

Composition of Distributed GI Services: Use case in Latin American utilities

Gustavo Zárrate
Utilities Solutions
Colombia – South America
e – mail: gzarrate@yahoo.com

Abstract

Distributed computing environments offer a great opportunity to find, consume and share data, functions, and services in several knowledge areas. Geographical Information Systems (GIS) represent one of these areas with potential applications in different domains such as Earth Sciences, Utilities Services and Cadastral Organizations. In this context it is possible to find Geographical Information (GI) components which are loosely coupled services that can be executed on the internet with an independent platform approach. GI components have a technological, functional and business value by themselves, but the real power for most situations comes from the possibility to assemble them in chain of services to achieve user's requirements and to fulfill their expectations; these chain of services - often called composite applications - behave as workflows of functions and data defined in an specific area of knowledge.

The composition process involves a number of important elements. First, a knowledge domain has to be included in order to have a common understanding of the process involved in the solution; this requirement is fulfilled with an ontology which is in charge of defining a common vocabulary in a domain area and the relationship between concepts. Description Logics, used as the framework to define an ontology, and the Web Ontology Language (OWL) through the Resource Description Framework (RDF) are commonly used to define and manage ontologies.

Another important required element in the composition process is service description: each GI component must have a description of its capabilities and interfaces in order to be used in a service chain. Organizations such as World Wide Web Consortium (W3C) have proposed standards for service descriptions; so far the Web Service Definition Language (WSDL) is used for this task; however WSDL defines only the format of messages to be sent and received by GI components: it only deals with component's syntactic scope.

Complementary, it is necessary to have a mechanism to combine GI components; some models have been proposed such as goal-driven ones (with OWL-S), automata-based models, and flowchart-based approaches. The Business Process Execution Language for Web Services (BPEL4WS or BPEL for short) is a XML standard proposed by main technology leaders to achieve the flowchart approach of component orchestration and is considered the most promising language to integrate multiple components in larger composite applications.

On the applied arena, Utilities companies in Latin America have a big social responsibility with the people, a big challenge with the investors, and big financial pressures from the governments through taxes and contributions for state plans. National organizations in charge of controlling utilities companies' management and technical quality levels such as

customer care, energy losses, service continuity, and financial feasibility, have reported that companies in Latin America are facing a challenging period to survive. Privatization plans, international competition, and deregulation laws have a big effect on these companies' possibilities as well.

Nevertheless this complex reality offers the possibility to surpass current shortcomings, increase value creation, find high technology solutions and apply new management proposals. Clients and their relationship with utilities companies together with network infrastructure for transmission and distribution are core elements of the service.

Key words:

Interoperability, business processes, semantics, WSDL, BPEL, ontology, service chaining, composition, execution.

Chapter 1. Introduction

Interoperability is one mythical issue that has requested and received several approaches to be solved. The first approaches went for standard formats and connectors between heterogeneous systems applications; this alternative solved some initial expectations, but its costs of maintenance and support were so high. Actually the proposals are dealing with the understanding between software components; in this context the components are self assembled in a dynamic way. The aim of all these proposals is to fulfill the user's requirements and business' expectations; finally all technological advances have to be business and user driven.

Service chaining is a way of organizing a set of functions and data sets in order to get a stronger solution; it is a synergetic approach: the combined interaction of two or more elements produces a stronger result than their individual efforts. There are two main proposals about composing: the business work flow approach with BPEL, which has support from the main technological companies, and the semantic web approach with OWL, which is a continuation of Artificial Intelligence (AI) and is still under research. The present study makes a proposal to join the strengths of these two approaches in the composition and execution stages of the service chaining. A prototype to join strengths is analyzed and designed; a practical application on energy and gas knowledge area applied in Latin America is presented as well.

Chapter 2. Interoperability and standards

Interoperability is the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units [OGCa, 2003].

Interoperability is the ability of a system or components of a system to provide information sharing and inter-application cooperative process control. As figure 1 shows, two systems X and Y can interoperate if X can send a request for service R to Y on mutual understanding of R by X and Y, and Y can return response S to X based on the mutual understanding of S. [Bishr, 1997].

