Building an Agile Self-Deposit Thesis Service for Institutional Repositories

Mara Nikolaidou, Marios Bekatoros, Tsadimas Anargyros,
Harokopio University, El. Venizelou 70, 17 671 Athens, Greece,
{mara, bekatoros, tsadimas}@hua.gr

Abstract

In this paper we discuss the experience obtained by developing and deploying a self-deposit thesis service for institutional repositories. Rather than integrating the service in an existing repository platform, it was developed as an autonomous software system using open-source workflow technology integrated with the Repository based on SOA principles, e.g. through web service interfaces. Mandatory electronic self-deposit thesis process is usually well-defined by internal regulations or procedures and involves a number of stakeholders besides students or IR curators, as for example Admission Office personnel. Thus, it should be more efficiently treated as an administration process independently of the underlying repository. The self-deposit thesis service is deployed in Harokopio University Repository, currently based on DSpace platform.

Keywords

Repository Services, Self-Deposit Process, Web Service API, Case Study

1 Motivation

Mandatory deposit of student thesis and dissertations in a digital form is a common policy adopted by Universities worldwide to enrich their Institutional Repositories [1]. The submission process should accommodate the overall lifecycle of a student's thesis or dissertation, from initial submittal to the thesisoffice, through the iterative review and approval process, to the final publication in a digital repository [2] and should be performed with the mediation and supervision of the library personnel [3]. Thus, it should be handled as an administration process involving different stakeholders and integrating autonomous information systems, rather than a simple repository service offered by a repository platform, such as searching or content viewing.

In Harokopio University of Athens the self-deposit of student thesis and dissertations is mandatory. These documents are maintained in digital collections hosted in the Institutional Repository, called *Estia*, based on DSpace Repository platform. Students should submit their thesis in the Library, if and only if they have successfully presented their thesis and obtained a grade, as certified by the Admission Office. If these conditions are met, the Library may accept their thesis. Estia supported a simple deposit process, enabling the students to upload and describe their thesis. This information was reviewed by a Librarian, who might corrected metadata information and accepted the submission, making the thesis available through the Repository search service. To ensure that students have presented their thesis, the Library personnel demanded that they should go to the Library in person and submit a corresponding certificate prior accepting their submission, while in cases where a problem occurred regarding the digital version of their thesis, students were contacted by e-mail. To fully automate the submission process, it was decided to implement a self-deposit service using modern workflow technology, facilitating its description as a complex business process. Both students and administration personnel, employed either in the Library or the Admission Office participate in the process, while the integration with external systems such as the University User Catalog (LDAP) and the Admission Information System should be explored.

There are numerous efforts to develop self-deposit services, usually as part of an integrated institutional repository architecture, as for example LIBRA [4]. In such case, the API provided by the corresponding repository platform (either DSpace or Fedora or EPrints or other) is used to develop the service or more often an external custom module is build for that purpose. In the case of Harokopio University, we explored to integrate the self-deposit service in the IR by properly

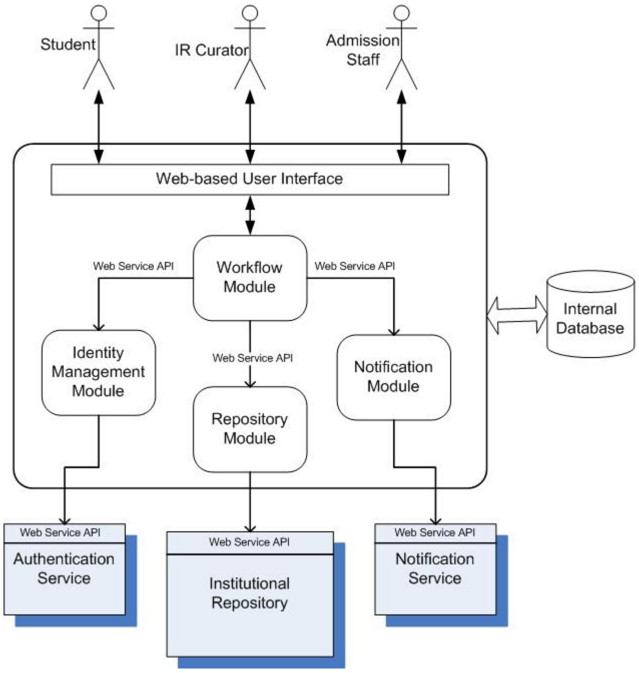


Figure 1: Self-Deposit Service Architecture

extending existing DSpace API. Though, as we studied the self-deposit process, we realized that although the process outcome (e.g. output data) is the addition of the student thesis in the repository, the process itself was not tightly depended on it. In practice, it could be treated autonomously, while the IR was one, probably the most important one, of the external systems the process should had to use to completed. Thus, it was decided a) to explore the development of an autonomous system to support the self-deposit process using currently available open-source technology and b) to integrate it with Estia repository and other external systems related to it. The most prominent way to achieve system integration is using standard SOA technology as web services, thus a Web Service API of the Repository platform should be used, rather that OAI-MPH or OAI ORE. Furthermore, the self-deposit service should be agnostic concerning the Repository platform, while it should be easily customized with no programming effort to accommodate different metadata schemes and workflow patterns. Thus, obtaining agile characteristics was one of our concerns when designing the system.

