Revista de Estudos Politécnicos
Polytechnical Studies Review
2006, Vol III, n.os 5/6, 073-085 ISSN: 1645-9911

MAID – Multi Agent for the Integration of Data

Vieira-Marques P.[•], Cruz-Correia R.[•;••], Costa P.[•], Palhares E.[•], Ferreira A.[•], Costa-Pereira A.[•;••]

pmarques@med.up.pt, rcorreia@med.up.pt, pcosta@med.up.pt, ernesto@med.up.pt, amlaf@med.up.pt, altamiro@med.up.pt

Abstract. Non-existence of a global overview or integration awareness leads to large and generalized heterogeneity of applications and technological approaches. Development of isolated systems has a serious and disruptive impact on daily practice and on clinical economic management. In this paper we describe a multi-agent system for the integration of heterogeneous clinical data sources and present the latest results originating from its use during the last year. The system aims to address several issues concerning the Portuguese healthcare information systems reality where it is common to find the lack of global strategy and planning related to the storage and management of patient's data. 814.000 reports were integrated from nine departmental applications during 2005. Around 500 doctors use the system per month. Our multi-agent approach has an independent and collaborative profile. It is designed to tackle, in a secure way, problems originating from existing systems heterogeneity and dynamic data production

Keywords: electronic health record; computer agents; integration and interoperability

[•] Biostatistical and Medical Informatics Department, Faculty of Medicine, University of Porto

[••] Center for Research in Health Information Systems and Technologies – CINTESIS, Faculty of Medicine of University of Porto

1. Introduction

Use of information and communication technologies (ICT) has been growing and spreading throughout all sectors of public service, healthcare institutions are no exception. In fact, due to the dissemination of ICT many hospital departments or even individual health professionals have acquired medical software or created their own computer databases in order to store and manage records containing relevant data from their patients. Unfortunately, due to the lack of a common strategy many of these information systems were not designed to be interoperable leading to inefficiency on current use and share of healthcare. Furthermore, the multiplication of these unarticulated systems generates redundant or contradictory data and the lack of standard terminologies or even the lack of single patient identifiers do not contribute for an easy integration making less viable a successful access to all available data.

Inefficient use of healthcare information is likely to be a major problem in large and complex health organizations. This is especially relevant when patient data, which is produced in heterogeneous environments, at various places and by different health professionals, needs to be available for authorized individuals at any point of care. This situation usually leads to huge costs associated with human and technical resources needed for non-automatic data collection, storage or integration. Also, the lack of an efficient information flow implies a delayed management of clinical report updates, mainly for some laboratory results, and an increased length of stay or delays in outpatient consultations and surgeries (Berg, 2001; Schmitt & Wofford, 2002).

Therefore, it is not surprising the growing demand to create integrated electronic patient records that would facilitate the communication process between health professionals (Lenz & Kuhn, 2001; Halamka & Osterland & Safran, 1999). Nevertheless, health institutions or individual health professionals may not be willing to give up their current stand-alone information systems, as they fear loosing the control over the data or loosing some system functions customized to their needs (Wyatt, 1995).

A virtual electronic patient record based on pre-existing information systems could help the integration process and facilitate the communication among them preventing loss of existent data or interfering with future software developments (Malamateniou & Vassilacopoulos, 2003). In order to build such a system it is necessary to deal not only with complexity present on the daily information flow but also with heterogeneous data sources, multiple communication technologies and implementation approaches. Multi agent paradigm together with interfacing technologies presents it self as a strong candidate to cope with the specificities of such scenario. It has powerful constructs for building self-sufficient autonomous, social and intelligent agents that can address complex distributed problems and to naturally map the actual work flow present on a healthcare institution regarding data collection and delivery (Weiss, 1999, Moreno A, Nealon, 2003). Some examples like iPointer , (Oliveira I. C. et al, 2002) and AIDA (Abelha, A et al, 2004) address this scenario using agent technology in order to integrate data from multiple heterogeneous sources. These systems aim to achieve integration at the data level. This approach can lead to more disruptive procedures and difficulties for their adoption. In order to avoid these obstacles we propose a system for the integration at document level.