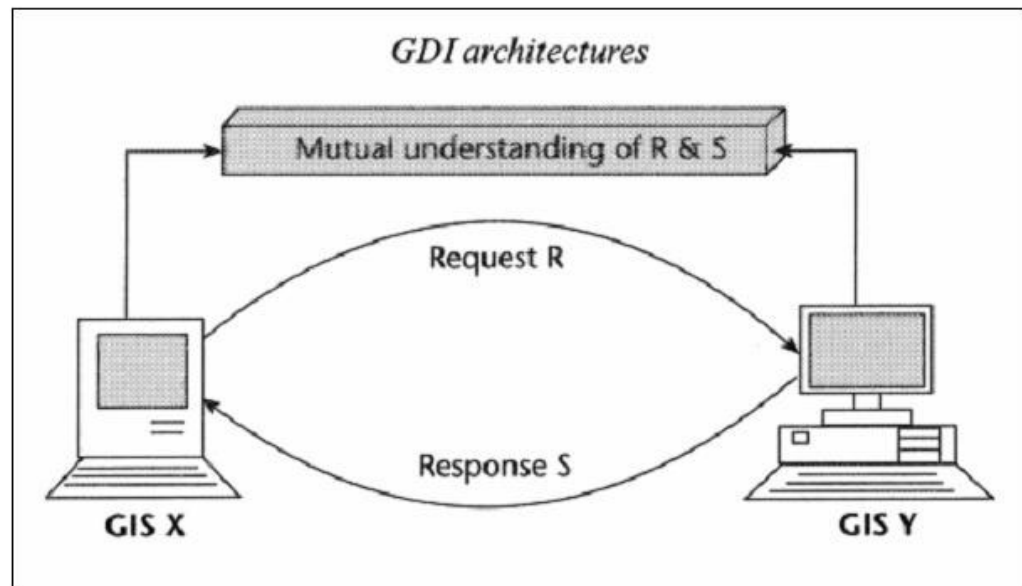


Figure 1. Interoperability requires mutual understanding of request and responses (From Bishr, 1997).

Standards facilitate the sharing of information and computer resources within and organization, and between organizations. At a practical level, the adoption and use of standards can save money and time. The value of wisely chosen standards for geospatial information users is reflected in three primary themes: portability, interoperability and maintainability [Croswell, 2003].

2.1. Models and data models

Models are abstractions of the real world. Different models are used to represent the views from the different stakeholders. While the users like to see the functions that they can perform with the system¹, the architect prefers to see the main components of the system, the designer want to see the detail inside of the system (subsystems, interfaces, classes), the testers look at the test cases and data sets to prove the system, and so on.

According to the Rational Unified Process², the architecture of a system is given by the following models: use-case model, analysis model, design model, deployment model, implementation model and test model. The Unified Modeling Language (UML) is a language designed for documenting and writing models. UML gives us a set of models to choose from. Selecting the right models is one of the main tasks of the project manager.

¹ This is called the use cases

² See more on the Rational Unified Process at: www.rational.com

In spite of the fact that all models have great relevance, the focus here is on the analysis models, including data models³. Data models are not the most exciting part of designing and implementing a software application, but they are one of the most essential. A formal model is an abstract and well defined system of concepts. A data model defines the vocabulary that can be used to describe and reason about things. A data model details how to take real world objects and make them understandable to a computer system [OGCb, 2003].

UML has been rapidly recognized as de facto standard in the information technology community and it is becoming widespread in the GIS community. Data models are usually done in workshops by a team composed by domain analyst (experts in the subject of discussion), and a group of skilled designers, who use modeling languages to document the results. In UML, data models are described using class diagrams. Although UML is being used for building the models, software vendors, government organizations, and industry organizations have defined many application-specific data models.

2.2. Geographic data models

All Geographic information Systems (GIS) are built using formal models that describe how things are located in space [ESRI, 1999]. In the geospatial world, the focus is on depicting things in the real world as points, lines and polygons (the geometry of the object) and their attributes (additional information about those objects). When linked together, a pair (geometry and attributes) representing one or more real world objects is called a feature.

