

# Development of an Educational Portal to Enhance the Data Accuracy of the Institution Repositories

Ali H. Al-Badi<sup>\*,1</sup>, Ali O. Al Majeeni<sup>2</sup>, Pam J. Mayhew<sup>2</sup>, and Abdullah S. Al-Rashdi<sup>3</sup>

<sup>1</sup>Information Systems Department, College of Commerce and Economics, Sultan Qaboos University, Oman

<sup>2</sup>School of Computing Sciences, University of East Anglia, Norwich, UK

<sup>3</sup>Center for Information Systems, Sultan Qaboos University, Oman

---

## Abstract

*Organisations store large amounts of data in their different databases; this data may be considered part of their wealth, and should be used for organisational benefit. Data integration may be done through new software applications, which contribute to planning by using innovative information products. There are two views toward the integration of databases; first is the data requirement of the organisation's business application; second is the technical capability to integrate various data enabling software applications to operate and achieve expected objectives. Thus, technology facilitates the use of new techniques, but creates a more technically complex environment; especially when information technology is based on layers and components structures. Building applications to access more than one type of database requires knowledge of many technical issues, including database structures. Proficiency is needed to apply the right mechanisms to connect to databases, and fetch and manipulate the data recovered. This paper explores the integration mechanisms between a corporate portal application, and various information system databases. Oracle Portal is used to demonstrate the integration techniques used to group data from different databases producing new information products. Case studies show how the portal helps to validate and enhance data.*

**Keywords** Portal, transformation, integration, information product, components structure

---

## 1. Introduction

Portal technology allows organisations to combine data from a diverse mix of databases to better serve business needs. The technology permits data to be dynamically fetched from multiple sources, consolidated, and through a set of tailored interactive interfaces, creates an interactive information exchange environment [1, 2]. Of course, this type of information delivery cannot be done without clear business objectives, clearly understood by portal designers and developers. Developers also need extensive knowledge of necessary tools and the various systems and database structures which will need to be accessed.

The readiness of the technical infrastructure in the organisation such as networking, Internet, security, and databases must be considered when addressing the technical feasibility of constructing a portal. Back office processes to integrate data from different databases are the backbone for information delivery through the portal's interface [9].

There are various portal technologies, such as Oracle Portal [3, 4], Microsoft SharePoint [5], IBM WebSphere Portal [6] and U-portal [7, 8]. Organisations select the portal technology according to their preferences. These preferences are determined by the organisational environment and infrastructure, so can differ from one organisation to another. The important factor for portal solution selection is that the existing applications are ready to integrate with the selected portal solution, and the organisation ability to build and support the solution internally and externally. This leads to two main concerns; delivery of the solution, which consists

of building the solution perspectives, and achieving the objectives and continuity after implementation, by enhancing services and supporting the solution.

Portal creates an active and interactive environment because it is integrated with various solutions. This requires keeping an eye on any changes in various systems in the organisation. It also requires the continuous updating of services, as organisational solutions grow to cope with changing business requirements.

This paper is intended to illustrate cross-organisational information sharing and dissemination by building a portal. The knowledge presented is based on extensive experience from the authors on an academic portal built at Sultan Qaboos University (SQU) using Oracle portal software.

## 2. Solution Development Perspectives

The size of an organisation, and its management culture, play a major role in the success of portal adoption. An organisation with a multi-database environment may have several different database custodians and a single interface screen may consist of data fetched from multiple sources. Therefore to build a portal, proper cooperation will be needed between different parties before it will be possible to fetch data from these databases, especially given that security criteria and data sensitivity levels may differ from one system to another. In the case of the SQU Portal these issues were determined as being potential obstacles.

Apart from the traditional electronic solution development life cycle, the authors found, through their portal adoption experiences, that portal projects require seven stages for portal as a whole, or for each service to be added as part of the portal. This is required in order to achieve continuous support for portal adoption, data integration, and information sharing. These stages may follow the iterative approach as follows:

**Convince:** staff that the portal serves all users within the institution/community, integrating information drawn from more than one data source. It is also critical to make users/custodians aware of the portal's benefits, gaining their support and approval, and securing integration with their databases. This includes

assuring them of the security, confidentiality, integrity and privacy of their data.

**Identify:** the data elements from the different databases that the portal needs to access. These have to be identified and requested from database custodians. This ensures that access to the data is only provided at the security level essential for the user and as specified by the data custodian.

