

Towards the integral management of library collections at the Technical University of Catalonia (UPC)

Roser Gómez roser.gomez@upc.es , **Marta López** marta.lopez-vivancos@upc.es , **Mercè Mestre** merce.mestre@upc.es , **Jordi Prats** jordi.prats@upc.es , **Anna Rovira** anna.rovira@upc.es

Technical University of Catalonia (UPC) Library.

Abstract:	1
1. Introduction	2
2. New roles for the library.....	2
3. Document archives at UPC libraries.....	2
4. Topics to consider.....	5
5. Integration model.....	5
6. Conclusions	7
7. Bibliography	7

Abstract:

The University Library at the Technical University of Catalonia (UPC) currently manages several collections with different automated systems. These collections include the bibliographic collections of libraries, the Digital Library (which includes commercial products, such as books, journals and databases) and the documents generated by the University (such as dissertations, teaching materials, research results, projects by students, digital videos, etc.).

The production processes of the digital materials that are generated by the University, which are highly specific and complex, follow an established workflow that traditional management systems are unprepared for. Increasingly, the University requests the Library to actively participate not only in the management of the final documents (within the library) but also during the entire life cycle of the documents generated. We should keep in mind that digital documentation has its own specific publication and editing requirements, which are sometimes very sophisticated, and therefore the management of different kinds of materials with one overall system is difficult.

The University Library has developed different systems to suit the characteristics of each collection, which has also spawned different support products that meet specific needs to be developed for the management of these collections. These products are therefore very efficient; however, they have led to a disparity in the procedures for document retrieval and, furthermore, the system's development and maintenance has become very expensive.

Faced with this situation, the University Library is analysing the advantages and disadvantages of using a unique management system or, alternatively, of continuing to employ different systems, as the previous stage to the design of a new strategy for the management of university production.

Keywords: Institutional repositories, OAI-MHP, Data Interoperability, digital collections management, university digital libraries

1. Introduction

In 2000, the UPC Library set its third strategic plan in motion. This plan, called Paideia, enunciated the need to launch a digital library project that would respond to newly-arisen necessities in the fields of learning and research.

The new digital library project builds upon the library's experience through its earlier web site, the most important aspects of which it maintains: the free-access online resource database, collections of electronic journals and commercial databases. The new digital library is not merely aimed at being the University Library's web site, but rather to serve as a gateway to the knowledge that UPC generates, and to give it greater visibility.

Managing this knowledge entails the difficult challenge of working the digital library into the procedures that come into play in the lifecycle of documents, and with the understanding that in many cases guaranteeing that a document will be preserved means taking responsibility for its publication. The plethora of creation and distribution channels, along with little standardisation and different publication formats make managing these sorts of documents even more difficult.

The aim of this paper is to give a summary of the current situation and to show a model for integrating all the University's collections into one single retrieval system that will allow global access that is not limited by the nature of the documents retrieved.

2. New roles for the library

The University Library manages scientific and technical documents that are gathered as support materials for teaching and research as well as those that are generated by the University itself, the collections of which have historically been fragmentary or unsystematic.

The introduction of information and communications technologies in the publication process has streamlined the process of gathering, managing and preserving documents, despite their numerous different formats.

Increasingly, the processes of gaining access to documents and producing and publishing them—tasks which before these technologies were clearly differentiated—are bleeding into each other and losing distinctness. In this environment, the library not only facilitates access to and the preservation of documents, but is also involved in the processes of creation and publication through working closely with the actors involved, such as lecturers and university publishing services.

3. Document archives at UPC libraries

Here we summarise the research and teaching document archives that are managed by the UPC Library; while some of them have been set in motion others are still in phases of development.

The Teacher's Mailbox

This project entails a collaboration between the University Library and the Institute of Education Sciences (ICE), which contains teaching material produced by lecturers: notes, transparencies, tests and presentations, all of which are labelled with very simple metadata (author, title, format and optional description) and other administrative metadata (school, qualification and subject).

The goal of the archive is to offer lecturers and students a server on which to save and publish teaching materials online in an organised and integrated manner. The archive guarantees access, maintenance, security and the usability of the materials.

Currently, the Teacher's Mailbox has been integrated into UPC's Digital Campus, which will facilitate access to its content for students and lecturers in all subject areas.

Examinations

This archive contains digitalised exams so as to offer students quick and easy access to past exams, as well as to preserve them and make them available to the public.

The format that is used is PDF. The documents are tagged with very simple metadata (author, title, file format and optional description) and other metadata are assigned for administrative purposes (school, subject and exam date).

The process of data entry begins in the libraries, where the paper exams are collected and subsequently scanned and uploaded to the server.

The archive has to be compatible with the rest of the servers that contain teaching material, such as the Teacher's Mailbox, and one must be able to consult its content from different teaching platforms.

UPC Books

This database contains electronic books written by UPC lecturers, which have been underwritten by a publishing council and marketed by *Edicions UPC*. These documents can be accessed free of charge from *Bibliotècnica*, the UPC Digital Library by the university community.

Each book has an identification code (ISBN or other) and is made up of a number of folders.

