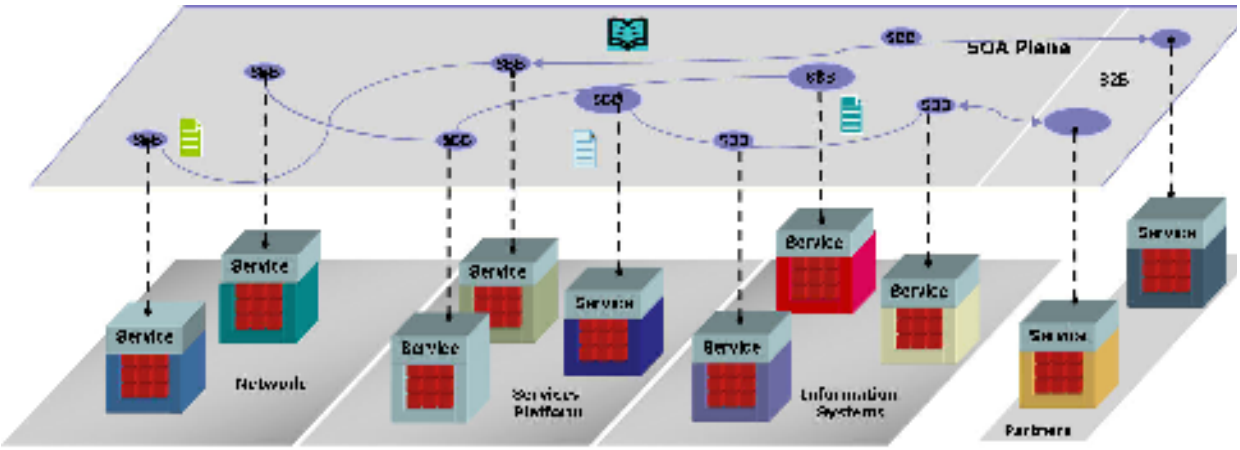
A scientific information and data sharing web portal based on service-oriented architecture

Umut Tutal



Erasmus School of Economics

Master Thesis Economics & Informatics

Economics & ICT programme

Studentid : 179387

Supervisor : Prof. Dr. Ir. Uzay Kaymak

2nd supervisor : Prof. Dr. Paul Wouters

Oktober 2009

ERASMUS UNIVERSITEIT ROTTERDAM

ACKNOWLEDGEMENT

I would like to use this opportunity to thank everyone who has contributed to the preparation of this report.

In particular, My sincere gratitude also goes out to my supervisors to Prof. Dr. Ir. Uzay Kaymak for supporting and supervising during my graduation period. Also special thanks to Prof. Dr. Paul Wouters from Royal Netherlands Academy of Arts and Sciences for supporting and providing feedback.

Furthermore, I would like to thank all my colleagues in Oracle for their supports and ideas.

A word of thanks to my family and especially my mother who have supported me unconditionally throughout the duration of master thesis.

Finally I want to thank all those who I forgot to mention. Please forgive me for forgetting.

Umut Tutal

Den Haag, Oktober 2009

ABSTRACT

The amount of information and resources on the Internet have increased tremendously in the last decennia. Scientists and university research communities are making use of several benefits like communicating with each other about research ideas, to share their knowledge interactively and to collect scientific information from the World Wide Web.

Web portals provide resources, web services and applications to support the researchers in finding useful information about several studies. The communication is also supported by web portals in order to share this information with the researchers. On the other hand, the complexity of different web portals, various interfaces between web services, overload of information and unstructured architecture of the web portals provide difficulties to research information effectively and efficiently.

This thesis is about designing an information and data sharing web portal for social sciences, based on service-oriented architecture to address the involved difficulties. This web portal will provide efficient communication between researchers and will maintain effective information searching.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

1 Introduction

This chapter is an introduction to the topic and research. There is a context, and the research question and methodology followed. Finally, the structure of the thesis will be outlined.

1.1 Context

The amount of information and resources on the World Wide Web[1] has increased tremendously in the last decennia. Scientists and university research communities are using WWW to communicate with each other about research ideas, to share their knowledge interactively and to collect scientific information from the WWW. Web portals provide resources, web services and applications to support the researchers in finding useful information about several studies. The communication is also supported by web portals in order to share this information with the researchers. On the other hand, the complexity of different web portals, various interfaces between web services, overload of information and unstructured architecture of the web portals provides difficulties to research information effectively and efficiently. (Chau, et. al., 2006). The following issues provides difficulties to manage information:

- Difficult to access and find information because multiple applications and content from disparate sources can be presented. If you don't know where to look how do you find it?
- Several different applications are to integrate in the portal, different logons instead of single sign-on, different user interfaces and different organization can lead to user dissatisfaction.
- Content is quickly out-of-date or inaccurate.
- Difficult to manage the technology, infrastructure of the web portal is very fragmented and expensive.

[1] WWW will be used throughout this paper for abbreviation of World Wide Web.

- Integration extranets and intranets can cost more time.

According to Eduventures (2006), universities are often operated as highly decentralized enterprises, with faculties allowed considerable autonomy to choose their information systems, business rules, and operating guidelines. In a decentralized environment, university IT managers may find themselves supporting, at relatively high cost, several operating platforms and applications, each with its own programming language and tools. Several universities and science communities are facing aging legacy systems which are difficult to be integrated. These legacy systems are forming several obstacles to seamless cross application computing capability, easy-to-navigate interfaces, and real-time enterprise wide access to accurate data. Scientists and university research communities are frustrated about multiple logins, different interfaces, losing time to find needed information.

Universities are realizing that legacy systems centric silos no longer make sense in the Internet age and are focusing more on a computing environment that, from the perspective of the university's many users, more closely like popular web-based commercial businesses. Web services are seen as the solution for integration. Scientists and university research communities will perform their task quickly if the information which they need, can be accessed easily. (Eduventures, 2006)

This thesis is about designing a information and data sharing web portal for social sciences, based on service-oriented architecture to address the involved problems. My study Economics & Informatics at Erasmus University covers aspects like automating and innovating business processes. Designing a information and data sharing web portal for social sciences will support automating and innovating business processes for social sciences. Before going to the following section, social science and supporting social science with information technology solutions will be explained.

Webster's New World Dictionary of the American Language, defines social science as

> "*the study of people living together in groups, as families, tribes, communities, etc.*" (Gordon, 1991)

The focus of this definition people and his environment. This definition can be formulated also as social science is one of the branches of science which is related to social life of people in their social environment. The interactions between people is also a part of this study. Some of sub branches of social science are anthropology, economics, geography, philosophy, law and history.
Social scientists are researching with how people as individual and together are, they are collecting and interpreting data.

Science remains in motion: new specializations are created and existing boundaries between disciplines are getting smaller. In the future, many new insights will be available as the result of multidisciplinary research. As other sciences, social science community also are attempting to gain benefits of information technology solutions in order to build an e-infrastructure. With developing an e-infrastructure for Social Science can provide several benefits:

- to provide social scientists integrated access to a variety of research resources
- acting as e-library for searching research papers
- to encourage the readers to communicate directly with authors concerning their researches
- to support knowledge sharing between scientists
- to help scientists for daily research activities like collecting data, analyzing data or generation outputs for decision making.

In the following sections, the information technology which can support social science activities, will be explained.

1.2 Research question

My main research question is how to design a web portal which supports information and data sharing for social sciences based on service-oriented architecture.

In order to give an answer to the research question the following sub questions are formulated:

- How can social sciences gain benefits from a web portal?
  - What is a web portal?
  - What are the major functions of the web portals?
  - What is reference architecture of the web portals?
  - Which benefits can be provided by the web portals?
- How can SOA concept be placed in a portal environment?
  - What is SOA?
  - How does SOA perform?
  - What kind of benefits can be gained by applying SOA?
- What are the requirements for the web portal supporting information and data sharing for social sciences based on SOA?
- How can a web portal support information and data sharing for social sciences based on SOA?
- How can this web portal be implemented?
- How can this research be validated?

1.3 Methodology

In this section, the methodology will be described to design a scientific information and data sharing web portal, based on service-oriented architecture.

- Literature study about web portals, service-oriented architecture and web services to gain more information about background of these concepts.
- Examine how web portal concept works and how a web portal can be acted as scientific information and data sharing instrument.
- Research web services and the integration between web services to understand how web services are performing and how the integration between them are applied.
- Explore SOA and how SOA can be applied on a social science web portal.
- Interviews with university docents and researchers to indicate the functional requirements of the web portal.
- Design architecture of social science web portal, which can be applied as a guide to the implementation of the web portal.
- A qualitative validation will be executed to validate functional requirements of the web portal and a prototype will be implemented to validate the architecture of the web portal.

1.4 Document structure

The rest of the chapters is structured along the sequence of the sub questions.

Chapter 2 describes web portal basics. First, the history and definition of portals will first be explained. After that the classification and portal types are examined. Furthermore, the reference architecture will be addressed. Finally follows a section on portlets, benefits and challenges of the portals.

Chapter 3 explains SOA integration in portal environment. In this chapter, definition and how SOA performs, will be outlined. Finally, benefits and challenges of SOA application are discussed.

Chapter 4 outlines system architecture. First, the requirements of system will be described. Furthermore, key components like web services and standards for web services are explained. After that, overview of activities which can be applied for implementing web services in a organization, will be described. Finally, reference architecture will be outlined.

Chapter 5 introduces validation of the research. A qualitative validation will be executed to validate the functional requirements, after that a small prototype is implemented to validate portal architecture.

Chapter 6 presents conclusion of this thesis and makes recommendations for further study.

2 Basics of web portal

The emphasis of this chapter is to provide answer to the following questions:

- What is a web portal?
- How can social sciences gain benefits from a web portal?
- What are the major functions of the web portals?
- What is reference architecture of the web portals?

In this chapter, web portal concept will be discussed. First, the history and definition of portals will first be explained. Then the classification and portal types are examined. Furthermore, the reference architecture will be addressed. Finally follows a section on portlets, benefits and challenges of the portals.

2.1 History

The term "portal" is often used in information technology[2] world. However, there are many different definitions about portal. The term has its origin in ancient architecture; derived from the Latin word "porta". This word means something like gate, door, or entry and indicates a monumentally shaped entrance of a building. In IT context the term "portal" is developed gradually as a entry point to relevant information on the Internet. This term was quickly regarded and nowadays used by several vendors and organisations to promote their web site beyond the original intra- and extranets. (Priebe, 2005)

Intra- and extranets are both providing information to different parties. Intranets are aiming for sharing information inside the organization, on the other hand, extranets for attracting external components to the organization. Because the information was

[2] IT will be used throughout this paper for abbreviation of information technology.

static, not up-do-dated, the problem was occurred: who is responsible for keeping information up-to-date. The developers are tried to bundle and structure the offered information. Besides that search functionalities, categorization of the information and providing a central entry point to are added in this concept.

These first generation portals required a sign on which links to some information that is organized by organization. With personalization features, it was then possible to arrange information to a user´s individual needs and desires. The benefits of portals were soon discovered by many organizations. Many intranets are extended to the portals; the first enterprise portals are created. (Priebe, 2005)

2.2 Definition

Larry Bowden, IBM Vice President of Portal Solutions, defines a portal as

> "*a single integrated point of comprehensive, ubiquitous and useful access to information, applications and people.*" (Priebe, 2005)

Java Portlet Specification defines a portal on more technical way:

> "*a portal is a web-based application that commonly provides personalization, single-sign-on, content aggregation from different sources within a webpage.*" (Priebe, 2005)

From both definitions, a web portal can be defined as an interactive, dynamic, goal-oriented website that provides information to the users of the web portal or other users on internet or intranet. Besides that the web portals structure the information and extend search capabilities. The users can exchange information and data sharing with other users on the web portal.

What can a portal do?

The portals serve as a central point of access to all applications needed for a specific business process for the users.

The main functions of portals are:

- Search

  This is the basis functionality for the most public web portals. The portals should support efficient search for contents. Depending on the privileges of the user, the limited information should be shown to the user.

- Information integration

  The integration of the information from different sources should be performed. The several techniques are used for integration like Unified Content Application Programming Interface[3]. The Unified Content API supports all current tools for implementing web environments like JAVA, C++, ActiveX and Java beans.

- Personalization

  Personalization is about to provide appropriate information to the users depending on their privileges or roles. Personalization functionality can be performed on several types like navigation, content and menu structure.

- Notification

  This functionality are developed to send information the user without requesting it. The user can subscribe to active information sources.

[3] API will be used throughout this paper for abbreviation of Application Programming Interface.

- Task management and workflow

  Task management functionality is about managing formally defined business processes. Furthermore, workflow functionality supports the automation of business processes. A portal should be able to prompt its users if they have tasks to execute.

- Collaboration and groupware

  These functionalities provide the required information is stored in the right place and in the right mode. Groupware software assists in less formal collaboration than workflow tools. As with workflow automation, groupware increases the value delivered by many types of specialized portals.

- Integration of applications and business intelligence

  A portal can integrate and support several information systems like application service provider[4] applications and business intelligence functionality.

- Infrastructure

  This is about to constitute the fundament for the work environment - the other 7 functionalities mentioned above build up on this one.

[4] ASP will be used throughout this paper for abbreviation of Application service provider.

2.3 Classification

The portals can be mainly classified as two types:

- Horizontal portal
- Vertical portal

Horizontal portal

A horizontal portal is an Internet portal site that offers many wide range and variety of the content and services. These portals are also called “megaportals”. These kind of portals puts no focus on a certain target group or a certain topic, but on entire Internet community and several topics. Functionalities like search engines and personalization for user by offering various channels (i.e. weather, news, stocks, sports) are usually provided by these portals.

Yahoo and Lycos are some examples of horizontal portals which provide wide range information on many different topics.

Vertical portal

In contrast to a horizontal portal, a vertical portal addresses to specific topics like line of business of product areas, but covers in more detail and also the technology employed remains the same. These portals target at a specific type of user or a vertical community. For example, an education portal will provide educational information and services for educational community like students, teachers, etc.

There are several possibilities for establishing special vertical portals on the market. These several solutions can be divided into three major groups that partially overlap:

- Corporate portals: provide personalized access to the chosen information of the specific organization.
- Commerce portals: support business-to-business and business-to-consumer e-commerce.
- Pervasive portals: support access via Pervasive Devices such as PDAs.

2.4 Portal types

The analyst and consulting company Ovum has designed a detailed taxonomy of portal types, their hierarchical affiliations included. After short descriptions of the major portals types, the figure 2.1 will display these portal types.

Specialized Portals

This kind of portals are designed for a specific purpose. Application Service

Provider portal is a specialized portal that provides access its customers to applications via an extranet of the Internet.

Marketspace Portals

Marketspace portals support business-to-business or business-to-consumer e-commerce. These portals have ability to find and access rich information about the products, to support e-commerce transactions. Also organizing discussion groups with other vendors and/or buyers belongs to the major functionalities of marketspace portals.

Public Web Portals

Public web portals provide user-friendly and several services to the public. Some of the major functionalities of this kind of portals are to extend information search and to make various service offerings available. Usually these portals are seen as entry point for web surfing.

Enterprise Portals

These portals are also called corporate portals, that provide limited information of a particular organization. Enterprise portals provide virtual workplace for internal and external parties to support business processes of the organization. Business partners of the organization like suppliers, customers have individual access to regarding information, information systems and services needed to execute their jobs.

A company´s public website is not a corporate portal. It can become one if personalization and navigation functionalities are supported. Furthermore more advanced enterprise portals are reachable via mobile devices as PDA´s and handheld PC´s.

Workspace Portals

Workspace portals provide all the information needed by the users to perform their jobs. Via a user interface, the user gets the regarding information. The specialized portal is an alternative to workspace portal.

Knowledge Portals

Knowledge portals have aim to increase the effectiveness of knowledge users like scientists by providing access to the information, extending search capabilities, supporting collaboration services and sophisticated information discovery services.