Even a common type of geographic object can be represented in a GIS in a variety of ways. Our interaction with objects is diverse, and so we can model them in many ways. The same object, for example an airport, can be represented as:

- An physical area with a main building, hangers and runways
- A set of flight routes used by the aircrafts having flow direction, volume, air company and other attributes
- A group of people working on it, including passengers, flight attendants, flight crew, mechanics and other personnel involved in the service
- A set of aircrafts with different characteristics such as owner, brand, model and capacity

However, no model is better than other. The type of application, the context of the problem to be solved, and the detail of information to be captured, determines which model is more suitable. In the example above, if there is need to rebuild an area of the airport, the first model is more appropriate, while if a new flight company enters to the market, the second one might be more appropriate to measure the impact of the change. The geodata produced by a civil engineer and a market analyst is different, even if they use the same software, because they must define the features differently.

With many stakeholders involved in the creation of schemas, in different places and institutions working in the same field (or domain application), there are also differences in the way the models are structured. Two databases might adopt the same classification method and have different class and attribute structures. E.g. a particular feature may be classified under different object classes in different databases, or an object in one database can be

³ Also called: domain models, content models, conceptual models

considered an attribute in another. The classes, attributes and their relationships can vary within or across contexts [Bishr, 1997].

It also happens frequently that the field elements have different names for specifying the same thing [Trias, 2003]. For example in cadastral application domain the word <Deed> for one institution can have the same meaning than <Title> for other institution.

Heterogeneity problem occurs when different communities wanting to share their data with each other have to contend with different views of the real world features, different modeling schemas, and different tools to represent, store, process and manage geospatial data sets. Bishr (2002) summarized these heterogeneity issues as syntactic, schematic and semantic heterogeneity. Syntactic heterogeneity deals with differences in the thematic and the geometric representation as well as the topologic relationships of spatial objects. Schematic heterogeneity deals with differences in the class hierarchies and attributes structure of two independent database schemas. Semantic heterogeneity is the way the same real world entity may have several meanings in different databases. It occurs due to differences in context information.

The Open geodata model has promised to be a solution to these problems. The open geodata model is a comprehensive “universal” geodata model, an open programming interface that provides a basis for building interoperable interfaces between geomatics systems with different geodata models. Information communities, groups of geodata producers and users who share a set of geographic feature definitions, are working on them.

Chapter 3. Service Composition

Basic ideas about service composition are formulated by OGC while adopting the Open GIS Reference Model for Open Distributed Processing (RM-ODP): “The computational viewpoint is concerned with the functional decomposition of the system into a set of services that interact at interfaces. This viewpoint captures the details of these components and interfaces without regard to distribution.” [OGCa, 2003].

Current GI composition concepts have several shortcomings, but there are some interesting proposals and approaches in the general computational community. Nevertheless the common practice in real situations is to define service chains for each transaction type starting from nothing; this means that every time a new service chain has to be composed, people involved in the composition process do not take advantage of previous experiences with similar transaction type, both in their own organizations or available through internet. The concept of GI component has to be extended to include composite components which can be predefined service chains.

Previous experiences could have generated service chains which can be reused in these new situations, either as atomic components or complex sets of components. With this point of view the service description has to include not only isolated components available so far, but also descriptions of service compositions with clear scope of their functionality; these new complex components still have to be loosely coupled elements. This structure would generate a complete new set of reused components with great advantages of generating fast, reliable, and proven solutions founded on distributed services.

Service composition involves the discovery of services to be composed and the integration of these discovered services. One fundamental requirement for the services is to be loose coupled components. With this behavior it is possible to define a web service as a network resident software service accessible through standardized protocols; another working definition for web service is pieces of functionality (functions and data) accessible on internet via interfaces.

3.1. Ontological approach

The core concept of ontology and semantic web is the representation of knowledge in a machine and interoperable way. The current technique is the use of Resource Description Framework (RDF) which allows the representation of knowledge as set of subjects, predicates, and values.