2. Scenario

Hospital de S. João patient's information resources are based on SONHO a system developed by IGIF (Instituto de Gestão Informática e Financeira da Saúde) that manages data related to hospital inpatient stays, outpatient and emergency visits. Although this system has the capacity to manage some general clinical data (in a module called SAM – Sistema de Apoio ao Médico) it has reduced flexibility regarding the use of these data by health professionals and lacks specificity regarding the needs of different medical specialities. This may explain why SONHO is still mostly used for administrative (egg. consultations scheduling) and financial (egg. Diagnosis Related Groups) purposes rather than for the care of patients.

With no surprise, over the past ten years, nearly two thirds of its 24 major clinical departments have implemented or acquired at least one information system to record specific data focused on the daily management of their patients and on their own medical research interests. Unfortunately, these departmental information systems (DIS) where designed using different data structures, database management systems, ontologies, communication protocols, file formats for reporting clinical results and user authentication systems. In addition, many of these systems are connect to medical devices (egg. monitoring, imaging or lab instruments), which increase the overall complexity. The scenario includes unrelated applications managing relevant pieces of clinical information and duplicated data scattered over several databases. The vast majority of patient's data can only be shared using paper records, leading to an inefficient flow of information translated into high administrative costs both in time and staff.

In May 2003, the HSJ Directory Board asked the Department of Biostatistics and Medical Informatics to create a Virtual Electronic Patient Record (VEPR) system aiming at delivering the maximum information possible to health professionals through the integration of departmental information systems (DIS). Nine DIS were initially chosen. These systems contained many different types of data and had been implemented using various technologies.

3. General architecture

The architecture of this VEPR allows the collection, integration and availability of clinical reports providing an up-to-date overview of patient medical history at all points of care. Three major modules were designed (Figure 1): the Multi-Agent system for Integration of Data (MAID) module, which provides for automatic document retrieval, the clinical central information repository (CRep) and the visualization (VIZ) module, which shows patient data upon user request.

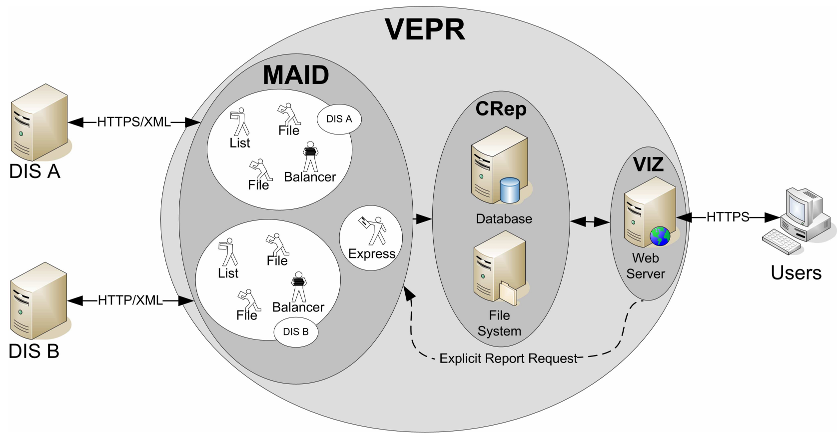

Figure 1 – General architecture for VEPR

As illustrated in Figure 1, MAID collects clinical reports from various hospital departments (e.g. DIS A and DIS B), stores them on a central CRep that consists of a database holding references to these clinical reports and a file system where reports are stored. After searching the database, VEPR users can access the integrated data of a particular patient through a web-based interface. When selecting a specific report, its content is downloaded from the central repository file system to the browser.

