privileged to a horizontal one (by patient). Today developers of health information systems try to make up for time but the rebuilding is difficult. A health information system contains several modules of which the number and the importance are a function of the knowledge at the creation time, the funds allocated and the partners implied in the IS.

– Administrative management of the patients: this part is always present and preceded without any exception the design of medical information systems;

– Nurse file: in this module several evolutions were noted; from simple diary of regulation to the control of the regulation;

– Patient file: this part never joins a design and content consensus;

The components of the MIS are similar to traditional IS. Only the trade orientation changes the contents of the various components.

Actors

In medicine, the human actors are multiple. From this multiplicity is born complexity. The wider the MIS is (MIS for ambulatory medicine *vs* MIS for networks of care), the larger the number of actors is and their trades are varied. We can count : doctors, paramedical professions, medical secretaries, non medical actors of health: auxiliary of life, social workers, administrative staff, technicians.

The majority of these categories are listed in the CPS card (electronic card for professionals of health). This card is used as entrance point with authentification ofpeople in the MIS. Each speciality has a particular knowledge which is difficult to share. It is thus very difficult to bind trades competences between them.

Hardware and software actors

We gather these two categories because they are closely dependent. The choice of the materials is directed towards the mobile pervasivity and systems: wireless technology is thus privileged and the mobile supports also: PDA, portable computers, etc. The software actor most important is of course the medical file. Consortia of standardization help with the formalization of the medical file by provision of the community of the preset classes to improve reutilisability.

Procedures of security

They are essential and prone to controversy. At the beginning of the computerization of health data, the principal handicap to which the scientists faced was the security of medical files, protection against illicit intrusions and the risk to see the patient's information medical seen and used by unauthorized people. However, it is quite obvious that it was easier to seize a medical file paper on a desk than to be introduced in an illegal way into an information processing system. It does not remain about it less than procedures of security are fundamental in medicine (as in any significant field). Several types of security are taken into account:

– the European Commission recently authorize the encoding of health data in the networks of care;

– the authentification with a coupling card CPS/Vitale card in France;

– the protection of networks with secure Intranet and firewall;

– the controlled access to medical data according to the degree of relevance and emergency.

Several other systems start to appear but are used still little because in particular of their dearness: biometric print, print of the iris.

Procedures of data transfer

These procedures will not be here detailed. They relate to the standardization of formats and data exchanges. The reader can refer to the paragraph 3 which shows a little part of data exchange for interoperability. Let us insist however on the fact that these procedures are inherent with the width of IS and the will of its creators "to open" their application.

Organisational procedures

The great difficulty for IS design will be today an organizational problem and no more a technical problem. That appears quite as true to us in health. The principal cause of countable deficit of health lies in the absence of optimization of material, software and human resources organization. The relevant and efficient analysis of organisational procedures will become a key factor of success in the future MIS.

Data flows

These flows are essential with the comprehension of the patient trajectory. They make it possible to include/understand intra-organization and interorganisations bonds. Only the modeling of flows intra-organization is succeeded. Within the framework of networks of care, the representation of flows is much more difficult: the process of care is nonlinear, the chronology brings only little diagnostic elements and many feedbacks are to be taken into account.

Physical flows

The problems arise in a more crucial way within the framework of networks of care where one externalise competences, care, administrative patient responsability and the material used. To transfer a patient to his residence supposes to create a hospital "out the walls" and thus to identify physical flows of drugs, disposable materials, resources, etc.

Financial flows

The major problem of financial flows raises a major antagonism: the dichotomy between the medical world and its financial support. The first systems separated the MIS from the management IS in hospitals. Today, information systems include medical and administrative data in order to be able to evaluate organisational, financial and countable procedures.

2.4. *Constraints for a good MIS*

2.4.1. *The first constraint: a good information*

The building of a real good HIS is totally inseparable from **capturing data directly from the source**. It means that other copies of the same data lead to mistakes and errors. The second way for good information is to imply health professionals with continuous feed-back, daily help in their job. A necessary adaptability of the computerized system to the health practitioner (and not the opposite) must also be done to simplify the interface and the communications between the machines and the users. On-line help system has to be added to the application. This help must be easily understood to avoid mistakes and consequently, annoyance, discouragement and then rejection of the system. Let us notice that the impact of interface quality is not so sizeable according to European countries. In France, it affects directly the use of medical computerized systems. To be really adopted (of course without any legal constraints), a computerized system must be very close to the natural language, the mental behaviour of the user and his meaning. Today systems come close to this constraint but are note totally human-aware. We propose some ideas in paragraph 4.1.

The good information constraint consists also in the **quality of structuration**. We will speak in depth about the structuring versus the non-structuring of medical data in a further paragraph. What is the real problem? To link the freedom of manual data capture and the necessity of structured data for automated treatment. The former ideas focused on the 'all-structured' to help for epidemiological studies, statistic reports, SQL queries and all types of computerized processing. We think that this way is really and definitely the only one to support quantitative queries. But the antagonism is then obvious. The real job of a medical professional is to care, nurse, relieve, help but surely not to capture every piece of the medical information he produces in formatted electronic documents. So, either the system automatically translates from paper-like document to receive structured data without a manual intervention, or the partition of the medical information must be so easy to understand that any real mental effort is needed. In any case, the structuring must be a final goal even if its not the former way for data storage.