**Authorise:** must be secured from the data custodians/owners identifying the required data and access levels. Approval has to be formally granted by the data custodians/owners, preferably in writing.

**Prototype/Present:** the structuring and presentation of the information needs allowing them to be reviewed and examined by the concerned parties. This will ensure that the stated business objectives are achieved.

**Acceptance:** of the interface through a proper security examination and assessment is required to ensure acceptance and approval from all stakeholders guaranteeing that outcomes meet requirements.

**Use:** must be encouraged. It is important to encourage people to use the portal facilities bringing additional benefits to them by integrating various databases. This requires creating an interactive environment of information sharing and dissemination.

**Support:** must be available, especially during project implementation. Continuous support for data custodians and users will help to maintain integration, and allow the updating and maintenance of the portal features. Keeping the portal up to date with accurate and relevant information will give users the encouragement necessary to rely upon the portal for their data needs.

All these seven perspectives lead to successful business support for data integration and dissemination. Failure to achieve any of them may create obstacles in the data integration process. Notwithstanding, it is difficult to achieve full support from all parties for all parts of the portal components and features. This depends on the knowledge, awareness, and commitment of stakeholders due to influencing factors which might include their needs along with top management pressure. These seven stages are illustrated in the figure below.

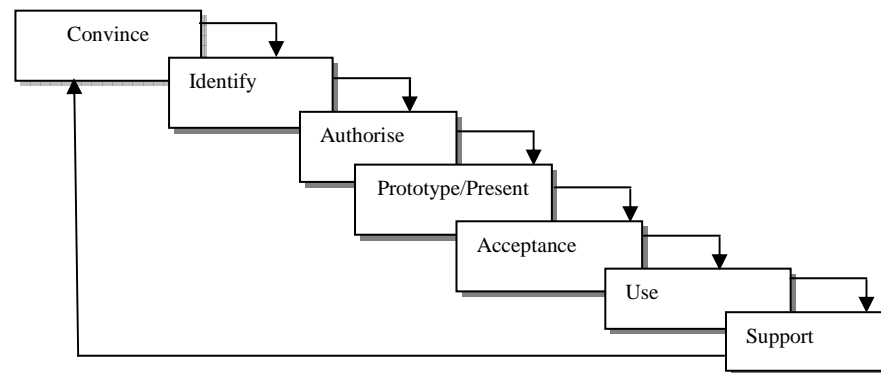


Figure 1: Portal Development Stages

As well as providing simple and easy to use access to existing resources and systems the Portal can also act as a tool whereby its owners can move towards fully integrated electronic management and administration. It can do this by leveraging the information and facilities found in existing on-line systems with E-mail and other modern communication tools to facilitate personalised services. Such services can include many regularly performed activities including matters like leave requisitions, pre-booking transport, printing requisitions and so forth.

To be able to implement the internal procedures the development team must understand them. This means it must coordinate with the concerned units, study what is being done, and produce an electronic analogue of these procedures. This can raise issues, since sometimes the actual procedures to be implemented are not in themselves fully recorded (documented), and the understanding of what must be done may vary between individuals. On the other hand the procedures may be properly laid down, however, staff may be unaware of them, or choose not to follow them correctly.

### 3. Technical Data Sharing and Integration Perspectives

Portals are usually designed to be user focused. That is to say they are designed to deliver integrated information to a particular user that is custom tailored to that individual's needs [2]. Specific issues need to be addressed. Firstly the user must be identified, since his or her experience may not be the same as anyone else's. Then it is important to provide a secure working environment for that user, since his or her Portal pages may contain confidential information. This means that the Portal must implement an Identification (who) and Authentication (security confirmation) process.

The common protection method to access a software application (but not necessarily the most secure) is through a username and password. One of the significant features of a portal is a "single sign on" process through this mechanism, which provides access to a wide variety of resources, each which may sit behind its own "login" process.

At the SQU the most common method of obtaining access to a username and password is by obtaining information from the standard directory source, in this case Microsoft's Active Directory (AD). However, because the SQU Portal is built using Oracle Portal, and uses Oracle Internet Directory (OID) for its identification and authentication processes, it is necessary for AD and OID to work together delivering efficient Portal sign on.

Organisations like the SQU provide a wide variety of software applications, and when someone signs onto the Portal it is necessary that he/she be given access only to those resources he/she is allowed to access (and at the appropriate privilege level). For example, students need to access information in the Student Information System (SIS), but staff access information in the Human Resources System (HRS). Similarly, faculty members access both SIS and HRS. This means that the Portal must be able to understand more than simply "this is a valid user". It needs to be able to determine the status of an individual (staff, faculty, student, other), it may also require information about his/her unit affiliation, his/her E-mail address, extension number, and a wide variety of other details.

