7th International Conference on DEVELOPMENT AND APPLICATION SYSTEMS
Suceava, Romania, May 27 – 29, 2004

USING XML TECHNOLOGIES FOR INFORMATION INTEGRATION WITHIN AN E-ENTERPRISE

Sabin-Corneliu BURAGA[1], Marius CIOCA[2]
[1] *"Al. I. Cuza" University of Iaşi, Romania – busaco@infoiasi.ro*
[2] *"Lucian Blaga" University of Sibiu, Romania – marius.cioca@ulbsibiu.ro*

Abstract. *The paper proposes a high-level approach for the e-enterprise modelling, able to respond to the new performance criteria, to the consideration of the whole life cycle of a product, process orientation, etc. The proposed model is based on GERAM methodology, UML modelling language and XML-based technologies. At the implementation level, we propose different solutions based on open standards and software applications.*
Keywords: *e-enterprise, integration, XML, modelling.*

Introduction

Even with the actual advances in e-business field, in the present the enterprises still are human-centric. From the IT (Information Technology) point of view, in any enterprise there are certain "IT-oriented islands" – legacy applications, relational database management systems (RDBMS), client programs, electronic services, etc. – difficult to be integrated in a coherent manner. The integration of enterprise with its suppliers and customers or with entities from different geographical regions is more complex.

Due to the difficulties above mentioned, for a good integration of the business we must take into consideration the internal and external factors that act upon the organization. Certain internal factors include the regulations, partners, customers and suppliers of an enterprise. Among external factor we mention economical and technological conditions, social-cultural and political national conditions, or international conditions (especially those imposed by the European Union).

The purpose of our research is to create a flexible enterprise model [8, 9, 10], able to respond to the new performance criteria, to the consideration of the whole life cycle of a product, process orientation, etc.

The model is based on GERAM reference architecture. In the design phase, the considered model uses UML (Unified Modelling Language). The information exchanged by various software components of the enterprise is modelled by using different high-level RDF (Resource Description Framework) constructs, which can be easily transformed into UML diagrams [2]. Using this approach, our model is closer to the actual research activities on the Semantic Web [11].

At the implementation level, the main criteria to choose and use the most adequate technologies are: the reliability and the maturity, the price, the easiness in working with them and also the platform-independence [8]. Because the actual and future e-enterprise – including Enterprise Resource Planning (ERP), Human Resource Management (HRM), and Client Relationship Management (CRM) systems – are more and more orientated to Web, it had been opted the following open-source technologies [7]:

- *for the data storage and management*: MySQL or/and PostgreSQL relational database servers,
- *for the logic of the application*: PHP and/or Java languages and their respective programming environments,
- *for information exchanging and interoperability*: XML family of meta-languages,

The following sections of the paper will present the methodologies and technologies used for business integration within an e-enterprise system and our proposals regarding different

flexible and real solutions for the implementation of such a Web-based system.

Reference Architecture

We choose the *GERAM* architecture developed by IFAC/IFIP, starting from the evaluation of the most representative reference architectures regarding the enterprises integration (e.g., CIMOSA, GRAI/GIM and PERA) [9, 12].
GERAM architecture (see figure 1) offers a description of all the elements recommended by engineering and integration of the enterprise. The main goal is to establish a standard for collecting the instruments and the methods after which the enterprise will benefit to successfully sustain the designing of initial integration and the changing processes that can take place during the life cycle of enterprises operating.
GERAM do not impose the use of specific tools and methods, but it defines the criteria that have to be satisfied by such an ensemble of adopted technologies.
Another reason why it had been opted for the GERAM architecture is its adaptability. GERAM is the easiest method to be adapted in the context of the Romanian industrial requirements. On a 1 to 5 scale, GERAM is on the top (near PERA method) from the adaptability point of view [8].

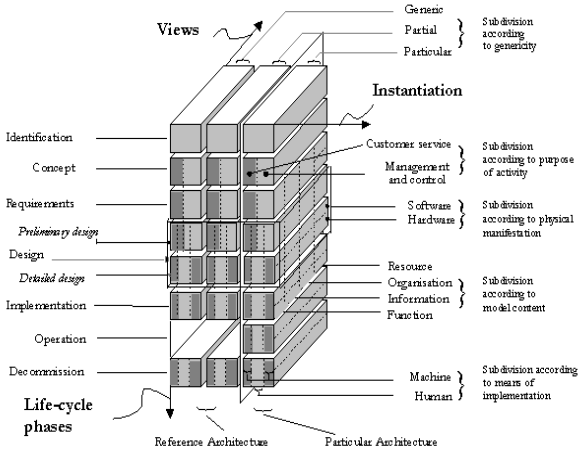

Figure 1. GERAM Architecture

Modelling Languages

In order to model the activities [12] (e.g., document flow) within an enterprise, we primary compared different modelling techniques, such as *CIMOSA (Computer Integrated Manufacturing Open System Architecture)*, *IDEF0 (Integrated Definition Language)*, *IEM (Integrated Enterprise Modelling)* and *UML (Unified Modelling Language)*.