The VEPR system server runs on a Pentium 4 (1.8GHz), with 768Mb RAM and a Linux RedHat Fedora Core 1.0 operating system. The central repository file system, which contains the clinical reports files, is located on a HP StorageWorks SAN, which is mounted in the VEPR server using the NFS protocol. The database, which contains the patient's identification and references to the clinical records, is stored on two HP Server RP5740 RISC computer cluster running an Oracle database management system.

MAID module is built on JADE (Java Agent Development Framework) (Bellifemine & Poggi & Rimassa, 1999) as a multi-agent FIPA compliant development platform, for agents' management and deployment.

4. Multi-Agent System for the Integration of Data (MAID)

The design of MAID inherits characteristics from multi-agent systems paradigm as it embodies an independent, autonomous, social, scalable and reliable behaviour. Its main entities, agents, were designed to cooperate and undertake the necessary actions in order to build a virtual electronic patient record making existing information available within a reasonable time frame.

The automatic document retrieval process is achieved by two independent actions performed by a set of agents: Firstly, the retrieval of report references consists on the questioning of each DIS for new clinical reports and the eventual retrieval of their references (list agents); secondly, the retrieval of report files consists on the actual retrieval of the correspondent clinical reports files from the DIS to the central repository (balancer and file agents). A particular case of this last action correspond to the immediate retrieval of a clinical report that is not yet available centrally but has meanwhile been requested by a user (express agent).

The described actions are mapped into behaviours that define agent's individuality. These behaviours describe the necessary steps that an agent must undertake in order to accomplish its purpose. All agents act on an asynchronous and autonomous way without the need for human intervention.

Inorder to interface DIS, MAID agents make use of a set of plug-ins available for network communication (e.g. HTTP, HTTPS), database communication (e.g. Oracle,), logging and digital signing.

Visualization module

The integrated clinical information in the VEPR is available through a VIZ website residing in the Hospital intranet. After authentication, healthcare professionals can access patient data using several search methods (patient name or in-patient consultation and emergency numbers). After patient search a list of all reports collected in the past 24 hours is presented (Figure 2), helping users that are just looking for a recently requested report. The remaining clinical information can also be accessed, allowing medical doctors to choose from departmental source (DIS) or chronological oriented views. Due to the version control rules only the last version of the report is available to the user. In addition, a chronological bar is displayed following the clinical reports list, giving a perspective of past patient hospital events (Figure 2).

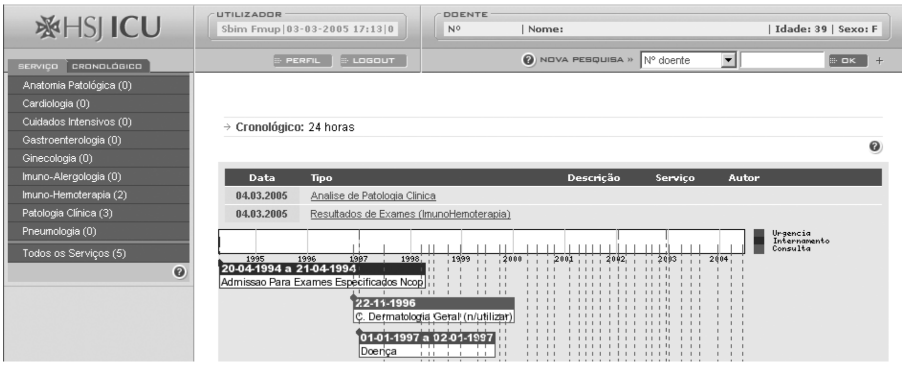

Figure 2- VIZ interface with chronological bar

Security issues

One of the main security issues is the trust put into the information withheld by the stored patient reports. Digital signatures are security mechanisms that provide the integrity of a report by enabling the detection of unauthorized modifications. If the digital signature does not match with report contents then this report is marked as not valid. Regarding availability equipment and power redundancy, backups and system monitoring were all put in place to guarantee availability of the system at all times.