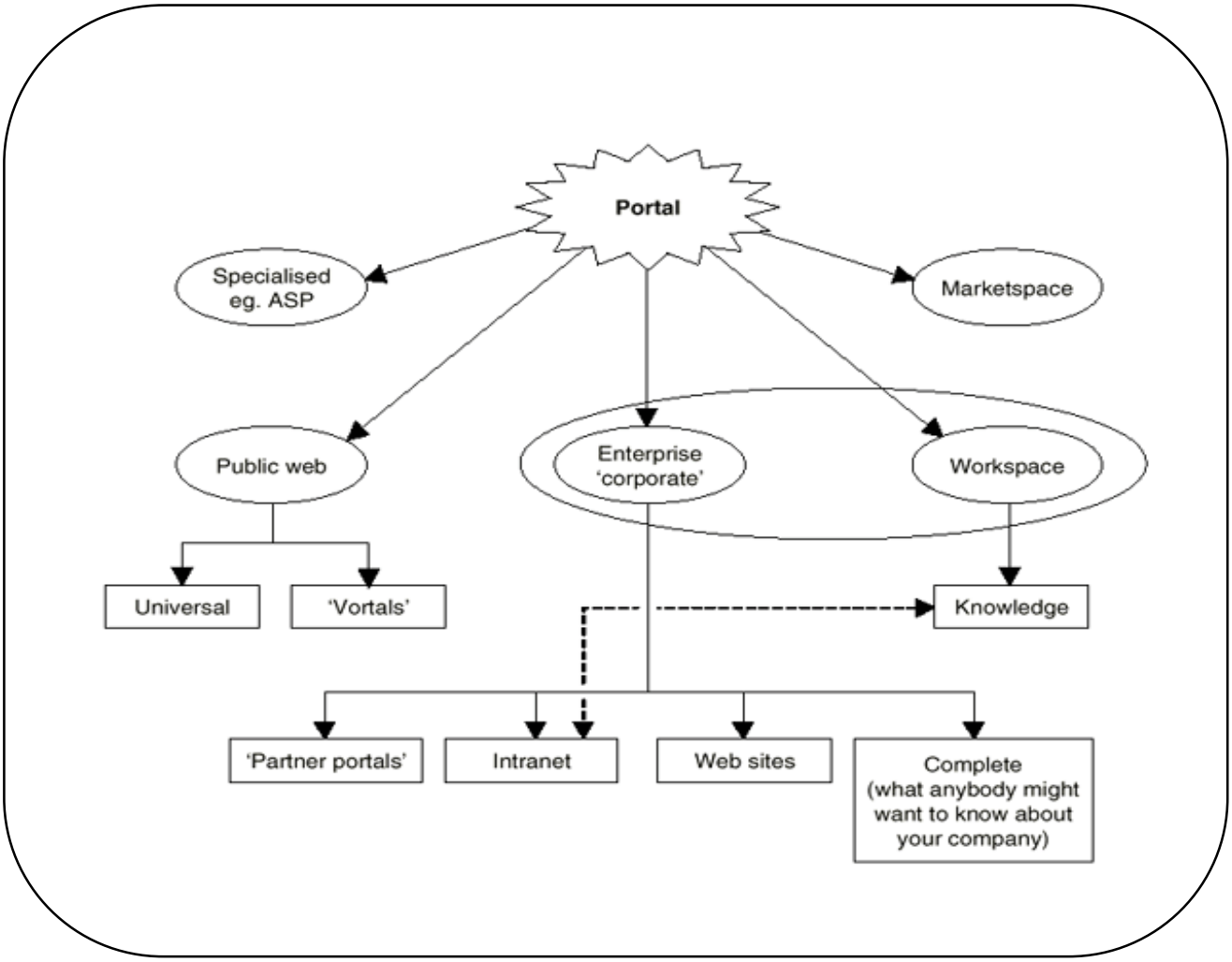


Figure 2.1: Major portal types

2.5 Reference architecture

The portal architecture is based on the classical 3-tier paradigm with presentation, application and data layer. There are also 2-tier and n-tier paradigms available. 2-tier architectures is one of the oldest architecture which contains 2 layers: client and server. The business logic is included inside the user interface on the client or in the server layer. Sometimes, the business logic can be separated between these layers. Expensive connections and limited connections are some of the problems which can occur in this architecture. A n-tier architecture is using different layers for allocating the responsibilities of an application. Middleware is usually used to connect the separate tiers.

3-tier architecture is more often used to refer to portals. In 3-tier architecture, the business logic, also called "application layer" is established in the middle tier. The processes can be maintained separately from other layers, thus a more manageable architecture. This architecture supports also integrating data from multiple sources.

On the presentation layer in 3-tier architecture, the client devices can be found, often a web browser. The application layer is acting as the core of the portal system. For several vendors, the portal is dependent on an application server. Delivery services have tasks to provide the portal pages to the presentation layer. Portal applications are represented by application portlets are also found in this layer. These portlets are accessed by a software interface called Portlet API. Via this Portlet API, the application portlets can use portal base services. These base services provide the core functionality of the portal:

- Layout management: Layout management is responsible for rendering and combining portal page before it is presented to the web browser
- Structure management: Structure management defines the structure and navigation between the different portal applications.

- Content management: this service brings content, document and knowledge functions in rules. Collection and structuring of the knowledge are also here provided. External document and content management systems can also be integrated for supporting authoring and administration.

- Personalization and security: these services deal with the management of user accounts and privileges. Role management provides individual portal views for users. For personalization, the user should be registered and authenticated when they want to access to the portal.

- Global search: the portal provides searches across the data sources within the portal, e.g., content, documents, databases, corporate applications and external sources.

- Collaboration and communication: these include email, calendars, discussion groups, forums; usually organized by sending emails.

- Process support: this can be done by integration workflow functionality into the portal in order to automate the page flow and lead the user to several applications.

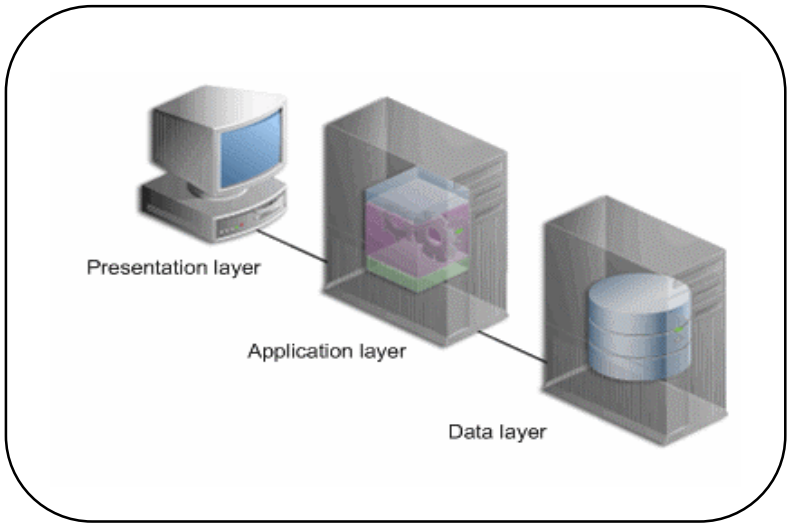


Figure 2.2: 3-tier architecture

On the data layer, back-end systems like operational application systems (ERP), analytical (business intelligence) application systems, data sources like relational databases and external content management systems can be found. Integration and transaction services provide the interfaces to these back-end systems. For some vendors, the integration and transaction services are part of the application layer.

On the following figure, portal reference architecture will be displayed in more details.

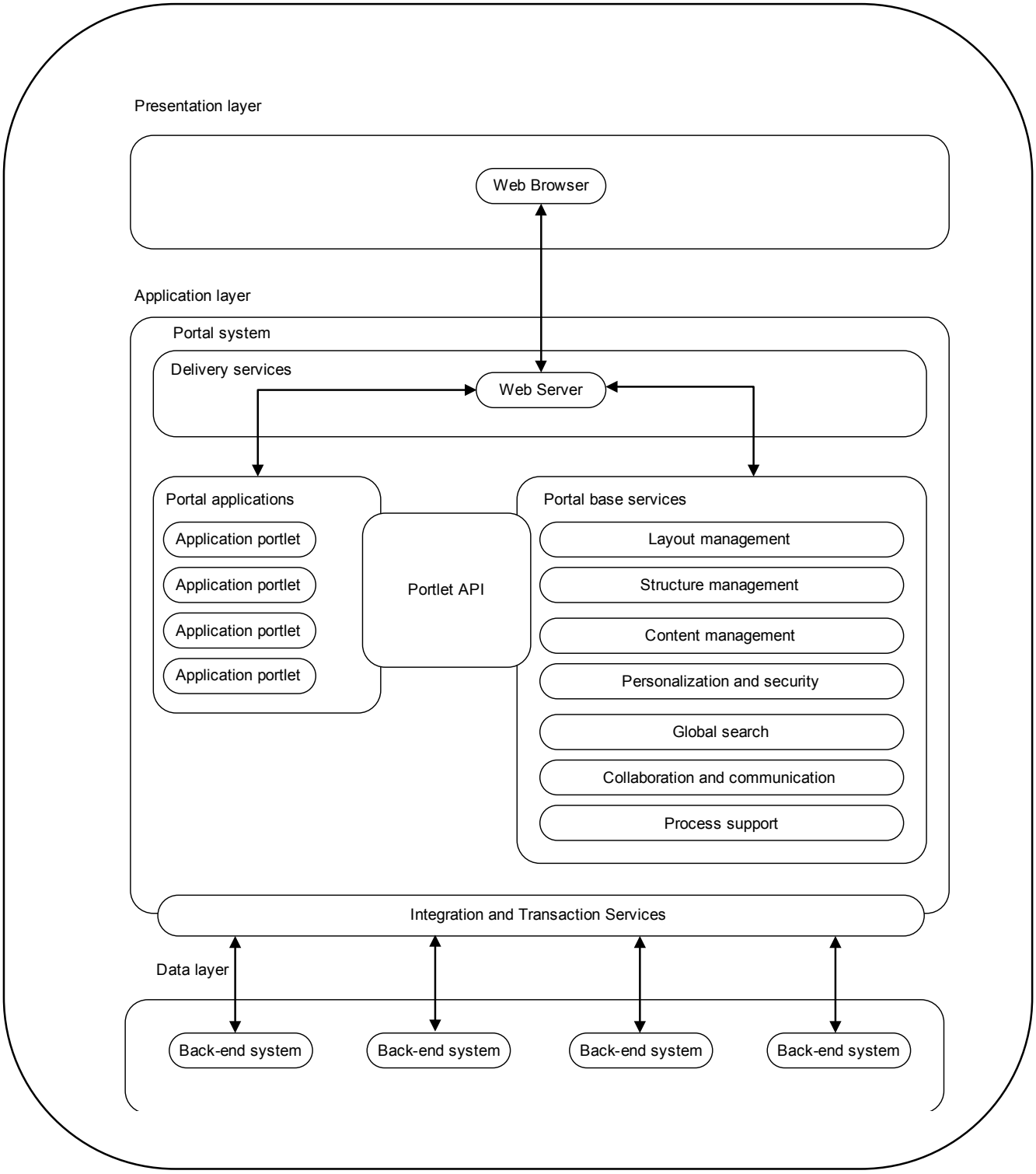


Figure 2.3: Portal reference architecture

2.6 Portlets

A portlet is a software code that manages of one section of a portal, can range from simple static HTML to custom applications accessing several systems, dynamically providing and manipulating data. It can handle several actions like connecting to a database, invoking a web service, downloading file. Implementing of portlets can be in any language that can generate HTML and XML.

What can a portlet do?

- Providing a interface into another application.
- Working together with other portlets to provide various services.
- Providing several actions like displaying data, accepting and validating user input, getting and setting data in database.
- Providing users with customized tools, services and information.
- Increasing portal flexibility by allowing integration of applications, tools and services into the portal.

Portlets are located in portlet containers. Portlet containers are components of portals and responsible for managing of all portlets of the portal. Portlets are the key to portal for reusing. The web services can be used by several portlets.

2.7 Benefits and challenges

In this section benefits and challenges of portals are explained.

Benefits

The portal concept is popular. Many industrials branches, education sectors, business vendors, healthcare, government departments have their own portals. There are several opportunities of using portals in corporate environment:

Reducing time of business processes

Processes will become more transparent and improve to follow and less errors will be made. The workload will be reduced work and lead-time will become shorter. Using self-service applications will reduce cost of support visits and calls. A higher employee satisfaction contributes to this also.

Reducing the costs

Bringing of all user interfaces of several applications together to one standard user interface will lead to lower support costs, maintenances costs and learning costs. The system integration costs will be reduced because of adaptable infrastructure and reusable services. Centralized data control leads also to lower costs.

Increasing performance

Portal will ensure high performance. The portals are usually role based and have a consistent navigation. The employees can locate and manage their work more easily. Furthermore, employees get to more autonomy, because they can regulate more itself. Besides that, employees can search better and faster information about daily facilities.

Sharing knowledge

Cooperation between employees, divisions and departments will be improved by using portals. Online messaging, using wikis, video conferencing are also provided by portals.

Acting as central point

Portals can also be used as central point for daily activities in the organizations. Daily operations like checking emails, task management, agendas, sending documents are mostly supported by portals. Several business processes can also be executed in portals. The employees don't need all the information and applications. Providing concerning information and applications will reduce the traffic on the portals.

Scientific organizations, universities and research organizations are also showing their interests more and more for portals. The Institute of Photogrammetry and Remote Sensing of Vienna University of Technology has developed a Scientific Information Portal to provide soil moisture data for science and real-time applications. Soil moisture is a key state variable of the global energy, water and carbon cycles and is important for a wide range of scientific and operational applications. Agriculture, forest ecology, civil engineering, water resource management, climate research, weather forecast, ecosystem modeling are some examples of these applications.

The Institute has decided that the data of the Global Soil Moisture Archive is published in the WWW on the following address "http://www.ipf.tuwien.ac.at/radar/ers-scat/home.htm". This scientific portal offers many opportunities such as:

- Providing the data to several scientific communities.
- Creating new possibilities for science and real-time applications in various fields.

- Exchanging the data between several applications to validate several processes.
- Easy access for the relevant communities.
- Making scientific journals available to share information.
- Informing interested users about soil moisture and offering them to discuss with each other about the possibilities.
- Acting as a basis for further developments about soil moisture.
- Easy to integrate business applications.

The main target, to attract several scientific communities and researchers, has been achieved by using scientific information portal as publishing platform to provide information on the WWW. Many scientists and researchers show interests after the visiting the web portal. (Trommler, et.al., 2002).

Challenges

Using portals are facing several challenges.

Like every new solution, implementing a portal would be facing different challenges, first of all users need to invest time to get used with it since they were working before in a traditional way: working alone or in project teams, face to face meetings instead of using portal's communications services.

Implementing a portal is considered an investment, it implies acquisition of new hardware and software, time and resources allocated to participate in implementation. After implementation it needs to be maintained and scientists using it need to be trained.

The portals can be overloaded with a many features and functionalities and users especially in the beginning can get confused if they are looking for something.

3 Service-oriented architecture integration

The emphasis of this chapter is to provide answer to the following questions:

- How can SOA concept be placed in a portal environment?
- What is SOA?
- How does SOA perform?
- What kind of benefits can be gained by applying SOA?

In the last years, IT systems have grown exponentially, this leads to more complex software architectures. The current traditional architectures have reached the limit of their capacities, while the needs of the organizations are growing. The organizations still need to respond to new business requirements, reduce the costs of IT and integrate applications faster and better. Furthermore, the current architectures provide little or no flexibility. Adjusting systems or integrating standard applications is difficult because everything is linked with everything. ICT is seen as obstacle for changing in the organizations. The source of the problems is caused by designing of the information systems. Many systems are custom designed for the organizational unit or the business process that it should be supported.

Separating of these applications is often difficult: the construction of the information systems is tied up with the design of the business process. Furthermore, there are numerous links between systems, changes could have a major impact on the related systems. The systems should be renewed systematically to provide more flexibility. For the most organizations, this is an impractical solution in terms of time and costs. An alternative solution is searched that tries to continue using existing systems as long as possible, but also creates the desired flexibility in the information systems.

The solution consists of two steps:

- Separating process specific parts in the software,
- Providing this business to other systems in the form of services

Services are small software programs that provide access to the underlying functionality or data from information system. International standards are used to define and to call services. There is a virtual layer over all existing systems. On this layer, the services are defined in a standardized way, which can be accessed regardless of programming language, location of the software or the platform.