The drawback of this approach is the complexity of generated semantic networks and the sub-optimal reasoning behavior. Description Logics offers an alternative restricting the meaningfulness, and the final efforts are generating ontology languages such as OWL and top-level ontologies such as OWL-S.

The current challenge is to get tools for editing OWL-S; research teams have developed some such as the plugin for Eclipse CODE from CMU (http://projects.semwebcentral.org/forum/forum.php?forum_id=227), OWL-S Editor from SMI, which is a plugin for Protégé Plugin (<http://owlseditor.semwebcentral.org/>), and OWL-S Editor from University of Malta (<http://staff.um.edu.mt/cabe2/supervising/undergraduate/owlseditFYP/OwlSEdit.html>). They offer an option to develop ontologies, and are still on initial versions, so their main obstacles are to install a stable release and to import or transfer pre-existence ontologies.

3.2. Workflow approach

The functionality of a web services is described with a specification in WSDL, that specifies the syntax of inputs and outputs; these messages are simple syntactic descriptions without a meaning [Srivastava, 2003]. The order of this type of messages has to be defined separately in a kind of workflow, which indicates the sequence, exceptions, decisions, and synchronism of the elements involved in the solution.

BPEL is cataloged as a current good approach to define this workflow and some software companies have released visual BPEL designer that facilitate the construction, edition and administration of service chains. One example is presented by Oracle with its BPEL Designer [Oracle, 2004a], which facilitates the development of SOA based applications by composing synchronous and asynchronous services into an end-to-end BPEL process flow.

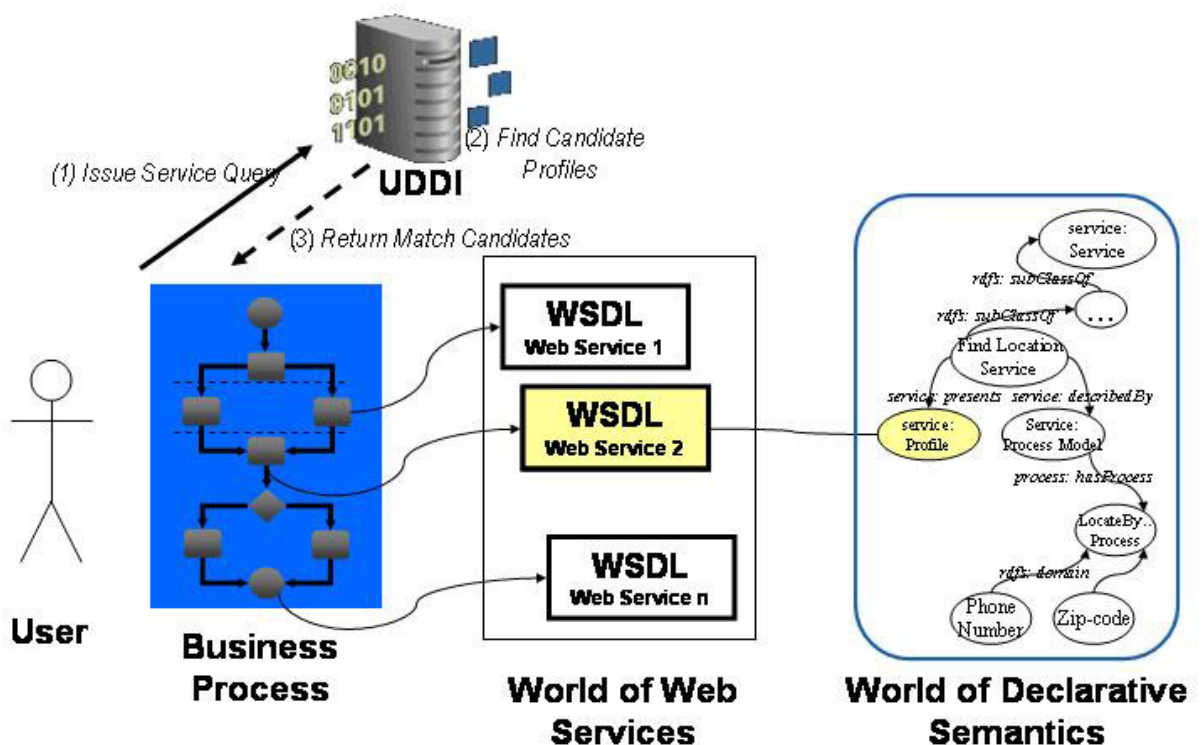
WSDL and BPEL are technical conventions which enable the exchange of information and procedures in a standard way. Nevertheless they do not tackle the semantics issue of the information; the agents in charge of processing the WSDL can deal with the messages defined as inputs and outputs, but they can not understand the meaning of the messages involved in the transaction.