In a well-planned Active Directory schema this information is available, and can be easily accessed during Portal sign on. In the case of the SQU, for historical reasons, there is a very simple

schema. This allows users to be identified as being staff, faculty or student. Using this schema the SQU Portal classifies the users and provides an environment customised for that class of user. However, it is not possible to provide a tightly focused environment for each user. This impacts the applications and systems being run, since it is not possible to grant privileged levels of access due to the lack of status information. Because of this it is planning to move to a more sophisticated schema allowing a rich set of data to be made available through a properly thought out and well developed directory structure. With such a schema it will be possible for the Portal to include a wider variety of focused resources designed to be appropriate to the specific needs of individual users. In the meantime users do have the ability to modify the screens they see to make them more specific to their needs, and this does address the deficiency in part.

#### 4. Integration Difficulties

Like many mature institutions, the SQU has a wide variety of data processing needs. This has led to the development of a rich assortment of different applications, each designed to address specific requirements. Some of these applications are large scale, such as the Library Information System and the Human Resources System. These large scale systems are usually hosted by the IT Centre itself on industry strength equipment, and are supported by professionally run facilities. Other systems can be quite modest, and may be run in a single personal computer under the control of a single SQU department.

Obviously providing access to these widely differing databases of application software, and different databases of support software, data, and so forth, represents a significant problem. Clearly it is not possible for the IT Centre, and the Portal, to provide access to small-scale systems where the quality of data, and operational status of the installations is questionable. It has therefore chosen to only provide access to systems that are properly structured and fully supported. However, even providing access to these services presents a number of particular challenges. For example, although Oracle is the database solution of choice for the SQU, and is well used on campus, the databases space, tables, and field names, follow different conventions dependent on the application. Frequently internal used coded values (for example, country mnemonic) do not follow internationally agreed

conventions, or, for that matter, those used in other systems. Then, of course, there is the issue of user identification and authentication; where some of these systems predate the deployment of Active Directory at SQU, so other directory solutions are used.

The result, therefore, is that the Portal team faced significant challenges in attempting to provide unified access to systems where the databases, user interfaces, data representations, and other issues, constitute some of the obstacles that impede such integration.

SQU portal offers services such as: checking library loans, salary payment, leave information, on-line searching for staff details, student/faculty timetable, and examination results, etc. All these services have been well received and are popular with SQU community.

#### 5. Integrating Microsoft Active Directory (AD) and Oracle Internet Directory (OID)

##### *5.1 Identification and Authentication Information - Overview*

As mentioned earlier, the SQU Portal has three categories of users: students, faculty and staff. Information about staff and faculty is stored in the Human Resources (HR) system; information about students is stored in the Student Information System (SIS). In addition, information related to users also can be retrieved from other systems, such as the Library Information System (LIS). Thus the Portal has to connect to the databases used by all these systems.

Most systems identify employees and students by a number, their ID. To retrieve data from the systems which store user information the ID number need to be acquired and passed to them. Different mechanisms are used to do this depending on whether the user is an employee or student.

When an employee signs into the Portal he/she identifies him/herself with his/her username, which is known to Active Directory (AD). Unfortunately the employee username does not include their ID number, which is needed to respond to user requests. The ID number is not included in AD as such, but rather under a different field name, in this case, under "description", which is mapped into the employee number (ID number) in the Oracle Internet Directory (OID). OID is where the user

profiles are kept, and eventually used to validate all accesses to various databases.

For students, the username does include their student ID number, however, this number needs to be extracted from the username. For example, when a student user named u024020 connects to the Portal, the number part of the student ID (24020) is found and passed to the database in order to retrieve the relevant information for the student.

Some services, which are not based on existing database systems, use the username as a key, not

the ID. Moreover, with some features such as the calendar service, there is no need to contact either the OID or manipulate the username.