The number of reports daily retrieved from each DIS is automatically compared to what is expected (estimated normal distribution values) and the number of sessions of different users is monitored. Any deviation from expected values triggers an alert message to the system administrator (Ferreira et al 2004).

By controlling access to information and protecting it whilst in transit along network communications infrastructures assures confidentiality. Access control policies were put in place as defined by the hospital administration after a proposal from an assigned committee defining roles and levels of access to VIZ. These policies were implemented using Role-Based Access Control (RBAC), an access control model used for large organizations (Ferreira et al 2005).

4.1. Report reference retrieval action

List agents

An individual list agent was assigned to each of the nine DIS. These agents regularly survey the assigned DIS looking for new clinical reports. For each of these reports a reference is obtained and stored on the CRep database. The list of references is retrieved exchanging XML messages with each DIS using network plug-ins.

Each list agent XML request is identified by an element composed by two attributes, a unique identification and a time interval. The first is used for message management and control while the second attribute is used for selecting clinical reports generated during a specific period of time. Depending on the nature of the problem, unsuccessful requests are stored for a later request (e.g. network unavailability) or audit action (e.g. database unavailability).

On the other hand, each DIS XML reply is composed by a management element (including the request unique identification) and a list element containing references to reports produced during the requested time interval. Each report reference element includes, among other attributes, patient identification, author, type and location (URL) of the report. List agents store these attributes on CRep database. List agents also update CRep database using patients' administrative data from SONHO.

The time interval of the last successful reply is stored for each DIS and used when MAID needs to be restarted (e.g. after a system crash). List agents assume parallel recovery behaviour, using the stored time interval to retrieve the list of reports references produced during inactivity of MAID.

4.2. Report file retrieval action

Balancer agents

To each of the nine DIS an individual balancer agents was assigned. These agents regularly survey the CRep database looking for new clinical reports references that need to be retrieved. When new references are available, a FIPA-Contract-Net negotiation protocol (FIPA, 2000) is initiated with a number of file agents in order to optimize work distribution. The negotiation currency is the current workload of each of the subordinated file agent. On protocol completion, the status of each report retrieval operation is returned to the balancer. If an error is reported a log entry is registered. Finally, the CRep database is updated by deleting the references of retrieved reports.

File agents

As described above each file agent is assigned to a specific balancer. File agents are designed to accept a list of report references after a negotiation protocol with the balancer. The list is processed sequentially as reports files are requested to the DIS.

For each single patient there is a directory on the CRep File System. The directory path is determined by splitting the last two pairs of digits of an internal patient identification number and concatenating the entire number in the end (eg. internal patient number 123456789 is stored at /67/89/123456789/). This operation guarantees a uniform distribution on the directory tree and easy maintenance operations.

After retrieval using available network plug-ins, the report is then digitally signed and stored.

Version control of the reports is performed using as fingerprint the SHA1 digest (NWGCS 1999) of the location attribute (URL). A new version of a particular report makes the older version unavailable for users, though securely preserved for auditing purposes.

Express agent

The express agent was designed to act in response to report unavailability on CRep file system. This happens when a report reference already present on the CRep database is requested by a VIZ user but hasn't already been downloaded by the report file retrieval action. This agent has a reactive nature to VIZ events. The evaluation of its activity may be an indicator of the system ability to anticipate user requests.