Chapter 4. Combination of strengths

WSDL and BPEL solve the technical issue of web services: the syntactical understanding of components, including the input and output messages. The combination of the workflow and ontological approaches deals with the feasibility of extending the WSDL content in order to include a “link” to a related ontology; the goal is to include a semantic annotation in the WSDL code of the components used in the service chain.

Some authors have proposed the extension of WSDL definition [Martin, 2001], and have declared that WSDL is only related with grounding of the service, or their proposed extension generates no valid WSDL files. Nevertheless our proposal is to include a kind of pointer to ontology that includes the definition and meaning of components discovered to do the service chain. It is not the inclusion of the ontology in the WSDL description; ontology descriptions are powerful by themselves and the objective is to join strengths.

An inspiring approach is presented by Peer [Peer, 2002] who proposes the use of Meaning Definition Language (MDL) in order to achieve the synergy between semantic web and web services. Nevertheless, his proposal has not been tested, so the challenge is to develop a combination of technologies and methods that changes so fast that requires an solid design and a fast implementation.



[Figure 2 \(Scenario to model\)](#) presents the outline of a real situation faced by a user; the process starts with that user who defines his requirements in a business work flow. This user knows that in the universe of web services he can find his needed components to solve his

requirement; so he looks in the UDDI for candidate services available in the internet; the UDDI searches in its repository for candidate services and sends back the corresponding service description. So far, the user has to know the meaning, scope and business use of the candidate services' descriptions. The language used to describe services or the business process itself is not declarative and does not facilitate symbolic manipulation.

The key element to enrich with semantics is the description of the service: the WSDL file; in this context each component will have an added-semantic information to its current syntactical definition. This extension will ease the composition process, because the user, or an automatic composer agent, could have a better interpretation of the objective, use, and scope of each component. This semantic link lets a higher level of flexibility and accuracy, and faster and more reliable compositions.

4.1. Function behavior

The functional behavior of a service composition includes a business model, which can be defined with BPEL, and the use of extended WSDL definition with semantic annotation. In this stage the composition task is strengthened with the extended meaning of components; the publisher of services can include a better definition of his services and the user can identify by himself the usefulness of them.

