Non-invasive light-weight integration engine for building EHR from autonomous distributed systems

Pere Crespo Molina[1], Carlos Angulo Fernández, José A. Maldonado Segura, David Moner Cano, Montserrat Robles Viejo
BET Group, Technical University of Valencia, Spain

Abstract. Pangea-LE is a message oriented light-weight integration engine, allowing concurrent access to clinical information from disperse and heterogeneous data sources. The engine extracts the information and serves it to the requester client applications in a flexible XML format. This XML response message can be formatted on demand by the appropriate XSL (Extensible Stylesheet Language) transformation in order to fit client application needs. In this article we present a real use case sample where Pangea-LE collects and generates "on the fly" a structured view of all the patient clinical information available in a healthcare organisation. This information is presented to healthcare professionals in an EHR (Electronic Health Record) viewer Web application with patient search and EHR browsing capabilities. Implantation in a real environment has been a notable success due to the non-invasive method which extremely respects the existing information systems.
Keywords: Systems integration, health records, electronic patient records, medical records, medical records linkage, hospital records, EN13606, architecture and sharing of electronic health records.

1. Introduction

Healthcare is a very data-intensive sector, producing and consuming a great amount of data. In healthcare organisations, especially hospitals, the big amount of data gets increasingly obscure due to its widely decentralised organisation which allowed different departments to meet specific or local data requirements without the coordination and/or standardisation of information systems. This led to fragmented and heterogeneous data resources, which contains health data about patients, so called islands of information, making the access and aggregation of data across systems very difficult [1]. This situation has created a large gap between the potential and actual value of the information content of EHRs.
Closing this gap by making efficient use of the health data held by these systems, could improve significantly patient care, clinical efficiency and empower research activity. Having this in mind, it results difficult to imagine a healthcare organisation without any kind of information sharing among its information systems. Even though, many

[1] pedcremo@fis.upv.es

hospitals already have best-of-breed departmental information systems, only a few possess an integrate workstation, allowing health professionals to access to relevant patient's healthcare information in a single and unified view. One solution is the acquisition or development of large and centralised information systems where integration gets guaranteed, but specialty services loose their freedom to select the software that fit best their requirements. On the other side, we can find some integration systems which allow each clinical service to use the software they need while getting data integrated all along the organisation. Furthermore, the less invasive the integration system is, the less modification in the integrated subsystems must be done. Most common integration systems must perform routing mechanisms, manage system events and queues to emulate point to point data sharing, which make them too heavy for the only purpose of getting a unified view of data; we need a lighter solution.

We present here an overview of the Pangea-LE system, emphasising its flexibility and quick organisation deployment process. Briefly, Pangea-LE is a data integration system that provides an integrated and global view over distributed health data sources. In Pangea-LE, the global view is easily customisable to different user groups whose needs may change over time. On the other hand, data source adapters are as general as possible. In an evolving environment as healthcare, the local databases may change often. The databases are designed and maintained to meet local needs, and changes are almost made independently of the integrated global view. Applications connected to Pangea-LE global data view do not need to be updated after any local data source change. These features make the Pangea-LE system a very valuable solution to achieve fast, easy and secure data sharing and integration.

2. Materials and Methods

Pangea-LE acts like a piece of glue or mediator [2] between existing health data repositories in an organisation and health professionals. It allows the definition and management of a global, integrated and structured view of all the clinical records stored for a patient in an organisation. This view is presented to the professionals through an EHR Web application. Information views are created on the fly and are role dependant so that enterprise-wide access roles can be defined for different professional profiles with specific access rights levels corresponding to different EHR views (clinicians, nurse, management, administrative, etc.)
We must highlight the purpose of this concrete use case is only to display EHR views. Information is presented in human readable way and the clinician must provide his/her particular interpretation and meaning. Nevertheless, it is worthy to notice that it is also possible to keep the original meaning of shared EHR extracts for machine

interpretation purposes, guided by some EHR standards. This use case has been tested only for experimental purposes according to the European norm CEN EN13606 [3][4].
Pangea-LE can be classified as a generic middleware that integrates clinical information. It is important to notice that there are only a few basic requirements to be accomplished before Pangea-LE could be deployed into a healthcare organisation:

- ✓Unique Patient Identification throughout all the organisation local information subsystems, for instance, in the form of one central or distributed Master Patient Index. In case there is none, the organisation must provide a method in order to solve patient identification conflicts.
- ✓Organisation-wide user authentication and role assignment. This task is commonly executed by means of a directory service such a LDAP (Light-weight Directory Access Protocol) [5].

