Lecture Notes in Computer Science: Ktisis: Building an open access institutional and cultural repository

Alexia Dini Kounoudes, Petros Artemi, Marios Zervas

Library and Information Services, Cyprus University of Technology, Arch. Kiprianou 31, 3036, Limassol, Cyprus
{alexia.kounoudes, petros.artemi.marios.zervas}@cut.ac.cy

Abstract. The unique value of cultural heritage has long been recognized together with the need for accurate and detailed information in order to preserve and manage cultural heritage material. Any organization whose mission includes promoting access to information is aware of the value of digital collections. For the last few years, digital technology has become very familiar in cultural organizations, providing enhanced access to the content. This paper gives information about Ktisis (http://ktisis.cut.ac.cy), the institutional repository of the Cyprus University of Technology (CUT). Ktisis was developed by the Library and Information Services of CUT. The paper reflects on the technical issues that the Library had to face in the preparation of this project and the strategy that had to be defined in order to tackle them. Such issues, among others, include the file and metadata format, the design and implementation software, etc.

Keywords: Open source software, institutional repository, intellectual rights, open access, cultural

1 Introduction

As the Information Society is evolving in a very fast manner, the technological growth plays an important part in various sectors of the social life, one of which is culture. The cultural diversity of the existing material and the disposal for the exploitation of its value, in combination with the requirement to preserve the original work, constitute the motivation for many institutions to work on projects in order to digitize content of cultural heritage.

Since the early 1990s libraries have worked on digitization projects to provide access to and to preserve unique materials in their collections. These collections are accessible in an enhanced format that allows searching and browsing via the World Wide Web. Preservation of the original content and wide access to researchers and the public are the major reasons that libraries are undertaking digitization projects. Through digitization, the library is able to provide access to all sorts of materials – text, photographs, manuscripts, audio, and moving image materials. Digitization also allows for the preservation of rare, fragile, and unique materials. Another benefit of

digitization is that it raises the profile of the institution as user's worldwide access its collections remotely.

An important aspect of a digital library involves handling intellectual property rights. When libraries undertake a digitization project, they need to take into consideration whether or not the material to be digitized is protected by copyright law, or whether it is in the public domain.

Cyprus has a rich cultural heritage in the form of monuments, buildings, paintings, manuscripts, coins, etc. Most of these items are publicly available at museums or archaeological sites. However, there are many private collections of important items that are in danger of being damaged or lost. Examples of such items are private collections of old photographs, volumes of newspapers or magazines that are no longer published and whose content is historical or can be considered as a cultural treasure, etc.

In this manner, the Library and Information Services at the Cyprus University of Technology has defined that one of its major priorities is to collect, disseminate and preserve cultural heritage and to contribute to the cultural evolution of Cyprus. Based on this priority, the Library has designed and developed the first institutional repository in Cyprus, named "Ktisis".

2 The Ktisis Project

Ktisis is the institutional repository created by the Library and Information Services at the Cyprus University of Technology. Ktisis's main purpose is to collect and preserve the products of the research of the academic staff and researchers of the university and also to collect, digitize and disseminate cultural content.

One of the main priorities of the Library was to define a strategy for the collection and archiving of cultural heritage material (old photographs, newspapers etc). The objective of this project is to fulfill the Library's mission for collecting, disseminating and preserving cultural heritage.

When Ktisis was at the designing stage, the Library defined the set of goals that Ktisis needed to achieve. These goals were:

- To locate and archive together cultural heritage items from private collections.
- To guarantee long-term preservation and access to the data.
- To promote interest and involvement in the digitization process and preservation of cultural heritage.
- To promote open access at the Cyprus University of Technology.

In order to fulfill the task, Ktisis needed to accomplish these goals by organizing the data in an effective and coherent representation providing easy access to the public.

Ktisis is a member of a world wide system of open access institutional repositories, participating actively in the new shaped model of scientific communication. Currently Ktisis provides its metadata as a data provider applying the OAI-PMH protocol of

metadata harvesting to the following service providers: OAIster, OpenDoar, Openarchives.gr, Driver, Openarchives.eu, Scientific Commons, University of Illinois Data Provider

2.1 Data selection

Considering the huge volume and heterogeneity of information on the web, the selection and evaluation of the material to be digitized is one of the most difficult tasks when setting up a digital library. The selection of the material for a digital repository involves handling intellectual property rights, legal issues and copyright.

The Library and Information Services has collaborated with the Patticheio Historical Archive Museum and Centre of Studies for the digitization and archiving of cultural heritage material belonging to the Centre.