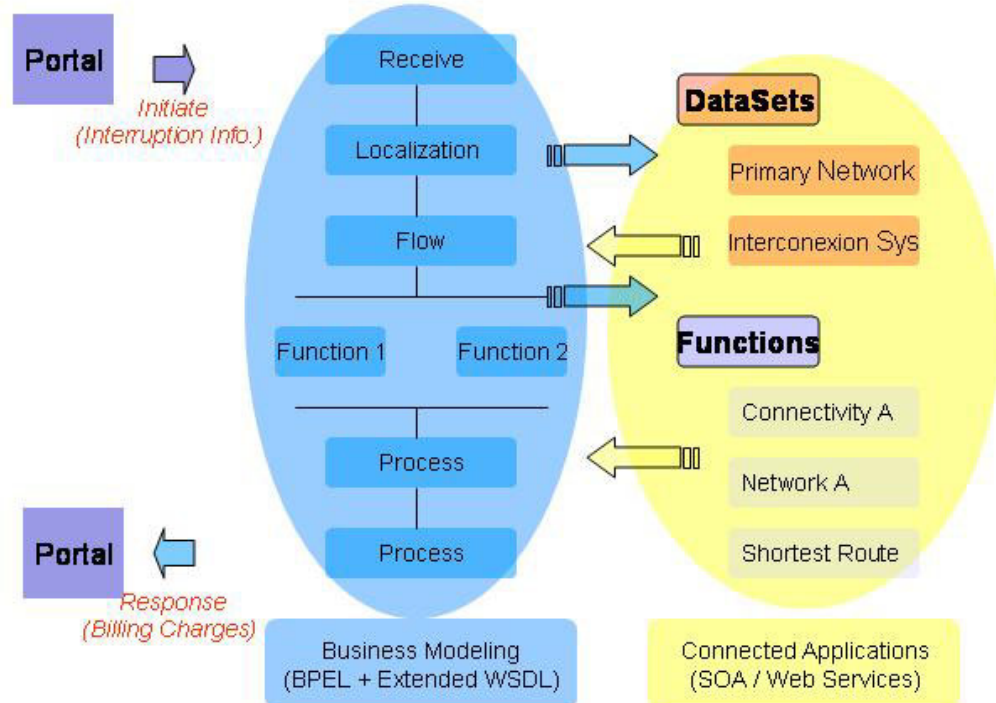


Figure 3. Service chaining functional behavior.

In the next stage, the user, a human or a machine, can consume the integrated process through the use of a portal; the called process follows the business definition and invokes the execution of each atomic service, both data sets and functions; the connected applications are SOA components, which interdependent results generate a final product for the user. The stage of execution uses the WSDL description to ground the used services, but the semantic extension is not useful anymore for its realization.

Chapter 5. Discussion on case study

From an empirical point of view, in Latin America a utility service is something that people realize about it when they don't have it: for example, generally it is supposed that when one arrives to his place, lights become on only with a switch contact; common people are not aware about the big infrastructure behind the scenes. Only people involved with energy and gas infrastructure know the complexity of planning, controlling and maintaining this kind of networks and pipelines, and the great efforts required to offer and maintain a good quality and continuity of service.

Additionally Latin American companies have the great pressure of public opinion, controlled by different stakeholders with a great diversity of interest, and privatization processes pushed by governmental policies of open markets. In this context utility companies have to define and implement in a fast way strategies to improve their weakness and implement ambitious plans to fulfill their stakeholders' expectations.

For our study the emergency use case has been chosen. It starts when the emergency line (e-line) officer receives the phone call. In that moment the e-line system gets the caller's phone number and with it the e-line does a search in the Customer Information Systems (CIS) to find the address of the caller. The address that the officer receives could be a precise one, a set of crossing avenues and streets, a postal code or any type that distinguishes a place.

The address given by the phone caller can be different from the emergency location. In that moment, the e-line officer has to confirm with the phone caller the real address of the emergency.

To find the location of the emergency a geocoding service is used. Once the e-line system finds the location, it shows the area in the screen in red color, and includes labels of main avenues, important buildings and geographical highlights that help to be sure about the place.

The questions that arrived for this use case are how to identify in an easier way the required components to solve this necessity? How to speed up the composition of service chains and how to register the resultant complex component for been used by future users?