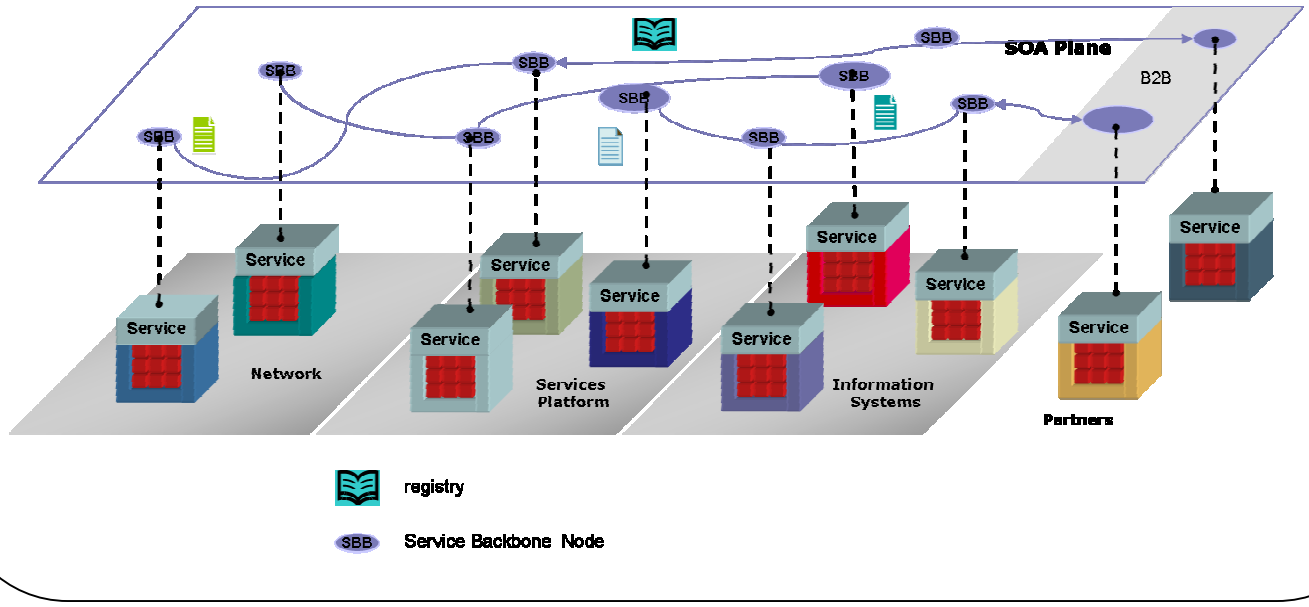


Figure 3.1: Virtual layer

Source: (Oracle, 2004)

3.1 Definition

The World Wide Web Consortium (W3C) defines SOA as

> *"a set of components which can be invoked, and whose interface descriptions can be published and discovered."* (Sprott, 2004)

The Organization for the Advancement of Structured Information Standards[5] (OASIS) defines SOA as

> *"Service Oriented Architecture (SOA) is a paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains."* (OASIS, 2006)

The entitities (people and organization) are able to create capabilities to solve or support a solution for their business problems. One person's needs can be covered by someone else's capabilities. In IT world, one computer agent's requirements is being covered by a computer agent belonging to a different owner. No one-to-one correlation between needs and capabilities is necessary. Any single need can be covered by combination of several capabilities while any single capability may address numerous needs. SOA provides a powerful framework for matching needs and capabilities and for combining capabilities to address those needs.

[5] OASIS is an international consortium that develops standards for security, e-business, and standardization efforts. OASIS has more than 5,000 participants representing over 600 organizations and individual members in 100 countries.

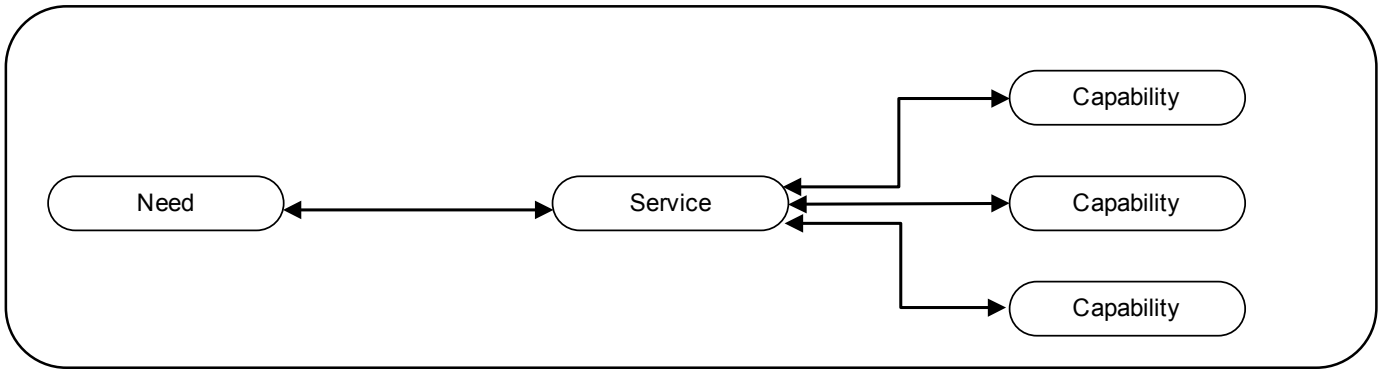


Figure 3.2: Need & Capability

In SOA architecture services are central. The relationship between needs, capabilities and services is as follows. A service is a mechanism that gives access to one or more capabilities. A service brings needs and capabilities together. This means that SOA itself doesn't contain solutions for problems, but provides the link between capabilities and needs.

An entity that offers a service is called a *service provider*. This entity can have a capability and can offer a service at the same time. A service can user one or more capabilities of other entities. The service provider does not have a capability.

The entities with a need to use a service called a *service consumer*. Service providers and service consumers are collectively referred to *service participants*.
In practice it is irrelevant whether a service provider itself has a capability, or uses other entities. A service consumer is not interested in how the service is performed, but only in the result.

In order to find services for entities from different domains, a *service registry* is constructed. Each service´s description can be found in this registry, which stated what the service does and how the interaction takes place. A service consumer can search and choose which service to its possible needs.

As seen in the SOA model, a service provider registers the service in the service registry. A service consumer can search the directory and read the service descriptions or service may be suitable for a need to fulfill. A service description also provides information to a service. A service consumer calls a service, where interaction is performed with the service provider.

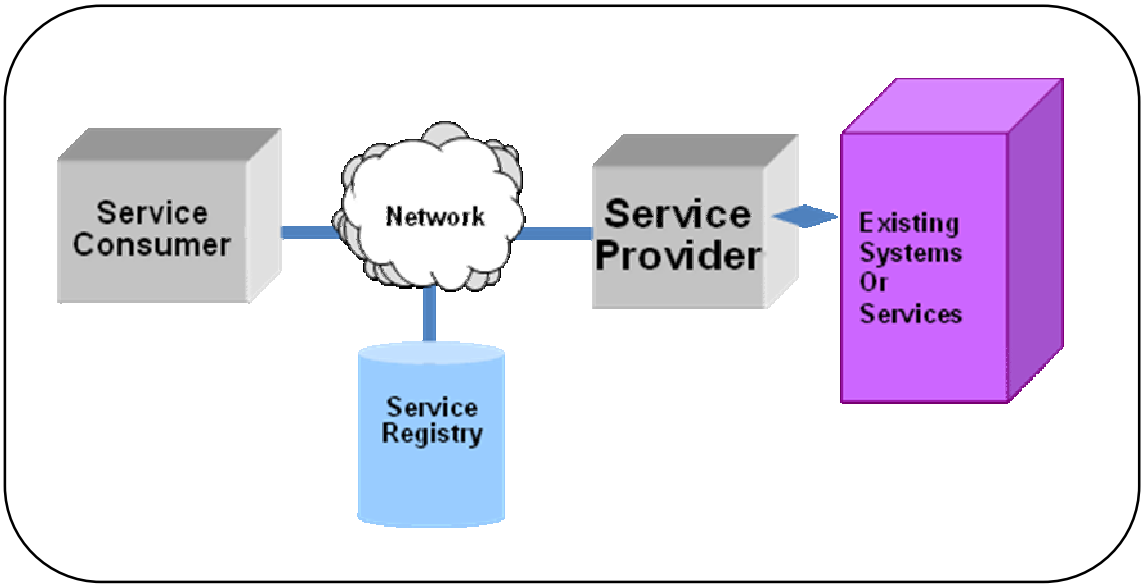


Figure 3.3: SOA model

SOA consists of a collection of web services which communicate with each other via single web portal interface. A web service can be described as an application component on internet which is provided by web protocols. A company process or a function is implemented and made available by a web service for concerning parties as other applications, web services or users. SOA explains the cooperation between the applications by means of these web services. These services can be applied independently or can be a part of activity, this means an activity can consist of one web service or multiple web services which depends on other web services. Authentication is generally required to access these web services.

The main components of SOA are illustrated in the figure 3.4.

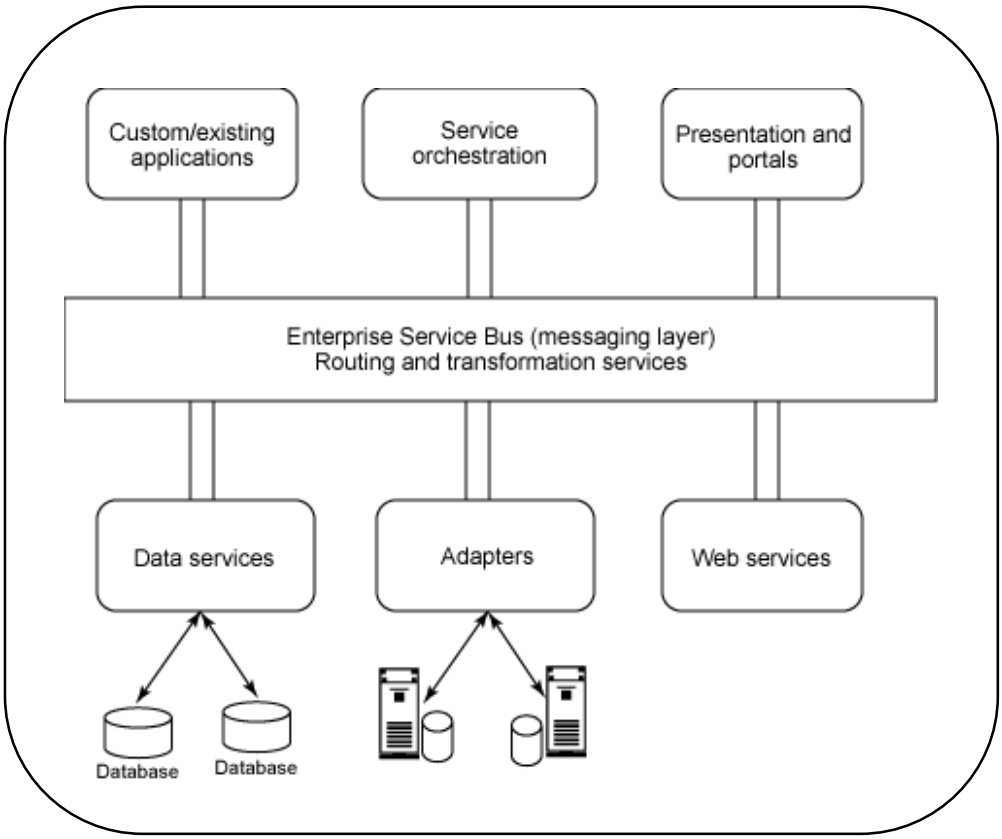


Figure 3.4: SOA components

- Web services: A web service is a small software which has functionality to call concerning application. Nowadays, there are many open international standards to define and use web services as WSDL, SOAP, XML and UDDI.

- Enterprise service bus: Enterprise service bus establishes the communication between service consumer and service provider to call the web services. Recently, many technological developments are implemented which are called middleware. These middleware's are aiming to provide the communication between software program's. CTG, RPC, JMS and MQ are some of these middleware's. Enterprise service bus will be described in more details in chapter 4.2.2.

- Service orchestration: Service orchestration engine is also called orchestrator. Orchestrator provides that concerning web services are called in a structured, designed order. On this way, the control of process can be arranged easily and flexible.

- Presentation and portals: The Presentation and portals provide the users a uniform access to the applications from multiple sources. Building a web portal can be implemented for individual users or group users. The information can be exchanged between the users on a structured way.

- Custom/existing applications: If we combine all these components of SOA, then we will create new applications which are composed for the users. These custom applications use the functionalities of the underlying/existing applications.

- Data services: Data services provide real time view of data from heterogeneous data sources.

- Adaptors: Adaptors have functionality to activate connectivity to leading improvements as packaged and custom enterprise applications.

3.2 SOA layers

In this section, Service-Oriented Architecture layers will be outlined. SOA consists of generally 4 layers: presentation layer, process layer, service layer and data layer.

Presentation layer:

The presentation layer is the user interface across the systems. This layer provides interaction with the user and hides the complexity of the underlying systems from the outside world. There are several ways for interaction, which are relatively easy to change, without the back-end systems must be modified. Portal is one of the most commonly used solutions.

Process layer:

Another layer in SOA is the process layer. This layer contains the business processes. Business processes can be composed of other business processes and of portfolio services. These processes provide how the business runs. A business process is an representation of the activities that are coordinated in an organization to perform a specific business function.

Business Service Orchestration or Business Service Choreography is located in this layer. The processes are represented as an orchestration of loosely coupled services. It is also responsible for entire life cycle management of the process orchestration. Processes address to the business specific requirements. Users, channels or other systems in the presentation layer use these business processes to invoke application functions.

There are several standards created for both the design and execution of business processes. Business Process Modelling Notation (BPMN) is a standard to be used for describing business processes. The major ICT suppliers which have developed web technology, have also developed standards for the execution of business processes.

Business Process Execution Language for Web Services (BPEL4WS) is at this moment the standard for choreography, because it is supported by major ICT suppliers as IBM, Oracle BEA and Microsoft.

Service layer:

In SOA it is all about services that provide the functionality of applications. This layer contains the portfolio services. This is also the central layer of the model. This layer supports the basic service feature of SOA.

The services have generally the following characteristics:

- Implementation independent: the services are implemented platform and tool independent. No knowledge is required by the requestor about platform, programming language or tool which the service is implemented.
- Business identifiable: during the design phase services are defined based on the business processes, business objects or business functions. The services can be large and small, the granularity of services depends on the need for functionality from the business process.
- Owner: there is an owner of the service. The interface of a service is provided by the owner of the service. New services can be added if the existing services are not performing well.
- Standard interface: for specifying the interface of services, international open standards are created. In this way a service can be published, found and called.
- Loosely coupled: services provide a service which may be used by multiple customers. To make services common useable, services should do their work independently from the context in which they are called.

- Location transparent: in ideal situation, a service consumer doesn't know where the service is located and how the service is implemented. It is sufficient if the service description is defined in a service registry and that during the calling of the service, a broker provides the message transport. In this way it is possible to use the service layer for the heterogeneity in platforms and to hide languages.

- Stateless: after the service is executed, no data will be held for a possible following request. This is called *stateless*. A service is stateless if it is not depended on the previous conditions of the previous services. The services can be switched easily in a different order to support a process.

- Use other services: in the handling of a service, another service can be used for the functionality.

- Use intern and extern: services can be provided to intern users and extern users, in a secure intern organization.

- Meets standard types: an appropriate way to increase the rates of generic services is to work with standard types of services. There are 3 major categories for standard service types.

    - Inspecting: to search and select information

    - Mutating: to add, update and delete information

    - Generating: to calculate information based on organization rules and data

The service standard types will be described in the following chapter.

Web services form in fact a subset of the total collection services, these are services which are reachable via internet technology. Internet technology is only one of the options to implement services, but it is increasingly applied. The concepts behind services are not new. Recent years, like CORBA, COM /DCOM, Remote Procedure Call (RPC) and Interface Definition Language (IDL) pass in review. The Internet

technology which is used by Web services, is rapidly growing to international industry standards.

Data layer:

This layer includes the databases where all data are stored. Besides that current information systems and external information systems of suppliers or partners are also located in this layer. The data is provided to the service layer.