The Library has also collaborated with the Cyprus Federation of Amateur Track and Field Athletics for the digitization of a collection of photographs. In addition to that, the Library is in contact with various other organizations or owners of private collections in order to review and record the existing material in order to include it in Ktisis.

2.2 Software selection

Following the directions of DRIVER (Digital Repository Infrastructure Vision for European Research, (http://www.driver-repository.eu), the Library and Information Services decided to set up Ktisis using the open source software DSpace (http://www.dspace.org), an open source software developed by the MIT Libraries and Hewlett-Packard Company that enables open sharing of content.

DSpace follows the national standards such as the Dublin Core (http://www.dublicore.org) and is compatible with the OAI-MPH protocol of metadata harvesting. It enables capturing the data in any digital format – text, video, audio, and data files. DSpace indexes the digital content, so users can search and retrieve the material. DSpace distributes the digital content over the World Wide Web and it preserves the material over the long term.

The long term preservation facility is supported by using three types of data formats: supported, known and unsupported types. For all three types, DSpace does bit preservation, i.e. the preserved file remains exactly the same over time, and not a single bit is changed. For supported types, DSpace does functional preservation as well. The file changes over time so that the material can me immediately usable in the same way it was originally, while the physical media and digital formats change.

DSpace is a community-based, open source software platform that can be downloaded free of charge and used to create a digital repository. Organizations and institutions can more easily share and preserve their scholarly collections with an archiving system that stores digital representations of books, theses, 3-D digital scans of objects, photographs, film, video, research data sets and other forms of content.

Because DSpace repositories are internet-based, it is easy for users to deposit content and browse collections from anywhere in the world. Materials in DSpace

repositories are distributed through the Internet and gain exposure through search engines, such as Google. These same items are permanently stored in a non-proprietary format, so researchers can continue to access its contents for decades to come.

2.3 Technical issues

In a digitization project, even minor decisions may have a potentially large future impact. At the beginning of the project, one of the first challenges was to determine the future size of the digital repository. In order to do that we had to count the amount of documents we had to digitize, to calculate the amount of space they will need at the host server and estimate the data scalability for the following years. After completing this demanding process we realized that we only have sufficient data to make rough estimations. There was no way to predict what kind of material will come into our hands in the future for digitization and preservation in Ktisis.

Ktisis is covering a wide range of originals: newspapers, journals, photographs, etc. The Library is trying to digitize these items in high quality, with 24 bits in colour depth and resolutions from 300 to 600 dots per inch (dpi). These images are registered in JPG format with sizes ranging between 30 and 800 KB.

2.4 Data processing

Using the appropriate digitization equipment that complies with the materials being digitized is essential for a successful digitization project. Therefore the Library has been equipped with a high resolution scanner suitable for scanning photos and textual material. The scanned images are stored in the JPG image file format. Then the items go through a watermarking procedure in order to protect them and secure them from unauthorized use.

2.5 Ktisis data organization

After the items have been scanned and stored in a JPG format, they need to be submitted to Ktisis. The DSpace content is organized around communities which can correspond to administrative entities such as schools, departments, labs and research centers, or digital collections. Within each community there can be an unlimited number of sub-communities and an unlimited number of collections. Each collection may contain an unlimited number of items.

The Ktisis content is organized in two communities, the Academic Publications containing the produce of the academic and research activity of the members of the university, and the Digital Collection containing the digital heritage material that this paper focuses on.

Currently, the Digital Collections community consists of three collections:

- "Satiriki epitheorisi", containing approximately 2800 items (number increases daily as new items are added continuously). This collection contains scanned images of caricatures from the cartoon newspaper "Satiriki Epitheorisi" ranging between 1964 and 1985.

- Photos from the Ports of Cyprus, containing 163 photos. This collection includes photographs taken during the construction of the Famagusta port during the period 1871-1933. The documentation took place in collaboration with the Patticheio Historical Archive Museum and Centre of Studies.

- Photos of Olympia Gymnastics sports club, containing 202 photos. The Olympic Gymnastics Sports Club is the oldest athletic club of Cyprus and the second oldest in the Greek area after the Panellinios Sports club. This collection contains photographs from the Cyprus Track and Athletics games. The whole history of the Cyprus track athletics can be seen through this collection. The photographs were taken from the Cyprus Federation of Amateur Track and Field Athletics photo album "Golden Cyprus Winners" as well as from the photographic archive of the Association of Veteran Track Athletes of Limassol. The documentation of the material was accomplished using the album "Golden Cyprus Winners".

2.6 Data submission and registration

The stage of data submission and registration comprises mainly of the creation and registration of the metadata for the resources, and the submission of the resources to the repository. All the items submitted to the repository are structured according to the Dublin Core metadata schema.