Every subsystem involved in an integration project is a potential container for a set of clinical concepts. Pangea-LE only offers read-only views. Content of local databases is maintained autonomously and locally; therefore, information quality must be assured by the original sources.
Pangea-LE architecture has four basic components:

1. Adapters: Heterogeneous data source access is reached through a set of JDBC (Java Database Connectivity) drivers for commercial and/or specific databases and through a set of configuration files to parameterise each integrated data source. The majority of data sources come from relational universe and can be accessed using one of the currently 221 different JDBC drivers available [6]. Other types of sources, like file systems, ftp, XML [7], messaging systems or even a Web Service JDBC Wrapper can also be defined and accessed in a homogenous way.

2. EHR Extract Definitions (EED) are the primary components in Pangea-LE since they describe the clinical concepts that can be shared among the different subsystems involved in an integration project. Neither whole system nor finest atomic data but only contextualised data, containing clinical information, should be shared through Pangea-LE. This selected information is organised and specified in a clinical EED entity. Several EED's can be defined for one subsystem, each one representing a different clinical concept. Each EED can only be shared as a whole. In other words, the minimum unit of information that can be shared between two subsystems in Pangea-LE is generated by an EED entity. EED entities are defined using XML files where some fixed descriptor elements define their behaviour:

- ✓Elements that specify a valid adapter configuration for accessing the information that the EED conforms.
- ✓Elements specifying data to be extracted from the data source.

- ✓Input parameters accepted by query processor to execute filters on data sources.
- ✓Valid calls (input parameter combinations which are allowed in order to build an EHR Extract from EED).
- ✓Elements specifying data pre-processing which allows source atomic data to be combined or transformed before XML generation.
- ✓Elements that describe the labelling and nesting format that constitutes the resulting XML document, etc.

EED entities can be instantiated by a Pangea-LE exposed Web Service (WS) [8]. This WS accepts XML petition messages pointing to the desired EED together with the required parameters (e.g. patient identification number). When this petition arrives to Pangea-LE and it is checked as a valid call, Query Processing Module (see Figure 1) can construct the appropriate SQL statements to fill with data the EED entity. This process involves obtaining a set of data and building meantime a XML response conforming the EED labelling and nesting rules defined [9] in the EED. Finally, if the petition requires XSL transformation, this is applied and the response is sent back to the requester application.

3. XSLT Transformations: Each EED can specify one or more XSL transformation files to be applied to the response XML message. This method allows Pangea-LE to adapt the output to different application message formats or devices (PDAs, Tablet PCs, etc.). Moreover, the same clinical entity can be formatted in different styles according to specific user access roles. Transformations can be applied in the server side or alternatively in the client application if it is enabled to perform this task. A data instance of an EED entity with no transformation applied is obtained by default.

4. EHR Browsing Trees: One of the most remarkable Pangea-LE features is the ability to organise the retrieved clinical information in a very flexible data tree structure. Once information has been extracted, different views can be configured to organise EHR in a clinical history manner (ordered by date), or according to information origin (emergencies, consults, explorations), etc. An EHR tree is also an XML document that defines how to structure, organise, aggregate, summarise and guide the extraction process of the clinical information associated to one patient. At run time, each EHR tree constitutes a health history view for a particular role and is mainly composed using attached EED entities. When one user accesses to the clinical information of a patient, the module that interprets EHR trees retrieves firstly only the minimum information needed to shape a summarised view of the patient's health history. Subsequently, user can interact with each particular tree node to get a more detailed view of the particular clinical subject described by the linked EED entity. All these possible interactions are

defined and also controlled by the EHR trees. Each structural component of the EHR tree data model corresponds to nodes in the visual EHR tree control used in the EHR viewer Web application. Thus, EHR tree nodes are user interactive and EHR can be build on demand. EHR browsing trees are designed for visualization purposes only and it is possible to define a different EHR tree for each different access role.