2 Self-Deposit Service Architecture and Components

The architecture of the system and its main components is presented in figure 1. The self-deposit service was developed in a parametric fashion as a collection of web services, while it can be properly instantiated to customize:

a) Thesis description using custom metadata fields, according to the schema supported by the underlying repository
b) The roles and users participating in the workflow
c) The sequence of self-deposit process steps.

Basic services provided by external systems are grouped into discrete system modules. The system architecture is based on the black-box, white-box approach. Each system module is responsible for a discrete activity, as user authentication, participant notification or repository management and provides a standard web service interface to the workflow engine according to the services it should provide (black-box view of the module). The way each module services are internally implemented (white-box view of the module) depends on the external system it co-operates with to provide the desired functionally. Based on this architecture different repositories and repository platforms can be used by the system, while the same may be applied to other modules as well, for example the notification mechanism, which may be a e-mail service or message posting in a social network. The only prerequisite is that all external systems provide a web service interface, preferable a RESTfull one, which is easily fulfilled by most systems anyway.

Repository Module The Repository Module provides the necessary interface for basic actions to accommodate thesis management, such as uploading, describing, storing, updating or deleting thesis. The provided web service interface provides operations that : a) create, delete, withdraw and reinstate items, b) create, alter, delete and retrieve metadata and c) add and delete files from items

These calls constitute the Repository Module API, while their implementation is based on existing Repository web-service API. Thus, the integration of different repository platforms, such as DSpace or Fedora, only affects the internal implementation of the Repository Module. The implementation of the corresponding web services consist of two layers: the first one implements internal functionality, while the second one consists of invoking external repository calls, masked to be integrated in web service implementation. To support a different repository platform, only the second layer must be modified to properly invoke the available web service repository API. Currently an interface for both DSpace and Fedora has been implemented.

Notification Module It provides direct communication between the self-deposit process participants, utilizing an e-mail or messaging system.

Identity Management Module The Identity Management Module facilitates both user identification and user authentication methods. The Authentication service provides the necessary user authentication to avoid unauthorized access, utilizing existing authentication infrastructure. This can be offered directly by a LDAP system or a single sign-on system like Shibboleth. Identification process is also based in information stored in an internal database. A user can be authenticated through the LDAP server but if the user is not stored in the database as a permanent or temporary user, the identification process will fail. This makes enables to initiate the submission process only for students permitted to do so, according to each Institution's internal processes and regulations.

Workflow Module The Self-Deposit Workflow module orchestrating the process and was developed based on J2EE architecture. It consist of a workflow engine, a SQL data management system and web service client libraries. The definition of self-deposit process steps and the metadata used for student thesis description is performed though a simple web-based administration interface. System configuration and external system interface is also described through the administration interface.

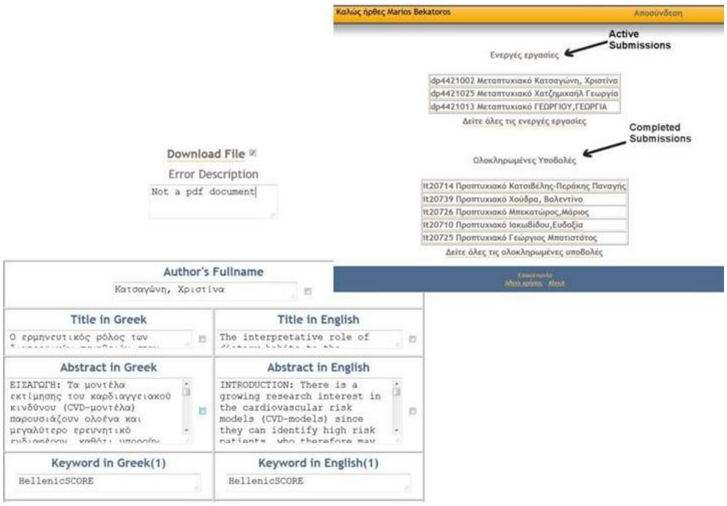
Configuration information is maintained in a set of configuration files in XML format. These files includes information of all external servers used by the Service, login information, metadata fields used for thesis description and notification messages sent to the users. At least 5 external servers should be defined, namely the Institutional Repository, Authentication and Messaging Server, the Application Server the self-deposit service is deployed and a Database server it may use. The configuration files also include metadata information in the following format: *Schema, Element, Qualifier, Language, Obligatory Element, TextArea or Input (HTML Form)* . A except of the configuration file is presented in the following:

```
String mailserver="planet.hua.gr";
String appserver="http://83.212.240.97:8081";
String metadata[][]=
{//Name,schema,element,qualifier,language,obligatory"1"=yes "0"=no,textarea="1"
input="0"
    {"Title in Greek","dc","title","","el","1","0"},
    {"Title in English","dc","title","alternative","en","1","0"},
    {"Abstract in Greek","dc","description","abstract","el","1","1"},
```

The agile implementation of the self-deposit service enabled us to a) easily integrate different repository platforms and numerous repositories in the submission process and b) facilitate the customization of the submission process by the Library personnel without actually writing any code. The project is developed and distributed under the Open Source License GNU Affero General Public License (AGPLv3) and all sources are freely available.

