Jean-Raoul Scherrer, Christian Lovis, François Borst

Hospital Informatics Centre,
Geneva Canton University Hospital,
Geneva, Switzerland

Review Paper

DIOGENE 2, a distributed Hospital Information System with an emphasis on its Medical Information Content

Abstract: DIOGENE 1 has been a mainframe-based centralised HIS with a star network of communication operating on a daily basis with 120 nursing ward units since 1978. Together the limited and costly growth capabilities of such a system with its extreme difficulty in cooperating jointly with other heterogeneous medical systems, with the need for faster networking expansions, led to the new design of a distributed architecture called DIOGENE 2. In 1989, a migration process between DIOGENE 1 and DIOGENE 2 was initiated and is now on the verge of being achieved. During the time of this new expansion of the HIS, it has been easy to cooperate with the decentralisation process of the new hospital organisation as well as facilitating the integration of new functionalities like i.e. new WIS architecture, medical office patient histories, integration based upon PCs with UNIX based client/server platforms. That approach combines the handling of paragraphs structured patient records with the use of medical natural language processing and semi-automatic encoding as well. Amongst these new functionalities the PACS are associated with image manipulation platforms called OSIRIS for X-Ray images as well as other tools devoted to molecular biology and genetics up to the ExPASy server on Internet using WWW / Mosaic which is accessible from all over the world. The distributed architecture appears well suited not only for the integration of these new functionalities but to keep them growing as smoothly as possible.

Keywords: Hospital Information Systems (HIS), distributed architecture, computer-based patient record, PACS, image manipulation, WWW/Mosaic

1. Introduction

Since the initial design phase from 1971 to 1972, the DIOGENE hospital information system at the University Hospital of Geneva has been treated as a whole and has retained its architectural unity, despite the need for modification and extension over the years. [1] In addition, to having a centralized patient database with the mechanisms for data protection and recovery of a transaction-oriented system, the DIOGENE system, like other large hospital information systems [2-8], supports many applications within the hospital [1.9]. The system was introduced for administrative applications, including personnel management, in 1974. Invoicing was added in 1977, patient admissions in 1978, and general accounting in 1979. The system was progressively extended to all the wards in 1979 and 1980. Extensions to the radiology department and laboratories occurred from 1979 to 1982. Interactive medical encoding by interns and residents was added in 1985. The system soon included new emergency laboratories (1986) monthly and yearly medical statistics on patients (1987), the pharmacy (1987), integrated outpatient clinics in surgery (1987) and medicine (1988), and the medical bacteriology laboratory (1988). At this stage, the main lines of the implementation were:

. A centralised database
. A pool of operators
. A telex printer network [10]

Centralisation of information used to be the main vehicle for integration. The pool of operators made possible a fairly high acceptance of the HIS by its users and eased the introduction of new applications.

In 1988, the DP division employed 100 persons with about 30% of the work force assigned to software development and maintenance tasks. The instrumentation consisted of a CDC mainframe (Cyber 855) backed up by a Cyber 840 used for development and large administrative batches.

Now in 1995, DIOGENE 1 is 17 years old, counting from the first transaction made on the system. At this moment, DIOGENE 2, a federated HIS, [11] is reaching its mature stage and is ready to take over. It should be noted that the deadline was set up as a result of an estimate of the time needed to build a second generation HIS with more growth capabilities being able to integrate new functionalities impossible before, like images, not on more objective factors affecting the DIOGENE life cycle.

At this time there were still a few published plans for a second generation of HIS; the most notable and promising, being at the design stage, were those of the Johns Hopkins Hospital in Baltimore [12].

For the years to come the success or failure of the DIOGENE 2 development project will not only rely on technical grounds and DP department internal skills and strengths. Ongoing research on information systems is going on to clarify the numerous relationships and interactions between information systems and their medical and patient care providing environment. One of the hopes of the DIOGENE 2 project is to leave more time for the social and political aspects of system development and operation.

On the technical side, there are still open problems to be solved in order to fulfill the promises of a highly available distributed system [13]. While open system's interoperability was a reality during a certain period at the end of the 80's, there is still a lack of high level tools and facilities to implement fail-safe distributed communication and computations.

If the Hospital Information Systems of the 90's are built on a set of interconnected components, a natural trend in this area is already the standardization of interfaces between such components. This effort can only be done by international bodies and even now will probably require a long time before significantly useful standards will be brought to the HIS builders community.

The leaders among these businesses should already be in a position to derive significant gains from the implementation of distributed computing to set up the paradigm of the distributed hospital and health model. Distributed systems will then facilitate the convergence of business and technology; open systems will be deployed for the benefit of both users and suppliers, as opposed to vendor defined systems, constraining the user to depend on a single supplier. Focus should be on a limited number of areas:

. high-level (conceptual) modeling of distributed information systems, where Europe has a strong standing
. systems integration and interoperability, to support systems federation
. management of federated systems, where quality of service is at risk.

Modeling information systems at a high level and using this model as the basis for implementation has proven successful for centralised application design. A distributed system as a whole should be modeled, constituting an architecture, as should the parts of a system, specifying functionality of components, the interaction between components, as well as interaction between the system and its environment.

A key part of the architecture of a distributed computing system is the specification of the interfaces and interactions between components of the system. Various models are used to describe interaction at different levels of a system.

Advanced interfaces should adapt to the behaviour and the characteristics of the individual users to provide an active support in the identification and retrieval of the information of interest from the various data sources.

Users have to operate within constraints, which are to be taken into account, including:

. interoperability with existing systems
. portability over different platforms, fault tolerance
. security
. standardised access methods and interfaces for data access.

In many cases, especially in the hospital medical school, users have a set of existing applications running on heterogeneous computers: there is thus a strong need for methods, tools and guidelines to help integrate newer applications with earlier "legacy" ones. This activity implies to deliver tools and methods for the integration of legacy systems, and migration which corresponds to an essential user need.

Distributed systems provide direct support for decentralised business units, making use of their own local processing, while being given access to hospital-wide information systems. A federation of systems complies with flat organisation structures (lean management). They promise to maximize the use of networked resources, services and databases, leading to a minimisation of costs.

HIS distributed architectures provide an environment where innovative off-the-shelf applications of varied heterogeneous sources can be developed and distributed, provided appropriate measures are taken to reduce their cost of entry, and to broaden their potential area of action and services. Such applications should be seen to be independent of the physical structure of the system. This need is fulfilled by the transparency of distributed systems.

