Integrated Data Management
A Case Study in Heterogeneous Data Sources in Brazilian Government

Sergio Assis Rodrigues[1], Marlon Alves[1], Allan Freitas Girao[1], Ricardo Silva[1],
Miriam Chaves[2] and Jano Moreira de Souza[1]
[1]*COPPE – Computer Science Department, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil*
[2]*Ministry of Planning, Budget and Management, Brasilia, Brazil*

Keywords: Data Integration, Heterogeneous Data Sources, Web Services, Government Management.

Abstract: Typically, governments deal with huge masses of data generated by a variety of agencies at different hierarchical levels. Generally, this information is organized into different data sources resulting from legacy and heterogeneous systems. This lack of integration leads to management difficulties, since without a set of consolidated information; decisions are made lacking the proper grounding. Therefore, this work presents two initiatives of integrating data, based on service oriented architecture, from heterogeneous sources in the Brazilian Government aiming at better management and information usage.

1 INTRODUCTION

Government agencies maintain many databases that commonly contain data related to planning, budgeting, costs, work progress, and other relevant information. Generally, these data are spread over heterogeneous data sources like databases, files or other forms containing unstructured and semistructured data (Domenig and Dittrich, 2000). In modern e-government operations, all these pieces of information need to be consolidated for the appropriate management and evaluation.

However, this information generated by different government agency levels and stored in heterogeneous data sources makes management a challenging task. Even within one simple agency, a variety of databases are maintained by different administrative departments representing the same kind of data.

Depending on the data collection and representation done by a government agency, the values can be differently measured hindering the data consolidation and even leading to inconsistencies. The rapid changes taking place in globalized world and the evolution of technology has brought the need for an e-government development (Fei Ye et al., 2011).

The focus of e-government construction has changed gradually from just business ideology to the incorporation of information resources integrating and sharing ideias (Fei et al., 2011). The key point to the e-government success is how to realize this information sharing and systems integration, with an effective data exchange, through several heterogeneous data sources.

This article presents case studies about two initiatives within a Brazilian program to promote the data integration inside the government agencies. The first one deals with data that represent the agencies hierarchy and responsibility while, the second, with data related to the cost and progress of government works. The main goal of this study is to develop tools to achieve better information control and consolidation for management improvement.

2 BACKGROUND

2.1 e-Government and Integration Benefits

e-Government normally refers to the use of information and telecommunication technologies by government agencies to offer services to citizens and enterprises, focusing on information exchange.

This interaction between the government and the country stakeholders has become a relevant way of economic growth and human development (Rahman,

Assis Rodrigues S., Alves M., Freitas Girão A., Silva R., Chaves M. and Moreira de Souza J..
Integrated Data Management - A Case Study in Heterogeneous Data Sources in Brazilian Government.
DOI: 10.5220/0004557203160321
In *Proceedings of the 15th International Conference on Enterprise Information Systems* (ICEIS-2013), pages 316-321
ISBN: 978-989-8565-59-4

2007). This can be so impactful that the gap between countries that have access to information and communication technologies (ICT) and those which do not, was named global digital divide, and reflects great educational, cultural and income differences.

However, to achieve a management level of efficiency and flexibility as same as developed countries needs a great use of ICT's and a planned way for information exchange. The experience in developed countries shows that this is possible if governments decentralize their responsibilities and processes, in the same time they integrate their data for a consolidation management (PCIP, 2002). Therefore, in order to promote e-government, data integration and information exchange are considered key features for success.

2.2 Web Services

With the advent of internet and telecommunication technologies, the world has interconnected itself and dramatically increased the amount of information exchanged. In this scenario, web service is a technology used to facilitate the automated integration between distributed and/or heterogeneous systems, aiming to support the information exchange, collaboratively tasks execution and business processes interconnection (Benslimane et al., 2005).

Thus, the use of web services as a way to integrate applications and data is quite broadcasted in the world and has changed the way organizations project their systems and data. With standards used by large organizations, componentization and reuse features, favoring the cost reduction and rapid feature composition, web services are an integration technology from simple scenarios to highly complex cases.

2.3 Model Driven Integration Strategy

The development of web services was envisioned as the architectural solution for modularisation and integration both between and intra companies, promising that the service-oriented architecture (SOA) would offer easy integration by different and independent services over the internet. However, its construction ends up not being automatic and requires a large human effort (Brambilla et al., 2007).