Semantics of data is also a guarantee of quality. The huge of medical information systems, electronic records, Web-based documents pose two main problems: data transfer -how preserve data semantics during the transfer from a system to another?- and data representation -how to create the data model before the programming step? According to us, semantics represents the main point of HIS. The degree of confidence in the data obtained (to care) directly from a database or from several data sources, raw data or treated data must be 100%. An information must be always exact, exhaustive, complete and confident. The query result: 'the patient Jack has a diabetes' is true to 95% can't be accepted. The medicine is not an exact science but to increase the uncertainty is inconceivable. This reality can partially explain the rejection of some medical information systems (particularly in France). So semantics maintenance of medical data must be a real priority in the design of information systems.

To plagiarize many authors, we can say that: **too much information kills information**. If semantics is essential to correctly understand medical information, the volume of this information must be studied. Relational databases limited the risk to loose the user among a huge a data. The important rising in the use of the Web increases the risk considerably. Query systems must be improve to fit this new challenge. What does it mean? 1) The definition of new query systems (SQL-like systems are only really efficient for structured data) even if some SQL-modified languages try to settle some problems linked to Web data retrieval; 2) The use of data semantics and domain ontologies to filter data; 3) the definition of users profiles and 4) the taking into account of time.

2.4.2. *The second constraint: the good person*

This aspect of quality consists of two main objectives: to find the good medical partner and to verify his ability to taking care of the patient.

The first aspect deals with the identification of the health professional. Traditionally, in hospital or at home, a patient is referenced by a doctor. But to maintain the continuity of medical action, it is really necessary to track the patient care. So to connect several health professionals between them can be helpful for the patient care. This problem is really acute when we consider home-hospital networks and particularly home care. This type of care will be developped in the further years because of economic reasons (it should be less expensive to care people at home rather than in hospital) and psychological reasons (it's often better to be at home for a patient rather than in hospital). The home care supposes that many persons are close to the patient for caring, helping, advising and globally participating to the patient health trajectory. To couple different jobs to better identify the health process and produce a better care following, these different jobs must be linked and each person must have a small knowledge of the other jobs. We can divide the jobs in the care process of two types : the medical professionnals and the non medical persons (social workers, household help, auxiliary of life). The recovering of knowledge is important to allow the care following. So to identify the job and the knowledge level of every person in the care process is essential to connect different medical actions. To identify the good person with the good information, it is necessary to adapt the today human interface to include non professionnals.

The second aspect concerns the ability to access medical information. Thus, connected to the nature of information, access rights must be taken into account (with biometric measures or login and password). This part of the security acess is difficult because of the multi-partners of the IS and the huge data of the medical record.

2.4.3. *The third constraint: the good time*

To receive information too late or too early is similar and negative. The time must be considered with a multi-faceted view.

What must be considered first is **the time to distribute** information. This time is often difficult to analyze because the care to patient is health state-linked. A patient having a suspicion of myocardial infarction must be dealt with in less than one hour. It supposes that information absolutely needed is immediately available (allergies, essential medical past history, etc.). An information system must be useful for that and adapted to emergency. Let us notice that this situation is possible when the patient is 'captive'. It means that the patient is knew in the hospital, his medical record is up-to-date and querying the system give an immediate good answer. For nomad people, the situation is not the same, more crucial and more difficult. Pervasive information systems should give a better answer to that particular problem (particularly with data grids). But let us consider the current problem. A today solution can be found in health cards. In France, the Vitale card exists and contains the administrative part of the social security information. A further version should integrate some essential medical information. We must add that some particular health cards already exist for particular chronic disease or health symptoms (cards for penicillin allergy for example). The nature of medical information, we just felt that, is conditioned to time. Emergency information must be obtained as soon as possible. Treatment or information of monitoring could be researched with a more long time and synthetic information is not time-acute. So, how can we reconcile the time-adaptability with a patient health fluctuation?

The second view concerns the **representation of time** in information systems. If static models are well-known and -used, dynamic aspects stay hard to model. UML proposes a whole of models tools to describe the organization (Booch *et al*., 1998). Some models like behaviour diagrams are useful for dynamic representation but most difficult to implement especially in medic area. Let us explain in a few words. The medical time is really difficult to represent because of its great modularity. Some papers have been written about medical time (for serializing data), about time-based database but not really, according to us, to the design of time in medical information systems. Why? Firstly because of the variability of the time-notion. The time-measure of the blood pressure depends on the patient state: a non-pathologic patient (without hypertension) can have a blood pressure measure twice a year. At the opposite, a hypertensive patient measure must be done once all 10 minutes. To represent the time with a month/year/day format is insufficient. Consequently, the comparison between time-data series is not easy. Approaches concerning timescales, temporal databases are developed and give some interesting results with Allen algorithm and closely related algorithms.

