PYXIDA : The Institutional Repository of the Athens University of Economics and Business (AUEB)

Antonios Mourikis
Librarian
Athens University of Economics Business

Chrysostomos Kapetis
Laboratory Teaching Staff
Athens University of Economics and and Business

Abstract : The present paper serves as a case study which demonstrates the project of implementation of the Institutional Repository by the AUEB's Library. The project was co-funded by the EU and the Greek government in the framework of a broader program of digitization. The paper describes the main stages of the creation and implementation of the repository including the adoption of a digital repository system, the digitization of selected printed items, the documentation of the digitized material and the definition of deposition policies.
The ultimate purpose of this article is to highlight the crucial factors for the successful implementation of the repository so as to serve as an added value for the University.

Keywords:PYXIDA, academic repository, copyright, digital collections, digital repositories, digitization projects, MOPSEUS digital repository system.

1. Introduction

Within the years 2013 – 2015 the Library of the Athens University of Economics and Business (AUEB) undertook a project of developing contemporary digital library services, the core of which was the creation and implementation of an Institutional Repository. The project was co-funded by the EU and the Greek government in the framework of a broader program of digitization called National Strategic Reference Framework (NSRF). The main objective of the Repository was to gather the extensive scientific, educational and research work produced by the Athens University of Economics and Business. The Repository also includes both old and rare material in the "Digital Collections of the Economic and Social Sciences", such as old textbooks, teaching manuals and rare monographs from the first decades of the twentieth

Received: 31.3.2016 / Accepted: 30.4.2016 ISSN 2241-1925

century. The digital collections include as well the "Economic and Social Sciences Archive" published by Professor D. Kalitsounakis. This journal has recognized historical value and is comprised of scientific articles of economic and social content from the fifty-year period of 1921 - 1971. The implementation of this project included the adoption of a digital repository system, the digitization of selected printed items, and the documentation of the digitized material and the definition of deposition policies.
The Repository was named PYXIDA (compass).[1] Using a collaborative method to pull financial and human resources, the Library finally accomplished the main project goals: workflow, copyright risk assessment, documentation of the digitized material and the definition of deposition policies. The two most important issues are that, first, the digitized material is now available not only for the academic community but also for the general public and, second, the Institutional Repository serves as an added value for the University (Chang and Hopkinson, 2014).
Once undertaken the responsibility of the project, the AUEB's Library made the decision to create a team of specialists in order to manage and accomplish the whole procedure. To begin with, a faculty member was assigned by the University administration as a project manager. Secondly, the manager took over to form a team consisted of professional experienced librarians and computer scientists. The team was assigned to contemplate on and decide for the aforementioned issues and to submit its proposals. Subsequently, the main strategic policies on the project were determined and each member of the team was designated for a particular part of the procedure. The project was divided into three subprojects: the adoption of a digital repository system, the digitization of selected material and the documentation of digitized materials, and the formation of deposition policies.
In what follows we present the main stages of the project.

2. Digital Repository System

With regard to the software system, the team of computer scientists determined the technical specifications according to the international standards, bearing in mind that a repository system should be customizable and fulfill the needs of the Institution. The process of selecting a digital library system involves creating a set of requirements that meet the needs of the organization and, consequently, comparing these requirements to the specifications of the varied solutions/products. Certainly, the way a specification is implemented plays a significant role.
The basic requirements in terms of the digital library software in our case are:

[1] The selection of the name "Compass" gives the Depository and the Digital Library the symbolic meaning of orientation and guideline-driven Excellence University disciplines cultivated.

Regarding the cataloguing, the system must be able to use different metadata schemas within the digital library, to utilize profiles for each metadata schema in order to reduce the time required for cataloguing, and to quickly pick terms from taxonomies within the cataloguing module. As to the Information organization, the system must be able to have a graph structure (not a simple hierarchy) in order to represent the organizational structure of the institution and to assign a digital object to multiple categories/containers. Finally, with reference to the digital preservation, the system should be able to store the entire lifecycle of each object and to have a robust preservation model.

All the above considered and through a public competition, the MOPSEUS digital repository system was selected (Gavrilis, Angelis, Papatheodorou, 2010), based on fedora software. The solution offered by the MOPSEUS digital library system addressed the above requirements by utilizing a flexible cataloguing subsystem that allows the administrator to define any metadata schema, assign attributes and properties and link an element with a taxonomy, by associating each digital item with its own metadata schema, by associating each digital item with multiple containers (using rdf properties), adopting PREMIS (OCLC, 2005) as a standard for digital preservation and finally by adopting fedora-commons (www.fedora-commons.org) as a data store and taking advantage of the versionable datastreams and preservation facilities it offers. The MOPSEUS digital library system builds on top of fedora-commons comes with its own relational database and provides access to the repository's resources through an API. Its front-end has been custom developed in order to meet the needs of the institution.