To change this scenario, the semantic web services concept emerged, i.e., web services centered on the web semantic ideology, where one glimpses a web not only for humans but also designed for automatic interaction between machines. Thus, research (Brambilla et al., 2007; Bensaber and Malki, 2008) has sought ways in Software Engineering to raise the abstraction of these mechanisms' construction in order to make its creation faster and simple and facilitate their reuse.

In this context, the Model Driven Engineering theory aims at reusability, portability and interoperability through the separation of architectural concerns between the system specification and implementation. Thus, on this kind of approach, the focus is on the model creation based on industry standards such as XML and UML, representing the requirements, processes and information flow to be managed by an application.

In this paper, considering the great mass of information produced inside the Brazilian Government agencies, we looked up and used a methodology based on models just to contribute to building web services that increase the data availability and integration.

This methodology, called MDArte, is based on the MDA architecture and is supported by UML - Unified Modelling Language (OMG, 2003). From the use of a standard language such as UML scale is gained, spread and supported as there is a diverse set of tools for UML modelling.

Addressing the web service development issue, the first step is to establish the service principles, i.e., what it will provide; the procedures involved, the return patterns, as if a contract was being established. Figure 1 shows an example of modelling a web service for managing employees.

Figure 1: Developing web services using MDArte.

In the example, an operation is available to get an employee's list. For this specification, the web service is modelled as a class with "Service" and "WebSrv" stereotypes, where the first refers to the construction of a standard service used for code separation into a business layer, where the business rules reside, while the second tells the engine that this service should be available as a web service.

With this stereotype addition, the approach is in charge of building and inferring the needed components and dependencies, besides making its composition, organization and packaging, removing this responsibility and effort from the developer. Analogously, the web service operation parameters are also modelled with a "WebServiceData"

stereotype like Figure 2 shows.

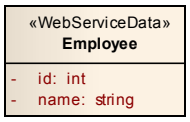

Figure 2: Developing WS parameters using MDArte.

This parameter information will be used by the web service operations, so that the approach can generate all artifacts needed for the service call and execution, like description, stub classes and etc.
On the other hand, to build a mechanism to consume an existent web service, the approach helps by simply modelling again a class using stereotypes. However, the stereotypes used are "Service" and "WebService-Client" as Figure 3 presents.

Figure 3: Configuring a web service client using MDArte.

Furthermore, through the tagged value named as *@andromda.web.service.wsdl.location*, it is specified where the service WSDL is that one wishes to create a client. From this, the MDArte is responsible for interpreting the service description and generate the classes and operations necessary to use the web service transparently to the developer.

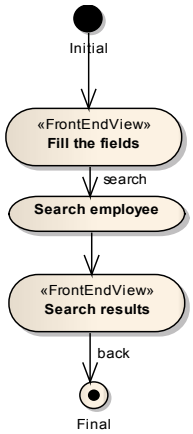

Figure 4: Modelling for generating pages in a web application with MDArte.

The MDArte approach extends its activities to building web applications in general, following the MVC architecture, being able to generate the entire structure of folders, classes, pages and libraries necessary for a web system implementation. For this, it takes use case diagrams and activity diagrams specifying the information flow through the system pages, as illustrated in Figure 4.

Again, stereotypes are used to specify behaviors, such as that used to specify activities that will become web pages. In this case, the stereotype "FrontEndView" represents that kind of behavior.

With the MDArte approach context, the next section presents two integration initiatives designed in the Brazilian Government context to support information management and exchange. Considering the size and complexity, the use of model driven web services with the MDArte's approach support aims accelerating and reducing team development effort on the integration mechanisms.

3 INTEGRATION INITIATIVES

This work seeks to improve the relationship between the government and the society that surrounds it. The integration techniques and tools described above were applied aiming to achieve a more efficient management of government information.

Noting the complexity of Brazilian government and the vast integration needs, this work focused on two initiatives in areas of large mass of data and relevant information to the government administration.

The first initiative deals with data on the definition, creation, extinction and competence of various bodies that comprise the Brazilian government. The latter, in turn, comes to tracking and monitoring the progress of government projects.

3.1 Ghelos Integration System

The publication of the decree 6944 of August 21, 2009, which institutionalizes the Organizational Information System of the Federal Government - SIORG, defined it as responsible for establishing information flows between agencies forming part of the Brazilian Government, aiming to support decision making processes, coordination of government activities and administer the registration of agencies and entities of the Federal Government.