The third aspect concerns the **delay** to take in charge a patient. This time is linked to the first aspect, the time for delivering good information. This facet is highly correlated to at-a-distance medical care and telemedicine. Information technology, networks, and all telematics-based applications open a new challenge in patient care. Current medical information systems are used for inpatients. The enlargement of places of care, the need for at-home care, for any-place-of care, lead to think about another approach of information systems with so-called communicating information systems. New tools help in creating at-a-distance information systems because new technologies exist: wireless networks, satellite transmission, medium waves transmission, shared medical record, dynamic and static design models, and so on. Main problems are the cost of the modification of current IS, politics' will and finally a real large consultation to connect people.

2.5. *Reasons for relative failures*

All job-wide firms have IS to manage data, to improve strategy, to make decision easier and better. The use of IS is based on consensus and could not be discussed. The attempts of IS in health have until now failed. If some parts found solutions, the whole is not powerful. The reasons are plethoric and of different nature. Most important seem to be the great difficulty of representing a complex world in perpetual evolution. Complexity can be apprehended on several levels:

– The fickleness of medical information: the data “loss of visual faculties” does not produce same information according to whether the patient is diabetic or not;

– The absence of standardized vocabulary: except the International Classification of Diseases(ICD 10 – WHO), no agrement arrived on job-oriented classifications;

– The difficulty of representing the semantics of medical information through multicriteria research;

– The absence of consensus on the structure of medical record: opposition between problem-oriented structuration and chronological structure;

– The difficulty for removing the no more relevant information from the medical record: the relevance of medical information decrease with the time, except for specific medical history;

– The search for information useful for the medical follow-up of the patient: made difficult by the bursting of the medical record between various places of care;

– The representation of various jobs in which each one can cover different specificities: liberal nurse and nurse of surgical units;

– The difficulty of defining IS borders and its universe of discourse.

These some factors, non exhaustive, show at which point this field is delicate to represent. Data processing likes structured data and little changing but health represents the opposite.

Let us take again the above mentioned factors. Each point concerns a particular sector of scientific research in data processing and it is undoubtedly the principal consequence of the quasi-failure of IS which always considered only one facet. The factor (1) concerns management and the capitalization of knowledge, factors (2) and (3) relate to ontologies and representation of data, points (4) and (7) can be solved thanks to modeling tools, factor (5) is a basic problem of data and point (6) relates to information search, adaptive interfaces and "intelligent" data query. So how to imagine the construction of a medical information system which does not take into account all these aspects ? The major difficulty thus consists in federating these elements to offer an information system oriented-trade where each partner will be able to find information he is looking for whereas it is located in the larg Web environment. The concept of patient care process is emerging which is called patient's trajectory. Let us say that some authors see a difference between care process and patient's trajectory. We use the both terms because it represents for us the interval between the patient enters the health system and the date he leaves. So medical information systems must imperatively provide a relevant visualization of the care process.

Let us turn some pages of history. In first applications of systemic to IS, the functional company division prevailed because it was representative of the idea that one was done of this world cutting in engineering and administrative departments (accounting department, store, direction departments) or/and in functions (financial or accountancy function, purchase function, etc.). This Taylorian cutting made it possible to identify the borders of each company's subsystem and thus to apply the traditional profitability and evaluation indicators. The more complex organization of current companies, their geographical, thematic, functional and job bursting change completely the organization of IS. The 2000s' years mark an evolution of companies'IS perhaps as important as the technological revolution born after the second world war. The health domain copies the same characteristics and the same evolution. If in hospitals, always prevailed a cutting in services (for political or medico-economic justifications) and event if this phenomenon must go on, the way of apprehending the MIS must be deeply anchored on the concept of care process i.e. transversality (or patient's trajectory).

Another reflexion must be carried out which also impacts the functional architecture of traditional IS. We have traditionnaly copied medical information on the human body. Classifications of diseases, medical records were created according to the organic cutting of the man: heart, lung, kidney, etc. This categorization of medical information is explained easily by the medical training: each medical discipline is specialized in a body part (brain for psychiatry and neurology, heart for cardiology, kidney for urology and nephrology, etc.) and the covering of knowledge between different medical specialities is very limited. The disadvantage in this way of organizing medicine is however counterbalanced by the facility of the training and the knowledge of the field. Organic cutting was thus naturally reflected at the software level. Computerized medical records, heart of the medical IS, were organ-oriented (cf figure 2).

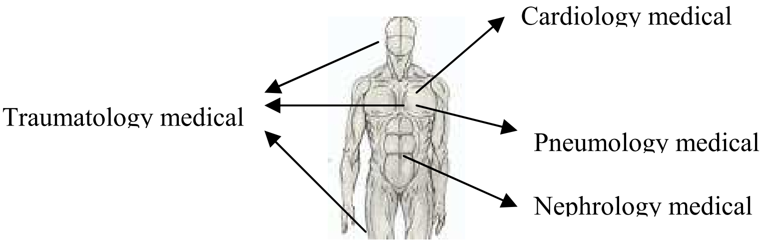

Figure 2. *The patient = a whole rather than a sum of body's parts*

We obtain inevitably intersections between this various softwares because of the necesseraly duplication of information. This representation raises three essential questions:

– How to identify the patient as a whole whereas he is represented only with a sum of body parts;

– How to identify the care process since we add to the geographical dispersion of the care, a anatomical distribution of medical information?