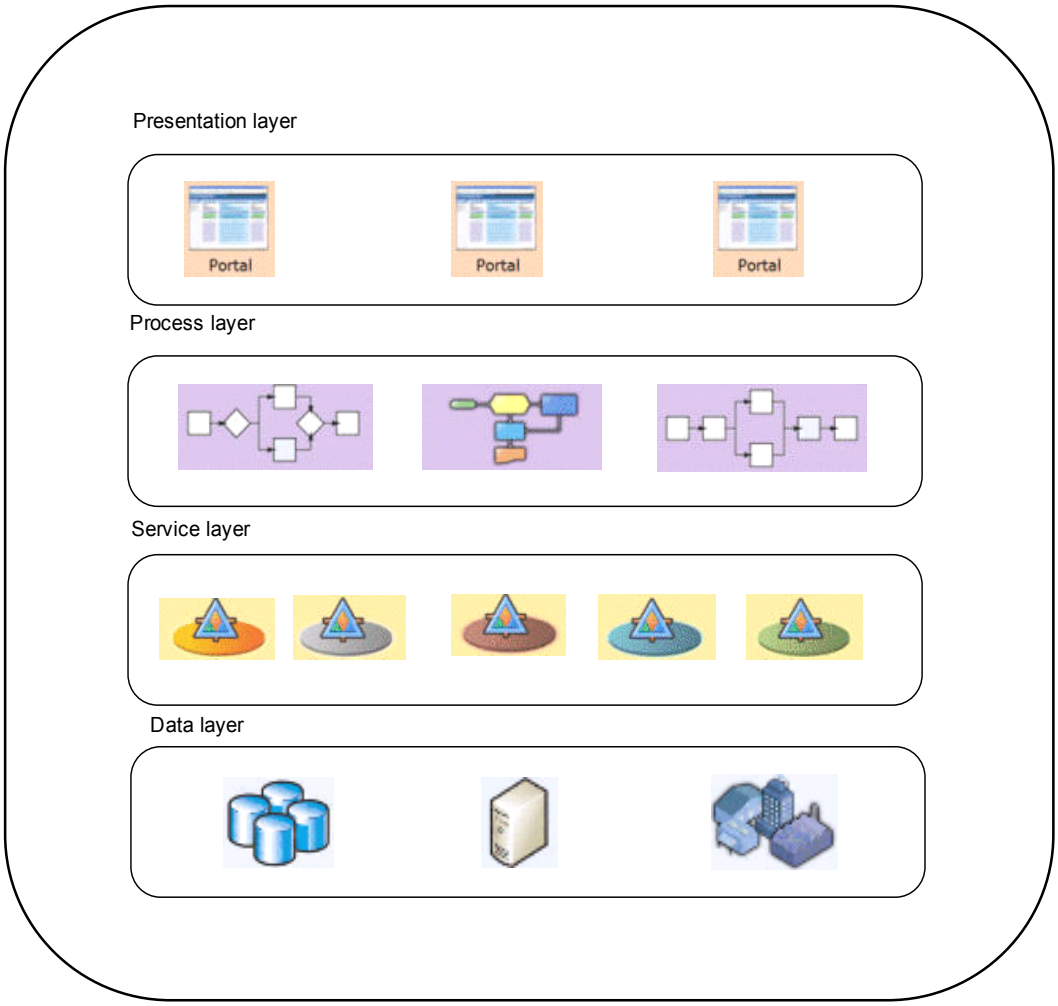


Figure 3.5: SOA layers

3.3 Benefits and challenges

In this section benefits and challenges of SOA implementation are described. Like using portals for scientific communities and researchers, SOA implementation can make several benefits available to support the organizations in different ways.

Benefits

SOA provides generally several benefits in the following categories: reducing integration expense, increasing asset reuse, increasing business agility and reduction of business risk. These core benefits present return at many different levels and parts of the organisation, regarding business problems. In addition, organisations which have implemented SOA, experienced the technological benefits like flexible architecture, integration of current applications and advance data integration. A web portal based on SOA provides increasing productivity in the processes of searching scientific information. Besides that, by collaborating web services, speed of knowledge generation will be reduced. Several web services can cooperate in order to support data sharing. This would take time if these processes will be executed in traditional development methods. Eventually, optimizing all these processes by SOA implementation will increase effectiveness of scientists in the knowledge economy. (Eduventures, 2006)

SOA approach supports organizations in reducing complexity. Reusability and flexibility aspects provide several benefits. The business processes converting into web services and connecting these services to each other in a modular manner will increase the efficiency of the business processes. In addition, this will also decrease development cycle of the software. The flexibility approach offers improving of the competitive position and creating new organization objectives such as providing these web services as business line for external parties. Incorporating the business processes and improving the efficiency of their operations by using existing and new

applications beside each other are also performed by flexibility aspect. Eventually, SOA will reduce the development costs, the maintenance costs of applications and future investments costs at additional innovations.

Challenges

Besides these benefits, SOA implementation is facing different challenges. The architecture is distributed requiring high availability and scalability. Because SOA implementation mostly consists of collaborating web services, SOA specific design patterns should be implemented. The possible interruptions between collaborating web services can lead to revoking of the process. SOA is complicated to manage because of interdependencies of the web services, distributed web services and several interfaces.

Cooperation

Web portals provide entry points to the WWW, offering structured and managed information to the users of internet or intranet. The presentation of the information is also structured. SOA provide web services to perform different kinds of processes. This can occur between several applications using standard protocols.

Web portal and SOA are complementary. SOA aims at the functional aspects, on the other hand Web portal at the presentation aspects.

At this moment, Web portals are acting as a more central spot within which several sources of information are provided.

4 System architecture

The aim of this chapter is to provide answers to the following questions:

- What are the requirements for the web portal supporting information and data sharing for social sciences based on SOA?
- How can a web portal support information and data sharing for social sciences based on SOA?
- How can this web portal be implemented?

The main objective of this section is to define the requirements of SciencePortal architecture that can be applied as a guide to the implementation of the portal. Furthermore, key components like web services and standards for web services are explained. After that, overview of activities which can be applied for implementing web services in a organization, will be described. Finally, reference architecture will be outlined.

Several interviews are held with university docents and researchers to indicate what are their daily activities, how the information technology is used to perform these tasks. Investigations and discovery sessions were organized in order to define their needs related to software use in research activities and how this can be linked with the portal solution. Valuable information about requirements was collected from these interviews.

4.1 Requirements

In this section, the requirements will be described. The requirements in this section will help to design social science portal. The requirements are based on the following categories:

- Functional requirements
- Security requirements
- SOA requirements
- Other requirements

Furthermore, each requirement will be indicated by a requirement priority to show the relative importance. There are three levels: High Priority(HP), Medium Priority(MP) and Low Priority(LP).

- High Priority: The requirement which uses 'HP' is an absolute requirement for system.
- Medium Priority: This kind of requirements may be implemented, provided the timetable or other conditions allow.
- Low Priority: These requirements extend the system with extra features, not necessary for the core operation of the system.

4.1.1 Functional requirements

Based on the interviews with university docents, researchers, scientists and examined literature, the following functional requirements are defined.

Users

- Each person who is part of the science community is a user in the portal.

  Priority: HP

- Each user has a profile which consists of user information like username, first name, last name, password, etc. Furthermore, username should be generated based on community protocols.

  Priority: HP

- A user profile will be created if a user is signed up.

  Priority: MP

- A portal user profile should be assigned to a role, has a default role “member” with default privileges.

  Priority: HP

- The portal should provide member, administrator as roles. Adding new roles also should be arranged.

  Priority: HP

- The portal provides following sections:

    - Profile section
    - Document section
    - Agenda section
    - Community section
    - Forum section
    - Project section
    - Research section
    - Search section
    - Support section
    - RSS section
    - Wiki section
    - Admin section

  Priority: HP

- These sections will be assigned to the roles, for instance a member role can have access to document section, but not admin section. Only administrators have access to admin section.

  Priority: HP

- The portal should have ability to authenticate the users who are authorized to access the portal, also should disallow unauthorized access.

  Priority: HP

- The portal should provide single sign-on authentication. Single sign-on allows user to log into portal once, access to several resources without being prompted to log in again.

  Priority: HP

- All transactions concerning authentication should be registered in logs.

  Priority: MP

- The portal should provide personalized interfaces for users.

  Priority: HP

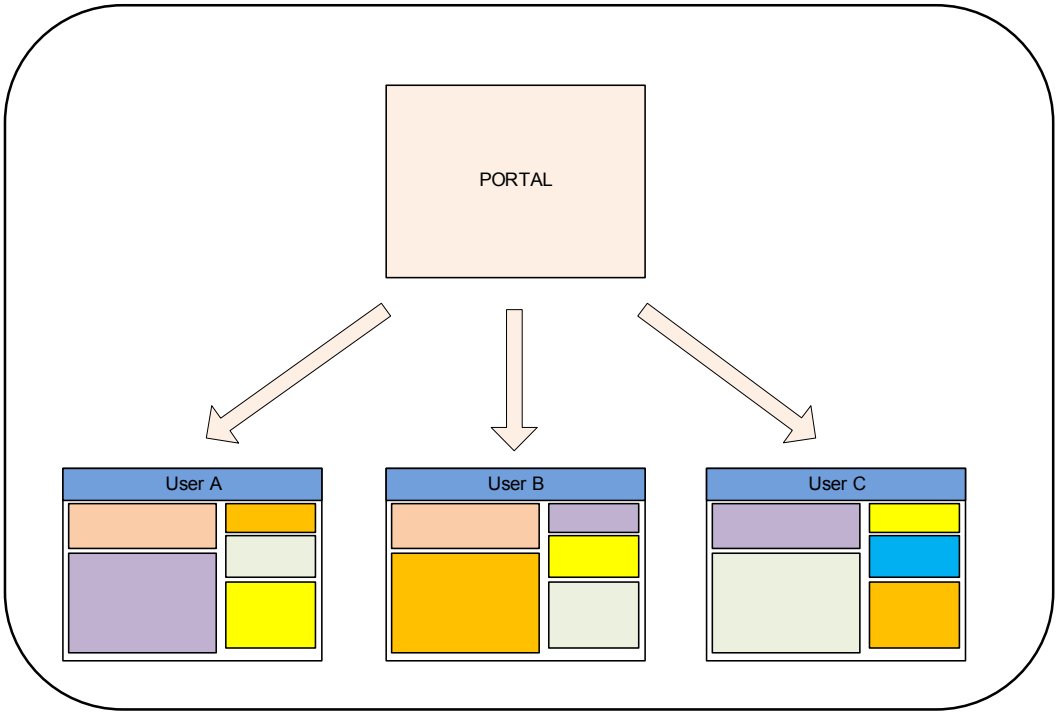


Figure 4.1: Personalization

Profile section

- A user can manage his/her profile page.
    - Topics
    - Notes
    - Agenda
    - Projects
    - Communities
    - Favorite documents
    - Requests to administrators

  Priority: HP

- The users should be able to manage their profiles. They can change their data in this section.

  Priority: MP

- Search agents can be implemented to organize searching documents. These agents can be applied in a certain periods like once a day or twice a day. Depending on these periods, users will receive emails about search results.

  Priority: LP

- Topics can be managed on profile section. Subscribing and unsubscribing can be requested by users.

  Priority: MP

- The users should be able to make notes.

  Priority: MP

- Managing communities should also be provided on this section.

  Priority: MP

- Users can preview documents which they already have added to their favorites.

  Priority: LP

- The users should be able to register possible problems regarding functionality of the portals or possible other requests. These requests will be handled by portal administrators. The users can trace status of their requests on their profile section.

  Priority: MP

- The users should be able to customize portal pages. They can add new information portlets like weather information or news information portlets on their profile section.

  Priority: HP

Documents section

- Users with proper privileges can check in new documents to portal.

  Priority: HP

- Depending on user privileges, upload limit per user can differ.

  Priority: MP

- A user should be able to create a workflow for check in process. Depending again on privileges, different users can review uploaded documents, other users can confirm or reject uploaded document.

  Priority: MP

- Favorite operations like tagging and linking can also be applied by users. Users can add multiple tags and links to a document. With tagging, uploaded document can be associated with a word or phrase; the document can then be found by searching for that tag. Linking provide also other resources or web sites which can be referred to the uploaded document.

  Priority: LP

- Most common formats used by social scientists are Microsoft Office products like Word, Excel and PDF of writing or reading documents. The portal should offer the possibility to work with other format types like LaTex, SPSS, Stata.

  Priority: MP

Agenda section

- The portal should support agenda functionality, the users can manage their agendas.

  Priority: HP

- The portal agenda should be integrated with possible other agendas to prevent duplicate data entering.

  Priority: HP

- Possible alerts regarding events on event can be configured by users. Emails or/and alert messages via mobile can be provided by portal.

  Priority: LP

Community section

- Portal should offer collaboration, which will make remote scientists from several disciplines more visible to one another and should allow the scientists to share their common interests and concerns that can lead to future collaborations.

  Priority: HP

- Communicating with other users should be implemented, this can be arranged via community, forums, blogs, video conferencing, chats or sending messages.

  Priority: HP

- The portal should provide online/offline member status.

  Priority: HP

- Communities can be created by users with proper privileges.

  Priority: MP

- Users can join communities to share information about their tasks, current activities or latest news.

  Priority: LP

- Social science researchers should have the option to post and check the announcements, conferences and job opportunities. Besides that, the researchers should be able to give comments to these sections.

  Priority: LP

Forums section

- Users with proper privileges should be able to create forums. Forums are representing a fundamental platform for social networking and information sharing.

  Priority: HP

- Users can join to forums which they desired to.

  Priority: MP

Project section

- Users with proper privileges should be able to create projects for several social science domains and also able to assign project members to these projects from different communities.

  Priority: HP

- The project section should include separate divisions for internal projects and external projects developed for third parties like industry organizations, information and communication technology companies, public institutions, hospitals and other organizations.

  Priority: MP

- The portal should offer the possibility to third parties involved in projects to be able to log in and check/modify information depending their privileges about projects, project planning, research results, resources allocated, budgets, etc.

  Priority: MP

- The users should be able to keep track of project accomplishment, planning, tasks, resources, budget, information.

  Priority: MP

- Effective communication and collaboration within project teams should be provided by the portal.

  Priority: LP

- The portal should be integrated with other project management systems.

  Priority: MP

Research section

- The portal should be act as an entry point for researchers to perform daily tasks. The existing information systems need to be integrated with portal.

  Priority: HP

- The portal should support multi languages.

  Priority: MP

- The users should be able to get access to desired information or data depending on their privileges. Sharing data, information, experience and expertise can facilitate scientists' researches.

  Priority: HP

- Operations on this data or information should be implemented as services so that the users can request these services to do their work.

  Priority: HP

- Scientists should be able to access datasets created by researchers working on theoretical models and simulations.

  Priority: LP

- The users should be able to generate workflows and assign web services to these workflows.

  Priority: MP

- The research services should be implemented as portal services for the users.

  Priority: HP

- The users should be able to use portal research services.

  Priority: HP

- The portal should provide requested structured information from database to users, the information could be filtered by users.

  Priority: MP

- The portal should provide all social science data resources from different providers. All types of data should be available through the portal. Social science data archives include surveys, censuses, administrative records, direct observation, diaries and etc. Besides that data sets, programs, and any information on empirical analysis, experiments and simulations should also be provided by the portal.

  Priority: HP

- Data management is an expression used by social scientists to encompass the tasks associated with linking related data resources, with coding and recoding data in a consistent manner, and with accessing related data resources and combining them within the process of data analysis. (NCeSS[6], viewed in 2009) The portal should provide activities which will cover a series of applied case studies in social science domains which require significant data management (such as in the processing of specialist information resources on occupations, educational qualifications, ethnicity, social care needs, and e-health databases).

  Priority: HP

- Documentation about data should be provided by the portal in order to maximize use by social scientists. The researcher should be able to use statistical analysis program to produce analytical results and use data documentation program to interpret the results. These programs should be provided as web services. Besides clear documentation; online videos, workshops should be provided by the portal.

  Priority: HP

[6] NCeSS will be used throughout this paper for abbreviation of National Centre of e-Social Science.

- The portal should provide mapping viewers to provide interactive and user-friendly interface to portal users. Different kinds of data which are based on researchers topic, like geospatial data, satellite images, patient images, etc should be displayed on the portal. The integration with external applications like Google maps should also be implemented.