5.1. Service chain composition process

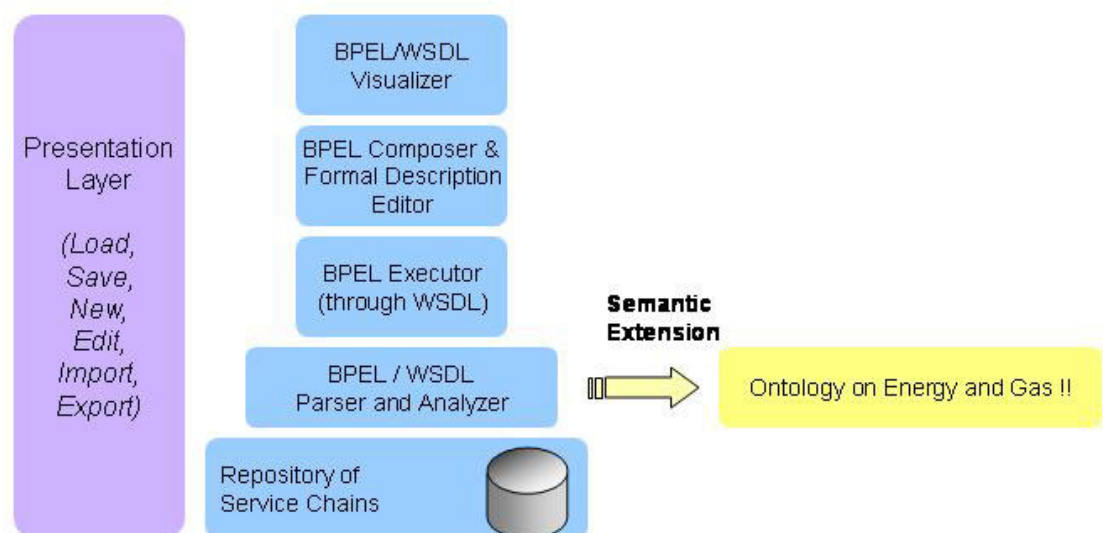
The first approach to build the required composition of services included the use of Model Builder incorporated in ArcGIS 9; this product offers an standard representation for the definition of workflows with inputs, functions, and results; the great limitation is that Model Builder works only with ArcToolbox functions; this means that the composition can be

very easy and fast if you know the scope, capabilities and scope of these functions, and if you have Arctoolbox as well. So the universe of web services is closed from this perspective.

Another option was the use of a WSDL and BPEL editor; the expectation was to implement, compose, or use a product with the capabilities of creation, edition, and, if it was possible, execution of service chains. One found proposal is BPEL Process Manager and BPEL Designer by Oracle; the BPEL Designer is in version 0.8, which means that it is very new and with many potential improvements. Nevertheless the available capabilities are sufficient to implement a set of synchronous and asynchronous service chains in order to solve the emergency use case.

A big challenge to solve is to generate a repository of WSDL descriptions with the semantic extension. So far, a local storage component has been proposed with the interaction of the BPEL and WSDL editor, which means that a local UDDI with extended descriptions is populated with the components found for the service chain. This local UDDI has to be aware about the validity of the generated WSDL files.

5.2. A composed prototype



[Figure 4 \(Proposed architecture\)](#) has an outline of the proposed prototype; it includes a presentation front end for the user interaction. Behind this user interface, one can find the repository of service chains, which can be atomic, composite, or simple processes, with the semantic extended WSDL descriptions.

On top of the repository there is service to parse and analyze BPEL and WSDL definitions. This is the place where the link to the related ontology has to be done, and the extended WSDL are generated. Finally over this layer, there are three functions for visualization, composition, and execution of service chains. The aim is to generate a visual composition and execution of the chains.

The prototype started as a development with a big programming component, but after checking the efforts around the same topic, it has become in an assembly of software components: one for editing, another for visualization, and another for execution. The programming effort is the construction of links and transfer of control between them. Now, the prototype is a new composition.

5.3. Outlook and to do's

The combination of strengths between web services and semantic web, through the linking of ontology in the service description file, offers an interesting and promising alternative to achieve faster service compositions right now, and semi automatic or totally automatic ones in a near future. The ideas proposed in this report have the aim of offering real applications for a specific knowledge area, but they can be applied in many other knowledge domains.

The project is still in progress, and this means that the final result could have some differences with the ideas proposed in this report. More difficulties in the implementation could be found, but so far there are already identified problems: scalability and availability issues; additionally the tool deals with redundant messages passing between the tools and the invoked components; this situation causes a quite inefficient use of the bandwidth, that in the case of GIS is a severe problem due to the size of the large geographical files.

Chapter 6. Recommendations

Further work needs to be done in the field of semantics in service chaining including links from WSDL definitions to RDF documents. The cooperation between web services and semantic web may produced interesting new and innovative techniques to take advantages of business driven solutions such as BPEL and AI approaches such as knowledge representation through OWL. The huge amount of business processes already defined and migrated to BPEL or another related language for web services, combined with semantic declaration methods, would offer an unprecedented opportunity for automatic composition, edition and validation of service chains.