– How to replace the patient in the middle of the medical IS "whereas he enters the health system part of body by part of body"?

The patient represents an entity and not a sum of body parts. We see it well, it is no more possible today, considering the wideness (extend) of the medical knowledge, the number of medical treatments and the increase of the population targeted to consider this functional design.

3. Related works

The structuring of the medical file is probably one of the most difficult exercises. The major difficulty is due to several phenomena. Medical information is complex, changeable according to various criteria (place, time, for example), related to a given trade (an urologist does not use same information as a general practitioner), retroactive and evolutionary. We can add to this list the absence of precise definition of the medical file between ambulatory medicine and hospital medicine or within the same structure between various doctors. Contents of the medical file are specific to each doctor. The absence of real obligation of sharing medical file, directives and other constraints involved a paer-based solution which was at the same time perfect for the doctors and completely unusable for a data-processing traanslation (the textual support does not make it possible to make research on the data). The impact of health networks made become aware the importance to share data, so to structure and model the medical file. Teams worked on these problems. We can quote works of J Cimino (Cimino, 1996; Barrows, 1994; Cimino, 1994) on the modeling of classifications and their inclusion in the medical file. We can also quote works of J.R. Scherrer with the systems Diogene and Galen (Borst, 1995; Alpay, 1995), (Scherrer, 1995; Bréant, 2000) in Geneva hospital.

French teams were also interested in this work: P. Degoulet's team in Paris and M. Fieschi's team in Marseille. These two teams regularly collaborated: (Soula, 1997; Burgun, 1995; Joubert, 1994; Degoulet, 1997; Engelmann, 1995; Lavril 1994).

The second part of the research works about medical information systems has concerned interoperability. Many computer-based medical applications were created for particular needs in closed environments. The first consequence and undoubtedly most serious was the total absence of communication between heterogeneous systems. In a completely antagonistic way, the need for creating shared medical files made essential the cooperation between heterogeneous information systems.

Works concerning the structuring of the medical file and interoperability between the medical systems became very important. Great projects were started to work in this field and to try to find a structure adequate and consensual medical file and related applications. We arrived at the middle of the Nineties with a spectacular profusion of research projects in this field. Among these projects, we can notice the European project AIM (Advanced Informatics in Medicine) which proceeded between 1991 and 1995 and take the first steps towards the computerization of medical files and to propose European and viable solutions. Several projects started from AIM consortium.

The goal of HELIOS project (Lavril, 1994) was to conceive an environment of development intended to computerize the medical applications. The principal idea of the project was to offer a set of tools ensuring interoperability between the applications. The awaited goals were: (I)to create an opened and modular environment of development to facilitate the production of medical applications, (II)to use a unified mechanism to integrate the medical application software; (III)to primarily target the environment towards the automation of the operating room; (IV)to support the reutilisability of medical applications. HELIOS was conceived in oriented-object.

The goal of project GEHR (Good European Health Record) was to define an architecture standardized for the computerized medical files (electronic medical records). It included storage, communication of medical information between doctors, hospitals and analysis laboratories. The discussion thread of the project was to ensure a perfect portability in order to abstract itself from problems of coding, language or operating system. The result of project GEHR was concretized by a multi-media architecture of data for the use and the sharing of medical files by taking into account clinical, technical and medico-legal functionalities.

The goal of the Menelas project was to carry out a system of access to medical files based on natural language. The idea of this project came from the observation of medical behaviors. Medical data are seized in the form of free text in medical or operational summaries. The Menelas project thus had two essential goals: to provide an access simplified to medical files in free text and to improve the multilingual access to medical nomenclatures.

Other projects were born at the international level: we can quote project W3C-EMRS and project HELP.

Boston's hospital has created a common medical file (W3C-EMRS) which specifies the types of data patient having to be transferred in the common file to ensure the continuity of care. The principal idea was to ensure a longitudinal sight of the patient care and to ensure a coordination of the medical actions. To achieve this goal, the authors translated the specifications into HL7 messages.

HELP system (Kuperman, 1991) is an information system based on knowledge. It included all the elements of taking into account and follow-up of the patient and integrated an engine of decision-making aid. The decision-making aid was done in an interactive and asynchronous way starting from data and from the management of effective time. HELP system was based on a basic architecture of data which was used also for the structuring of patient files.

4. Medical information system for the future

Many research tasks are to be carried out to thwart the negative impact of heterogeneous systems. A possible solution which is the subject of many publications consists in building platforms of data exchanges which are useful of "translator" between the various information systems. Other solutions, which we privilege, consist in using existing systems and moving the problem of interoperability on the level of the interface. It is thus a question of creating interfaces able to go to seek medical information in the data sources and to rebuild on the doctor computer the medical file which he needs to work.

4.1. *Iinterfaces adaptation to users*

4.1.1. *Generality*

The heterogeneity of medical information systems shows a double antagonist consequence: the impossibility of changing well-used legacy systems because of cost-, and time reasons; and the necessity of sharing medical file to carry out the patient care. Many difficulties can be noticed: heterogeneity of the data structure

(XML, RDB, OODB), heterogeneity of the content (num_pat vs n°Patient), duplication of data, lack of standardized vocabulary (ICD, ICPC, others thesauri). These difficulties must find different solutions according to each problem: ontologies to thwart heterogeneous vocabularies, specific algorithms to detect a given patient (with his name, first name and adress), mediator to translate heterogeneous data sources in a pivot language, etc.