  Priority: HP

- There is a great diversity in the range of methods which are used in social sciences. The following methods should be supported by the portal.
  - Qualitative Interviews
  - Desk research
  - Case study
  - Surveys
  - Participant Observation
  - Ethnography
  - Historical
  - Simulation
  - Experiments
  - Formal modelling
  - Webmetrics
  - Clinical

  Priority: MP

- Scientists are using several kinds of software tools in order to execute their researches. These mostly used software tools should be integrated in the portal.
  - Qualitative Interviews
  - Webmetrics
  - Geographic
  - Visualizing
  - Simulation
  - Content Analysis
  - Integrating
  - Video Analysis
  - Database
  - Quantitative
  - Qualitative

  Priority: MP

- Sharing and analyzing real-time (financial) news should be supported. This should be integrated to help social science researchers to test and evaluate theories about the nature of (financial) news and to support decision-making process depending on these theories.

  Priority: LP

- The portal should be flexible to support research innovation and research infrastructures.

  Priority: HP

- The users should be able to present media such as animation and videos.

  Priority: HP

- The portal should allow users to obtain different kinds of statistics and information. Furthermore, users should be able to generate models based on these statistics and information.

  Priority: HP

- Analytical tools should be integrated in the portal in order to make analyzes effectively of collected data and transform in valuable information for social science. The social science researchers should be able to use these tools collaboratively. The researchers should be able to generate several outputs from these analytical tools.

  Priority: HP

- Text-mining technologies offer the possibility of processing large amounts of textual data systematically. These technologies should be provided by the portal also in order to reduce human errors and to save time.

  Priority: MP

- The several data mining techniques should be provided by the portal for users to research activities.

  Priority: MP

Search section

- The users should be able to browse for documents depending on roles or privileges. The documents should be provided under topics.

  Priority: HP

- Searching for different resource types like images, sounds, etc. should be also arranged.

  Priority: MP

- Furthermore, search functionality should be supported by auto complete or search suggestions functionality of the portal. Besides that, searching activities of portal users will be registered in the database. Users can preview search history logs.

  Priority: LP

- Searching results will be displayed with current number of results and total number of results.

  Priority: LP

- User should be able to specify maximum number of results to be returned, maximum number of results per page to be returned.

  Priority: LP

- User can configure results view by sorting concerned fields.

  Priority: LP

- User should be able to add desired documents to his/her favorites. Via emails, user can recommend desired documents to others users.

  Priority: MP

- User should be able to access scientific journal articles online, books, chapters and thesis.

  Priority: MP

- The portal should be integrated with the most commonly used online search engines like Google, Google Scholar, RePEC and the Social Science Research Network.

  Priority: MP

- Scientists should be able to give ranking to the documents depending on the number of citations and number of downloads. These functionality will indicate them the importance of the documents for other users in the community.

  Priority: LP

Support section

- All requests to administrators can be traced in this section.

  Priority: MP

- All requests are associated with a status like pending, rejected, solved, closed, etc.

  Priority: MP

- The user of the request and administrator will receive emails when the status of the request changes.

  Priority: MP

RSS section

- This section is about collecting information from other sources.

  Priority: HP

Wiki section

- Wiki section should be used to build a knowledge base.

  Priority: HP

- The users should be able to export wiki pages on several forms to reuse then content in other contexts.

  Priority: LP

Admin section

- Only portal administrators have access to admin section. This section includes following sections.
  - Email configuration
  - Role management
  - User management
  - News management
  - Language management
  - Web design configuration

  Priority: HP

- Email configuration should be managed in these sections. Email message templates can be edited by administrators.

  Priority: HP

- Creating new roles, modifying privileges of the roles should be managed in admin section.

  Priority: HP

- All configuration and managing tasks regarding admin section will be handled.

  Priority: HP

- Possible system errors and warnings which may occur, should be registered in logs. System administrators should be able to trace system errors and warnings.

  Priority: HP

4.1.2 Security Requirements

In this section, security requirements about the portal will be given. The portals provide content to several parties like user groups, users and public Internet users. The authorization and authentication is important to provide confidential information to right users.

The following security requirements should be applied:

- The users should be assigned to a specific role.

  Priority: HP

- The portal should provide confidential access to the users to the documents.

  Priority: HP

- The database, files and applications should be located in a secure environment.

  Priority: HP

- The firewall should be applied to provide secure data storage and transfer.

  Priority: MP

- The encryption should be implemented to secure data transfer and authentication.

  Priority: MP

- The authentication should be secured with Secure Sockets Layer (SSL) and Transport Layer Security (TLS). These cryptographic protocols provide security and data integrity for communications.
  Priority: MP

4.1.3 Other requirements

- The portal should be user friendly. Scientists should easily understand how the portal works and be able to execute their tasks without much effort.

  Priority: HP

- The portal should be able to operate on different platforms.

  Priority: HP

- The users can work offline if needed.

  Priority: LP

- The documentation about designing and implementing of services or components should be done based on a methodology to give a clear overview to the concerning parties as designers, developers, testers, etc.

  Priority: MP

- Back-ups should be performed on a regular and automatic basis.

  Priority: HP

- A clear explanation about the components and their relationships should be available along with all other relevant documentations.

  Priority: MP

- Before the system goes live, all bugs and designs issues should be clear. It also should be clear who is responsible for bug fixing.

  Priority: HP

- The portal should react on the requests in a short time as possible.

  Priority: LP

- The portal services and resources should be 7/24 available for the users.

  Priority: HP

- Scalability is also relevant point. The capacity of the portal could be extended by adding new services or resources.

  Priority: MP

- The portal should be monitored regularly, the possible problems should be solved in a short time.

  Priority: LP

- Upgrading hardware and software should cause no problems.

  Priority: LP

4.1.4 SOA requirements

- Supporting existing systems: the existing systems shouldn't be removed, these systems should operate further. With the new architecture, the existing systems should be integrated.

  Priority: HP

- Integration: this is about all integration from different topics
  - User interaction integration to provide a simple user experience.
  - Applications integration to provide flexible communication between applications.
  - Process integration to regulate applications and services.
  - Information integration to manage the data.

  Priority: HP

- Incremental implementations: this is about step for step implementation from developing the architecture. Many integration projects have not succeeded because of the complexity and costs.

  Priority: MP

- Using standard framework: For reusing and structured implementation, a standard framework should be used. With a standard framework, the implementation can be done efficiently.

  Priority: LP

- Providing a clear and documented API to allow external applications to use data from the portal.

  Priority: LP

4.2 Web services

Web services are a promising implementation of the service-oriented architecture, intended to provide a standard means of interoperating between different software applications, running on a variety of platforms and/or frameworks. This interoperability is gained through the use of asset of XML-based open standards for defining, publishing and using web services. (Protogeros, 2008)

In this section, web service types which can be applicable on SciencePortal are outlined. Furthermore, standards for web services which are parts of the architecture and support the interaction between a web service requester and provider, are described in this section.

A service offers a recognizable business logic or data. What is recognizable? The applications provide enormous differences in syntax and semantics of data via the service layer. Something which is called a company in a application, is called a organization in the other application. An attribute in one application is defined as a numeric field of 15 characters, in the other an alphanumeric string. Introducing a service layer means that these differences are showing up and that translations are necessary. The modern Enterprise Service Busses can manage this. It provides an enormous challenge, but also the chance to work slowly towards a uniform, company-wide conceptual framework in which future services can be defined. Enterprise Service Bus uses XML for messaging. Common object model is registered in a XML Schema Definition (XSD). Then the translation of an application to this model is implemented through a tool based on an XML or a programming language.

Service types

According to LigtHart, There are three major standard service types. (LigtHart, 2005) New standard service types can be added in this structure. Integrating service is added under Mutating. After this standard service types, SciencePortal related services which will support portal, will be described.

Inspecting:

- Search: based on a fixed set by the owner is a collection of information will be obtained.
- Select: selecting of object based on selection parameters. This service returns a list of objects which meet the criteria.

Mutating:

- Register: by this service type, a object will be registered and initial data will be stored for example in the database.
- Update: it is about changing of the data of a certain object and possible relations.
- Delete: deleting is about to removing definitely from the database.
- Finish: this service will terminate the active life cycle of the object.
- Transform: this is about changing the status of a object. This is also called *transition*. The transition name is usually a part of the name of service like acceptingOrder.
- Integrate: integrating service is about integration of other services. Two or more services can be linked together.

Generating:

- Generate: it is about generating a instance of object, but it is not persistent, so it is not stored in the database. This instance will be returned to calling consumer. It is like generating default values for a object.
- Select value: this service provides a limited set of possible values for a certain attribute.
- Validate: it is about different kinds of validations like checking whether the object exists in the database.
- Calculate: this service calculates data.

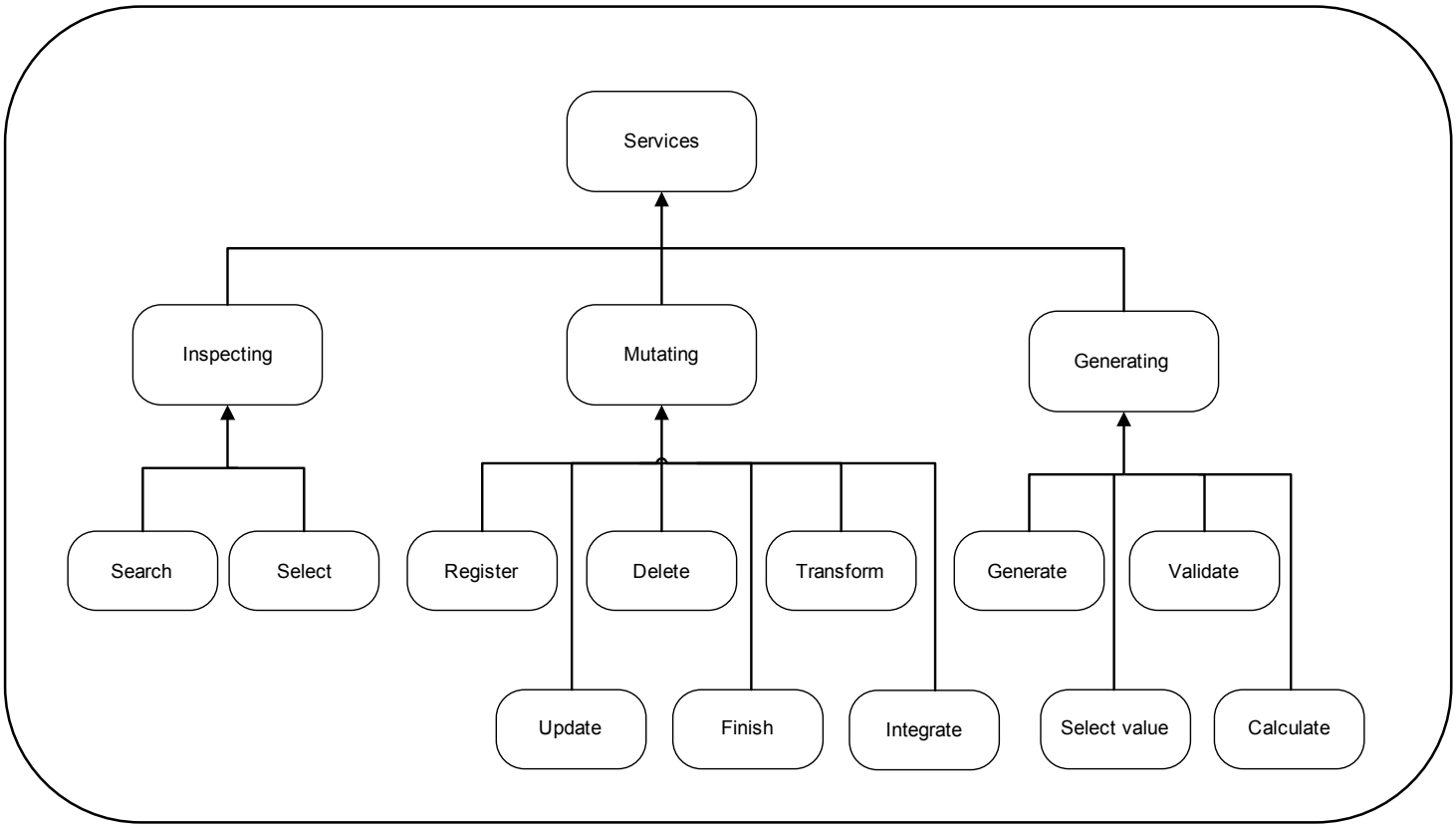


Figure 4.2: Standard service types

4.2.1 SciencePortal services

Depending upon requirements which are defined in the previous section, following SciencePortal services should be implemented to provide desired functionalities as services.

4.2.1.1 Portal base services

These portal services are meant for common use for different users of the portal. As displayed in Figure 2.3, these services provide the core functionality of the portal.

Authentication service

The authentication service should have ability to authenticate the users for accessing to the portal. Based on single sign-on authentication, a user can log into portal once, access to several resources, without log in again for other resources. All transactions regarding authentication will be registered via this service in database.

Personalization

This service should provide appropriate information to authenticated users. Based on roles and privileges of users navigation, content and menu structure will be provided by personalization service.

Search

This search service should provide search functionality for portal users. Depending on roles and privileges of users the limited information should be returned to the portal. This service can include searching content, documents and databases.

Document

Content, document and knowledge will be provided by this service. Collection and structuring of knowledge are also here arranged. External document and content management systems can also be integrated for portal users.

4.2.1.2 Collaboration and communication services

The collaboration and communication services will include desired agenda, community, forum, blogs and email functionalities. Integration is key component for these services, integration with other systems or portals will be implemented in these services.

4.2.1.3 Data services

Data services will provide access to content in databases. Based on search reference, data services will return concerning data. Data services often maintain indexes to support quick finding process of resources by name or other attributes of the resource. Data services will also support several data types like images, sounds, etc. Again, integration with multiple databases, external systems and portals should be provided.

4.2.1.4 Research services

Research services address to daily tasks which users of portal are performing. Several services conform research activities like analyzing, visualizing data, etc. will be placed under research services. Furthermore, data viewers to display several kinds of data like geospatial data, satellite data, patient images, etc should be implemented as services. Besides that, generating models based on information and statistics should be provided as a service. As mentioned in the previous services, integration between other systems should be applied.

4.2.2 Standards for (web)services

The technology behind web services is only one of the possibilities to implement the services. There are also technologies which might be used like RPC's, message-based integration, Java Remote Method Invocation (RMI) en Microsoft .NET remoting. However, web service technology is leading in great tempo.

For standardization in the field of Web services a global consortium is established, called Web Services Interoperability Organization (www.ws-i.org). This consortium consists of almost 150 companies from different industrial zones, have aim to promote web services by making generic protocols for messaging and stimulating the implementation of web services.

The core of the Web services architecture consists of specifications, such as XML, SOAP and WSDL that support the interaction of a web service requester with a web service provider. The potential discovery of the web service description can be realized as a universal description, discovery and integration (UDDI). (Protogeros, 2008)

The following technologies are considered in relation to web services.

Hyper Text Transfer Protocol (HTTP)

HTTP is the most used protocol for the transport of web services requests and responses. Because HTTP is standard transport protocol for internet, every web server and every client on web browser uses it. The latest version is stabile and under administration of World Wide Web Consortium.

Extensible Markup Language (XML)

XML is a key requirement for a standardized, platform independent, flexible and extensible language which is derived from Standard Generalized Markup Language (SGML). XML is also under management W3C. XML become actually the language for transferring data via internet, also for web services.

Simple Object Access Protocol (SOAP)

SOAP is a XML based protocol to provide structured messages in a distributed environment. It describes the envelope of a message and the rules, how they should be handled by the receiver. SOAP provides the communication with web services. In addition, a transport binding framework is also defined by SOAP for exchanging envelopes using network protocols as HTTP, SMTP, FTP, RMI/IIOP. SOAP is also under management of W3C.

Web Services Definition Language (WSDL)

WSDL describes web services so that the possible consumers know what the web service does and how this web service should be called. A WSDL description is a first step to identify all characteristics of the web service. The message which is exchanged between requester and provided, is defined abstractly and then link to physical deployment information to provide an endpoint. The messages are usually bound to SOAP and HTTP protocol. Because WSDL uses XML based notation to describe message formats, the service which is described, can be mapped to any platform or messaging system.