This decree also established the SIORG as reference for other structural systems (Integrated Human Resource Management - SIAPE, Integrated Data Budgetary - SIDOR, Financial Management System of the Federal Government - SIAFI, Integrated General Services Administration - SIASG, Management Information System and Planning - SIGPLAN, Flights and Daily Award System - SCDP, Computer Information Resources

and Administration System - SISP and all systems of corporate use of the Federal Executive sphere which will be established) that need to use the registry of organs and administrative units.

In this situation of centralization and unification of the organizational units' registration and management in the Federal Government, the need for a single view between the SIORG's structural reference and the multiple structuring systems arose.

To achieve this goal, an integration system named GHELOS (Management System for Organs Harmonization) provides the possibility of a single view of systems to be integrated in order to establish the relationship of the various systems that are not mapped to the SIORG's organizational structure.

Using the MDArte technology, described in section 2.3, service oriented architecture (Figure 5) was developed and three layers can be detailed:

- **Integration Layer:** in this step, the information from the structure systems are evaluated and treated in a special apuration. Then, this data is loaded into the GHELOS harmonization database. At the same time, data from the SIORG structure system is loaded using a web service into the harmonization base;
- **Management Layer:** this module is responsible for consolidating all the organizational information from SIORG and the structure systems in a single view with their mapping;
- **Client Layer:** different stakeholders like agencies directors, politicians and government employees can consult the consolidated information and use them for decision making.

Agencies users will access the GHELOS system through a Web Application and obtain information about the mapped and unmapped organs through interfaces supported by the application.

Figure 6 illustrates the mapping bulletin functionality, where the GHELOS system shows the amount of mappings carried by the government organizational units. This information can be filtered by the government unit administration type, like direct and indirect or the power sphere, as executive, legislative and judiciary.

Other feature of this system is the mapping view where the organizational map between SIORG and the structure system is visualized aiding the stakeholder to relate the organs and unifying data.

Periodically, GHELOS sends notifications to government agencies representatives informing the

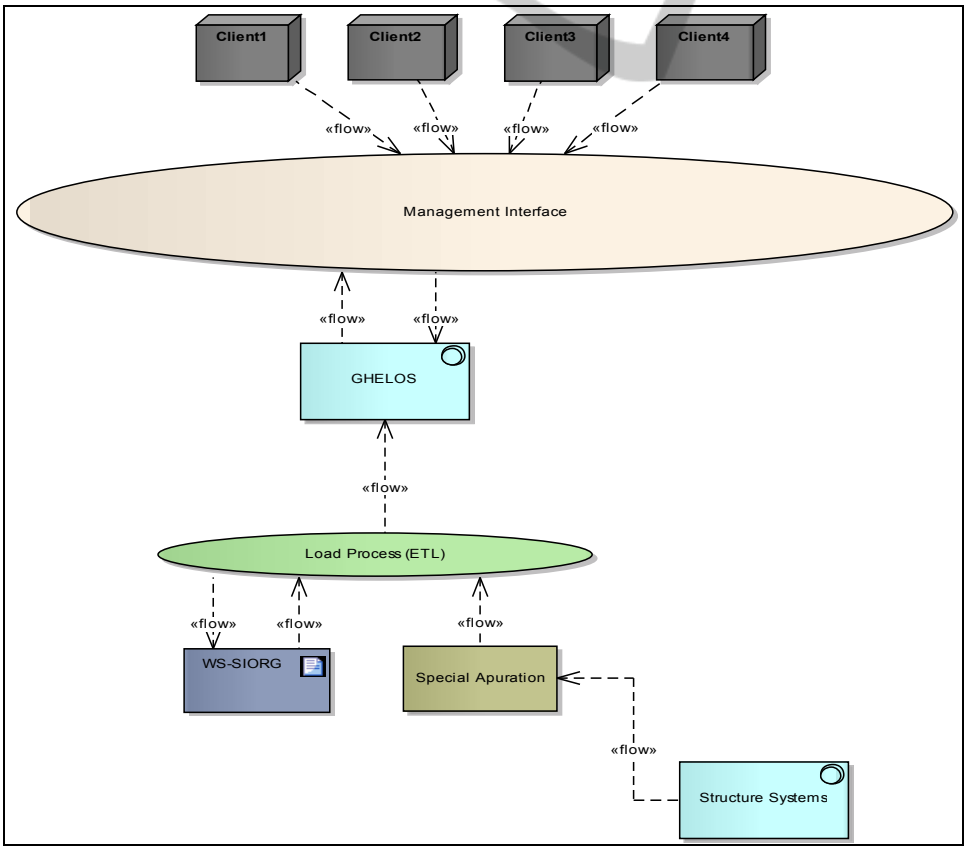

Figure 5: GHELOS architecture.