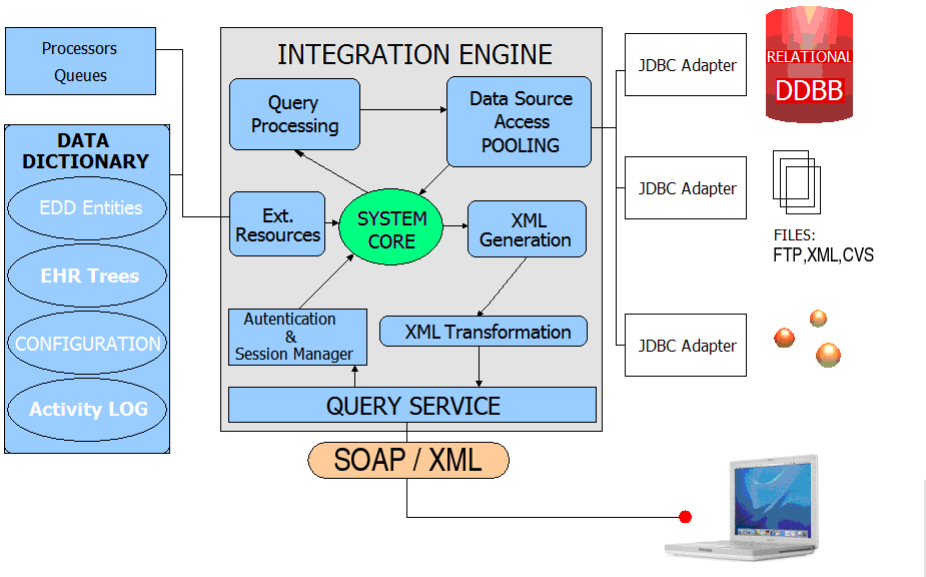

Figure 1 - Pangea-LE Architecture

Pangea-LE system core has been developed using technologies from a wide variety of free software components [10]: Eclipse as development framework, Ant for compiling and packaging, JDBC drivers for relational database source access, Struts framework and Prizetags for EHR Web viewer, Xerces, Xalan and Jdom libraries for XML parsing and transformation, Tomcat as application container, AXIS for Web Service development, JMeter for performance and functional testing, Log4j and HSQLDB for log support. The whole system has been developed in Java, so it is multiplatform. Two different clients have been Java developed; one version as a swing desktop application and the second as a Web version; both sharing the same capabilities.

3. Results

Pangea-LE deployment in a real environment such as CHGUV (General University Hospital of Valencia) has been a very worthy experience. Currently, CHGUV has 592 beds, 21 surgeries, 470 doctors and serves a population of 350.000 inhabitants.

The deployment process itself has been progressive and strictly coordinated. The most important systems have been the first to be integrated. At the moment, the majority of patient information is available electronically through Pangea-LE, including: Alerts, emergencies, inpatient episodes, outpatient consultations, laboratory results, biopsy and cytology study results, magnetic resonance reports, mammography, endoscopies and most of the discharge summaries from specialty services.

However, there are still some information sources that have not been engaged with Pangea-LE because of their lack of patient identification conformity.

A multidisciplinary work team has been created in order to take integration project to success. The team is composed of representative specialists from clinician, management, clinical documentation and informatics departments. They are in charge of the coordination of resources, agenda and priorities during the project deployment time. Easy access, completeness and immediate retrieving of information have made EHR viewer application rapidly extends organisation-wide. One year after implantation, its number of users is 472 (approximately the half of users in the organisation), 180 of them access daily to the EHR viewer. At present, more than twenty local systems have been integrated and 36 EED have been defined. The system has served, on average, 12000 petitions per month.