WSDL is also based on XML and under management of W3C. The web services are described based on the following components:

- Interface information about all public functions
- Datatype information about all message requests and responses
- Binding information about the transport protocol
- Address information to locate specific service

Universal Description, Discovery and Integration (UDDI)

UDDI provides a standard method for registering and publishing web services. In other words, UDDI is the common registry directory where the web services are registered and published. This method is queried by SOAP messages and provide access to WSDL documents.

The organizations can register their web services and search for other web services of other organizations. UDDI registries can be deployed in three ways:

- Public
- Extra-enterprise
- Intra-enterprise

Intra- and extra-enterprise are private registries deployed by an organization or business partners. But public registries are acting as a resource for Internet-based web services and test environment for developers of web services. Big vendors like Oracle, Microsoft, SAP and IBM are providing public registries. UDDI is under administration of OASIS-sonsortium.

Enterprise service bus (ESB)

Traditional middleware is not sufficient to support SOA applications. Although the products like RPC's, ORB's and Message Oriented Middleware (MOM) perform well, the lack of industry-wide standards make this technology insufficient suitable. The Enterprise service bus (ESB) offers a better functionality in comparison with previous middleware products.

An ESB is a service technology-based infrastructural facility that supports direct intelligent communication between loosely coupled en organization components. In SOA, all communication takes place in the form of messages. The term 'mediation' is often used, make clear that ESB is a intelligent software that can assist the messaging around handling service requests and responses. In a ideal situation services are invoked in a uniform way, the requester does not need to have knowledge of location, platform and technology of the called service. This is all hidden by ESB. ESB set communication where different patterns should be supported (such as asynchronous and synchronous communications, request / reply, fire / forget mechanisms) and different protocols can be applied (such as SOAP, HTTP, JMS, RMI). Furthermore ESB provide quality guarantees for security, guaranteed delivery of messages, performance and transaction integrity.

4.3 8 steps to web services

The following 8 steps will provide overview of activities conform implementing web services in a organization. Actually, the current situation should be transformed to desired situation in a structured way. Dividing whole process into a number of steps will also support decision making between these steps. It depends on the situation in which an organization is, which problem can occur during the implementation, how long it will take certain steps or the steps which are not relevant for the organization, can be omitted. (Noordzij, 2004)
With these 8 steps a manager of an organization is able to give clear instruction to what needs to be done, while a direct reporting structure is created to monitor progress.

Figure 4.3 shows eight steps which is necessary to implement web services in a structured way. Each step is described after this figure.

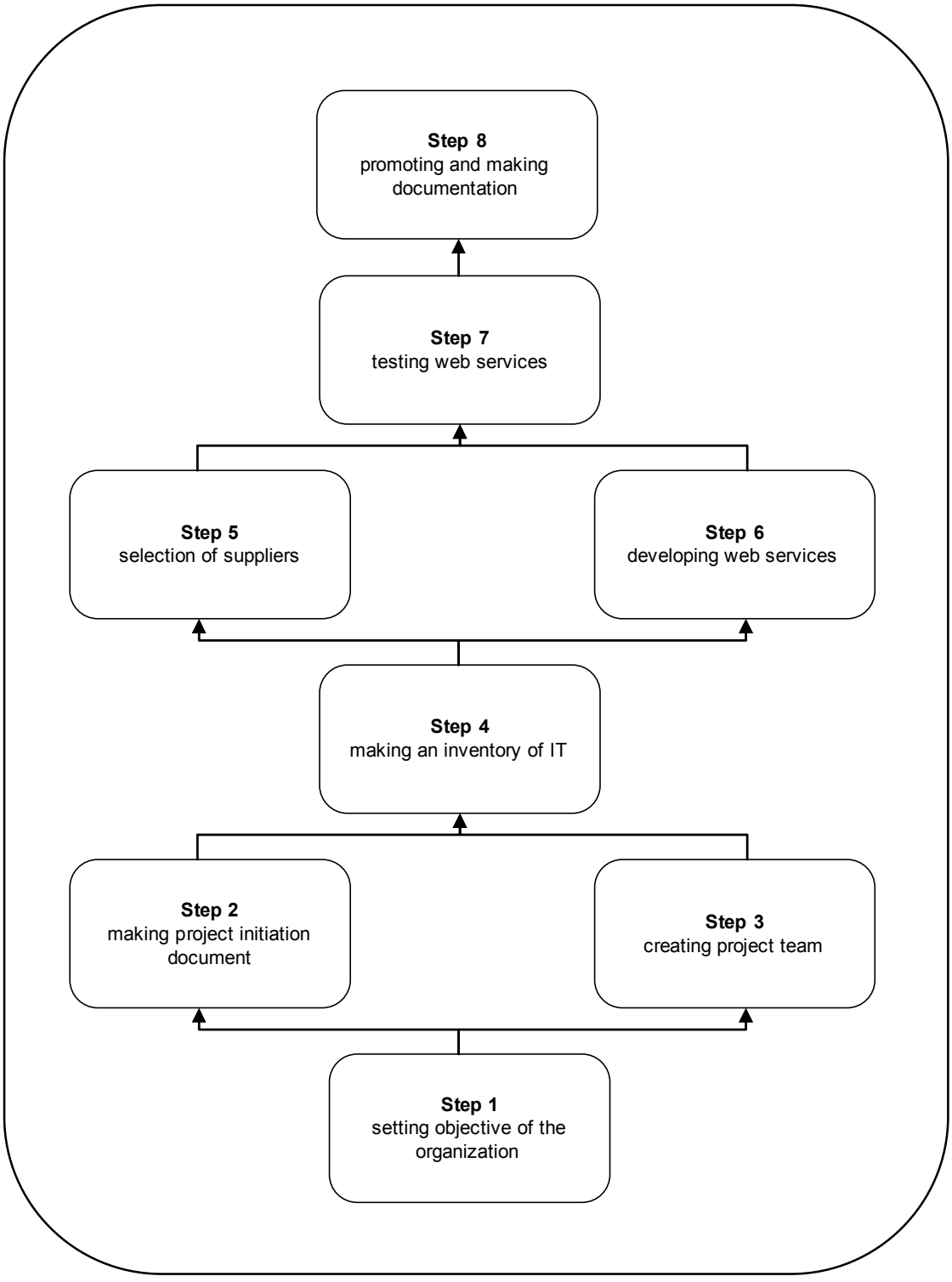


Figure 4.3: 8 steps

Step 1: setting objective of the organization

Setting objective of the organization, one for short term, one for long term, provides why web services are applied and also how they are implemented. The preparedness of the organization should be determined for the right selection to perform: in which sections of the organization web services will be applied, which costs should be covered and finally exploit advantages.

Step 2: making the project initiation document

Project initiation document consists of several steps and describes essentially how the implementation of web services will be applied. Also which improvements are expected, are explained. In this step, choice of a method which will be used for designing, developing and implementing, is provided. It can be also chosen for a custom method for these activities. Schedule for planning will be outlined.

Step 3: creating project team

Just before the start of the project, the team members will be selected for project team. Furthermore the organization should be informed about the activities which will be started. The composition of the project team can be arranged in different ways, besides, in addition to a good guidance, a powerful start is always welcome.

Step 4: making an inventory of IT

The first task of the project team is to determine which IT facilities are available and how web services should be applied with these facilities. Also the migration and implementation scenario is here chosen.

Step 5: selection of suppliers

There are several suppliers in the market each with its own product range with specific advantages, disadvantages and areas of deployment. In this step a matrix will be built based on the previously stated objectives in which the various deals are made.

Step 6: developing web services

The web services are developed in this step. The project team will develop web services in a prototype. It is not only about only developing but also experience which is gained, should be recorded for later use. The development of web services should be based on a specific framework or method.

Step 7: testing web services

This step is involved to all development projects. Testing of developed web services should be done carefully also based on a test plan.

Step 8: promoting and making documentation

Promoting and documentation will provide the organization to make profit of advantages of web services, while the web services platform remains manageable.

4.4 Reference architecture

SciencePortal reference architecture is shown in figure 4.4. This reference architecture contains four different classes of services that support the requirements of the portal.

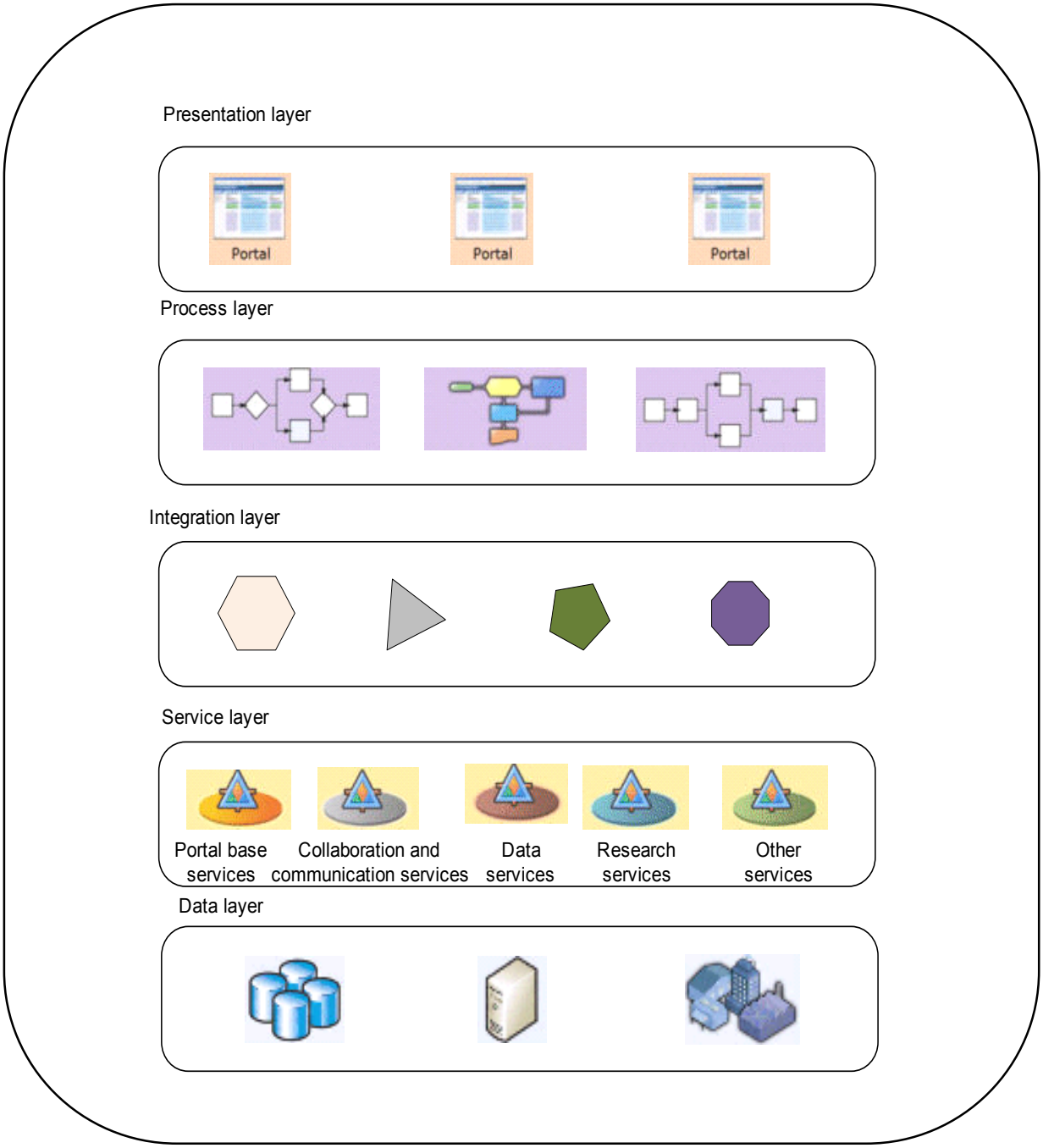


Figure 4.4: SciencePortal Reference Architecture

Different form SOA reference architecture, SciencePortal Reference Architecture includes a extra layer, Integration layer. This layer supports the integration between process and service layers. From this layer, the services can be provided in the service layer.

The integration layer is very important as an intermediary to and between the applications. In a complex environment with many different business applications integration is indispensable. When a user wants to use a service, it can through the integration layer, which then provides the link with the applications.

Service layer provides actually the functionality and data from the data layer. Services are provided by the business components and ensure that the internal operation and structure of the business component is encapsulated. This means that a third-party may user business services without the internal operation of the business components or knowing structure of the business object.

5 SciencePortal research validation

The emphasis of this chapter is to provide answer to the following question:

- How can this research be validated?

In this section, qualitative validation which is executed to validate the functional requirements, will be explained. After that SciencePortal prototype will be described.

Qualitative validation is aiming to obtain reliable information about the functional requirements. This type of validation provides more information about demands, needs and opinions of the social science researchers. This validation is executed with a social scientist in a university. Based on the validation, the functional requirements and priorities regarding them were confirmed and updated where it was necessary. The details regarding validation can be found below.

- Single sign-on authentication is preferred, but without the limitation of the resources. In the past social scientists were experiencing difficulties in accessing several resources if they were choosing single sign-on authentication. They didn't get access to some resources because single sign-on authentication was not integrated to those sections. It is important that single sign-on authentication should be integrated with all sections in the portal.

- Priority for support activities like registering possible errors and sending requests to the portal's administrator was updated from LP to MP. Social scientists consider that the problems regarding portal's functionality should be solved as with higher priority.

- The portal should provide all social science data resources from different providers. This requirement can be obstructed by licensing policy. The social scientists from different countries of the world have access to the portal and each country has own protocols to regulate licensing. This can lead to problems when sharing data resources.

- Security requirements are very important for social scientists. The portal environments should be very secure, all security options should be provided. If the social scientists experience the portal unsecure, then they will not use the portal.

A small prototype is implemented to validate SciencePortal Architecture. “Register User” and “Display Dutch Painters Location” process will be explained with activity diagram. Furthermore the implemented web services will be outlined. This process will be referred to SciencePortal Architecture. The portlets of this prototype will be shown in Appendix A.

The following technology is used for implementing this prototype:

- PHP
- MySQL
- HTML
- CSS
- AJAX
- Javascript
- XML
- Google maps

A user wants to register self via a portlet. This will be displayed with a activity diagram. The activity diagram outlines how the process is followed. This diagram describes a certain activity form the initial state till the end state.

First, user requests “register user portlet”. Web server serves a portlet with user default values. These default values can be generated by parameters like IP address of user. After that, user will enter required data and requests register process. In web server, the entered data will be validated. Depending on business logic, a xml will be generated and served back to portlet. Via AJAX, the portlet will display warnings. If the entered data is validated, the entered data will be stored in database. Furthermore, “send emails” process will be executed, which sends emails to the concerning parties like user and portal administrators. Eventually, user can activate his/her account.