Federation, in this respect, is seen to be the key structuring principle to combine components of a system. Finally worldwide interconnection is likely to have a considerable impact on the development of various types of distributed systems, as anticipated by the European Union White Paper on "Growth, Competitiveness and Employment" (Bulletin of the European Communities, supplement 6/93) [14], and the United States government policy [15].

2. From the Organisational Model to the Conceptual Design

The hospital or the health system traditionally may be described as constituted by *resources* (structures, human ware, technological), *patients* requiring care and the dynamic information attached to them and the *organisation* (that is to say the way resources are arranged to provide effective/efficient services to patients) [16].

Such a health system is a set of interrelated organisation subsystems presenting the contingency dynamics of the organisation over time, navigated by management. *The result is a matrix structure [17]. That appears particularly well suited for hospitals/health systems since they may easily be split into "functional or operational units" that look alike from the standpoint of the set of resources and functions.*

These "operational/functional units" in our large University Teaching Hospital now coincide with the new department faculty organisation. There are nine such units:
1. Medicine
2. Clinical pathology
3. Pharmacology - Anaesthesiology
4. Surgery
5. Paediatrics
6. Gynaecology/obstetrics
7. Clinical neuroscience and dermatology
8. Community medicine
9. Radiology.

For each of these units, the following functional areas should be considered:
1. Patient management
2. Medical care
3. Medical support
4. Nursing
5. Ancillary services
6. Administration and management.

In the DIOGENE migration process, one of the major moves was to migrate from a former centralized ADT application (A = Admission, D = Discharge, T = Transfer) to a more general purpose application dealing indifferently with inpatients and outpatients, pre-hospitalized waiting lists of patients, follow-up post-op consultations, that has been designated by IMPACT (I = Identification, M = Movement, P = Prestations / services, AC = Actions, T = Transactions). All the DIOGENE 2 application programs intereact with IMPACT in one way or another and even more so with the archive server called ARCHIMED, which is also suited for any interactive queries from the users following in this respect, the Clinquery model [18].

All the functional interactions between the operation units may be seen in a matrix form with "actors" (patients, physicians, nurses, consultants, ancillary personnel, clerks and managers) and "actions (medical care, patient management, nursing care, medical support, ancillary services, administration and management) [19]. Since there is a certain lack of homogeneity between the operational units, it is more convenient to consider more general sub-entities: the wards that could become a standard: "the ward information system" (WIS).

3. The Ward Information System (WIS) and the his Architecture

As mentioned by Degoulet et al. [20] a Software Engineering Environment (SEE) in medicine should gather tools for the development of medical applications. "It should include tools to represent information and knowledge about patients, diseases and medical procedure and tools to exploit the data and knowledge bases for various tasks like medical action, clinical research or health care quality evaluation". Such an SEE is an appropriate standardization process to develop an integrated standardized patient care WIS on a broad open/distributed architecture that was inconceivable on the previous centralised architecture. Indeed the workload of each WIS is close to using all the resources previously available only on a mainframe.

Nowadays, from many perspectives, clinical workstations are becoming increasingly central to providing access to computer-based patient records [21]. Hence, multimedia medical workstations represent the natural tool for accessing the hospital information system environemnt [22]. This new emphasis on WIS coping so well with the new multimedia tools available on the market and with the distributed/open architecture, sheds light to better understand the perfect match between the health decentralisation process and the availability of the new computer architecture. In a way the migration to distributed systems is ward-driven and introduces quite a number of new functionalities.

Faced with an increasing number of patients, as well as a rapid improvement of technology, the ward is today constrained to increased computer assistance within the hospital. The need to alleviate the administrative work, to improve communication and the computer environment is crucial in order to maintain and enhance the quality of medical care. The ward staff has to assess the soundness of their medical decisions, they must get more insights about any possible medical malpractice, especially in a situation of continuous knowledge growth, and finally they must appraise the utilisation of resources.

Today HIS [23-24] successfully solve problems within two areas of the hospital:
1) administrative management and
2) communications.

Administrative tasks are not specific to hospitals, they are currently in use in many enterprises, and they are commonly available in many hospitals. However, communication has to be understood here with the meaning of integration of entities, the goal being to make these entitites work together and share common information.

In this regard, communication tasks are more difficult in large institutions, and the requirements of the hospital are quite diversified: laboratory requests, x-ray examinations, patient appointment drug prescriptions, orders for care material, etc. The true value of the HIS is certainly in the more or less complete integration of these tasks so strongly interacting with each other, rather than the simple enumeration of them. Such a goal is rarely fully achieved within the present HIS, especially those which are commercially available or still centralised.

All the above-mentioned tasks are considered as part of the HIS. Even when they are performed by the ward staff, HIS transactions have to be separated to a conceptual point of view from WIS transactions. HIS transactions are constrained and oriented by different requirements, not all of them being originated in the ward. For this reason, a limit has been drawn between the WIS and the HIS, as represented in Figure 1.

What is meant here is that the HIS is responsible for communications with any ancillary area outside the ward. The WIS is encapsulated into the HIS as well as the WIMS into the WIS. The WIS is free from the burden of communications with the rest of the hospital. This statement is certainly grounded on the existence of real HIS. HIS more or less act as insulators between the ward and the hospital.

The WIS is considered here in the environment of a HIS, which fulfills all the communication tasks outside the ward. The WIS is responsible for handling data specific to the ward, pertaining to the patient, the source being the nurse or the physician. The WIS user provides and exchanges data with a surrounding HIS. A necessary frontier - though a somewhat abstract one - exists between HIS and WIS data.

The DIOGENE 2 distributed hospital information system may be briefly described in the following way:

. The experience of the migration process (initiated in 1986) of a former mainframe-based legacy system (DIOGENE 1) onto a full scale distributed/open architecture (DIOGENE 2) has been achieved successfully. The scope of applications is ADT for in/out patients, Laboratories, Radiology information system, Rx reports, Image manipulation platform and PACS, Teleradiology for satellite hospital, Pharmacy, patient care team management, integrated medical office automation, archiving and medical documentation, personnel management, management applications and tools.