We select UML [17], because this methodology is an object-oriented technology which model an enterprise in such a manner that is very close to reality. Enterprises deal with events and occurrences and all occurrences can be easily modelled as objects. There is a close relation between the real life occurrences and the objects in the model (often a one-to-one relationship). The semantic gap between reality and the model is a small one.
Object-oriented technology is also highly applicable to modelling organizations and their business processes. If the same technique is used to model a business as is used to build the supporting information system, the transition between the two activities will be easy and distinct.
UML is a language for specifying, visualization, building and documenting the software systems products, and also for modelling of enterprise or other non IT systems.
UML, among other facilities, offers a simple and expressive language of visual modelling, extensible and specialized mechanisms which allow the expressing of the base concepts and the independence from the actual programming languages and processes/methods of developing.
UML sustains a variety of upper level developing concepts like collaborations, frames, configurations and components. These concepts are very useful in the context of e-enterprise modelling.
Also, UML allows describing the various tasks or internal processes, which every business process consist of, as well as the way in which these internal processes interact to offer to a given actor a service or a product.

The UML has four distinguishing characteristics in comparison to other modelling languages: it is a general purpose language, broadly applicable, with good software support – in fact, UML is an industry standard.

XML Technologies

The *XML* (*Extensible Markup Language*) [3, 18] language is a recommendation of the World Wide Web Consortium for a meta-language to define mark-ups (annotations) for content publishing especially on the WWW space. The goal of the XML meta-language is to give some benefits not available in HTML, such as arbitrary extensions of a document's elements (tags) and their attributes, support for documents with complex structure, and validation of document structure with respect to an optional document-structure grammar, called a *DTD* (*Document Type Definition*). Also, instead of DTD, an object-oriented method for validation of XML documents can be used: an *XML Schema* [18]. The last manner of validation has many advantages in respect to UML.
As a standard recommended by the Web Consortium, XML is considered as the data format for information interchanging between various Internet and Web applications. The XML popularity is mainly due to its flexibility in the representation of many data types. The uses of mark-ups give to the XML language the possibility of self-description, and its extensible nature makes possible the definition of new document types, with a special destination (e.g. business rules, multimedia, data-flow etc.).
Using XML, the semantics and the structure of the data exchanged by diverse Web business applications is preserved. One of the key advantages is that the data can be organized as in an object-oriented database. As XML is format-independent, there is possible to generate multiple – XHTML, SMIL, WML or XUL – outputs effortlessly by transforming XML documents via *XSL* (*Extensible Stylesheet Language*) constructs [5]. Similarly to the *CSS* (*Cascading Style Sheets*), the XSL documents separate the content from representation.

Since 1998, XML has grown into a great family of standards integrating key technologies from three previously independent domains: documents, databases, and the Internet. Several examples of XML-based languages are [3, 18]:

- *MathML (Mathematics Markup Language),*
- *SMIL (Synchronized Multimedia Integration Language),*
- *RDF (Resource Description Framework),*
- *XUL (Extensible User-interface Language),*
- *WML (Wireless Markup Language),*
- *BRML (Business Rule Markup Language).*

In order to shift towards the Semantic Web [11], there were developed a series of XML-based languages specialized in the modelling of knowledge – for example, *RDF (Resource Description Framework)* and *OWL (Web Ontology Language)* [11, 18].
From our point of view, these languages could be the best solution to store and process various information about the products/services, customers and suppliers of an e-enterprise.

Modelling Enterprise Information with XML

In order to interchange and exchange information between various software components of the e-enterprise, we propose an XML-based approach. The e-enterprise must provide useful tools for management of the (semantic) information [8, 9].
We have to suggest a Web-based environment to help and assist the employers (especially the managerial staff) to create task-oriented information spaces from various (raw) collections of heterogeneous data.
There are main four stages in the information management cycle:

1. *information gathering* (using multiple heterogeneous sources);
2. *information analysis* (applying UML-related techniques to model the data components of the enterprise);
3. *information organization* (converting UML diagrams into collections of semantic data, by using XML/RDF assertions) – for details, see [2];

4. *information visualization* (using XSL transformations, XML documents of the enterprise can be easily manipulated and browsed on any platform – mainframe, desktop computer, mobile phone, wireless terminal, etc. – within the intranet or publicly on the Web) [5].

Also, XML language can be used for serialization [1] of the information interchanged by the e-enterprise's various components.

Implementation Solutions

For the implementation phase, we choose different open available technologies (see also figure 2), integrated into a flexible and modular manner.