Both worlds, web services and knowledge administration, have to be linked in a way of taking advantage of both universes. It not the idea of absorbing one into the other; it is the opportunity to let both of them continue with their research and development, but thinking in a bridge between them.

The generation of domain-specific ontologies linked to sets of web services with a shared conceptualization would speed up the process of composition and validation of service chains. The aim is to get dynamic and flexible solving methods.

Chapter 7. Conclusions

This report has focused on the use of semantic extensions of WSDL description in the field of service chaining. Foundations of interoperability, heterogeneities and service chaining have been discussed, and an identification of shortcomings of ontology and workflow approaches has been included. Further the report also addresses a case scenario on energy and gas in Latin America, and the possible areas of future research.

References

Bishr, Y., 1997: Semantics aspects of Interoperable GIS. ITC Publication Series. ISBN 906164 1411. Enschede, The Netherlands. Chapter 3 (pp. 27-43)

Bishr, Y; Radwan, M, 2002: GDI Architectures. In: Geospatial Data Infrastructure. Concepts, cases and good practice. Edited by: Groot, R, McLaughlin, J. Enschede, The Netherlands

Croswell, 2002. The role of standards in support of GDI. In: Geospatial Data Infrastructure. Concepts, cases and good practice. Edited by: Groot, R, McLaughlin, J. Enschede, The Netherlands

ESRI, 1999: Modeling our world. The ESRI Guide to Geodatabase design. ESRI Press, Redlands, California. Chapter 1 (pp. 1-21)

Martin, D.; et. al, 2001: DAML-S: Semantic markup for web services. White paper.

OGC 2003a: OpenGIS Reference Model (v. 0.1.2). Reference Number: OGC 03-40. OpenGIS Consortium Inc., URL: <<http://www.opengis.org/docs/03-040.pdf>> Accessed on November 24, 2004

OGC, 2003b: Data models and Interoperability. An Open GIS consortium (OGC) white paper: URL: < <http://www.opengeospatial.org/press/?page=papers>> Accessed on November 15, 2004

Oracle, 2004a: Orchestrating Web Services: The Case for a BPEL Server. An Oracle white paper, June 2004. Available in URL: <http://www.oracle.com/solutions/integration/BPEL_whitepaper.pdf> Accessed on September 17, 2004.

Oracle, 2004b: Oracle BPEL Process Manager. Product webpage. URL: <<http://www.oracle.com/technology/products/ias/bpel/index.html>> Accessed on October 28, 2004.

Peer, J, 2002: Bringing together semantic web and web services. In ISWC 2002, edited by Horrocks I and Hendler J., LNCS 2342, Springer-Verlag Berlin Heidelberg, pp. 279 -291.

Srivastava, B; Koehler, J, 2003: Web service composition – Current Solutions and Open Problems. An IBM India Research Laboratory white paper available in URL: <www.zurich.ibm.com/pdf/ebizz/icaps-ws.pdf> Accessed on September 3, 2004.

Trias Aditya, K.M., 2003: Semantics and Interoperability in Geo Web Services. ITC, The Netherlands. pp. 25.

Authors/Presenters Biography:

Gustavo Zarrate received his University degree in Information Systems engineering and Computation at the Universidad Distrital in Colombia. He received his certification in Marketing and Sales Management at the Universidad Industrial de Santander in Colombia and his specialization certificate in Business Management at the Central University in Colombia as well.

Actually Mr. Zarrate is doing his research in geospatial sciences in the International Institute of Geo-Information Science and Earth Observation (ITC) in the Netherlands. Through his research Mr. Zarrate is working in modeling and designing the integration of billing databases with geospatial data with the case study focused on energy and gas companies; with that purpose Mr. Zarrate is working on composition and execution on distributed geographical information components.

Mr. Zarrate has worked in Latin America for 14 years and he has been focused in public utilities services such as electricity and gas. Mr. Zarrate has designed and implemented solutions for billing processes in energy and gas companies in Colombia, Ecuador, and Argentina, using Oracle products in Unix and Windows environments