3. Digitization

As regards the digitization, the professional librarians undertook the responsibility of choosing the material, defining the metadata and deposition policies and copying with the copyright issues. Furthermore, in collaboration with the computer scientists, they defined the digitization standards and requirements. A private company took on the technical part of the digitization following a public competition, and this part of the project was carried out within five months. At the same time,while the digitization was in progress, the metadata and deposition policies were defined by the professional librarians. In this stage of the project, the team of the professional librarians decided to upload to the Repository the metadata of the documents digitized in the past such as the «Economic and Social Sciences Archive» by Kallitsounakis, and by this way a first control of the documentation procedure was carried out; this served as a model for the rest of the digitized documents.

Initiating the project, the professional librarians were charged with the duty of locating, selecting and assessing the items for digitization. In the first place, the team had to define the selection criteria for the documents and to specify the special reasons that rendered the need of the digitization of each collection imperative. Based on these criteria, the team decided to propose the following collections: postgraduate theses, old textbooks, rare books, university yearbooks

and archive material. Furthermore, the authors of the current paper mention the reasons for each collection in detail so as to highlight the importance of this selection.

3.1 Postgraduate thesis

The documents of this collection consist of theses conducted by the postgraduate students of the University from 1993 until 2007. The context of this material included specialized and original scientific knowledge; there was only one printed copy available for each dissertation which, because of the frequency of its use, was inevitably suffering physical damage. Last but not least, the collection is quite extensive (almost 2000 unique copies), kept growing and due to limitation of space, it had to be digitized.

3.2 Old textbooks

The second collection selected to be digitized was that of the old textbooks. This collection consists of books which were taught in the University from its foundation in 1920 until the decade of 70's. These books, whose authors had been professors of the University, are in Greek language and are considered to be rare since their publication year dates back to the beginning or the middle of the 20th century. Some of them had never been published as formal editions, but had been used as course notes. These books are of historical as well as of practical value for the academic community, considering the information that can be derived on the University curriculum and the development of teaching. Taking into account their physical damage, the team decided to suggest this collection for digitization.

As we have already mentioned, the items of this collection were subjected to copyright control and the team had to plan an effective strategy in order to avoid taking risks or to decrease the possibilities of risks. First of all, the items were divided into five categories according to the international directives and practices (Shepard, 2013 and Smith, 2012). In the first category were placed books the copyright of which had expired, as seventy years after the author's death had passed, so we were not obliged to ask for digitization and deposition license. In the second category were placed books on the author – or authors – of which we were not able to find any information, therefore they were placed in the category of orphan works. The third category was occupied by the books whose authors have been deceased but their first-degree relatives had been located (sons, daughters, etc.). The fourth category was made up of books on which we obtained information by second or third-degree relatives. The last category consisted of books whose authors were alive so it was possible to address them. Finally, the books from the first-two categories belonged to items in public domain and we worked on the rest of them meticulously. During the search for information we used a variety of sources so as to locate the authors themselves or the authors' relatives. The most frequently used source was the

internet, through which we managed to discover much information or other sources which finally led us to the point. Another source of information was the University archive where plenty of data was found and enhanced our searching. The most helpful source was the University personnel and the University professors themselves. Through personal interviews, we managed to collect extremely important information which we would not otherwise be able to gather. The next step was to prepare the appropriate procedure so as to get the permission for the digitization. Legal advice was requested by the law department of the University and the procedure was completed when the copyright holders signed the relevant documents.

3.3 Rare books

The third collection for digitization consists of books published from the late of the 19th century until the decade of 40's. These are, mainly, economic books in the Greek language, which refer to the financial and social life of those years. Through these items we are able to retrieve important information not only on Greece but also on the rest of Europe and, finally, we can ensure comprehension of the socioeconomic conditions of those times. We have to mention that the digitized books selected from this collection are not subjected to copyright control since the seventy years' period after the author's death has expired.

3.4 Yearbooks

The University has been publishing yearbooks from its foundation till today. Some of them are rare given that they had been published until the decade of 1960, there are only one or two copies available and certainly there is the issue of physical damage. These yearbooks have both historical and practical value as through them the establishment, the development and the general history of the University is presented.

3.5 Archive material

The last category involves general archive material from the Secretariat senate and the University's schools. The digitization and preservation of this material was deemed mandatory as the information included is of great importance and in some cases crucial. Through this material we are able to find information on significant and critical meetings that took place in the past and resulted in important decisions.

In parallel work with the first part of the project, the team of professional librarians had to locate, write down and assess the material selected for digitization. The team was divided into three subgroups, each of them comprised of two persons and taking over specific collection: the first group undertook the collection of postgraduate theses, the second group the collections of old textbooks and rare books and the third one the yearbooks and the archive

material. Each group worked on these collections by writing down and signalizing problems and particularities. For instance, in the collection of old textbooks there are copies handwritten or of different scale from the normal A4, either smaller or bigger. Furthermore, regarding the archive material, some of its parts were difficult to handle as they included massive volumes; in that cases, the digitization team had to unbinde them, to continue with the digitization and then to bind them again. In addition to that, in collaboration with the company, they had to decide which items could be digitized within the university facilities and which not, because, in the first case, bulky machines had to be transferred from the company's facilities, which was not always feasible. Overcoming these issues, the groups prepared a list of all the items available and the process of digitization begun. The material was gradually carried to the company, starting from the postgraduate theses and continuing with the old textbooks and the rare books and finally with the yearbooks and the archive material. Each movement of the material inside or outside the University building was noted down and after the end of the digitization, a control not only to the physical but also to the electronic material was conducted by the members of the team.