For data storage, various relational database management systems can be used. We propose *MySQL* and *PostgreSQL* servers, for their good performances and support for different platforms and programming languages (e.g., Perl, PHP, C/C++ or Java). These servers are freely available on Linux and Windows platforms.

Different information – e.g., user preferences, configuration files, etc. – can be stored directly as XML documents or XML native databases. We propose *Apache XIndice* [14] as an XML native database management system. Other solutions to be considered are the *Tamino*, *eXists* or *Socrates XML* native database servers.

For the logic of the application, on the server side one of the best solutions is the use of the PHP [3, 4, 15] application server, because of its good connectivity to database systems and built-in facilities for XML processing, using both *DOM (Document Object Model)* and *SAX (Simple API for XML)* models. PHP offers a simple easy-to-use object-oriented interpreted language, similarly with C and Perl.

Using PHP, we can easily transform the information stored within databases into XML documents in order to transform it in client mark-up (e.g., HTML and XHTML for classical Web browsers, WML for wireless devices or other XML-based mark-ups for Web services).

Other approach can consider Java [13, 16] as the programming language and environment for developing the software components of the e-enterprise. Among important advantages we can mention independence of platform, object-oriented orientation and an excellent support for database and XML technologies (processing, Web services, multiple source integration etc.).

Also, the Web user-interface can be entirely built by using XML technologies – for more details, see [5].

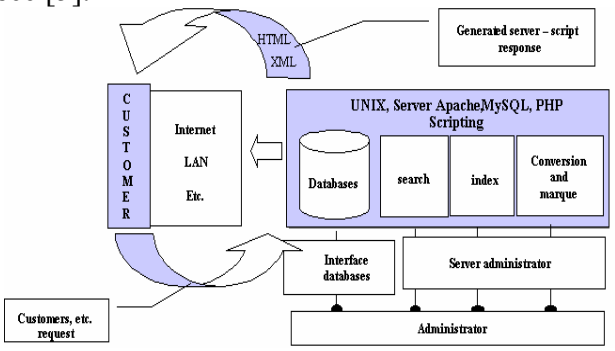

Figure 2. Modules of the e-enterprise

E-Enterprise Integration

In the realization of the model, we structured the activity into four levels of information [7], regarding the considered methodologies and techniques:

1. *System integration*;
2. *Application integration*;
3. *Business integration*;
4. *Enterprise integration*.

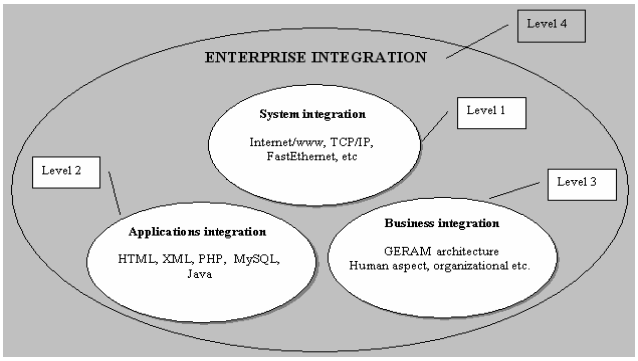

Figure 3. Levels of e-enterprise integration

Figure 3 presents these four levels of e-enterprise integration. The level 4 includes all precedent levels of integration.
More details can be found in [7].

The full process of modelling and data integrating for an e-enterprise is depicted in figure 4.
If we take into account the extranet of an e-enterprise, we must consider a supplemental level (the fifth) – *inter-enterprise integration*.

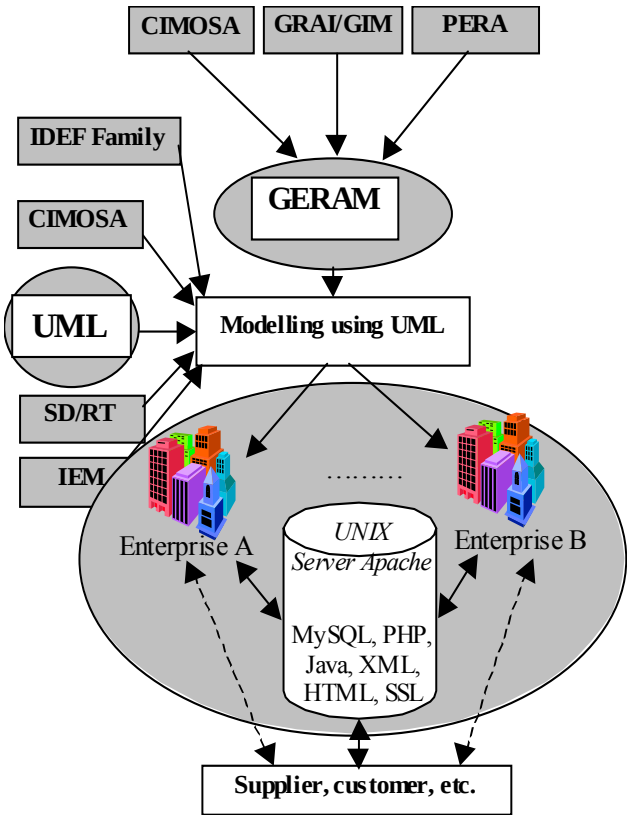

Figure 4. Enterprise integration using GERAM architecture, UML language and Web technologies