Edicions UPC is the owner of the database, but the process of data entry and maintenance is under taken by the UPC Library staff.

UPC Books is supposed to be integrated with the teaching platforms, *FènixDoc*ⁱ and the UPC Library Catalogue.

UPC Digital Video Archive

This project is a collaboration between the Institute of Education Sciences (ICE), the Technical School of Castelldefels (EPSC) and the University Library, whose aim is to create a digital archive of videos of UPC's institutional and teaching activities.

The data feeding process is undertaken at the Teaching Resources Factoriesⁱⁱ located at the different UPC library locations.

The heterogeneity of its content means that access can be gained to the Digital Video Archive from a variety of different University platforms, such as from the UPC web site (as a link), or the Digital Campus.

Final Theses (PFC/TFC)

This archive comprises the final theses of those students who consent to having them made accessible online. It is preferred that the theses be submitted in PDF format. The objective of this project is to preserve the work done by students before obtaining their Degrees as well as to make it more visible.

The author and the University Library's interns are responsible for introducing the text and the librarian revises the thesis's metadata before its final publication.

Research Archive

Herein one finds research reports and preprints/post prints produced by lecturers so as to keep a record of the results of their research as well as of the journals published by UPC. The aim is to preserve and publicise the results of UPC's teaching and research staff.

The documents are tagged using the Dublin Core standard according to the guidelines set by the University Libraries Consortium of Catalonia (CBUC) alongside additional administrative metadata.

Research reports and preprints/post prints are self-archived by their authors (lecturers and researchers). Journals are introduced by University Library interns and the metadata are revised by librarians.

The Research Archive has still to be integrated with *FènixDoc* and the UPC Library Catalogue.

TDX - Doctoral Theses Online

This database contains digital copies of doctoral theses defended at Catalan universities in order to provide free access to them on the Internet. Since 2000, the TDX has participated in the NDLTD Project.

Its aims are to broadcast the results of university research around the world, to increase the accessibility and visibility of its authors and to promote the electronic publication of research work from Catalan universities.

Authors send complete texts and abstracts in the format indicated therein and sign publishing contracts that authorise UPC to publish their theses.

This online resource has no need to be integrated with any other archive. TDX is publicised through aggregators that use the Open Archives Initiative (among which figures OAIster).

4. Topics to consider

The management models followed by the organisers of each archive project entailed developing differentiated products and focusing on specific objectives for each project. In this way, each archive could develop in its own way without being bound to the growth and evolution of the other collections.

One of the goals that has been sought through the use of this model is the integration of the Digital Library with other information technology systems at the University; the Teaching Intranets have been integrated with the Teacher's Mailbox and the publications of UPC Books with *FènixDoc*. This integration entails an increase in costs in terms of the development of the archives, which is vital to the Digital Library's not becoming isolated from the rest of the University's activities.

A second idea to consider in the development of the archives is unifying them by giving the databases used to manage them a single, common basic structure. This would then unify most of the metadata that describe their respective contents, which would guarantee a great degree of coherence from one archive to the next.

Nevertheless, in some cases, such as that of the Examinations Archive, agility in the publication of content and ease of management have taken precedence over the quality of document tags or the development of more sophisticated data retrieval systems, which has had direct repercussions on the quality of the metadata assigned to them.

Such a system does not contribute to a vision of the collective of archives as a whole. Other aspects, such as copyright control or indexing, have required complementary projects in order to be carried out. An example of this is the *Bibliotècnica By Subject* portal, which was developed to facilitate access to specific Digital Library content by thematic area. The integration model proposed in the following section follows in this line of work.

5. Integration model

The management of the data that make up the Digital Library makes it possible for one to gain access to them in a number of different ways. Though we have mentioned the advantages of there being different systems of managing data in each of the archives, we should also focus on how to integrate them into a single data retrieval system that will in the future offer users global access.

Up to the present, the strategy that has been followed has been that of cataloguing doctoral theses on the TDX and the online books published by *Edicions UPC* on the UPC Library Catalogue and the University Union Catalogue of Catalonia (CCUC). Though cataloguing following the AACR2 standard and the MARC format has enabled a highly effective integration of these data with other resources, it has doubled the amount of work due to the fact that data are first entered into the archive and later again, manually, to the catalogue. This model is perfectly applicable to doctoral theses, but it would be difficult to apply it to other archives in which collection expansion is highly variable and a constant workload is difficult to maintain.

The Z39.50 standard does not seem to be a viable solution for the integration of all the collections. Its complexity makes the development of the tools needed for its implementation prohibitively expensive.

One should also bear in mind that the University Libraries Consortium of Catalonia (CBUC) is currently studying changing the automation system used by the University Union Catalogue of Catalonia (CCUC), which would also entail changing the UPC Library Catalogue. Furthermore, the incorporation of tools such as meta-search engines and DOMs has also been looked at as a way to take on new projects and to exploit the resources being developed or subscribed to in conjunction with others.