### 5.2 Synchronising Active Directory and Oracle Internet Directory

All users' information is maintained in AD in one way or another. The OID has to read or update the user details from Active Directory in a synchronised manner, manipulating description and username fields to obtain the user ID as mentioned above. The following diagram shows the synchronisation process:

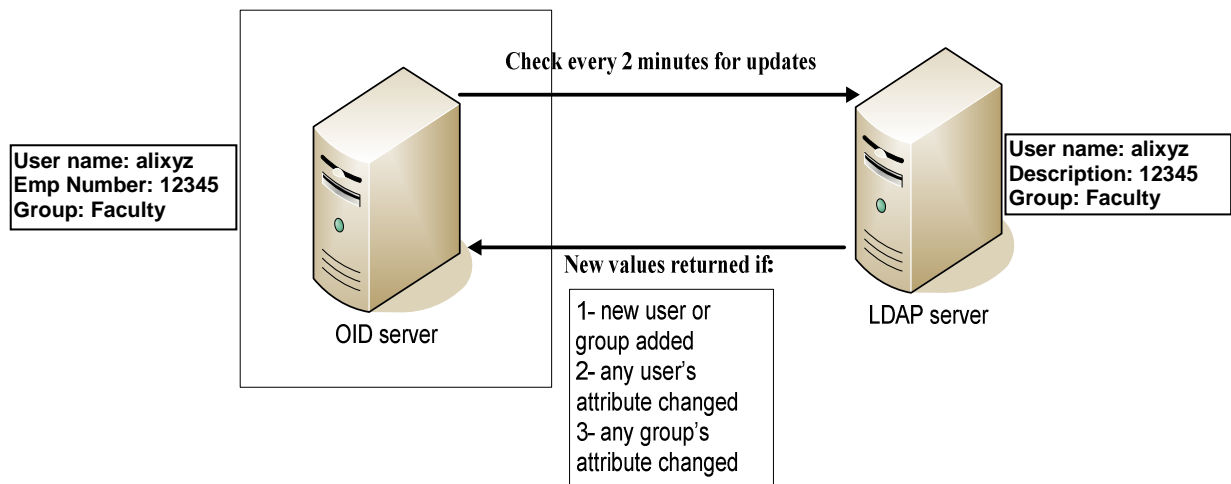


Figure 2: Synchronising Active Directory and Oracle Internet Directory (for an employee)

After the synchronisation processes, which is done periodically at the back-end, the OID will be sent a complete user-profile, and there will be no need to access the AD when user requests a service. For example, when a user (faculty/staff) connects to the Portal a program is executed to

read the ID from OID. The program searches the same group and returns the employee ID number (employee ID number is the key in all the existing databases, that is, SIS, HRMS, Library System). The figure below illustrates how the ID is extracted.

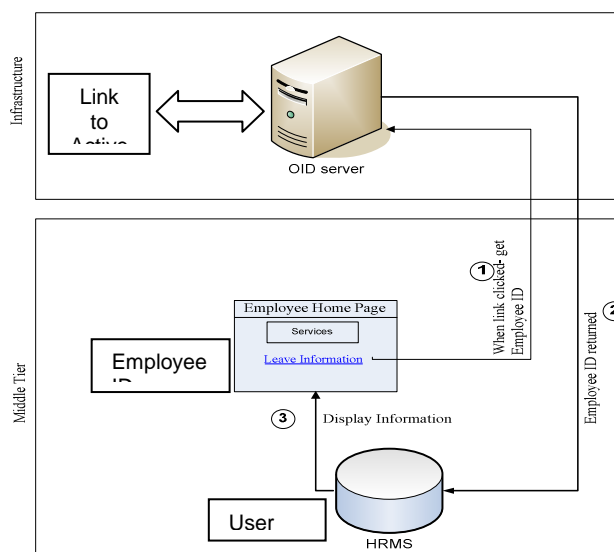


Figure 3: Illustration of the process of requesting a service through the portal

AD – OID synchronisation is driven by a set of rules, which allow the difference in the naming convention in both directories to be resolved. This process is discussed in the next section.

### 5.3 Creating the Domain and Attribute Mapping Rules

Oracle OID tools are used to integrate AD and OID, this requires domain and attribute mapping rules to be defined; the rules are defined in a mapping file. This file has two sections, Domain Rules and Attribute Rules, these specify the edits to be automatically performed.

- "Domain Rules" tells the system the location of users and groups in the AD server-side and where to make those changes in the OID server.
- "Attribute Rules" tell the system which attributes on the AD server-side will be mapped to which attributes on the OID server side.