. The initial network has been redesigned and recabled according to the following standards: Ethernet TCP/IP, FDDI backbone, prelimi-

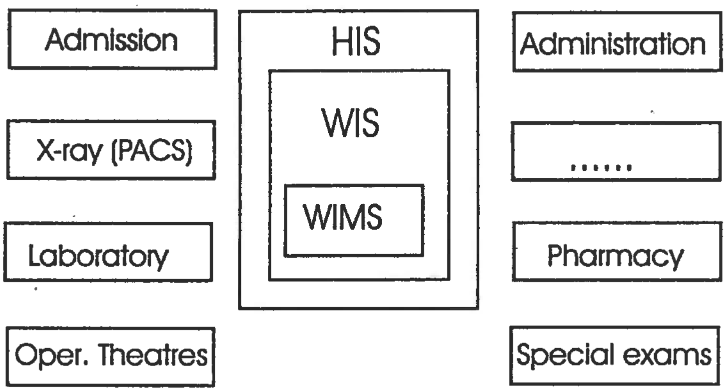

Figure 1: The WIS, the HIS, the WIMS (Ward Image Manipulation System) and ancillary areas.

nary efforts towards ATM. The network is open to the outside world using Internet (including WWW/ Mosaic), X25, PDS to ISDN.

- The new distributed production is Openview based that is with network node manager, operation center, etc.
- The 32 servers of the client-server architecture:
  - there are 3 database servers (2 RS 6000-990 and 1 HP9000-170) plus one additional DB server that is expected in the following months (SUN-Sparc-Center-2000E, dealing with SOLARIS operating system).
  - 5 mixed DB servers/applications (all from the HP-9000 family)
  - 5 application servers (all from the HP 9000 family)
  - 19 servers devoted to basic common services requester, directories, remote printing facilities, e-mail, X25 Internet gateways) that also includes development platforms.
- The application machines (this includes a Mosaic Server) 39 Sparc 10 or 20 plus an integrated PACS.
- At the moment 2000 PCs are connected to the network with 789 PCs/ 80486 under Windows. In addition to that there are 183 integrated Macintosh and 1080 printers.
- The software standards and main streams are UNIX, SOLARIS, X-Motif, C and C++, INGRES, Oracle, Windows, Windows NT, etc. Consequently, the field of expertise could be stated this way:

Experience of a migration of a large legacy system; Distributed/open architecture practical expertise with an emphasis on how to integrate heterogeneous systems with respect to de facto standards.

4. New Structures for the Data

The corpus of data on which the WIS is based is the *Medical Record.* The rationale of the WS is to integrate data from outside as well as inside providers into cross departmental views [25-26]. The computerized medical record is designed in order to replace the traditional paper based medical record [27-31]. In our architectures, the total Computerized Patient Records (CPR's) are handled through a devoted server called UNIDOC [32] that interacts with the WIS.

This computerized medical record is a real and substantial improvement for the ward. Indeed, a major deficiency of any paper based medical record is the fact that data can only be displayed in the format in which they were originally recorded. Nevertheless, the main advantage of the electronic media is that data needs only to be entered once (thus avoiding errors) and can be presented in an almost unlimited number of contexts [33]. The third and fourth objectives of the WIS are the following:

Information is entered into the WIS from various sources or actors, inside and outside the ward. Integration of data is an essential quality. The incorporation of a knowledge base and the use of inferential capabilities are advocated.

Moreover, besides the medical record, the benefit of a computerized environment is the integration and addition of different tools and systems, which can perform automatically or interactively. The computation of the date of the next biological test in the case of a periodic treatment is an example of basic data treatment. Decision Support Systems of many kinds are also available. They can also be sophisticated knowledge-based systems requiring high level artificial intelligence tools [34]. For instance, diagnosis support software or treatment planning software require tools for temporal reasoning [35]. A number of references are available, but will not be enumerated here. However, the perspectives for integrating such tools are of considerable interest for the WIS. This leads to another objective:

It is possible to schematize the WIS processing by the schema of the following figure. Both sets of data, collected from HIS and from WIS as well, combined with knowledge-based modules, define the information involved in the WIS. Mere computations - like recording and displaying which are supplied by software acting directly on the medical record - should be distinguished from more complex computations - like decision making about patients - where quite a few accesses to both data and knowledge base appear to be essential.

5. Medical Office Automation Integrated into the Distributed HIS

The UNIDOC system is a client-service platform that has been developed and made operational within DIOGENE 2. It is based upon a fully standardized and distributed open systems architecture and has been described in detail [32]. It is shown that UNIDOC illustrates a feasible marriage of the two technologies, UNIX, MS-DOS / Windows. It constitutes an integrated office automation system.

Indeed, a patient history and a medical record are mostly composed of text, i.e. the patient's profile, which includes all the clinical findings, and the discharge summary, follow-up notes and specialized consultant's reports [36]. Although most of these texts are still usually written on paper, an increasing amount of texts are already for some time in machine-readable form, due to the use of personal computers (PCs) and laser printers. Hence, the total patient record could be archived electronically if the infrastructure were able to handle it [37]. Therefore, the use of PCs to produce paper archives is debatable when archiving in machine-readable form is feasible.

Therefore, the gathering and archiving of these texts in machine-readable form has many characteristics of computer-based medical records. In Geneva, approximately, 2000 PCs are connected to the Hospital Information System DIOGENE 2, with the possibility of accessing all the functions offered by the system without losing any of their MS-DOS Windows word processing capabilities. The UNIDOC system, presented in this paper, takes all these features into account, a real marriage of technologies between the MS-DOS Windows environment and the UNIX-based distributed client-server architecture. The INGRES database management system supports the entire archiving process of the medical patient texts, structured by prelabelled paragraphs and automatically indexed. Both the quality and accessibility of the records are enhanced, while the archiving capacity is neither too limited nor too expensive being based upon a cheap SCSI disc expansion.

The present availability of the MS-Windows NT, either as "standard" or "advanced server" (challenging OSF/UNIX) as a new and powerful operating system platform (having "multitasking" and "multithread" capabilities with advanced security features that fulfil the C2 level of requirements of the US Dept. of Defense) for client-server computing makes evolution and portability of UNIDOC concepts far easier. On center stage is the WIN 32 subsystem. There are other subsystems for running OS/2 applications, POSIX (level 1003.1 only), and VDM (Virtual DOS Machine) for running MS-DOS as well as a 16-bit windows application. There are already facilities to execute "well behaved" DOS applications. Besides, Windows NT has its native RPC and also an SQL Server interfacing Oracle ^{R}, Sybase ^{R}, and Ingres (including Ingresnet). Furthermore, with the prospective use of Windows NT and PCs - input devices - may even become client machines as such.