5. Results

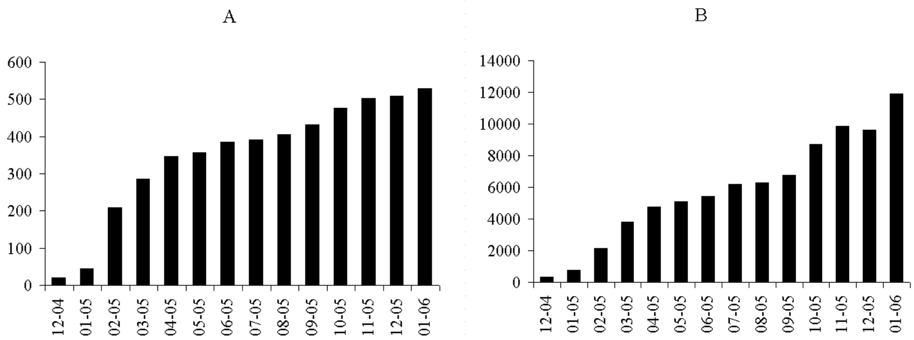

Figure 3 - Number of distinct users (A) and Number of sessions (B)

MAID is running since May 2004 and currently scans daily the nine DIS. At the present, documents are collected in HTML and PDF format. During the last year, nearly 814.000 clinical reports from 73.200 patients have been collected. On average, nearly 2.400 new documents are retrieved and stored each day. VIZ was made available for testing in October 2004 but only started to be known and routinely used since December 2004. The number of sessions grew from 315 in December 2004 to 11.927 in January 2006 (Figure 3). The number of different users using the VEPR grew in the same period from 21 to 530, respectively (Figure 3).

MAID provides within a reasonable time-frame, approximately 11 to 15 minutes, for report availability after its creation on DIS. On average it takes additional 10 minutes for reports download and storage. However if a user requests it first the download is immediate.

Figure 4 is a box-plot graph (created with SPSS 13) representing the variation of elapsed time between reference retrieval and user request on five different DIS. On Figure 5 it is possible to observe the hourly distribution of report creation by Clinical Pathology DIS. It contains results from two distinct weekdays (workday and weekend).

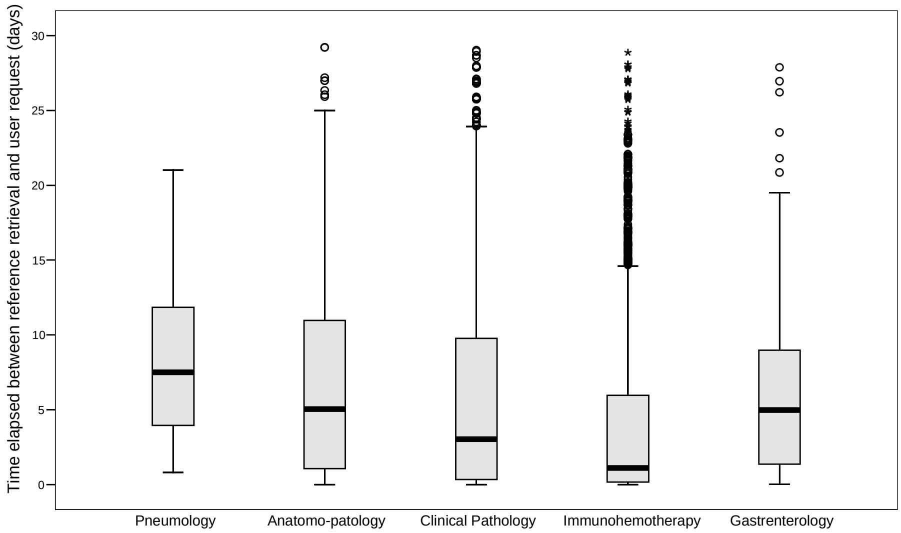

Figure 4 - Box-plot graph representing the variation of elapsed time between reference retrieval and user request on five different DIS (data collected from one year execution period)

Figure 5 - Average hourly report number distribution on Sundays and Mondays for Clinical Pathology Department (data collected from one year execution period)

6. Discussion

The session number growth assesses the interest on using an integrated visualization schema for clinical reports (Figure 3). Clinical practitioners can now have access to several types of reports within the same interface without the need for constant application interchange or relying on human intervention for report delivery. As additional DIS is incorporated on the VEPR a more complete overview will be available. The system has been widely accepted and even reports that are available trough local applications are being preferably visualized trough VEPR system.