If we cannot change MIS because of previous reasons, we can use the different technologies to overcome the problem and to solve it at the interface level. Doctors are not computer scientists and to move the heterogeneity problem from the DB level to the interface does not mask the problem. So an interface-based approach has to mask the difficulty to the user and propose user-friendly and ergonomic graphical user interfaces. Three concepts have to be considered: the presentation of the interface as a web-based application with easy navigation through web pages, the adaptation to the user profile (according to access rights –very important in medicine-, medical specialty, etc.) the real-time adaptability to the user browsing (MIS learning, increasing of the knowledge, etc.).

We said that to move the heterogeneity problem to the interface does not mask the difficulty to retrieve, remove and share medical data: it is a crucial problem. Mediators and exchange platforms propose solutions to translate relational, object-oriented and XML data into a pivot language or into XML language. Solutions exist and are correct because it is less difficult to transform very-structured data (like DB) into semi-strutured or unstructured data than the opposite. The trickiest problem is the semantic one. To rebuild the medical record of a patient whose cares took place in different places of care is really difficult because the identification of the patient is not obligatory or different from a place to another, so to be sure to link medical data of a unique patient and not to mix data from different patients is very difficult. Many research projects work on this problem. We can cite for example, what is done in (Quantin, 2004; Kelman, 2002). Numerous works are led about semantic web concerning particularly ontologies, interoperability, facetted databases and context-based semantic. Among these works, we can cite (Widhalm, 2003; Laks, 2003), (Ross, 2004) and (Zhu, 2004).

We proposed a new solution based on a coupled Topic Maps and Description Logics approach to overcome the semantic data integration and a graphical interface based on cartography to simulate the web navigation.

4.1.2. *Concepts cartography*

General remarks

The reconstitution of the medical history of a patient becomes then a challenge except to build *ad hoc* systems where each element is verified again (at the conceptual and semantic level). For these same reasons, distributing medical information from heterogeneous sources and leading epidemiologic studies according to stratification criteria is not possible without manual intervention.

These two problems led us to think of another mode of interrogation. Since the *ad hoc* rebuilding is impossible (except to carry out expensive handling), it seems preferable to build a system downstream with a complete freedom for the user to question these data without being concerned with constraints enumerated above.

For that, we built a system of interrogation/navigation based on Topic Maps. Data structure and values are represented in a single format representing navigation cartography. This double system makes it possible to handle medical information according to two different manners: adapted to the user profile and adaptive in real time with navigation. It produces two results:

– A centred-patient navigation allowing rebuilding his medical history;

– A centred-population navigation allowing incorporating statistical data for epidemiological studies.

Topic Maps approach

Topic Maps represent a formalism of management, representation and organization of knowledge (Sigel, 2000). It is used to formalize human knowledge with an aim of facilitating and optimizing the operations of creation, management and search for this knowledge. Topic Maps represent knowledge in a related graph composed of nodes bound by semantic relations. Topic Maps thus make it possible to represent knowledge but also to organize resources. The basic concepts of Topic Maps are very simple (ISO, 1999). Topic Map is represented by a graph of topics, which constitutes subjects, linked by semantic associations. Topics associated resources called occurrences. To exchange Topic Maps, TopicMaps.Org Authoring Group standardized in 2001 a XML-based grammar for serializing Topic Maps. This grammar is called XTM (XML for Topic Maps).

Thus, Topic Maps constitute a very powerful formalism of representation. It makes it possible to represent relational databases, object-oriented databases or XML files with the same flexibility. The great force of the formalism led to its antagonism: it is not possible to represent hierarchies between the topics (in a universe where all is topic). Topic Maps are coupled with description logics which makes possible to obtain, at the same time, flexibility and simplicity of representation (Calvanese, 1999; Baader, 1999).

This information system, intended for the end-user, provides an interface of navigation/interrogation of multi-source data enabling him to free itself from the localization of data, of their structure and their nature. A session of interrogation/navigation is concretized by the elaboration of a document non persistent and formatted according to users' preferences. This document is called virtual document because (especially for legal reasons), it is not stored in the system.

Centred-subject navigation is applied to the interrogation of a subject. The subject can be anything: Joyce, blood analysis of Robert or the Boston hospital. From the subject chosen, the navigation/interrogation continues stages in stages and every link between two subjects contains the cardinality of the set: we speak about exactly instantiated graph. The result of a browsing is represented by a personalized and virtual document.