3 Integrating Self-Deposit Service in Harokopio University Repository

Self-deposit service installation in the Library of Harokopio University, called E-Thesis service, has been integrated with Estia Repository, utilizing the corresponding DSpace Repository API. Dublin Core metadata elements are used for thesis description. The University's LDAP server is used for authentication, while notifications are sent using the University's e-mail server. The service may also communicate with Greek National Dissertation Repository, though this feature is not currently deployed. The Self-deposit Service was deployed on a Glassfish Application Server and uses MySQL as internal database.



Καλώς ήρθες Marios Bekatoros

Αποσύνδεση

Active Submissions

Ενεργές εργασίες

dp4421002 Μεταπτυχιακό Κατσαγώνη, Χριστίνα
dp4421025 Μεταπτυχιακό Χατζημιχαήλ Γεωργία
dp4421013 Μεταπτυχιακό ΓΕΩΡΓΙΟΥ,ΓΕΩΡΓΙΑ

Δείτε όλες τις ενεργές εργασίες

Completed Submissions

Ολοκληρωμένες Υποβολές

lt20714 Προπτυχιακό Κατσιβέλης-Περάκης Παναγής
lt20739 Προπτυχιακό Χούδρα, Βαλεντίνο
lt20726 Προπτυχιακό Μπεκατώρος,Μάριος
lt20710 Προπτυχιακό Ιακωβίδου,Ευδοξία
lt20725 Προπτυχιακό Γεώργιος Μπατιστάτος

Δείτε όλες τις ολοκληρωμένες υποβολές

Download File ☑

Error Description: Not a pdf document

Author's Fullname: Κατσαγώνη, Χριστίνα

Title in Greek: Ο ερμηνευτικός ρόλος των

Title in English: The interpretative role of

Abstract in Greek: ΕΙΣΑΓΩΓΗ: Τα μοντέλα εκτίμησης του καρδιαγγειακού κινδύνου (CVD-μοντέλα) παρουσιάζουν ολοένα και μεγαλύτερο ερευνητικό

Abstract in English: INTRODUCTION: There is a growing research interest in the cardiovascular risk models (CVD-models) since they can identify high risk

Keyword in Greek(1): HellenicSCORE

Keyword in English(1): HellenicSCORE

Figure 2: IR curator Screen-shots

The service is operating on a pilot basis since September 2011 for the students of two Departments and full employed by June 2012. The submission process besides students and IR curators, also involves Admission Office personnel. In practice, an admission officer, responsible for verifying the student has presented the thesis and obtained a corresponding grade is the one that initiates the process. Instead of issuing a hard-copy certificate, the officer updates the expect thesis submission student list and a corresponding notification to upload the thesis is sent to the student via e-mail. After uploading the thesis and filling the corresponding metadata fields, the student may submit this information for review. Consequently, the IR curator reviews the thesis data and may either accept the submission and permanently add the thesis in the repository or make comments and send it back to the student for review. In any case, involved stakeholders are notified by e-mail messages, whenever they should participate in the submission workflow. A snapshot of the IR Curators pending list and the review form is presented in figure 2. Currently the only language supported is Greek. User feedback was encouraging. One of the main advantages of the service identified by IR curators was its simplicity and customization features.

4 Conclusions

To effectively accommodate well-defined institutional processes, such as the student thesis self-deposit process, depended more on internal regulations and procedures, rather than the Institutional Repository itself, it is more efficient to develop and deploy autonomous systems interacting with the repository rather than integrating them within the repository platform. In this way, an advanced degree of flexibility in service development is obtained, while repository infrastructure extendability and interoperability is promoted. To this end the standardization of repository platform web service interfaces is a necessity in a way similar to OAI-PMH or OAI ORE, to promote interoperability in terms of the functionality exposed rather than the data exchanged.

References

[1] Lena Veiga e Silva, Marcos André Gonçalves, and Alberto H. F. Laender. Evaluating a digital library self-archiving service: The bdbcomp user case study. *Information Processing and Management*, 43(4):1103–1120, July 2007.

[2] Adam Mikeal, James Creel, Alexey Maslov, Scott Phillips, John Leggett, and Mark McFarland. Large-scale etd repositories: a case study of a digital library application. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '09, pages 135–144, New York, NY, USA, 2009. ACM.

[3] Nicholas Joint. Institutional repositories, self-archiving and the role of the library. *Library Review*, 55(2):81–84, 2006.

[4] Meloni Julie. Libra: An unmediated, self-deposit, institutional repository at the university of virginia. In *Proceedings of the 6th International Conference on Open Repositories*, OR 2011, 2011.