The split phase execution of first collecting references and then downloading them emphasizes the priority of knowing about the existence of new patient information. The actual report is only important when a user requests it. As shown in Figure 4, the elapsed time between reference retrieval and report visualization varies according to different DIS. This reflects the distinct urgent nature of reports produced by each DIS. (e.g. Immunohemotherapy versus Gastrenterology). By using the hourly distribution of reports production (Figure 6) and visualization requests (Figure 4) it is possible to identify patterns that can be used to model future optimization procedures on agents' behaviours for reference retrieval and reports downloads which, in network or system constraint situations, can define scheduling priorities for each department or type of report.

From the architectural point of view MAID system was built without major changes to local DIS, it had no impact on department routine leading to less development costs and user adaptation effort.

Additionally, the integration of different DIS enabled the implementation of features that assess and ensure data quality. By cross checking several sources that provide for patient data identification it is possible to identify data inconsistencies. Automatic detection mechanisms were put in place and reports are generated with identified errors.

7. Future work

Clinical information is naturally distributed, not only in different departments but also between institutions. As patients visit several healthcare institutions their information records should be made available where needed independently of geographical displacement. Having this in mind further developments are planned for extending integration capabilities of MAID to other central information systems using mobile agents technology.

References

Abelha, A. M., Machado J., Alves, V,Neves, J.(2004) Health data management in the medical arena. In WSEAS Transactions on Computers. ISSN 1109-2750. 3:6.

Bellifemine F, Poggi A, Rimassa G. (1999) Jade a fipa-compliant agent framework. In *PAAM 99, 97–108*.

Berg M. (2001). Implementing information systems in health care organizations: myths and challenges. *Int J Med Inf, 64(2-3): 143–156*.

Ferreira A, Cruz-Correia R, Antunes L, Palhares E, Marques P, Costa P, Costa-Pereira A .(2004). Integrity for electronic patient record reports. CBMS'04; 4-9.

Ferreira A., Correia R., Antunes L., Oliveira-Palhares E., Farinha P., Costa-Pereira A. (2005). How to start modelling access control in a healthcare organization. Proceedings of the 10th International Symposium for Health Information Management Research.

FIPA (2000). Agent management specification. *Technical report, Foundation for Intelligent Physical Agents*.

Halamka JD, Osterland C, Safran C. (1999) Careweb, a web-based medical record for an integrated health care delivery system. *Int J Med Inform, 54(1): 1–8*.

Lenz R, Kuhn KA. (2001) Intranet meets hospital information systems: the solution to the integration problem? *Methods Inf Med, 40(2): 99–105*.

Malamateniou F, Vassilacopoulos G. (2003). An implementation of a virtual patient record using web services. *Stud Health Technol Inform, 95*.

Moreno A, Nealon J. (2003). Agent-based applications in health care. In Applications of Software Agent Technology in the Health Care Domain, Whitestein Series in Software Agent Technologies, 3–18.

NWGCS - Network working group of Cisco Systems (1999).US Secure Hash Algorithm 1 (SHA1). RFC 3174. Available at ftp://ftp-rfc-editor.org/in-notes/rfc2246.txt.

Oliveira I. C., Belo O.; Cunha J. P. (2002) iPOINTER: Um sistema multiagente para a integração de fontes de dados clínicos em ambientes hospitalares. Procedings CAPSI 2002.

Schmitt KF, Wofford DA. (2002). Financial analysis projects clear returns from electronic medical records. *Healthc Financ Manage, 56(1): 52–7*.

Weiss G. (1999). Multi-agent Systems – A modern approach to distributed Artificial Intelligence. MIT Press, Cambridge, Massachusetts.

Wyatt JC. (1995). Hospital information management: the need for clinical leadership. *BMJ, 311: 175–8*.