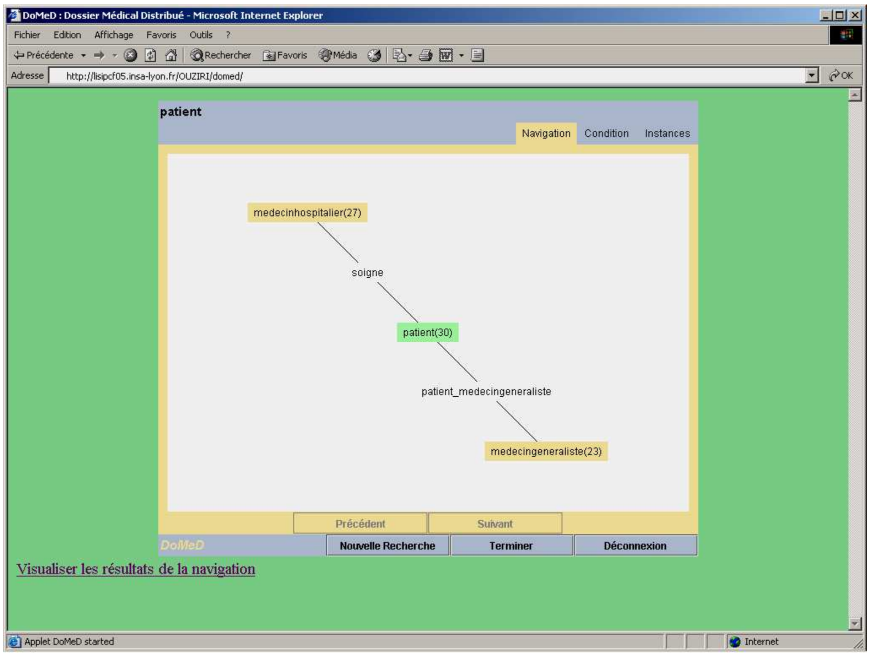

Figure 3. *Cartography centred-subject*

The second mode of navigation consists of dividing the whole of data sources in order to carry out statistical groupings. This interface is not used for navigation in a space of data. The user visualizes data analysis in a hierarchy of concepts which expresses semantic relations and data distributions. From the Topic Map, the user explores the concepts and specifies the analysis attributes. The system calculates the distribution and returns the results. This functionality leads to data partition according to a particular interest and with immediate result displayed on the screen.

4.1.3. *User profiles*

To improve our idea, we work on the concept of user profile in order to give fluidity to the inter-job communication. The first user-profile application is to remove not-relevant information according to the context (defined by the scope concept in Topic Maps). Indeed, all users of the MIS do not have the same waitings, knowledge, competences, centers of interest, etc... So it is necessary to define criteria or semantic rules for each user or user group. Two essential points must be taken into account: the construction of user profile and evolution and update of this profile. With the idea of interoperability, profile contents to be built will be characterized by a whole of open dimensions. Until now, we distinguished seven dimensions:

a) Professional data: they include information about identity, profession, main activity, place of work of the user. These data are relatively stable.

b) Preferences: this dimension represents specific preferences i.e., centres of interest (sport medicine, dietetics, etc).

c) Level of knowledge: this dimension provides information about the knowledge level of users. Two users who have the same activity do not have necessarily the same competence level.

d) Domain ontology: to allow the translation of a medical term in another simpler for a user with fewer competences (e.g., Myocardial Infarction → Heart Disease).

e) Authentication and confidentiality: this double dimension represents the access right to medical information (for each user and place of care) and data protection (cryptography).

f) Preferences feedback: the behaviour of the user is identified during the navigation or the interrogation of the system.

g) Miscellaneous information: this dimension represents information that can't be classified in other dimensions.

4.2. *Medical data grids*

4.2.1. *Definition*

Pervasive information systems are difficult to define. It probably concerns several information systems (not only once) and the definition is different according to the point of view which one adopts. If we try, we can say that a pervasive information system is a collection of communicating information systems that are able to produce a data for a good person at a good time in a good place on a wished computing device. According to E.J. Sol (Sol, 2001), people will use within 10 years, hundred interconnected computing devices around them, a majority of wireless connections and many embedded devices. It supposes that major data-processing problems in the years to come are: the security, the replication and the resilience. The security problem is double: to guarantee the sense and the semantic of the data during the transaction and to ensure that the data is only viewed by the authorized person.

A direct consequence of pervasive information systems is pervasive computing. In (Cherniack, 2001), the authors consider that pervasive computing make a computer so imbedded, so fitting, so natural that we use it without even thinking about. They consider that pervasive computing has to force the computer to go to people and not the opposite. These definitions meet on the crucial point: to bring computerized data at any moment at any place on any device to the person who needs it. Medical domain undergoes the same evolution. Some articles are dedicated to this evolution of health information systems. So, it is obvious that contrary to other domains where computer science has gradually evolved, health domain has a great jump to do to meet this challenge.

The main way to adapt pervasive computing to health area is the health care networks. A health care network can be defined as a link between health professionals for a disease. This link represents the whole data useful for diagnose, understand, care and follow a particular disease. Many health networks are created in France and in Europe to manage complementary health care around a particular disease: in France, are proposed health networks in kidney failure, pregnancy, measles and so on. The networks group together different health professionals and are realized with particular computerized systems. Many problems can be raised: every new networks develop its own structure but use similar conceptual entities, to make heterogeneous information systems is very difficult, and to be sure to link semantic-garanteed data is practically impossible to do. A solution can be found in data grids. More and more works deal with data grids. Main works concern image processing and retrieval on grids (Montagnat *et al.*, 2005).