Órgão	Qtd. de Unidades	Qtd. de Mapeados	% Mapeamento
46 - Advocacia-Geral da União	710	8	1,1%
1930 - Ministério da Previdência Social	335	3	0,9%
304 - Ministério da Saúde	1136	22	1,9%
42672 - Ministério das Cidades	97	0	0,0%
3159 - Ministério das Comunicações	188	0	0,0%
263 - Ministério das Relações Exteriores	760	113	14,9%
2852 - Ministério de Minas e Energia	98	34	34,7%
17125 - Ministério do Desenvolvimento Agrário	183	57	31,1%
3162 - Ministério do Desenvolvimento, Indústria e Comércio Exterior	195	22	11,3%
1945 - Ministério do Desenvolvimento Social e Combate à Fome	313	21	6,7%
36670 - Ministério do Esporte	117	5	4,3%
1927 - Ministério do Meio Ambiente	191	1	0,5%
2981 - Ministério do Planejamento, Orçamento e Gestão	736	337	45,8%
2846 - Ministério dos Transportes	218	0	0,0%

Figure 6: GHELOS mapping bulletin functionality shows the integration level of adherence.

mapping of their respective units, to remember the importance of this mapping constant review for the correct integration and data quality.

All development made in the GHELOS system, using the MDArte approach, took 8 domain classes representing the necessary entities to model the integration map between SIORG and the structuring systems. The normal flow involves around 60000 organs in the SIORG structure (by web service) while the structure systems like SIAPE generate above 120000 entries (by ETL).

With GHELOS, Brazilian government organs obtained from the tool assistance in the integration and unification process determined by decree 6944. In addition, the system offered means of monitoring the entire integration progress to those responsible, in order to support decision making.

3.2 Helos Integration System

Currently, there are two systems called PACINTER (PAC Integration) and SISPAC (PAC System), which deal with the same projects, but have different identification (project code). These systems deal with government projects developed along all over the country, from bridges, roads to airports and housing.

The proposal of this integration system is that these systems are synchronized, but for this they need the mapping of their ventures. In order to provide this feature, HELOS (Venture Harmonisation System) was created. This application has as main objective to map the projects managed by the PACINTER base with the SISPAC base.

Currently, this scenario makes it difficult to analyse these projects progress for the Federal Government, which frustrate preventive measures on their administration to be taken.

The application consists of four software elements that implement its functionality. They are: the Web Service for mapping (WS-DeparaPACInterSISPAC), the PACINTER base Query Web Service (WS-EspelhoPACINTER), the interface screens, the Web Services orchestrator. The Web Service was developed by making use of the standard architecture MDA (Model Driven Architecture) with the MDArte Framework. The screens through which users interact with the application were developed using JSP and were incorporated into the process model defined on the service orchestration environment.

Agencies users will access all these integration services for mapping the projects through a Web Application (the HELOS system) and fill the details of the mapping process through the application interfaces.

Figure 7 shows the HELOS architecture illustrating the layers designed to support the integration desired for the project management between PACINTER and SISPAC. The clients use their preferred browser (in computer, laptop, phone, tablet, etc) to access the web application layer that, behind the scenes, uses a service orchestrator to collect the mapping between both legacy systems.

The integration system offers a functionality to observe the general mapping status per project type. It is a relevant mechanism to analyze the integration level acquired at that moment.

The main feature, otherwise, is the possibility to make the mapping between data from both systems. An expert of the project's information can associate them using the integration system and, all this information, is disseminated beyond all government organizational units.

There is also a management view to describe all the ventures that have not had a map yet. An agency director can take advantage of this information to demand the responsible for that data to realize that map.