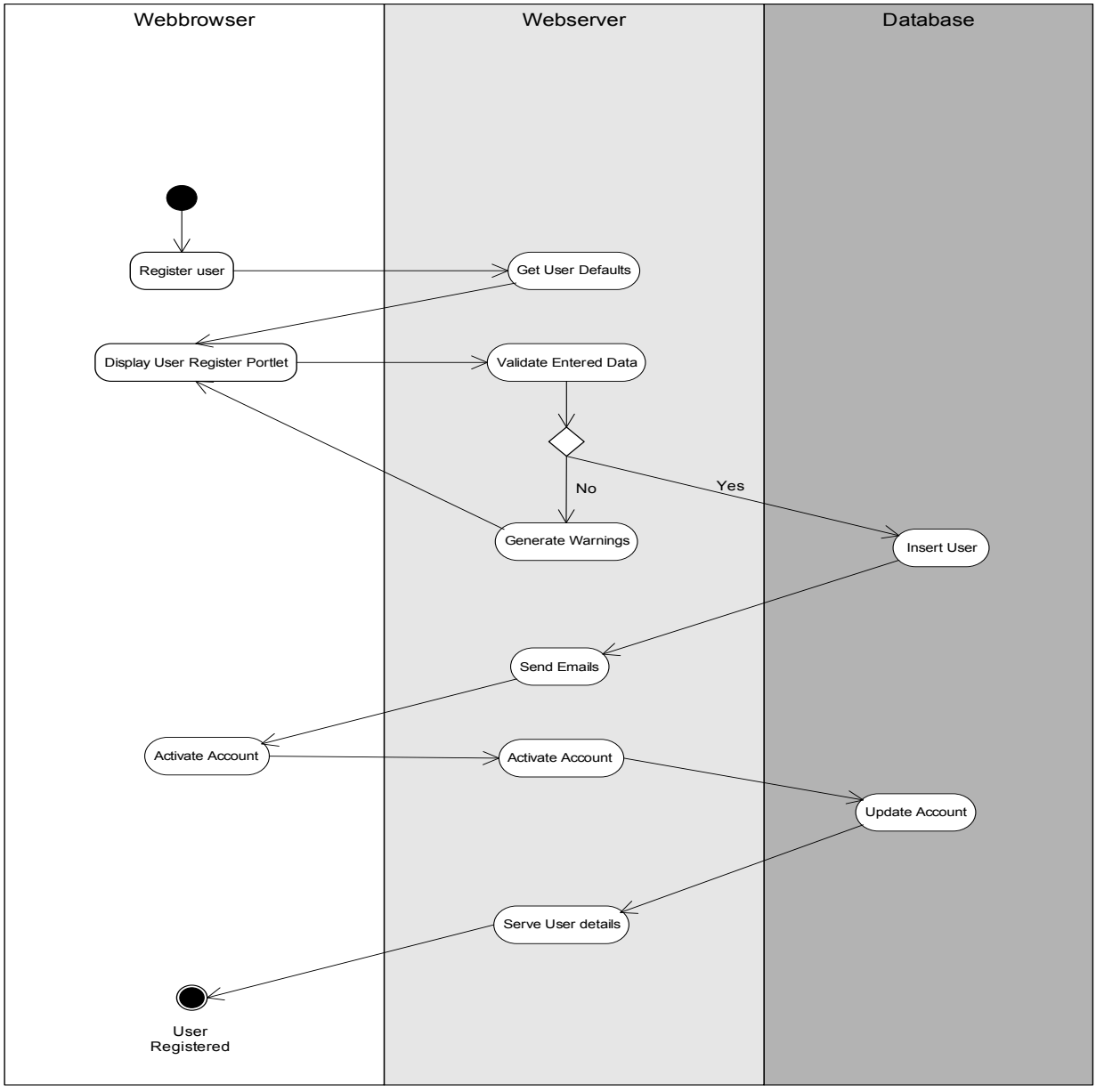


Figure 5.1: Activity diagram

The following web services are implemented to support “Register User” process.

Web service	Description
Get User Defaults	This web service generates defaults value for inserting user data.
Validate User	“Validate User” web service validates entered data which is sent to this web service. Depending on business logic, entered data will be validated on constraints like required, unique, numeric, email address validation. Output of this web service will be send to “Generate XML” web service to create a XML. In other words “Generate Warning” web service will be called, if validation doesn´t succeed.
Generate XML	A XML will be generated based on item and warning type. Item can be “cLastName”, Warning type: “required”.
Generate Warning	This web service is integrated web service to call “Validate User” and “Generate XML” web services. Depending on output “Validate User”, a XML will be generated.
Insert User	“Insert User” web service will store data in the database. This database can be located anywhere. The important point is that this web service should have access to the database.
Send Emails	This web service will send emails to the desired parties like new created user and portal administrator. The email configuration can also be retrieved from another web service. The attributes like subject, priority, content of the email can be manipulated.
Activate Account	This web service will activate the user account.
Update User	This web service will update user data in the database, stored when the user account is activated.

Table 5.1: Web services

The following figure displays SciencePortal Architecture for “Register User” process.

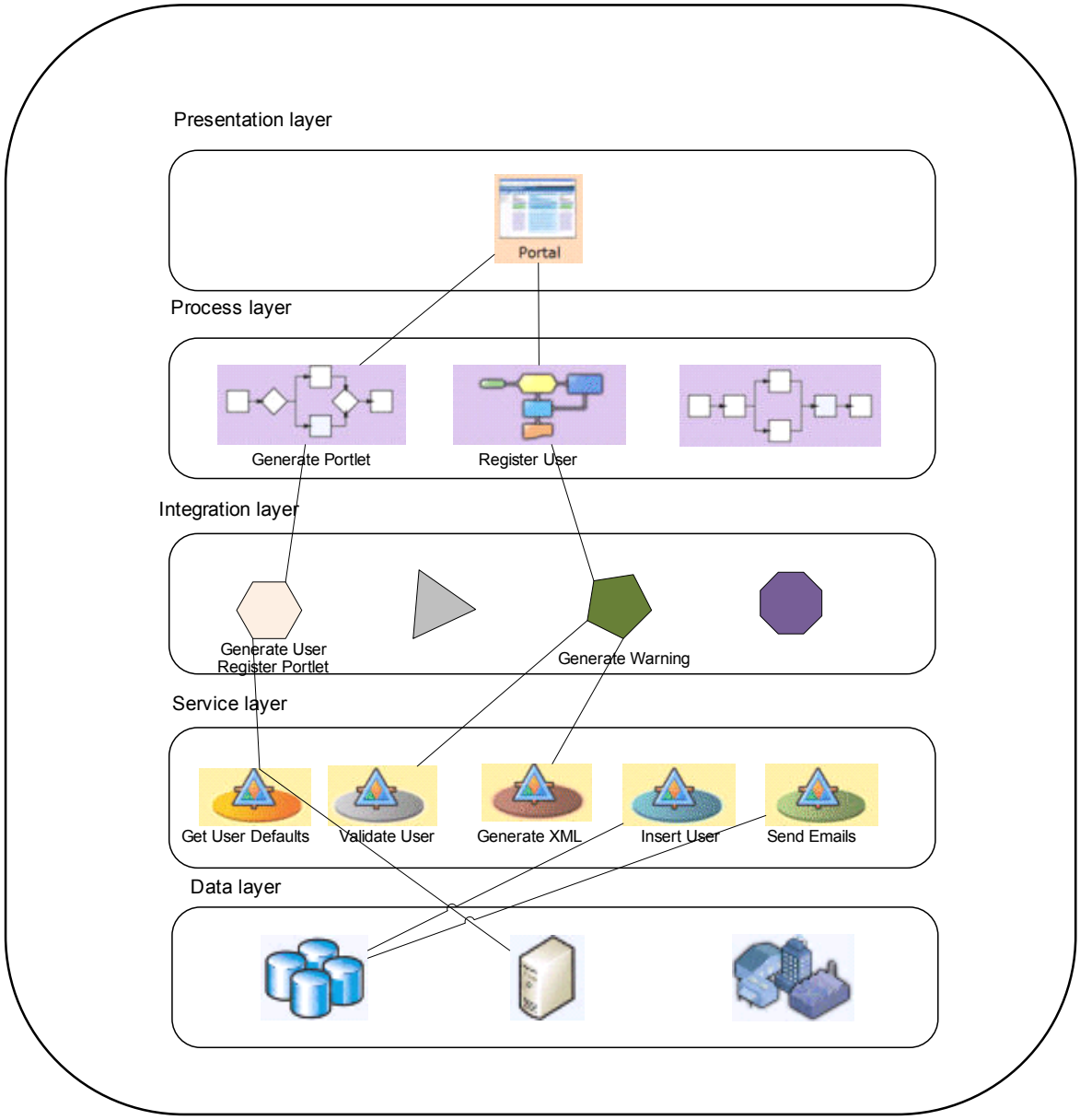


Figure 5.2: SciencePortal Architecture

A user wants to see Dutch Painters Location. The activity diagram will be shown to get overview of the process.

First, user requests “display Dutch painters location portlet”. Web server requests a dutch painters location information from the database. After that “Generate Map” web service will be requested. Furthermore, Dutch painters location information will be served to “Generate Map” web service. Finally, the user will see map of Dutch painters location.

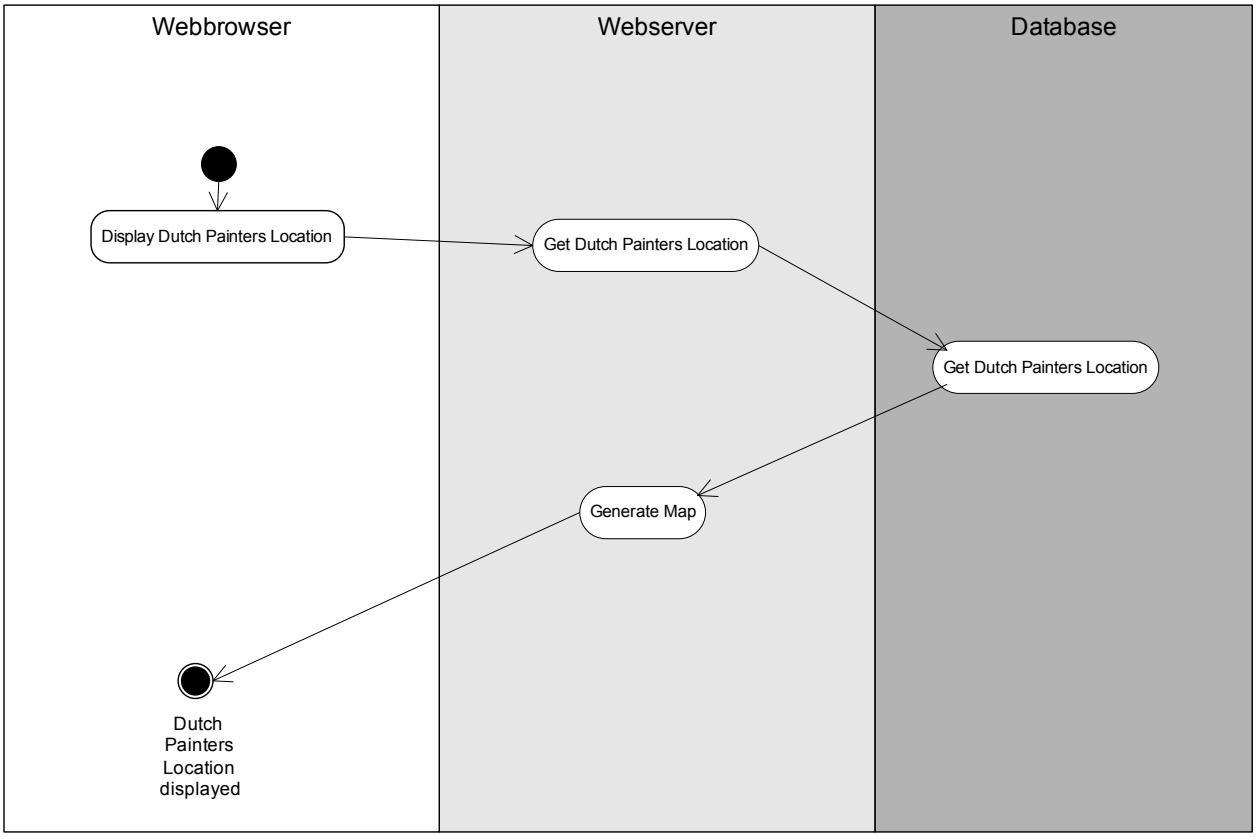


Figure 5.3: Activity diagram

The following web services are implemented to support “Display Dutch Painters Location” process.

Web service	Description
Get Dutch Painters Location	This web service retrieves Dutch painters location information from database and generates XML.
Generate Map	“Validate User” web service validates entered data which is sent to this web service. Depending on business logic, entered data will be validated on constraints like required, unique, numeric, email address validation. Output of this web service will be send to “Generate XML” web service to create a XML. In other words “Generate Warning” web service will be called, if validation doesn´t succeed.

Table 5.2: Web services

The following figure displays SciencePortal Architecture for “Display Dutch Painters Location” process.

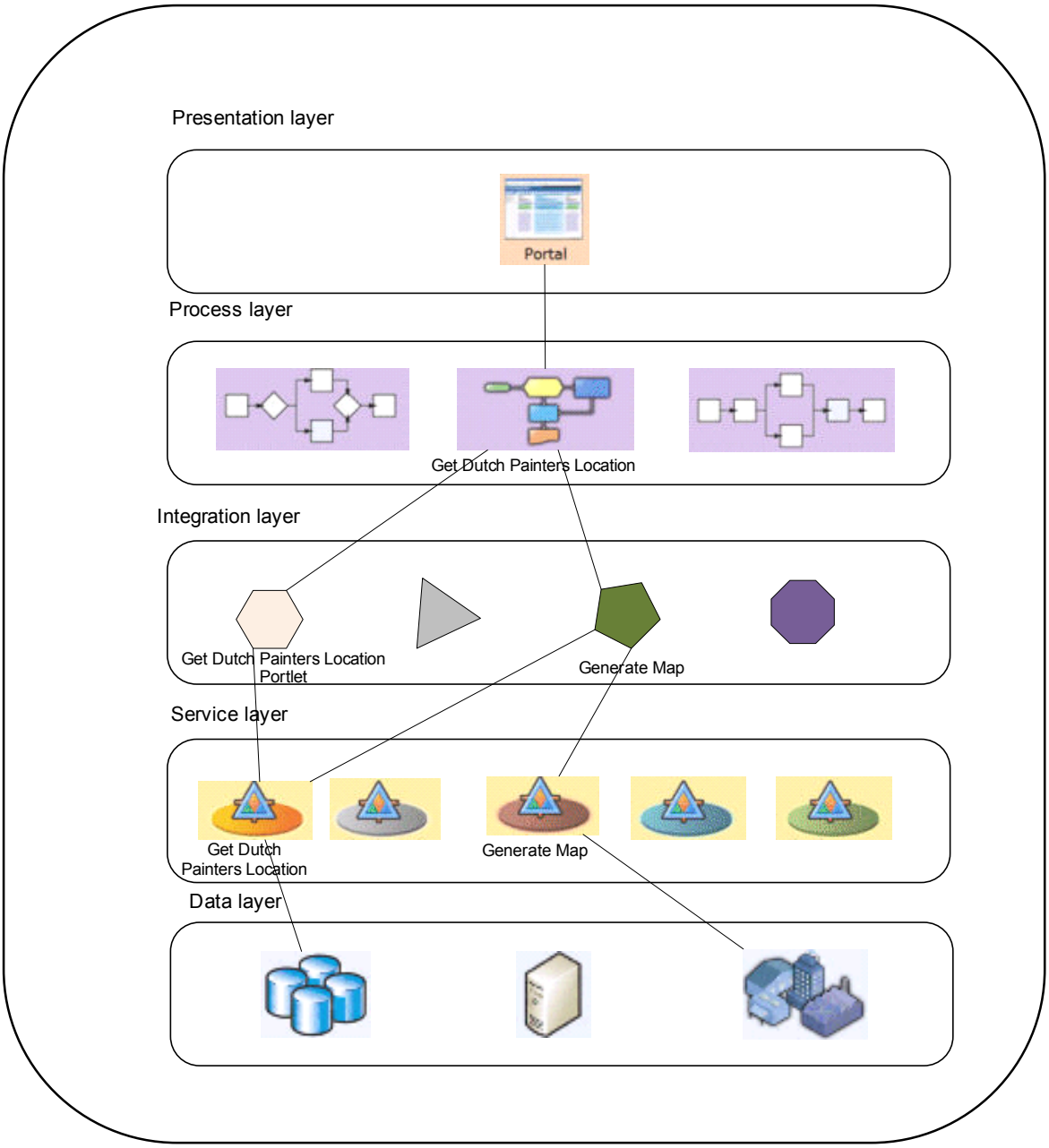


Figure 5.4: SciencePortal Architecture

6 Conclusion

The main aim of this study is to design web portal which supports information and data sharing for social sciences based on service-oriented architecture. During this study, the specific activities are applied to define reference architecture of SciencePortal.

First, extensive literature study about web portal, SOA and web services are done. How web portal concept works and how this can be applied as scientific information and data sharing instrument, is examined. Furthermore, SOA and web services as key component of SOA are researched in order to gain better overview regarding architecture.

Interviews were held with university docents and social science researchers to indicate the functional requirements of the web portal. The functional requirements of SciencePortal architecture are defined in a such way that can be applied as a guide to the implementation of the portal. Based on these requirements, reference architecture is designed. Besides that the thesis contains 8 steps guidance to an overview of activities to support implementing web services in a organization.