4.2.2. *Data access and patient identification*

Data access represents a specific problem in grids. The data are located in distributed information systems with no idea of place of storage. Some metadata are stored in the local cache memory but to find a data set coming from different data sources is difficult. In medicine the problem is more crucial because the identification of a patient is non existing or different from an IS to the other. So to rebuild the medical history of a patient without a correct identification number is very difficult. We try to adapt existing algorithm (Quantin, 2004) to grid problematic. The architecture of our system is the following:

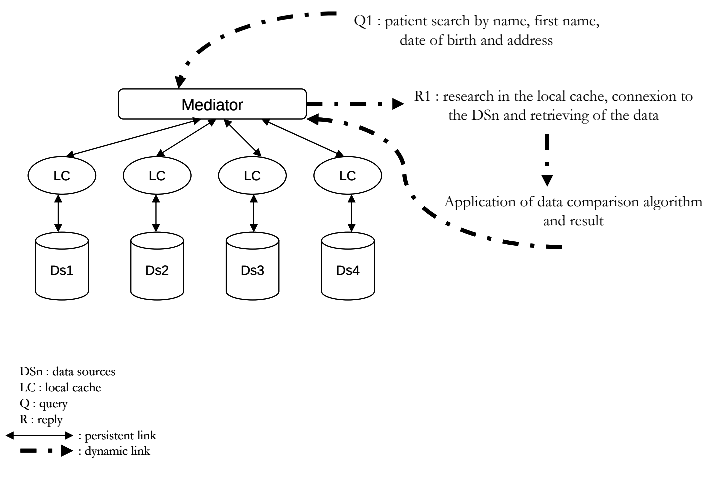

Figure 4. *Data retrieval based on grid architecture*

Query Q1 is sent to the mediator which role is to translate the query into a pivot language to be understood by local cache. Every local cache containing metadata corresponding to the query serach for the data in the local sources. After returning the reply, many data can be concerned with the query. Then, we apply a data comparison algorithm to compare all the character strings and give the good patient according choosen criteria. The reply is translated by the mediator in a pivot language to be used subsequently for any need (transmission of data, statitiscal need, etc.). All queries (Q1...Qn) are sent in a dynamic view to have a persistant and relevant result. This dynamic part is not discussed in the paper.

5. Conclusion

We have presented in this paper some discussion about the future of medical information systems. What we think about this is mainly the ability of new systems to deliver the good information to the good person at the good time on the good device.At one side, MIS can't be anymore static, totally structured, non flexible to each user, uniform and non changing. At the other side, MIS can't be easily modified to take into account new challenges. A solution is to tranfer the problem from the conceptual level to the interface level. At this level, two possibilities coexist: data access from heterogeneous information systems through data grids and adaptative interfaces to navigate in a data space and query data with a complete semantic guarantee. We have presented in few ideas basic research ideas of our team work. The reader can find more precisions in the following papers (Ouziri, 2005), (Ouziri, 2003).

6. Bibliography

Allen J.F., "Maintaining knowledge, about intervals", *Communication of the ACM*, 1983, (26), 11, pp. 832-843.

Alpay L.L., Lovis C., Nowlan W.A., Solomon W.D., Rush T.W., Baud R.H., Juge C., Scherrer J.R., "Constructing clinical application: the Galen approach", Medinfo 1995 ; 8 Pt1 : 65-9.

Baader F., "Logic-based knowledge representation", M.J. Woldridge and M. Velose eds, *Artificial Intelligence today, Recent trends and developments*, 1999, pp. 13-41.

Barrows R.C. Jr, Cimino J.J., Clayton P.D., "Mapping clinically useful terminology to a controlled medical vocabulary", *Proc Annu Symp Comput Appl Med Care*, 1994, 211-5.

Booch J., Rumbaugh J., Jacobson I., *The unified modeling language : user guide*, Booch, Jacobson, Rumbaugh eds, Addison Wesley, 1998.

Borst F., Griesser P., Bourdilloud R., Scherrer J.R., "Fifteen years of medical encoding in the Diogene HIS", *Medinfo*, 1995 ; 8 Pt1 , 43-6.

Bréant C., Borst F., Campi D., Griesser V., Son Le H., Junod J.M., « Expanding Diogene with a clinical information system based on a new hospital-wide clinical findings dictionary", *Int. Journal of Medical Informatics*, 58-59, 2000, pp. 167-177.

Burgun A., Botti G., Bodenreider O., Delamarre D., Leveque J.M., Lukacs B., Mayeux D., Brémond M., Kohler F., Fieschi M., "Knowledge acquisition from the UMLS sources : application to the description of surgical procedures", *Medinfo*, 1995 ; 8 Pt 1 :75-9.

Calvanese D., De Giacomo G., Lenzerini M., « Representing and reasoning on XML documents : a description logic approach », *Journal of Logic and Computation*, 1999, Vol. 9, No. 3, pp. 295-318.

Cherniack M., Franklin M.J., Zdonik S.B., "Data Management for Pervasive Computing", Tutorial, *proceedings of the 27th International Conference on Very Large Data Bases*, Sept. 11-14, 2001, Roma, Italy, available on http://www.vldb.org/dblp/db/conf/vldb/vldb2001.html

Cimino J.J., "Review paper: coding system in health care", *Methods Inf Med*, 1996, Dec, 35, (4-5), pp. 273-84.