At present, the secretariats of the departments of Surgery and Radiology handle documents in this way. Documents that are processed by the system are discharge letters sent to downtown physicians, operating room reports, imaging reports including reports for X-rays, CTs, ultrasound, MRI and quantitative angiography, etc. With this approach it is possible to accommodate all specialized reports of consultants. Besides, the department of General Internal Medicine is preparing a new standardized procedure for collecting a patient 's full history and all the follow-up notes by using the same UNIDOC system, which means an additional three PCs and one more laser printer for each ward unit. The text processing system that has been adopted for all applications is Winword. All documents which need to be collected are structured paragraph by paragraph. Each category of document has its own set of paragraphs and its associated setup of keywords for archiving and retrieval purposes.

Several of the paragraphs are automatically completed and indexed. Automatic natural language processing of the paragraph contents, and of their knowledge representation, has been successfully accomplished but still with a research restrictive environment [38-40]. However it demonstrates the feasibility to represent medical texts in a DBMS consisting of detailed date as well as inferences well suited for clinical use. Such information can also be matched with Medline.

All the detailed information regarding patients' identifications, insurance companies, addresses, social and professional backgrounds, are obtained directly and automatically from the patient identification server of the HIS DIOGENE without any extra data acquisition effort. The case history or documents for any patient may be retrieved by any physician or secretary who has the right of access thanks to the HIS network facilities.

6. Natural Language Processing and Semi-Automatic Encoding

The need for systems that are able to accept multiple European languages is of paramount interest, as language barriers can be a strong impediment for large-scale communication in Europe, in particular regarding telemedicine, the Natural Language Processing (NLP) component offers a large variety of medical services according to natural language free input. It allows the multilingual analysis of medical texts and the storage of the meaning of these texts under a deep knowledge representation that can be queried whenever it is needed. In addition, it provides facilities to handle knowledge source embedded into the conceptual typologies and into the dictionaries.

Such an NLP service is also a significant part of a WIS. It is in this respect that the NLP service is locked onto the HELIOS software bus well suited for being the basic software engineering infrastructure of the WIS [41]. Coming back again to UNIDOC [32], indeed all the data related to the patient either coming from the DIOGENE 2 identification server and/or from the PC-text acquisition, are merged according to predefined labeled paragraphs that can be retrieved according to their indexation. However, the contents of the paragraph remain black boxes.

Keeping in mind that the most relevant document of the patient record is the discharge letter to be sent to the patient's GP, it is the most appropriate way to handle the encoding from that document. However, since the coding has to be done first from displays of the selected, most usual terms of one medical service, the patient is mostly undercoded and the more complex the patient's medical record is, the more undercoded it is since besides the principle diagnosis, the associated complex commodities are partly left aside. See the illustrative Figure 2.

This is why a new tool had to be designed for the detailed semi-automatic encoding of paragraph contents, that provides a better exhaustivity of the coding besides being of far better quality this new approach is also far easier to use [42].