Most significant use of EHR viewer occurs during patient encounters when clinicians have a quick access to past patient encounters and explorations information, but not only Pangea-LE is helping in healthcare supplying but also in clinical documentation department where some discharge report must be processed to accomplish legal requirements: Diagnostics must be codified in ICD-x system.

4. **Discussion**

The use of Pangea-LE in CHGUV is restricted to an EHR viewer, that is, to extract clinical information regarding one patient by means of the defined EED entities. Current capabilities do not meet some interoperability requirements which might be satisfied in the future either by improving existing capabilities or adding new ones:

- ✓ Asynchronous communication mechanisms are needed to support petition / subscription methods, event control and process state management. Also flexible mechanisms for message transformation, scheduling and routing between integrated systems are needed as a basis requirement in order to satisfy clinical processes natural workflow.
- ✓ Pangea-LE's infrastructure may be a helping tool in ETL (Extract, Transform and Load) processes required in clinical / economical research. Pangea-LE may ease the always laborious load stage of data warehouses / data marts.
- ✓ Due to the wide range of clinical information that it manages and its easy use, the EHR viewer has widely extended throughout the organisation leaving the door open

to become the star dashboard application. It also may become a virtual medical desktop used by health professionals to access to his/her particular set of applications

- ✓ Extending Pangea-LE for inter-organisation communication is not a trivial issue since the Service Oriented Architecture must be yet defined and fine tuned. A restricted solution is being designed based on patient identification uniqueness zones and a hierarchy of server nodes, each one as an EHR zone manager.
- ✓ Nowadays, the more widely extended strategy on enterprise application integration is the Enterprise Server Bus (ESB), as a logical Bus where integrated systems are plugged and applications interchange messages. Primary resolved issue with ESB is the number of point to point application connections to make them interoperable. Pangea-LE is away from the functionalities of an ESB by far, since it does not implement routing mechanisms, asynchronous messaging, etc. Despite that, it is a flexible system that can progressively incorporate these functionalities. Anyway, we must focus on the nature of the processes in healthcare organisations, and note that, because of complexity reasons, they significantly differ from conventional organisations, for which ESBs were thought.

5. Conclusion

Pangea-LE allows a non-invasive integration mechanism, completely respectful with the already existing autonomous subsystems of a health organisation. Pangea-LE offers healthcare professionals an easy and secure way to access all the distributed patient information available in the organisation information subsystems through a single user interface using standard formats for information transmission.

High flexibility, fully scalable and without almost any preliminary requirement, Pangea-LE allows fast EHR deployment throughout a whole organisation. Our experience in CHGUV has shown that an integration project full specification can be done in only a few weeks, in fact, the main part of it was only a matter of a few days.

Acknowledgements

The authors would like to thank CHGUV for their always willing for help. This work was partially funded by the Spanish Ministry of Science and Technology (MEC – TSI2004-06475-102-01)

References

[1] Bouguettaya, A., Benatallah, B., Elmagarmid, A. *Interconnecting heterogeneous information systems.* Kluwer Academic Publishers (1998).

[2] Wiederhold, G. Mediators in the Architecture of Future Information Systems. Computer 25(3), 38-49 (1992).

[3] Maldonado JA, Robles M, Crespo P. *Integration of distributed healthcare records: publishing legacy data as XML documents compliant with CEN/TC251 ENV13606.*Proceedings of the 16th IEEE Symposium on Computer Based Medical Systems. IEEE Computer Society Press, pp. 213-218 (2004).

[4] http://www.centc251.org (last visited July 27, 2005)

[5] LDAP RFC: http://www.ietf.org/rfc/rfc2251.txt (last visited December 15, 2005)

[6] JDBC driver's list. http://developers.sun.com/product/jdbc/drivers (last visited Dec. 15, 2005)

[7] http://www.w3.org/XML (last visited December 12, 2005)

[8] http://www.w3c.org/2002/ws (last visited November 29, 2005)

[9] http://www.hitsw.com/products_services/xml_platform/allora.html (last visited December 11, 2005)

[10] Apache software foundation. http://www.apache.org/ (last visited December 15, 2005)