Out of the archives currently being developed by the University Library, at least five (TDX, PFC, the Digital Video Archive, Teacher's Mailbox and the Research Archive) have the goal of publishing contributors' work in open access. Providing these collections with a greater visibility is not made possible by simply granting open access to them, but rather through incorporating metadata that describe their content to external meta-search engines, which in turn enhance their potential use and accessibility. Since 2001, based on the initiative undertaken by the Open Archive Initiative through the development of the OAI-MHP protocol, achieving this goal has meant incorporating this standard into collection management and thus making the library a provider of metadata to external services. The CBUC has already integrated this protocol into the TDX archive.

The generation of OAI-MHP servers for the aforementioned archives marks the departure point for a project to integrate all of them; this in turn would facilitate access to metadata and would allow the creation of a single archive of metadata. The quick pace with which this protocol is being implemented in a number of environments will guarantee its applicability to future projects.

In the long term, the project to create a single archive of metadata foresees the following:

- Defining a common model for metadata for every archive based on the specifications drawn up by the CBUC for the implementation of the Dublin Core (DC) standard.
- Enhancing the basic structure with unique administrative metadata, such as levels of accessibility, which serve to enrich the information.
- Studying the degree of granularity of the objects in the archives, such as that of examinations. It will have to be determined whether it makes sense to retrieve individual examinations from this archive, or whether it would be more effective to link to the web page of the subject to gain access to online copies of all the related exams.
- Creating OAI-MHP servers for internal and external use.
- Studying, based on the Dublin Core standard, the indexing of other web pages that make up the Digital Library, and developing a parser to analyse pages and extract metadata from them.
- Developing an OAI-MHP server for the UPC Library Catalogue.
- Creating a single archive of metadata.
- Creating a data inquiry and retrieval system for the metadata archive.

This project already has a large number of tools in free code that will facilitate its implementation. Namely, the products developed by the University of Illinois have proved to be readily adaptable to our requirements, as they make use of programming tools that are already familiar to the University Library. Most archives are being developed using the same platform and follow the same data structure, which makes work easier and facilitates the application of unified solutions.

Finally, the experienced gained from working with this standard will go on to be used to enrich the retrieval tool by capturing resource metadata not directly managed by UPC.

6. Conclusions

The role played by university libraries in the process of transmitting and generating new knowledge continues to evolve. They are passing from being mere providers of scientific and technical information to becoming producers—and publishers—of documents.

The disperse geographical nature of scientific production may become a critical factor if UPC aims to present a unified image of its research activity. Fortunately, standards are continually being defined. In an environment such as the Internet, for which standardisation is made difficult by the sheer number of actors it involves and is often determined due to the commonly-accepted use of a given protocol, the high acceptance of OAI-MHP or the fact that Dublin Core has already been recognised as an ISO standard mean that streamlining projects along the lines of these standards is both reliable and profitable.

The accelerating evolution towards open-access systems for specialised documents also implies the need to guarantee access to external resources, which need not always be commercial ones, from the Digital Library itself. Again, the use of the standards mentioned above facilitates the integration of these resources into the Digital Library in a way that is clearly transparent to users. By adding metadata assigned by external providers to its cataloguing system, the Library expands its horizons as it reduces the dispersion and fragmentation of information.

7. Bibliography

Arms, William Y., et al. A case study in metadata harvesting: the NSDL. *Library Hi Tech*. 21 (2): 228-237

Arms, William Y., et al. (2002). A spectrum of interoperability: the site for science prototype for the NSDL. *D-Lib Magazine*. 8 (1)
<<http://www.dlib.org/dlib/january02/arms/01arms.html>>

Fang, Li. (2004). A developing search service: heterogeneous resources integration and retrieval system. *D-Lib Magazine*. 10 (3)
<<http://www.dlib.org/dlib/march04/fang/03fang.html>>

Gómez, Roser; López, Marta; Rovira, Anna (2003). A New Type of Library for the New European Space for Education:UPC's Learning and Research Resources Centre. *IATUL Proceedings (New Series)*. 13
<http://www.iatul.org/conference/proceedings/vol13/papers/GOMEZ_fulltext.pdf>

Hagerdon, Kat. (2003). OAIster: a "no dead ends" service provider. *Library Hi Tech*. 21 (2): 170-181

Open Archives Initiative. (2003). *The Open Archives Initiative Protocol for Metadata Harvesting*
<<http://www.openarchives.org/OAI/openarchivesprotocol.html>>

Prats, J., Rico, O., Calderón, P., Majó, R. (2003). Bibliotècnica: la biblioteca digital de la UPC. *Boletín de la Red Iris*. (66-67): 47:50

University of Illinois. (2003). *UIUC OAI Metadata Harvesting Project*
<<http://uilib-oai.sourceforge.net/>>

University of Illinois. (2003). *Implementation of a Scholarly Information Portal Using the Open Archives Initiative Protocol for Metadata Harvesting: Final Report to the Andrew W. Mellon Foundation*.
<<http://oai.grainger.uiuc.edu/FinalReport/TableofContents.htm>>

ⁱ FènixDoc is the portal to the UPC scientific and research production Information System (Fènix) and that offers access to the full-text documents that are available at the Digital Library (electronic journals, Doctoral Theses or UPC Books)

ⁱⁱ The aims of the Factories are to offer tools and specialised personal support to produce new multimedia teaching materials