First considering the domain rules, since there are three groups of users in different Organization Units (OU): staff, students and faculty, the operation must be done three times with different OUs. For example for staff:

```
OU=staff,DC=squ,DC=edu,DC=om:cn=Users,dc=squ,dc=edu,dc=om
```

The left domain, before the ":", represents users in the AD source.

```
OU=staff,DC=squ,DC=edu,DC=om:cn=Users,dc=squ,dc=edu,dc=om
```

The right domain, after the ":", defines the edits in the OID destination.

```
OU=staff,DC=squ,DC=edu,DC=om:cn=Users,dc=squ,dc=edu,dc=om
```

Now, examining a simple "Attribute Rules" definition:

```
cn: :container: cn: :orclContainer
```

The "cn" attribute value, which is part of the "container" object class at the source (AD), will be mapped to the "cn" attribute, which is part of the "orclContainer" object class on the destination (OID).

As well as the three Organization Units there is a fourth one, for groups. In Active Directory we

create the same OU in the OID and three mapping file for three synchronization profiles for each user groups and one for the group membership. The mapping file then allows the users and their attributes to be mapped from AD to OID.

## 6. Verification and Enhancement of Data Accuracy

As any IT professional will tell you, the information retained in on-line systems is only as good as its quality. To ensure the quality of this data the majority of systems implement some sort of verification and validation processes.

There are several ways that these processes can be performed. Probably the most common is to use machine based verification rules, but sometimes the use of this technique is not possible. In these cases one might handle the process procedurally, for example by having the data entry keyed twice (by different staff). Because of the difficulty of ensuring accuracy one always has some doubts as to data quality.

SQU retains significant quantity of information in its on-line systems. However, before the Portal was developed, the majority of campus users; faculty, staff and students; had no access to such data, since the systems were not designed for this purpose.

Consequently, although entered data had been through a checking process, it had frequently not been subject to, perhaps, the most important validation – screening by the data subject him/herself. The Portal, by exposing this previously masked data to the critical view of the data subject allows a final accuracy check to be made. Sample cases follow showing how this exposure has improved data quality.

### 6.1 User Validation of Data - Sample Cases

In this section, the authors present some cases where the Portal helped users to validate the correctness of data stored about them in various systems at SQU. Data validation is a process of ensuring that the data is valid. In a typical computerized system, data is validated by implementing business rules by applying them in a data entry field. For example a business rule might be:

- Knowing if an employee is on scholarship:
  - Checking the scholarship period,
  - Checking the employee's status (on study leave, in post, etc.)

## 7 Development of an Educational Portal to Enhance the Data Accuracy of the Institution Repositories

Below is a set of cases illustrating how the exposure of business data to user validation has assisted the SQU in improving its data quality.

*Case 1:* A user discovers that the phone number inserted in the system is incorrect.

Consequences:

- The organization needs to reach its employees urgently for some reason during the annual leave, or even after working hours, but the employees are not reachable since the contact number is wrong.

*Case 2:* A user notices that a training course he/she has taken is logged as being for one year, while, in fact, it was only for a few days.

Consequences:

- Misleading information when searching for employees. The search facility that has been provided in the portal showed that the employee is on training.
- Since there is incorrect information in the employee record, this may impact his/her training eligibility.

*Case 3:* The instructor's name assigned to a particular course is incorrectly indicated in the Student Information System. Being visible in the Portal, this can be seen by faculty and students and hence be corrected.

Consequences:

- Students will be instructed by instructor different from that shown on his/her timetable.
- Communication will be impacted owing to incorrect addresses.
- Student surveys will attribute performance to the wrong instructor.

*Case 4:* An employee finds that information regarding an official mission was entered wrongly, the visited country not being the one actually visited.

Consequences:

- Wrong information in employee and institutional records.
- Potential for embarrassing mistakes in follow-up activities.

*Case 5:* The Human Resources System does not properly represent the position of an employee or unit, the Portal user detects this incorrect data. For example, the employee is post as the divisional head is improperly identified, or the post is itself incorrectly noted.

Consequences:

- Systems which use the data to grant access rights will incorrectly allocate rights. This

may expose information to those without the right to access it, or otherwise grant access as a level not permitted.

- Automatically generated reports, for example, divisional staffing and structure diagrams, will show invalid data.

*Case 6:* The acquisition data for newly arrived books, as maintained by the Library System, is incorrect owing to a programming error (i.e. using client date rather than the server date). When Portal users request new acquisitions data they can easily detect this error due to the nature of the invalid information displayed.

Consequences:

- Users of the Library system do not receive proper up-to-date information regarding acquisitions.

### 7. Portal Cost Benefits

When considering cost benefits for the portal at SQU it is important to address a number of issues. The first is development, then there is implementation, support and maintenance.