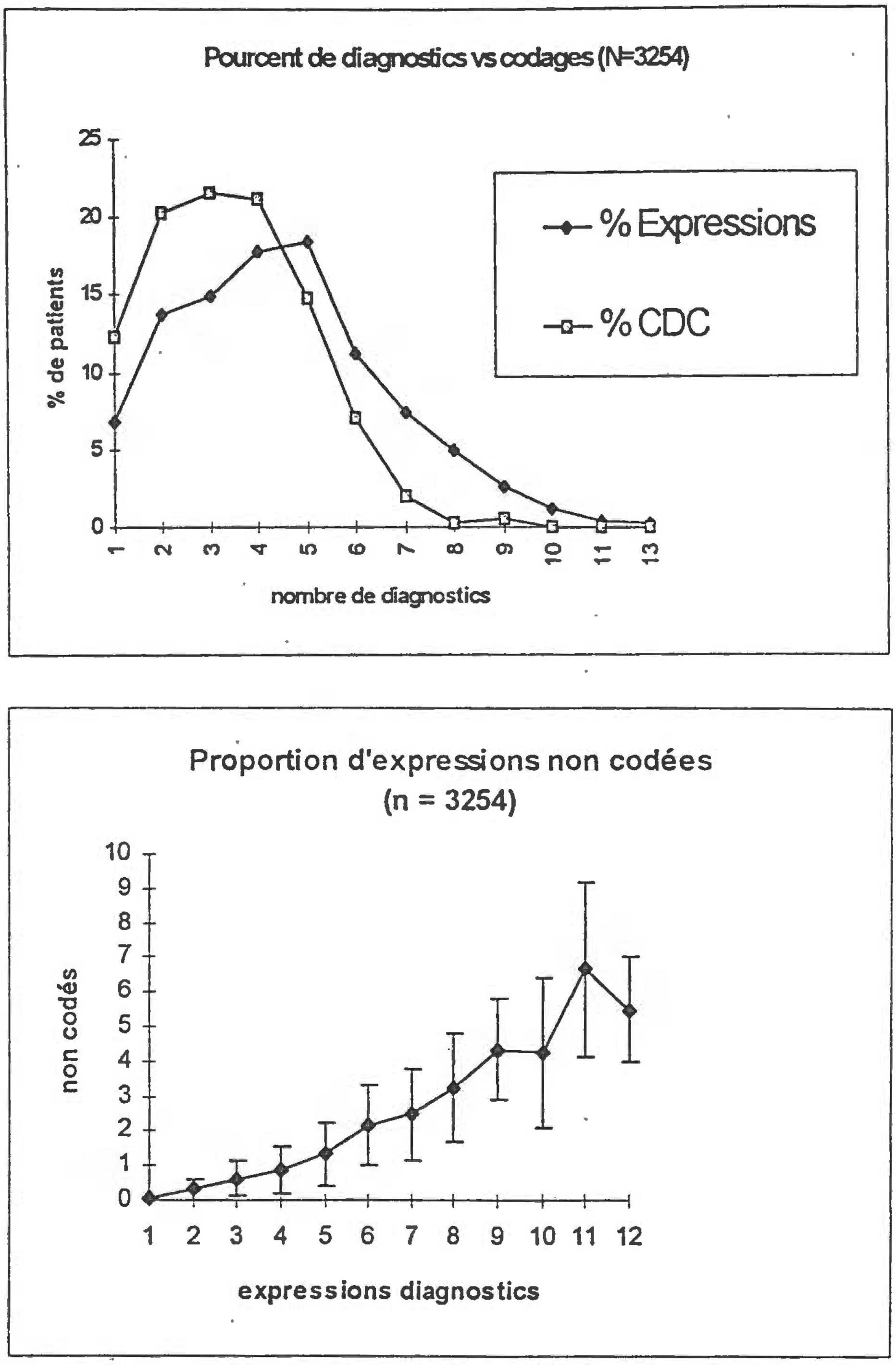

Figure 2

Although one can nowadays use many different classifications, some of them extremely large, like SNOMED International [43], there is a real lack of tools ensuring good quality and exhaustivity of diagnosis encoding. In this field, Chute et al, report that only up to approximately 60% of encoding fully matches the original diagnosis [44]. This may have important consequences if one thinks of all the economic decisions that are taken on behalf of these data. There are few stages between a clinical diagnosis and its expression as a code in a knowledge representation or disease classification such as ICD-9. At each stage, aspects of the semantic content can be lost and specific means must be found to minimise this loss of information. We now distinguish three main stages: a) the voicing of the diagnosis in the physician's usual language, b) the understanding of the coding scheme one has, and c) the projection of the diagnosis in the classification scheme.

The WHO International Classification of Diseases version 9 (ICD), possibly with the CM modificators, is the most currently used disease classification [49]. Our approach is based on literature [46-48], as well as on our hospital experience in coding [49] and in natural language processing [38]. We developed an informatic tool that largely simplifies the diagnosis encoding for ICD-9 and significantly increases the encoding quality. This tool has some very important characteristics:

a) it is user friendly, since it uses a graphical interface,
b) it is a low cost solution, as it runs on any Personal Computer with Microsoft Windows,
c) it is reliable, based on a large corpus of controlled expressions (more than 40'000),
d) it is physician oriented, our tool builds a set of controlled expressions centred on the diagnosis in French natural language.

Example: left heart failure due to atrial fibrillation after surgery

Quality:		
décompensation cardiaque (heart failure)	428.9:	insufficient
décompensation cardiaque gauche (left heart failure)	428.1:	good
fibrillation auriculaire (atrial fibrillation)	427.3:	insufficient
fibrillation auriculaire postopératoire (atrial fibrillation after surgery)	997.1:	good

Note here the position of 997.1 in the hierarchy and the illustration of exceptions.

Completeness:	428.1:	insufficient
	997.1:	insufficient
	428.1 & 997.1	good

In the previous example, each code matches with a specific and distinct part of the sentence. In this example, one can also note that expressions that seem near in terms of semantical content may be distant in the hierarchy of ICD-9 (like 427.3 and 997.1, both applying to atrial fibrillation). The first step of the analysis is linguistics. A second technique is used whenever a word is not recognised or not in the dictionary. This is the intraword processing step and we use the approach of morpho-semantems [50].

7. Digital Image Management and Communication in Medicine (pacs and image manipulation)

With the rapid development of digital imaging modalities in medicine, there is an increasing need for an efficient management and archival of medical images in digital form. Picture Archiving and Communication Systems (PACS) are becoming an essential component of medical imaging equipment, allowing for medical images to be accessed and stored directly in digital form. A paper by O. Ratib et al [51] describes a hospital-wide PACS currently under development at the University Hospital of Geneva, based on an open architecture, regrouping equipment from different vendors in a distributed topology. The image archival is organized in multiple locations geographically distributed in the hospital. The PACS database is fully integrated with the concurrent Radiology Information System (RIS) [52] and Hospital Information System (HIS) DIOGENE 2. A standard image storage format called the PAPYRUS [53] format was developed for the storage of medical images from a variety of imaging modalities. To provide a more uniform under interface on a variety of different workstations, a common platform for image display and manipulation called OSIRIS was developed [54].

The PACS developed at the University Hospital of Geneva is based on a distributed architecture with hierarchical archive of images and related data with multiple archive and display servers. This type of architecture seems to be, nowadays, the only viable solution for large scale PACS, where images from a variety of imaging modalities are accessed from a large number of consultation points. The alternative design previously implemented in some other institutions is a centralized archive, where images and related data are accessible from a single distribution point. A centralized architecture is not easily expandable to large scale PACS and requires very high performance hardware to maintain acceptable throughput and data access. A distributed architecture is more suitable for hierarchical storage, where duplicates of the images are temporarily stored on local display servers to be more accessible and insures high performance for a large number of workstations.

The Geneva PACS design corresponds to a second generation implementation strategy advocated by several authors during recent years [55-56]. The distributed PACS infrastructure is the basic design concept to ensure that PACS includes features such as standardization, open architecture, capability for future growth, connectivity, and reliability. We adopted a modular architecture for easy expandability and adaptability to the rapid evolution of computer technology. The concept of hierarchical management of image communication through "display servers" allows efficient dispatching of images to workstations and a better accessibility of the images through optimal routing scenarios. The complexity of such a system resides in the fact that a control mechanism is required to keep track of several copies of the images on different file servers. It must also remove files that are not being accessed for a certain time to keep enough free space on the file server disks for additional images. Such a management and control mechanism can be considered as a "distributed cache system". As such, it should maintain multiple copies of the image files on several storage devices on the network transparently.

Besides, in the late 1970s, and with increasing use of computers in clinical applications, the American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) recognized the emerging need for a standard method for transferring images and associated information between devices manufactured by various vendors. These devices produce a variety of digital image formats.

ACR and NEMA formed a joint committee (ACR-NEMA committee) in 1983 to develop a standard to:

1. promote communication of digital image information, regardless of device manufacturer;
2. facilitate the development and expansion of picture, archiving and communication systems (PACS) that can also interface with other systems of hospital information;
3. allow the creation of diagnostic information databases that can be interrogated by a wide variety of devices distributed geographically.

The first version of PAPYRUS file format was created in 1990 based on the data dictionary and data structure of the ACR/NEMA 2.0 communication standard [57]. This file format responded to a need for an image storage and communication format. Such a format should remain independent of any communication method used to transfer the images or any database structure used to handle these images. The PAPYRUS format was further extended and adopted by several European projects on digital imaging and teleradiology. The University Hospital of Geneva also provides a complete source code library called the "PAPYRUS toolkit" for facilitating reading and writing of PAPYRUS files.

In 1992 the ACR/NEMA committee issued some "pre-final" drafts of the completely remodelled DICOM Standard. Throughout the development of this new standard we have been in close contact with members of ACR-NEMA to update the PAPYRUS file format to match the new DICOM data structure.

Though the major impact of DICOM is expected to be on PACS it will certainly have a significant impact on image communication in general by simplifying many equipment interfacing tasks.