HELOS development, using the MDArte approach

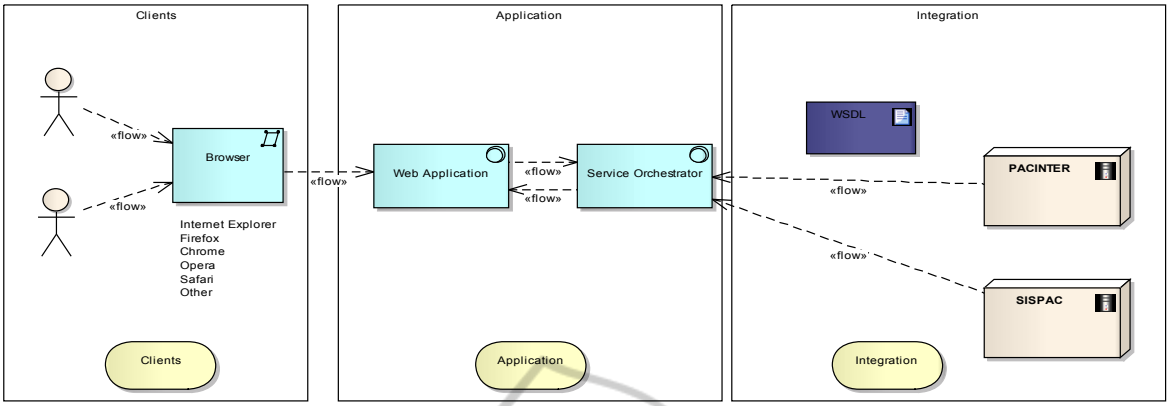

Figure 7: HELOS architecture.

approach, requires just the use of two web service clients that access the PACINTER base and the SISPAC base. While the SISPAC service already exists, the PACINTER web service was also developed using the MDArte tool leading to more than 58 domain classes that were exposed by a web service interface automatically generated through the model that contained their specification.

4 CONCLUSIONS

Brazilian Government has had several heterogeneous systems leading with the same kind of information which make consolidation and management analysis difficult. To improve this scenario, two preliminary initiatives were built to support better information environment with more integration and availability.

As any kind of integration initiative in a complex case, like the Brazilian one, leads to huge effort of development. Therefore, this paper presented a different approach based on a model driven architecture, where the focus is on the specification level to build integration systems that comprises the needs described with less effort.

Both initiatives, GHELOS and HELOS were developed with this technology and have proved to be a great tool for promoting integration in Government. This led the Brazilian Government to a better management environment where the consolidated information was available and allowed the stakeholders to make better decisions when required.

REFERENCES

Bensaber, D.A., Malki, M., 2008. Development of semantic web services: model driven approach. In NOTERE '08, *Proceedings of the 8th international conference on New technologies in distributed systems*, ACM New York, NY, USA, Article 40 , 11 pages.

Benslimane, D., Dustdar, S., Sheth, A.., 2008. Services Mashups: The New Generation of Web Applications. In *IEEE Internet Computing*, vol.12, no.5, pp. 13-15.

Brambilla, M., Ceri, S., Facca, F.M., Celino, I., Cerizza, D., Valle, E.D., 2007. Model-driven design and development of semantic Web service applications. In *ACM Transactions of Internet Technologies*, 8, 1, Article 3.

Domenig, R., Dittrich, K. R., 2000. A query based approach for integrating heterogeneous data sources. In CIKM '00, *Proceedings of the ninth international conference on Information and knowledge management* . ACM, New York, NY, USA, 453-460.

Fei Ye, Hao Li, Min Hu, 2011. The Construct of Data Integration Model of Heterogeneous E-Government System Based on Topic Maps. In *ICICTA, International Conference on Intelligent Computation Technology and Automation* , vol.1, pp. 263-266, 28-29 March.

Rahman H., 2007. E-government readiness: from the design table to the grass roots. In (ICEGOV '07), *Proceedings of the 1st international conference on Theory and practice of electronic governance*, Tomasz Janowski and Theresa A. Pardo (Eds.). ACM, New York, NY, USA, pp. 225-232.

Moore, R., Lopes, J., 1999. Paper templates. In *TEMPLATE'06, 1st International Conference on Template Production*. SciTePress.

PCIP, 2002. Roadmap for E-Government in the Developing World: 10 Questions E-Government Leaders should ask themselves, *The Working Group on E-Government in the Developing World*, *Pacific Council on International Policy*, Los Angeles, CA.

Smith, J., 1998. *The book*, The publishing company. London, 2nd edition.