Conclusion

The paper proposed several methodologies and techniques to be used in the context of electronic enterprise integration and modelling the internal business processes.
As the reference architecture we presented a very suitable methodology from our point of view: GERAM architecture. The UML language was preferred as modelling language, because of its excellent capabilities in specifying, visualization, building and documenting the software systems products. Also, UML offers a fine support for object-oriented programming, a crucial facilities in the implementation stage.
The XML family of languages is proposed to be adopted for information exchanging and interoperability. XML is considered the best solution for information interchange between various modules of the e-enterprise system and for semantic representation of data (knowledge about products/services, employers, customers, suppliers, etc.).
At the implementation level, we suggest the use of PHP or Java environments. Their advantages (availability, support for object-oriented programming, database connectivity, XML processing, platform independence etc.) are very attractive for the programmers, in the context of a viable and flexible implementation. All proposed methodologies and technologies are open-source and can be used without restrictions.
Further studies will take into consideration the use of agent-oriented technologies, following our research presented in [1] and [6].

References

[1] Alboaie, S., Buraga, S.C., Alboaie, L. (2003) *An XML-based Serialization of Information Exchanged by Software Agents*, Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics – *SCI 2003*, Orlando, USA.

[2] K.Baclawski *et al.* (2001) *Extending UML to Support Ontology Engineering for the Semantic Web*, in Proceedings of the 4th International Conference on the Unified Modelling Language, LNCS, Springer-Verlag.

[3] Buraga, S. C. (2001) *Web Technologies* (in Romanian), Matrix Rom, Bucharest.

[4] Buraga, S.C. (ed.) (2003) *Web Applications. Case-studies in PHP* (in Romanian), Polirom, Iaşi.

[5] Buraga, S. C. (2003) *An XML-based Approach in Designing and Building of Web User-Interfaces*, The 6th International

Symposium on Economic Informatics (Digital Economy) – IE 2003 Proceedings, Inforec Press, Bucharest.
[6] Buraga, S.C., Alboaie, S., Alboaie, L. (2004) *An XML/RDF-based Proposal to Exchange Information within a Multi-Agent System*, Concurrent Information Processing and Computing – CIPC NATO Advanced Research Workshop – Post-Proceedings, D.Grigoraş *et al.* (eds.), IOS Press (to appear).
[7] Cioca, M., Buraga, S.C. (2003) *New Tools for Human Resource Management in e-Business: Combining UML Language, Reference Architectures and Web Programming*, *CD-ROM* Proceedings of IEEE INDIN'03, Canada.
[8] Cioca, M., Buraga, S.C. (2003) *Instruments and Web technologies for Implementing Architectures and Integration Informatics Systems in Virtual Enterprise*, Proceedings of the 3rd International Conference on Research And Development In Mechanical Industry – RaDMI 2003, Herceg Novi, Montenegro Adriatic.
[9] Cioca, M., Buraga, S.C. (2003) *Integration Methodologies of Enterprises in "e-Europe" Utilizing Reference Architectures, Modelling Languages and Web Technologies*, Proceedings of the 6th International MTeM Conference 2003, Cs.Gyenge (ed.), MTeM, Cluj, Romania.
[10] Cioca, M., Filip, F.G. (1998) *Standards Support for Integrated Enterprise Design and Control*. International Symposium on Concurrent Enterprising, Sinaia, Romania.
[11] Davies, J., Fensel, D., van Harmelen, F. (eds.) (2003), *Towards the Semantic Web*, John Wiley & Sons.
[12] Kosanke, K. (1997) *Comparison of enterprise modelling methodologies*. In J. Goosenaerts, F. Kimura, H. Wortmann (Eds.). Information Infrastructure Systems for Manufacturing. Chapman & Hall, London.
[13] Tanasă, Ş., Olaru, C., Andrei, Ş. (2003) *Java from 0 to Expert* (in Romanian), Polirom, Iaşi.
[14] * * * (2004) *Apache (XML Support)*: *http://xml.apache.org/*
[15] * * * (2004) *PHP: HyperText Preprocessor (PHP)*: *http://www.php.net/*
[16] * * * (2004) *Java*: *http://java.sun.com/*
[17] * * * (2004) *Unified Modelling Language*: *http://www.uml.org/*
[18] * * * (2004) *World-Wide Web Consortium's Technical Reports*: *http://www.w3.org/TR/*