Probably of greatest significance is that the SQU developed its solution, from initial design study to implementation, almost entirely with its own internal resources. Development was done using the Centre's existing experienced staff, supplemented by training to build skills, and used available hardware, software, database systems, and infrastructure resources. Consequently, while some cost would have to be attribute to the use of these existing facilities, apart from specialised training, additional cost was minimised. The full development cycle took about two years, however it should be remembered that the costs accrued should be looked at in a strategic way, and written off over the lifetime of the resource. In this case at least five years, and probably ten.

Of course, for organisations without the internal resources necessary to support such a development program, this could represent a significant investment. There would be two ways of addressing this, the first would be to build internal resources, the second would be to float a development project. Whichever solution taken the process could prove costly.

Implementation meant scaling the solution for enterprise use. It requires significant investment in hardware and software, since with the development environment support was needed for only a few users, but the production environment needed to be capable of supporting

thousands. As the portal was built to be a Web based service users required no special facilities to gain access, and minimal special training to explain how it should be used.

Maintenance and support presents no special problems for the SQU, because the solution was based on an industry standard software solution (Oracle Portal), Web based user interfaces, and was developed by the University's own team. Having the skills set maintained in-house gives the SQU benefits of the Portal resource itself, it builds the expertise of IT staff, and minimises exposure to external agencies.

The Portal provides the SQU facilities that it simply did not have in the past, and it does so in an open and easy to use way. It also provides a platform upon which new software facilities and services can be built, eliminating the need for proprietary client front ends. Although it is difficult to place a monetary value on the benefits alluded to, all should be considered as being positive developments for the SQU and the campus community as a whole.

## 8. Conclusions

There are two critical factors that lead to the successful adoption of portal based software applications solutions; data migration and integration. For portals, integration is more visible as a critical problem, since it requires consolidation of different types of data from multiple sources. Failure to integrate means failure to adopt a portal to provide a complete set of enterprise based information services. Integration is a challenge because solutions use different technologies, and environments may include different enabling technological components.

In this project we saw several cases where successful implementation helped to validate and enhance the accuracy of stored data in various organisational databases. This was a great benefit that SQU gained by adopting this IT solution. Portal development, therefore, helps transform businesses by improving organisational effectiveness and efficiency.

It was noticed that as more services were added to the portal, more requests for other services were made. When the benefits of the Portal were understood, and people become more comfortable using it, they became more motivated to participate in development through the process of constructive feedback.

The Portal has generated an increased interest among users as well as information managers due to the technology's ability to improve the flow and exchange of information across the enterprise [2].

## Acknowledgements

The authors would like to express their appreciation and gratitude to Stephen Millmore and Khalil Al-Fulaiti for their valuable comments, suggestions during the writing up of this paper.

## References

- [1] Mateus B. Costa, Rodolfo F. Resende, Eduardo F. Nakamura, Marcelo V. Segatto, "Software Frameworks for Information Systems Integration based on Web Services", SAC '08, March 16-20, 2008, Fortaleza, Ceará, Brazil
- [2] Brian Detlor, "The corporate portal as information infrastructure-towards a framework for portal design", International Journal of Information Management 20 (2000) 91-101
- [3] Oracle Corporation, (2006a), "Oracle Portal center", accessed on 14/05/06, available at: [<http://www.oracle.com/technology/products/ias/portal/index.html>]
- [4] Oracle Corporation, (2006b), "Oracle Portal", accessed on 22/07/06, available at: [[http://www.oracle.com/appserver/portal\\_home.html](http://www.oracle.com/appserver/portal_home.html)]
- [5] Microsoft, (2006), "SharePoint Portal Server 2003 Product Information", accessed on 14/05/06, available at: [<http://www.microsoft.com/office/sharepoint/prodinfo/default.mspx>]
- [6] IBM, (2006), "WebSphere Portal", accessed on 14/05/06, available at: [<http://www-4.ibm.com/software/webservers/portal>]
- [7] JA-SIG, (2006a), "Collaboration: Sharing Ideas and Sharing Solutions", accessed on 14/05/06, available at: [<http://www.ja-sig.org>]
- [8] JA-SIG, (2006b), "Evolving portal implementations from participating universities and partners", accessed May 09/2006, available at: [<http://www.uportal.org/>]
- [9] United Nations e-Government Survey 2008: From e-Government to Connected Governance, United Nations, New York, 2008