4. Documentation and deposition policies

After the end of the first part of the project, the team continued with the documentation and the definition of deposition policies. The team meticulously studied the available metadata schemas in Greek and international academic repositories, noting down the best practices and choices applied in order to decide which of them would meet our needs. The final plan included the basic Dublin Core schema (www.dublincore.org) enriched with the Refined Dublin Core schema. The team also decided to form different schema profile for each type of material (theses, monographs, etc.).

Thereinafter, the library's professional team studied the appropriate practices on the deposition polices and how these could be applied to our system. The outcome of this procedure resulted in a framework of rules which is outlined as follows:

Every member of the academic community is entitled to submit his or her scientific work to the Institutional Repository through the self-archiving service. The depositor must have a password and login code in order to be able to submit his/her work. In case that there are more than one authors, the person who makes the deposit must have the consent of the rest of them so as to complete the procedure. The postgraduate and doctoral students are obliged to submit their theses to the repository. By submitting their work the depositor and the authors sign a document in which declares that they have not violated the copyright law, his/her thesis is approved by the scientific committee, he/she allows the Library to upload the work on the repository, without changing the context but only making a digital copy just for preserve and security reasons. The depositor holds the right to connect his/her works with a creative common license. After the submission, the library personnel makes a quality control of

the metadata especially for postgraduate and doctoral theses and the Library issues a certificate which is delivered to the student's department.

5. Future work and conclusions

In parallel with the implementation of the repository, a research team of the department of Informatics developed a Web archiving system from the sites of the AUEB (Plachouras, Kapetis, Vazirgiannis, 2010). This system is based on an open source software and supports two basic services: first, the archiving service based on the web crawler Hetitrix which has implemented in internet archive, and second, the search and information retrieval service based on wayback and NutchWax open source software.
First, the Hetitrix starting from a set of seed URLs, performs the collection of the Web content and recursively downloads URLs, following their links to discover new URLs. The crawler fetches any file available on the Web sites of AUEB. The crawler also downloads files which are linked from Web pages of AUEB, but do not belong to the domain aueb.gr, in order to allow a better browsing experience of the archives. The crawled Web content is stored compressed on hard disk in files containing several records.
The necessity of the preservation of Web resources lies in the admission that the information published
on the Web is very easy to get lost. A huge volume of information published is ephemeral as it is not reproduced in printed format. A future target of the AUEB's library is the integration of this tool in the Institutional Repository so that digital subject collections are created from the archived web content. We believe that the completion of these systems will considerably strengthen the role of the library regarding:

- The preservation of the history of the institute
- The use of historical data for future research
- The preservation and protection of cultural heritage and scientific production

Other future goals of the library are the formation of a plan in order to include all the policies required for the preservation of the digitized collections of the repository, as well as the definition of the procedure which must rule the operation of the repository in order to be the core of the electronic information for the university community.
To conclude, we would like to stress that the significance of an Institutional Repository mainly lies in the added value for the University. This value, also, lies in the uniqueness of the collection contained in the repository. Furthermore, is our strong belief that the close cooperation between library information specialists and computer scientists who are very well aware of the sources

available and the row of information is the key factor for the successful implementation of an Institutional Repository.

References

Chang, N. and Hopkinson, A. (2014). Institutional Repositories: libraries coming to the help of their institutions. *QQML*, 2 495 – 501

Dublin Core Medata Initiative. DCMI Metadata Terms. http://www.dublincore.org

Gavrillis, D., Angelis, S. and Papatheodorou, C. (2010). MOPSEUS – A digital repository system with semantically enhanced preservation services. *Proceedings of the7th International Conference on Preservation of Digital Object*, Austria, September 2010, 135-143

Fedora Object XML (FOXML), http://www.fedora-commons.org

PREMIS Working Group. Data dictionary for preservation metadata: final report of the PREMIS Working Group. *OCLC Online Computer Library Center & Research Libraries Group*, USA, 2005

Shepard, E. (2013). A digital collection collaboration in an Academic Repository: a case study. *Journal of Archival Organization*, 11 205-220

Smith, K. (2012). Copyright risk management: principles and strategies for large-scale digitization projects in special collections. *Research Library Issues*, 279 17-23

Plachouras, V., Kapetis, M. and Vazirgiannis, M. Archiving the Web sites of Athens University of Economics and Business. *Proceedings of the 19th Panhellenic Conference of Academic Libraries.* Athens, November 2010, 73-86