The Dublin Core set of metadata elements provides a small and fundamental group of text elements through which most resources can be described and catalogued. Using only 15 base text fields, a Dublin Core metadata record can describe physical resources such as books, digital materials such as video, sound, image, or text files, and composite media like web pages. Metadata records based on Dublin Core are intended to be used for cross-domain information resource description and have become standard in the fields of Library and Information Science and Computer Science. Implementations of Dublin Core typically make use of XML.

DSpace is installed and configured to use the Dublin Core metadata schema by default. Dublin Core is made up of elements, and qualifiers. The Dublin Core basic elements can be seen in Table 1.

Table 1. The Dublin core basic elements.

TITLE	The name given to the resource by the CREATOR or PUBLISHER.
CREATOR	The person(s) or organization(s) primarily responsible for the intellectual content of the resource; the author

SUBJECT	The topic of the resource; also keywords, phrases or classification descriptors that describe the subject or content of the resource.
DESCRIPTION	A textual description of the content of the resource, including abstracts in the case of document-like objects; also may be a content description in the case of visual resources.
PUBLISHER	The entity responsible for making the resource available in its present form, such as a publisher, university department or corporate entity.
CONTRIBUTORS	Person(s) or organization(s) in addition to those specified in the CREATOR element, who have made significant intellectual contributions to the resource but on a secondary basis.
DATE	The date the resource was made available in its present form.
TYPE	The resource type, such as home page, novel, poem, working paper, technical report, essay or dictionary. It is expected that TYPE will be chosen from an enumerated list of types.
FORMAT	The data representation of the resource, such as text/html, ASCII, Postscript file, executable application or JPG image. FORMAT will be assigned from enumerated lists such as registered Internet Media Types (MIME types). MIME types are defined according to the RFC2046 standard.
IDENTIFIER	A string or number used to uniquely identify the resource. Examples from networked resources include URLs and URNs (when implemented).
SOURCE	The work, either print or electronic, from which the resource is delivered (if applicable).
LANGUAGE	The language(s) of the intellectual content of the resource.
RELATION	The relationship to other resources. Formal specification of RELATION is currently under development.
COVERAGE	The spatial locations and temporal duration characteristics of the resource. Formal specification of COVERAGE is also now being developed.
RIGHTS MANAGEMENT	A link (URL or other suitable URI as appropriate) to a copyright notice, a rights-management statement or perhaps a server that would provide such information in a dynamic way.

DSpace provides the functionality to create different Dublin Core metadata schemas according to the type of the items belonging to a collection. For example, in

the case of Ktisis, the metadata describing a collection of photographs is different to the metadata describing a collection of PhD theses.

2.7 Access and intellectual rights

Anyone using the internet can access and view the contents of Ktisis without any restrictions. The users can find and retrieve information from the repository easy and fast as Ktisis supports interoperability providing access via multiple search engines and other discovery tools. However, due to copyright issues the items cannot be published or reproduced without the written consent of the owner.

Since its operation, Ktisis provides the capability to choose and apply one of the available Creative Commons licenses. Using the Creative Commons licenses is not obligatory. However, the option to use them is offered in order to be able to define in an easy way the rights that the owner of the item being submitted keeps and the rights that the owner discharges for the use of other creators. A Creative Commons license is a simple way to encourage others to share and reuse your digital content.

Our experience in using the Creative Commons licensing schemes so far has shown that if someone is publishing material of any variety on the web and is interested in spreading this work, then licensing this work under Creative Commons is an effective action as it actively encourages sharing the work with permission and attribution.

Also Creative Commons licenses bridge the gap between full copyright and the public domain, moving content from "all rights reserved" to "some rights reserved." Creative Commons licenses are not a replacement for copyright; rather, they allow you to modify your copyright terms so your work can be shared and reproduced legally. Additionally, everyone can benefit from the Creative Commons as adding such a license to your work can be seen as a marketing tool. Releasing content under a Creative Commons license gives the content creator full control over how the owner would like his/hers content attributed.

In a digital repository it is required that the rights of the creators are being defined precisely, together with the user rights. This information is precisely defined in the Creative Commons licenses. Currently, there has not been any work on adjusting the Creative Commons licenses for the Cyprus law, so Ktisis uses the English version instead. The Library and Information Services has been in contact with the relevant team of Creative Commons in order to start the procedure of implementing the Cypriot version of the Creative Commons licenses. This procedure is still at the early stages.

The licensing procedure in Ktisis consists of two steps: the first step involves the Creative Commons licenses, and the second involves the Library's license of submission and dissemination.