Finally, to validate the research, a qualitative validation is executed to confirm functional requirements of the web portal with a social science researcher. Besides that, SciencePortal; a simple prototype is implemented to validate the architecture of the web portal. The technology which is used for this portal is PHP, MySQL, HTML, CSS, AJAX, Javascript, XML and Google maps.

The portal concept is very popular. Many industrials branches, education sectors, business vendors, healthcare companies, government departments already have portals or they are in the process of implementing them. There are several opportunities of using portals in the organizations. The functional requirements which are defined in this thesis, can be extended with new functional requirements. All other requirements are defined to support functional requirements, these obviously can also be updated if needed.

The SOA concept and its application in organizations is still in progress and will still need time to prove its value for organizations. SOA requirements show how SOA should be integrated with the portal. Before the organizations start with SOA implementation, the 8 steps defined for supporting implementation of web services should be examined carefully.

Services should be implemented independent, this way they can provide decoupling of processes and applications. It will certainly take several years to get benefits of SOA implementations. The expected benefits include reusability, agility and interoperability. The expected challenges may arise in governance, costs related to implementation and maintenance but the benefits will be visible on a medium or long timeframe.

References

1. Allen, P. (2006) *Service Orientation*, Cambridge
2. Baresi, L. and Di Nitto, E. (2007) *Test and analysis of web services*, Springer
3. Barry & Associates, Inc., *Web Services and Service-Oriented Architectures*, website : < http://www.service-architecture.com/>, viewed in mart 2008
4. Channabasavaiah, K., Holley, K. (2004) *Migrating to a service-oriented architecture,* IBM
5. Chau, M., Huang, Z., Qin, J., Zhou, Y. and Chen, H. (2006) *Building a scientific knowledge web portal:The NanoPort experience***,** Decision Support Systems 42, pp.1216–1238
6. Demirkan, H., Kauffman, R., Vayghan, J., Fill, H., Karagiannis, D. and Maglio, P. (2008) *Service-oriented technology and management: Perspectives on research and practice for the coming decade*
7. Drew, C. J., Hardman, M. L. and Hart A.W. (1996) *Designing and conducting research*, University of Utah
8. Eduventures (2006) *Service-Oriented Architecture and Web Services: The Next Big Thing in University Enterprise Computing*
9. Edwards, J. (2004) *Leveraging web services*, Amacom
10. Fensel, D., Lausen, H., Polleres, A., de Bruijn, J., Stollberg, M., Roman, D. and Domingue, J. (2007) *Enabling Semantic Web Services*, Springer
11. Gordon, S. (1991) *The history and philosophy of social science*, Routledge

12. Hak Kuk, S., Soo Kim, H., Lee, J., Han, S., Park, S. (2008) *An e-Engineering framework based on service-oriented architecture and agent technologies*

13. IBM (2008) *Model and build ESB SOA frameworks*, website :

< http://www.ibm.com/developerworks/web/library/wa-soaesb/>, viewed in august 2008

14. Kandaswamy, G., Fang, L., Huang, Y., Shirasuna, S., Marru, S. and Gannon, D. (2006) *Building web services for scientific grid applications*, IBM Journal of Research and Development. Vol. 50, No. 2/3, pp. 249-260.

15. Ligthart, A., van Es, R., van Gerwen, N., van Rooij, R. and Graave, J. (2005) *Service-oriented architecture,* Ordina System Integration & Development

16. Line, M. B. (1980) *Design of information systems in the social sciences,* Bath University

17. Meyer, E. T., Dutton, W. H. (2009) *Top-Down e-Infrastructure Meets Bottom-Up Research Innovation: The Social Shaping of e-Research*

18. Noordzij, M., van Krugten, P. and Hoogenboom, M. (2004) *Web services migration and implementation*

19. OASIS (2006) *Reference Model for Service Oriented Architecture 1.0*

20. Pierce, M., Geoffrey, F. (2003) *Scientific Applications as Web Services: A Simple Guide*, Indiana University

21. Protogeros, N. (2008) *Agent and web service technologies in virtual enterprises*, Information science reference

22. Sample, J. T., Shaw, K., Tu, S. and Abdelguerfi, M. (2008) *Geospatial services and applications for the Internet*, Springer

23. School of Business, University of Applied Sciences Northwestern Switzerland (2007) *Accelerating Transition to Virtual Research Organisation in Social Science*

24. Shrum, W., Genuth, J. and Chompalov, I. (2007) *Structures of scientific collaboration*, Cambridge

25. Skinner, A.S. (1996) *A system of social science,* Oxford

26. Sprott, D., Wilkes, L. (2004) *Understanding Service-Oriented Architecture*

27. Torres-Vitolas,C., Blake, M. and Shipsey, F. (2008) *NEEO User Requirement Report*

28. Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E. and Olson, G. (2007) From Shared Databases to Communities of Practice: A Taxonomy of Collaboratories

29. The Open Group (2008) *The Open Group Architecture Framework*

< http://www.opengroup.org/architecture/togaf8-doc/arch/>, viewed in december 2008

30. Torsten, P. (2005) *Building integrative enterprise knowledge portals with semantic web technologies*

31. Trommler, M., Wagner, W., Scipal, K. and Kidd, R. (2002) *Scientific Information Portal in the World WideWeb: The Global Soil Moisture Archive 1992- 2000 from ERS Scatterometer Data*

32. Yang, X., Allan, R. and Finch, J. (2007) *The NCeSS Portal - a Web 2.0 Enabled Collaborative Virtual Research Environment for Social Scientists*

Used abbreviations

API	Application Programming Interface
ASF	Apache Software Foundation
ASP	Application Service Provider
BPEL4WS	Business Process Execution Language for Web Services
BPMN	Business Process Modeling Notation
HTTP	Hyper Text Transfer Protocol
ICT	Information Communication Technology
IDL	Interface Definition Language
IEEE	the Institute of Electrical and Electronics Engineers, Inc.
IT	Information Technology
NCeSS	National Centre for e-Social Science
OASIS	The Organization for the Advancement of Structured Information Standards
RMI	Java Remote Method Invocation
RPC	Remote Procedure Call
RSS	Really Simple Syndication
SOA	Service-Oriented Architecture
SOAP	Simple Object Access Protocol
SGML	Standard Generalized Markup Language
SSL	Secure Socket Layer
TLS	Transport Layer Security
UDDI	Universal Description, Discovery and Integration
W3C	World Wide Web Consortium
WSDL	Web Service Description Language

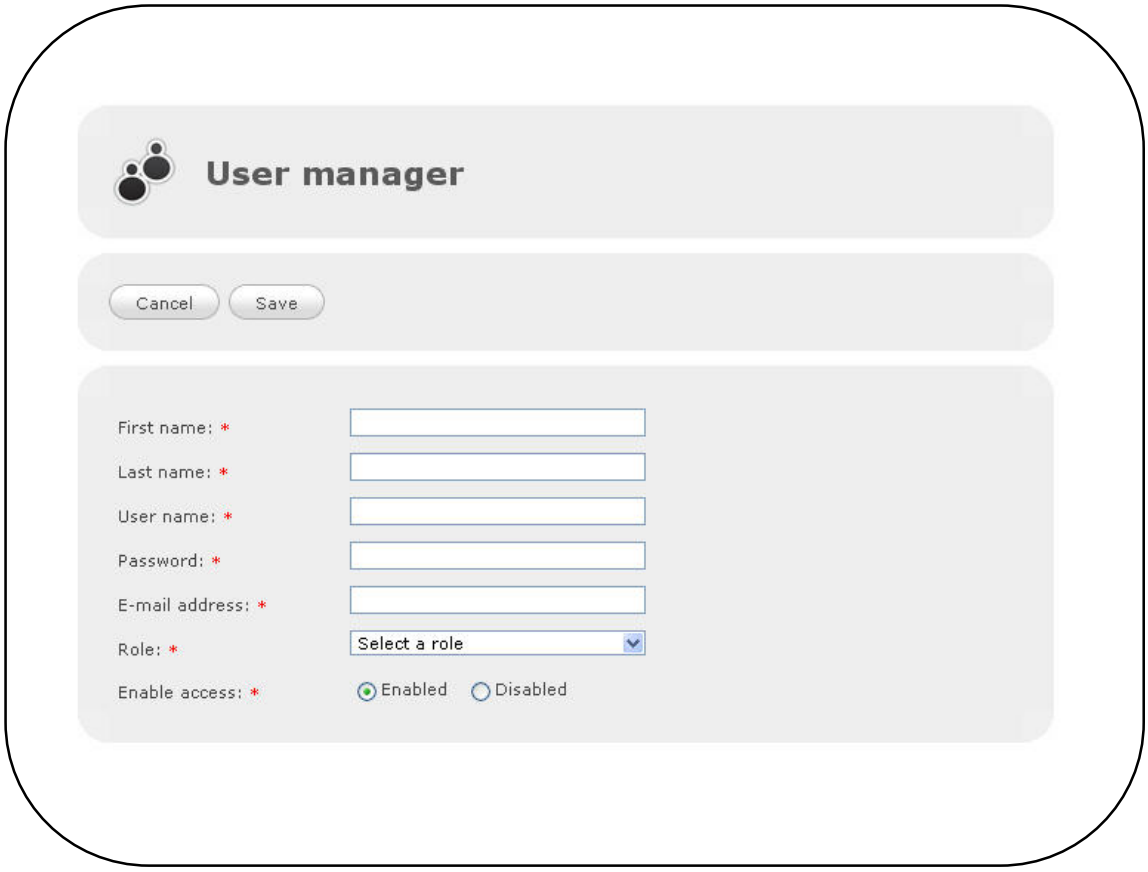
WWW	World Wide Web
XML	eXtensible Markup Language
XSD	XML Schema Definition

Appendix A – Register User

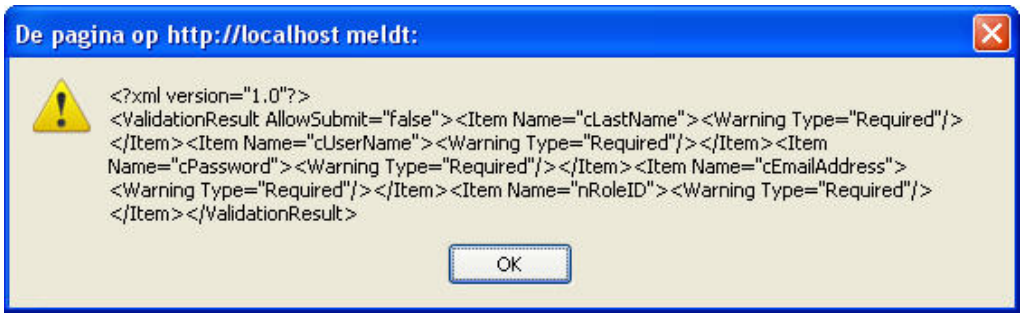
&

Display Dutch Painters Location

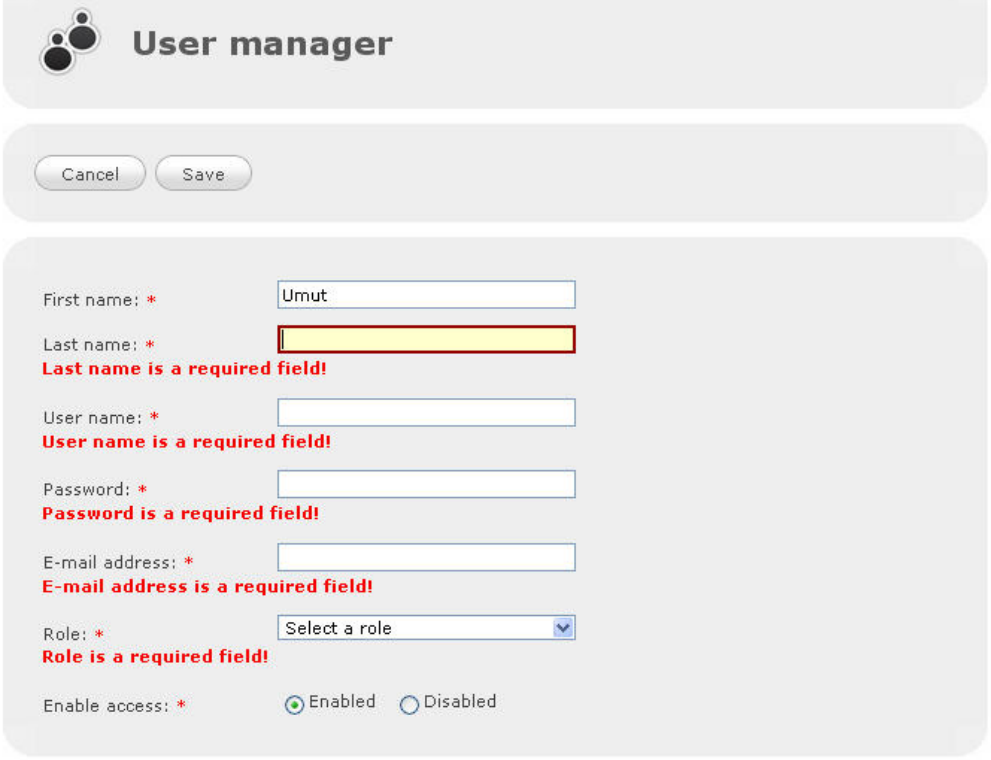
Portlets



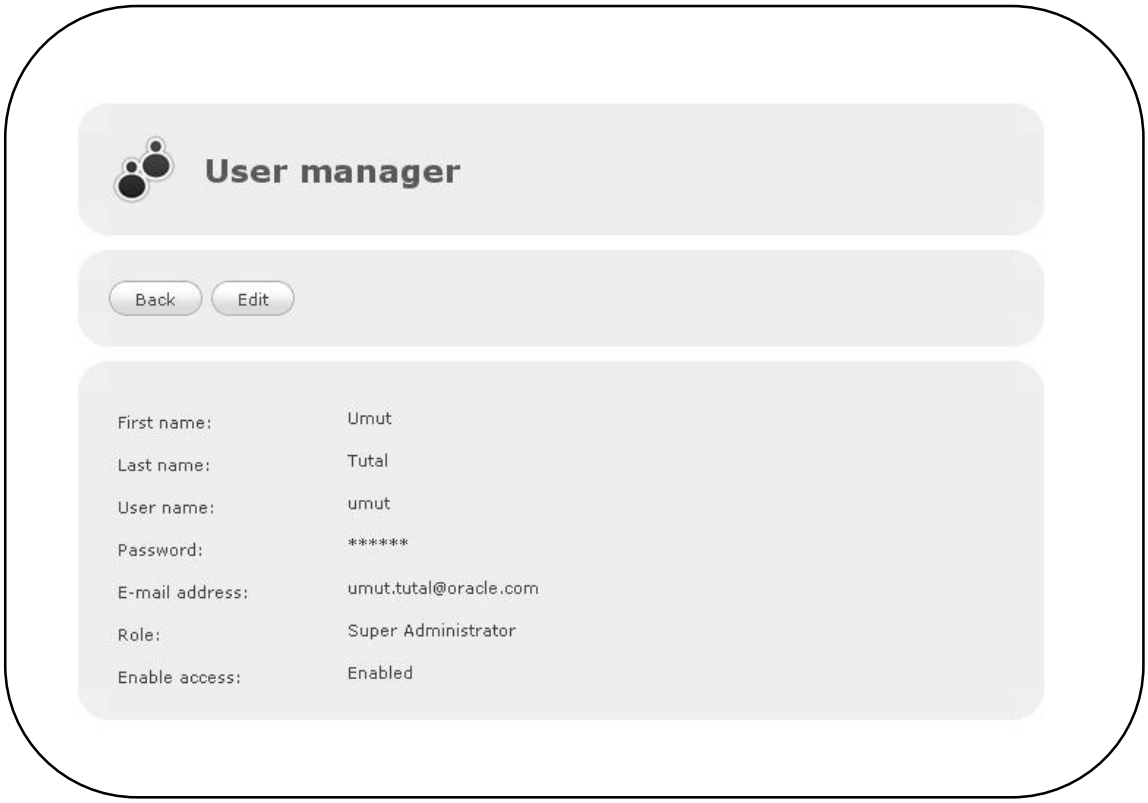
Register User Portlet



Register User Portlet



Register User Portlet with warnings



Register User Portlet confirmation

Display Dutch Painters Location Portlet