Many of the digital imaging techniques used in radiologic procedures (including computed tomography, magnetic resonance imaging, and even conventional radiography) are being converted to computed radiography, and a trend toward digital image management and communication systems can be observed. Developing tools for computer-based image display and manipulation that are appropriate for medical users is a challenge. Each digital technique may come with specific requirements that vary with the manufacturer. Commercial solutions are usually designed for a specific use or imaging technique and are not easily adapted to other uses. In a clinical environment where images from various sources are evaluated, a set of generic image-manipulation tools is needed.

At the University of California in Los Angeles (UCLA), an experiment with a prototype program, Calipso (California image-processing software), indicated a need for a portable platform with an interface that was friendly to medical users. A more ambitious project was then performed by O. Ratib et al. [54] closely interconnected with the new distributed RIS of DIOGENE 2 - HIS [52].

We designed a general-purpose computer program, Osiris, for the display, manipulation, and analysis of digital medical images. The program offers an intuitive, window-based interface with direct access to generic tools. Characterized by user-friendliness, portability, and extensibility, Osiris is compatible with both Unix-based and Macintosh-based platforms. It is readily modified and can be used to develop new tools. It is able to monitor the entries made during a work session and thus provide data on its use. Osiris and its source code are being distributed, free of charge, to universities and research groups around the world.

8. Digital Image Management in Moleculer Biology

Another fruitful medical domain of image handling is devoted to molecular biology and human genetics. Indeed, at the University Hospital of Geneva we are carrying out the MELANIE project on automatic diagnosis from two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). 2D-PAGE is a biochemical technique for separating proteins in a biological sample [58]. The proteins migrate in a polyacrylamide gel according to two important characteristics: their isoelectric point and their molecular weight. After staining the gel, one can observe spots which are spread over the gel according to these two characteristics. While the spot position indicates the isoelectric point and the molecular weight of the protein, the spot darkness is related to the protein concentration. The main property of the 2D-PAGE technique is its very high resolution power. Up to 2,000 proteins can be separated from plasma samples and other physiological fluids, while a 2D-PAGE image of a tissue sample can separate up to 3,000 spots. The main goals of the MELANIE project are the study of protein expression in diseases and the development of a "molecular scanner", that is, a tool which assists a clinician in establishing a diagnosis from a 2D-PAGE map [59].

We have developed a computer system (also called MELANIE, from Medical ELectrophoresis ANalysis Interactive Expert) to analyse 2D-PAGE images. The MELANIE system performs automatic spot detection and quantification and protein map comparisons. Other tools have been implemented to assist the clinician in analysing a 2D-PAGE image, particularly the ExPASy server providing hypermedia research facilities.

Recently, there has been a tremendous increase in the number, size and complexity of the databases used by molecular biologists. The great majority of these databases are available on the Internet, which links most major academic and industrial research centers. Various servers and tools exist to help the research worker retrieve information relevant to his/her work. Until the late 1980s, there were only two ways of accessing these databases over the net: electronic mail servers, which allow the retrieval of individual pieces of data relatively slowly (minutes to hours), and FTP servers from which whole databases can be downloaded so that they can be serched on a local machine. During the past few years, new transfer protocols, such as WAIS (wide-area information system) and Gopher, have been developed, which offer keyword search facilities that provide a flexible way of retrieving data interactively in seconds.

World-Wide Web

Finding the right piece of information on the net has been dramatically simplified by the WorldWide Web (WWW) Initiative. WWW, which originated at CERN in Geneva, is a global information retrieval system merging the power of worldwide networks, hypertext and multimedia.

The ExPASy WWW server

ExPASy is a WWW server set up at the University Hospital of Geneva and the Medical Biochemistry Department of Geneva University, and is dedicated to molecular biology with an emphasis on data relevant to proteins. The two main entry points on the server give access to the SWISS-PROT database of annotated protein sequences and the SWISS-2DPAGE database of two-dimensional gel electrophoresis images. SWISS-PROT can be searched by protein description, entry name or accession number, or referenced author name, as well as by performing a full text search on all the annotation fields.

From any SWISS-PROT entry, the user can access the corresponding entries in other databases by way of hypertext links. These links appear to the user as marked text that can be activated using a pointing device such as a mouse.

The ExPASy WWW server has links from SWISS-PROT to EMBL (nucleotide sequences), PDB (three-dimensional structures), Medline (bibliographic references), PROSITE (protein sites and patterns), OMIM (human genes and genetic diseases), FlyBase (*Drosophila* genomic data), REBASE (restriction enzymes), SWISS-3DIMAGE (images of three-dimensional structures of proteins) and SWISS-2DPAGE. Links to other databases will be added in the near future [60].

Conclusions

From our own experience of the migration of a former centralised HIS to a fully open-distributed architecture, homogeneity has turned out to be a dream, only heterogeneity is the fact. There are different platforms each with different evolutionary pace (workstations, UNIX, legacy OS, etc). Standards are not considered here as the "Unification" paradigm, but rather a means to guarantee openness. The evolution has on the other hand caused fragmentation, in terms of islands of information, to appear. Indeed, the technological scenario in an organisation consists of a set of already existing heterogeneous applications and the market is offering quite a number of cheap but isolated applications to support the specific needs of the users in the autonomous units of the health enterprise.

Distributed systems allow autonomy of individual applications and supply the necessary framework for integration and interoperability, inside and between organisations. The goal will be achieved by evolution and migration from existing to new structures of Information Technology in organisations. Migration strategies are seen to be of paramount importance, as are aspects of performance and gain for the user application. The usefulness of distributed computing is becoming increasingly better proven by successful implementations, with more ease for growth capabilities and hence for further expansion.