When submitting an item and before completing the submission procedure, the user can choose whether he/she wishes to declare the rights of the item being submitted. If the user wishes to use the Creative Commons license, then Ktisis connects with the Creative Commons website. The choices offered at this stage involve the definition of whether the creator or owner of the item allows the

commercial use of the item or not, and also whether the owner/creator allows modifications. Then, the available Creative Commons are being displayed where the user is prompted choose one of them.

2.8 Promotion of open access

Ktisis was developed as part of the Library's mission to promote open access at the Cyprus University of Technology. The main reason for universities to have institutional repositories is to enhance the visibility, access and impact of the research product of the university.

Researchers can be actively involved in the promotion of open access by submitting their articles in Open Access journals. Academics can also be more involved in the promotion of Open Access by educating the next generation of scientists in open access. They have to make sure that the new researchers understand the issues of open access and that it increases the impact of research articles.

2.9 Statistics of use

The Ktisis access statistics show that there is a continuous interest in the content with users visiting Ktisis from all over the world. Analytically, approximately 20000 visitors visited Ktisis in the year 2009. The average number of visitors per month is 1750 from which the 24% are returning visitors.

Approximately 1300 visitors accessed Ktisis directly, 5000 through referring sites and the rest of the visitors found Ktisis through search engines. Ktisis's visitors came from countries such as Greece, Cyprus, the United States, the United Kingdom, India, Germany, Canada, Australia, China, Iran, etc.

2.10 Future Work

As mentioned before, one of the Library's goals is to promote interest and involvement in the digitization process and preservation of cultural heritage. Therefore the Library aims to contribute to the KYPRIANA program(http://www.kypriana.eu/). This program is collaboration between various Cyprus organizations in order to contribute in a coordinated way in the growth of the EUROPEANA (www.europeana.eu) project. KYPRIANA aims to become the digital library of Cyprus.

Ktisis has been created using the open source software DSpace. The default installation of DSpace gives the capability to export the metadata in the Dublin Core standard using the OAI-PMH protocol. The problem is that these data cannot be imported into Europeana as Europeana does not support the Dublin core standard, but the Europeana Semantic Elements (ESE), an application profile based on the Dublin Core. The last year a lot of work has been done in Greece in the basis of the program EuropeanaLocal in order to overcome this problem. The outcome of this work was the creation of a DSpace plugin which has been created and is available to use free of

charge in DSpace installations. The Library aims to apply this plugin into the Ktisis installation making the Ktisis data compatible to the Europeana requirements.

2.11 Lessons Learnt

Today Libraries around the world are deeply involved in the task of preserving the intellectual and cultural heritage of their countries. The Library and Information Services of the Cyprus University of Technology is a pioneer in this field in Cyprus being the very first organization that has created an institutional repository.

The barriers that have come across the Library's work during the development of Ktisis have led the team to infer a number of lessons learnt that may be useful to others. First, we have realized that building an institutional repository requires engaging people with different skills and knowledge in the areas of project management, technology, cataloguing, metadata, etc. Additionally, the development of a repository is a procedure that requires a lot of time. The staff working on the development needs to have a balanced workload in order to reach the short and long term goals of the project. Libraries with limited resources need to be creative and persistent when creating a repository as there are many difficulties in the process.

Institutional repositories are still relevantly new but a number of important benefits have already become apparent. First of all, they support the open access movement. They provide a specialized information service by giving access to the academic and research work of the university and to the digital collections of cultural heritage material that the Library has digitized. Today institutional repositories address a very strong need in the academic world and therefore there will be a steady growth of repositories globally.

References

1.BORBINHA, J. An infrastructure for a national digital library.

2. Harnad, S. (2005). The implementation of the Berlin declaration on open access. D-Lib Magazine, 11(3)

3.Lawrence, S., Giles, C. L., & Bollacker, K. (1999). Digital libraries and autonomous citation indexing. IEEE Computer, 32(6), 67-71.

4.Lopatin, L. (2006). Library digitization projects, issues and guidelines. Library Hi Tech, 24(2), 273-289.

5.Lynch, C., & Garcia-Molina, H. (1996). Interoperability, scaling, and the digital libraries research agenda. Microcomputers for Information Management, 13(2), 85-132.

6.Samuelson, P. (1995). Copyright and digital libraries. Communications of the ACM, 38(4)

7. Suber, P. (2003). Removing the barriers to research: An introduction to open access for librarians. College & Research Libraries News, 64(2), 92-94.

8. Hutar, J., Melichar, M., Stoklasova, B. Czech National Digital Library and Long-term Preservation Issues, CULTURAL HERITAGE on line Empowering users: an active role for user communities,15-16 December 2009, Florence, Italy