Cimino J.J., Clayton P.D., Hripcsak G., Johnson J.B., "Knowledge-based approaches to the maintenance of a large controlled medical terminology", *J Am Med Inform Assoc,* 1994, Jan-feb, 1, (1), pp. 35-50.

Davis G.B., "Information systems conceptual foundations: looking backward and forward", *chapter 5, available at http://www.terry.uga.edu/~ekarah/davis.pdf*

Degoulet P., Fieschi M., "Critical dimensions in medical informatics", *Int J Med Inf*, 44, (1), 1997 Mar, pp. 21-6.

Engelmann U., Jean F.C., Meinzer H.P., Schröter A., Günnel U., Demiris A.M., Schäfer M., Evers H., Degoulet P., "Integrated image processing in clinical applications : the Helios approach", *Medinfo*, 1995 ; 8 Pt 1, 411-5.

Flory A., Verdier C., Mercatello A., "Conceptual modelling of medical information", *Health and System Science*, Hermès eds, Vol. 1, No. 3-4, Dec. 1997, pp. 263-90.

ISO/IEC 113250, Topic Maps, Dec. ISO/IEC FCD, 1999.

Joubert M., Fieschi M., Robert J.J., Tafazzoli A., "Users conceptual views on medical information databases", *Int J Biomed comput*, 1994 Oct, 37, (2), pp. 93-104.

Kelman CW, Bass J, Holman CDJ., "Research use of linked health data - a 'best practice' protocol", *Australian and New Zealand Journal of Public Health*, 2002, 26, (3), pp. 251-255.

Lakshmanan L.V.S., Sadri F., "Interoperability on XML Data", *Lecture Notes in Computer Science*, Publisher: Springer-Verlag GmbH, ISSN: 0302-9743, Vol. 2870/2003, pp. 146-163.

Lavril M., Dore L., Zapletal E., Jean F.C., Degoulet P., "A reuse oriented development database : the Helios object information system", *Comput Methods Programs Biomed*, 1994 Dec; 45 Suppl, S35-45.

Montagnat J., Bellet F., Benoit-Cattin H., Breton V., Brunie L., Duque H., Legre Y., Magnin I.E., Miguet S., Pierson J.M., Seitz L., Tweed T., "Medical images simulation, storage and procesing on the European Datagrid testbed", *Journal of grid computing*, Kluwer, Vol.2, num 4, 2005:387-400.

Ouziri M., Verdier C., Flory A., "Semantic indexing for intelligent browsing of distributed data", *13th Int. Conf. On Information and Intelligent System, IIS 2003*, June 2-5, 2003, Zakopane, Poland.

Ouziri M., Verdier C., Flory A., "Data integration and user modelling: An approach based on Topic Maps and Description Logics", *ICEIS 2005, 6th Int. Conf. on Enterprise Information Systems*, supported by ACM, Miami, USA, May 24-28, 2005.

Quantin C., Binquet C., Bourquard K., Pattisina R., Gouyon-Cornet B., Ferdynus C.; Gouyon J.B., Allaert F.A., "Which are the best identifiers for record linkage?", pp. 221-227, (7) *Medical Informatics and the Internet in Medicine*, ISSN 1463-9238, Publisher: Taylor and Francis Ltd, Vol. 29, Issues 3-4, Numbers 3-4/September-December 2004.

Ross K.A., Janevski A., "Querying faceted databases", *Second International Workshop on Semantic Web and Databases*, 29-30 August 2004, Co-located with *VLDB 2004*, Toronto, Canada, 30 August - 3 September 2004.

Scherrer J.R., "The hospital information system integrated patient records", *Comput Methods Programs Biomed*, 1995 Sep-Oct, 48, (1-2), 1-6.

Sigel A., "Towards knowledge organization with Topic Maps", *XML Europe 2000*, Palais des congrès Paris, 12-16 juin 2000.

Sol E.J., "Ambience Intelligence with the Ubiquitous Network, the Embedded Computer Devices and the Hidden Databases", *Invited paper, proceedings of the 27th International Conference on Very Large Data Bases*, Sept. 11-14, 2001, Roma, Italy.

Soula G., Delquie P., Vitton P., Bartoli J.M., Fieschi M., "Designing medical hypermedia : the Forum project", *Stud Health Technol Inform*, 1997, 43 Pt B, 712-6.

Verdier C. Flory A., « An information system for epidemiology based on a computer-based medical record », *Journal Methods of Information in Medicine*, Vol. 33, 5/94, december 1994, pp. 496-501.

Widhalm R., Mueck T.A. "Merging Topics in Well-Formed XML Topic Maps", *Lecture Notes in Computer Science*, Publisher: Springer-Verlag GmbH, Vol. 2870/2003, pp. 64-79.

Zhu H., Madnick S.E., Siegel M.D., Representation and reasoning about changing semantics in data sources", *Second International Workshop on Semantic Web and Databases*, 29-30 August 2004, Co-located with *VLDB 2004*, Toronto, Canada, 30 August - 3 September 2004.