References

1. Scherrer JR, Baud RH, Hochstrasser D, Ratib O. *An integrated hospital information system in Geneva.* MD Computing. 1990; 7(2): 81-89.
2. McDonald CJ, Murray R, Jeris D, Bhargava B, Seeger J, Blevins L. *A computer-based record and clinical monitoring system for ambulatory care.* Am J Public Health. 1977;67:240-5.
3. Pryor TA, Gardner RM, Clayton PD, Warner HR. *The HELP System.* J Med Syst 1983; 7:87-102.
4. Simborg DW, Chadwick M, Whiting-O'Keefe QU, Tolchin SG, Kahn SA, Bergan ES. *Local area networks and the hospital.* Comput Biomed Res 1983; 16:247-59.
5. Stead WW, Hammond WE. *Computerized medical records: a new resource for*

clinical decision making. J Med Syst 1983;7:213-20.
6. Bakker AR. *The development of an integrated and co-operative hospital information system.* Med Inf 1984; 9:135-42.
7. Bleich HL, Beckley RF, Horowitz GL, et al. *Clinical computing in a teaching hospital.* N Engl J Med 1985; 312:756-64.
8. Whiting O'Keefe QE, Whiting A, Henke J. *The STOR clinical information system.* MD Comput 1988; 5 (5): 8-21.
9. Scherrer JR, Baud R, Brisebarre A, et al. *A hospital information system in continuous operation and expansion.* In : Orthner HF, Blum BI, eds. Implementing health care information systems. New York: Springer-Verlag. 1989; 100-22.
10. Rougé A, Baud RH, Scherrer JR. *An integrated network of telexprinters within DIOGENE hospital information system.* In: van Bemmel JH, Ball MJ, Wigertz O, eds. MEDINFO 83 : Proceedings of the Fourth Conference on Medical Informatics. Amsterdam: North-Holland. 1983; 1159-62.
11. Rougé A, Baud RH, Brisebarre A, de Roulet D, Scherrer JR. *DIOGENE 2: A federated Hospital Information System for the 90's.* In: Barber B, Cao D, Qin D, Wagner G, eds. MEDINFO 89 : Proceedings of the Sixth Conference on Medical Informatics. Amsterdam : North -Holland. 1989; 326-9.
12. Tolchin SG. *Overview of an architectural approach to the development of the Johns Hopkins Hospital distributed clinical information system.* J Med Syst 1986; 10(4):321-38.
13. European Union. Industrial Working Groups for the Fourth Framework Programme - Software Technologies : Distributed Systems Technology and Practice. Draft Report (version 4.1), 17-18 February 1994, 8-9 March 1994.
14. Commission of the European Communities: *Growth, Competitiveness, Employment : the challenges and ways forward into the 21st Century (White Paper).* Brussels 5 December 1993.
15. Lindberg DAB, *Global information infrastructure.* Int J Bio-Med Comp 1994; 34:13-19.
16. Ferrara FM. *The Role of Methodologies and conceptual models in the standardization of Medical Information.* Timmers T. Blum BI. eds. *Software engineering in medical informatics: Proceedings of the IMIA working conference in Amsterdam.* 1991; Elsevier Science Publishers B.V. (North-Holland). 255-266.
17. Weinhall TD, Raveh YA. *Managing growing organizations. A new approach.* John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore 1983; 172-94.
18. Safran C, Porter D, Lightfoot J, Rury CD, Underhill LH, Bleich HL, Slack WV. *ClinQuery a system for on-line searching of data in a teaching hospital.* Ann Intern Med 1989; 111:751-6.
19. Baud RH, Scherrer JR, Coignard J, Lucas L, Degoulet P, Jean FC, Jaulent MC, Springub A, Engelmann U, Meinzer HP. *The concept of a Ward Information System.* In: Adlassnig KP, Grabner G, Bengtsson S, Hansen R, eds. Medical Informatics Europe 1991: Proceedings of the European conference on Medical Informatics, Lecture Notes in Medical Informatics. Springer Verlag, Berlin, Heidelberg, New York. 1991; 81-5.
20. Degoulet P, Coignard J, Jean FC, Jaulent MC, Lucas L, Ben Saïd M, Meinzer HP, Engelmann U, Springub A, Baud R, Scherrer JR. *The HELIOS European project on Software Engineering. In : Software Engineering in Medical Informatics.* Timmers T, Blum BI, eds. Elsevier Science Publishers B.V. (North-Holland) Amsterdam 1991; 125-137.
21. Foy JL. *Integrating non-proprietary work stations into proprietary systems.* Int J Bio-Med Comp 1994; 34:277-283.
22. Jean FC, Lavril M, Lemaitre D, Sauquet D, Degoulet P. *A software engineering approach for medical workstation development.* Int J Bio-Med Comp 1994; 34 : 249-260.
23. Collen MF. : *Implementing Hospital Computer Systems, General Requirements. In: Hospital Computer Systems.* New York, Wiley & Sons 1974; 3-23.
24. Bakker AR, Ball MJ, Scherrer JR and Willems JL. *Towards New Hospital Information Systems.* In : IFIP-IMIA Working Group 10 Nijmegen Conference, North-Holland, 1988.
25. Bergeron BP, Greenes RA. *Intelligent Visual Input : A graphical Method for rapid entry of patient-specific data.* In : Proceedings of the eleventh annual Symposium on Computer Applications in Medical Care, November 1-4. 1987, Washington D.C., W.W. Stead eds. IEEE Computer Society Press: 281-286.
26. Cimino JJ, Barnett GO. *The physician's workstation : Recording a physical examination using a controlled vocabulary.* In : Proceedings of the eleventh annual Symposium on Computer Applications in Medical Care, November 1-4, 1987, Washington D.C., Stead WW. eds. IEEE Computer Society Press: 287-291.
27. Degoulet P, Devriès C, Sauquet D. *Data and knowledge integration in hospital information systems.* In : Towards new hospital information systems. Bakker AR, Ball MJ, Scherrer JR Willems JL. eds. Amsterdam : North-Holland. 1988; 151-156.
28. Pryor TA. *The help medical record system.* Clinical Computing. 1988; 5(5): 22-33.
29. McDonald CJ, Blevins L, Tierney WM, Martin DK. *The Regenstrief medical records.* Clinical Computing. 1988; 5(5): 34-47.
30. Stead WW, Hammond WE. *Computer-based medical records : the centerpiece of TMR.* Clinical Computing. 1988; 5(5): 48-61.
31. Tournade MF, Assicot M, Clement G, Piquet JC, Lion F, Da Silva JL, Rajaona J, Reversat JP, Hugon D, Auffret G, Flamant E. *L'expérience IGR en matière de gestion des dossiers médicaux.* Informatique et Santé. Springer-Verlag eds. 1989; 1:110-116.
32. Scherrer JR, Revillard C, Borst F, Berthoud M, Lovis C. *Medical office automation integrated into the distributed architecture of a Hospital Information System.* Meth Inf Med 1994; 33; 174-9.
33. Langer RD. *TALMRS ; Time-Affinity Linked Medical Record System: In: Proceedings of the eleventh annual symposium on Computer Applications in Medical Care.* November 1-4, 1987, Washington D.C., Stead WW. eds. : IEEE Computer Society Press. 271-274.
34. Ogura H. Yamamoto K, Furutani H, Kitazoe Y, Hirakawa M, Sagara E. *Online prescription order and prescription support in an integrated hospital information system.* In : Medical Informatics, An international journal of information processing in health care. 1985; 10 (4): 287-299.
35. Morel-Guillemaz AM, Baud R, Scherrer JR. Le problème de l'analyse temporelle dans les dossiers médicaux. In : Intelligence Artificielle et thérapeutique, Informatique et cancer, Informatique et "économie" hospitalières. Journées francophones d'informatique médicale de Montpellier, 14-16 mars 1989, ENSP Collection ACTES. 1989; 256-363.
36. Scherrer JR. *Developments in medical care and nursing care.* In: Bakker AR, Ehlers CT, Bryant JR, Hammond WE. eds. Hospital Information systems; Scope-Design-Architecture. Amsterdam: North-Holland Publ Comp 1992; 3-9.
37. O'Desky RI, Ball MJ, Ball EE. *Computers in health care for the 21st century.* Meth Inform Med 1990; 29: 158-61.
38. Baud RH, Rassinoux AM, Scherrer JR. *Natural language processing and semantical representation of medical texts.* Meth Inform Med 1992; 31: 117-25.

39. Sager N, Lyman M, Nhan NT, Tick LJ, Borst F, Scherrer JR, *Clinical knowledge bases from natural language patient documents*. In: Lun KC, Degoulet P, Piemme TE, Rienhoff O. eds. MEDINFO 92. Amsterdam: North-Holland Publ Comp 1992; 1375-81.
40. Rassinoux AM, Michel PA, Juge C, Baud R, Scherrer JR. *Natural Language Processing of medical texts with the HELIOS environment*. Comp Meth and Programs in Biomed, 45 Suppl. 1994; 579-596.
41. Engelmann U, Jean FC, Degoulet P, eds. *Supplement : The HELIOS Software Engineering Environment. Results of the European AIM project HELIOS-2*. Comp Meth and Programs in Biomed 1994; 45 suppl. 51-52.
42. Lovis C, Michel PA, Baud R, Scherrer JR. *Use of a conceptual semi-automatic ICD-9 encoding system in an hospital environment*. AIME 95. Artificial Intelligence in Medicine Europe. Accepted.
43. Côté RA, Rothwell DJ, Palotay JL, Beckett RS, Broche L. *SNOMED international, systematized nomenclature of medicine. Introduction*. College of American Pathologists, Northfield, Illinois, April 1993.
44. Chute CG, Atkin GE, Ihrke DM, *An empirical evaluation of concept capture by clinical classification*. In: Proceedings MEDINFO 92. Lun KC, Degoulet P, Piemme TE, Rienhoff O. eds. North-Holland, Amsterdam. 1992; 1469-1474.
45. World Health Organization International Classification of Diseases, Version 9. Geneva: WHO. 1980.
46. Hersh WR, Evans DA, Monarch IA, Lefferts RG, Handerson SK, Gorman PN. *Indexing Effectiveness of linguistic and non-linguistic Approaches to automated Indexing*. In : Proceedings MEDINFO 92. Lun KC, Degoulet P, Piemme TE, Rienhoff O. eds. North-Holland, Amsterdam. 1992; 1402-1408.
47. Rossi Mori A. *Model for representation of semantics in medicine*. CEN/TC251/ PT003 N34. 1992.
48. Rossi Mori A, Bernauer J, Pakarinen V, Rector AL, Robbè P, Ceusters W, Hurlen P, Ogonowski A, Olesen H. *Models for Representation of Terminologies and Coding Systems in Medicine*. In: Proceedings of Opportunities for European and US cooperation in standardization in health care informatics. 1992; Geneva.
49. Frütiger P. *Prospective encoding, transcoding and pragmatic representation of medical language*. In: Computerized Natural Medical Language Processing for Knowledge Engineering. Scherrer JR, Côté RA, Mandil SH. eds. Elsevier, Amsterdam. 1989; 83-94.
50. Lovis C, Michel PA, Baud R, Scherrer JR, *Word Segmentation Processing: a way to exponentially extend medical dictionaries*. To be published: MEDINFO 95.
51. Ratib O, Ligier Y, Scherrer JR. *Digital image management and communication in medicine*. Computerized Medical Imaging and Graphics. 1994; 18(2) : 73-84.
52. DO QH, Conti S, Munoz J, Pillou E, Tacchino M, Ratib O, Trayser G. *UNIMAGE, a new RIS for the DIOGENE 2 environment at Geneva Cantonal University Hospital*. Med Inform 1993; 18 (2) : 143-148.
53. Ratib O, Hoehn H, Girard C, Parisot C. *Papyrus 3.0 : DICOM - compatible file format*. Med Inform 1994; 19 (2): 171-178.
54. Ligier Y, Ratib O, Logean M, Girard C. *Osiris: A medical image-manipulation system*. MD Computing. 1994; 11(4) : 212-218.
55. Huang HK, Taira RK. *Infrastructure design of a picture archiving and communication system*. AJR. 1992; 158 : 743-749.
56. Osteaux M. *A second generation PACS concept*. European Congress of Radiology, Vienna. European Radiology. 1991; 1 -68.
57. Ratib O, Appel R, Scherrer JR. *PAPYRUS : A multimodality Image File Format for PACS and teleradiology*. Chicago: RSNA. 1990; 177 - 320.
58. Pun T, Hochstrasser DF, Appel RD, Funk M, Villars V, Pellegrini C. *Computerized classification of two - dimensional gel electrophoretograms by correspondence analysis and ascendant hierarchical clustering*. Applied and Theoretical Electrophoresis. 1988; 1:3-9.
59. Hochstrasser DF, Appel RD, Vargas R, Perrier R, Ravier F, Funk M, Pellegrini Ch, Müller AF, Scherrer JR. *Clinical Molecular Scanner : The Melanie Project*. Hawaii. Jan. 1991; Proc HICSS-24, Hawaii International Conference on System Sciences: 626-635.
60. Appel RD, Bairoch A, Hochstrasser DF. *A new generation retrieval tools for biologists: the example of the ExPASy WWW Server*. Trends in Biochemical Sciences. TIBS 19. 1994; 258-260.

Address of the first author:
Jean-Raoul Scherrer,
Hospital Informatics Centre,
Genèva Canton University Hospital,
24 Rue Micheli-du-Crest,
CH-1211 Genèva, Switzerland