

# **Thematic digital libraries at the University of Porto: Metadata integration over a repository infrastructure**

Isabel Barroso<sup>1</sup>, Marta Azevedo<sup>2</sup>, Cristina Ribeiro<sup>3</sup>

<sup>1</sup> Faculty of Fine Arts, University of Porto; <sup>2</sup> Faculty of Food and Nutritional Sciences, University of Porto; <sup>3</sup> Faculty of Engineering, University of Porto/INESC-Porto

The University of Porto has a well-established set of specialized libraries serving the research and student population of its 14 schools. Thematic digital libraries can be valuable for organizing specific collections and for supporting emergent communities. This work focuses on two case studies, one in the area of the Fine Arts and the other in the area of Food and Nutrition. For building both digital libraries we propose to use the existing university repository infrastructure and to establish a metadata workflow that makes use of available descriptions in the library catalogues and in the university information system. We expect that such an approach, which takes into account the institutional context and resources, can be used in other collections at our university and inspire similar initiatives elsewhere.

**Keywords: metadata integration, institutional repositories, DSpace, digital libraries, description and cataloguing**

## **1 Thematic digital libraries at the University of Porto**

The University of Porto is the largest Portuguese university, serving a population of over 28,000 students and covering research areas that range from the humanities, social sciences, fine arts, architecture, economics and business studies to science, engineering, medicine, sports and law. The university currently has three information systems that handle bibliographic records. The online library catalogues, managed with the Aleph system, are used for cataloguing, acquisition, loan and administration. The university information system developed on the in-house SIGARRA technology [1] includes the publications module used by the research community to deposit scientific publications, which are then validated by the library services. The university repository, built on the DSpace platform [2], is intended to manage and give open access to the scientific production of the university.

Our goal is to create thematic collections of digital objects, in the form of digital libraries, taking into account the information management context in the university and reusing existing bibliographic descriptions. This will guarantee quality descriptions in the libraries, avoid duplicate description, and use the technological support of an existing repository service. We build on earlier work concerning integration platforms and the extension of institutional repositories [3,4] and create specific procedures for metadata integration.

## **2 The Fine Arts and the Food and Nutrition Digital Libraries**

The Library of Fine Arts holds an important collection with almost 230 years of history. The selected collection includes a variety of documents from the historical archive and the library: ancient books and magazines, engravings from 16<sup>th</sup> to 19<sup>th</sup> century, as well as minutes from ordinary and general conferences, letters, personal files of artists as students and as teachers and manuscripts. Researchers, students and teachers within and outside the University often request the collection. Significantly, none of these documents is accessible on the online catalogue. By providing the integration of collections from both the historical archive and the library, we will enable users to have full access to the collection and contribute to its preservation. As an example, a search for Silva Porto, a well known Portuguese painter, will retrieve all kinds of documents, including exhibition catalogues, photographs, his personal file, books written by and of him or even minutes where his name is mentioned, letters received and delivered all over his academic life.

The Food and Nutrition Digital Library aims to preserve and ensure access and reuse of the bibliographic works developed by the faculty community for scientific, academic, didactic and dissemination purposes. The prototype digital library will be based on the collection of digitized documents authored by one of the mentors of the institution, Dr. Emílio Peres. Dr. Peres, considered the father of Food and Nutrition Sciences in Portugal, is recognized for his seminal and unique work, for his brilliant communication skills and for his fresh scientific vision. The collection includes valuable manuscripts concerning the Food Education National Campaigns in the 80's, sets of slides used in his first lectures, the myriad articles published in local and national media, as well as radio conferences and TV interviews and programs.

Regarding both case studies, the main issue is that some of the objects are already described in the online bibliographic catalogue or in the university information system, but others will have to be described. The need to reuse existing quality metadata for describing digital objects presented a challenge to integrate metadata from different systems into the digital libraries.

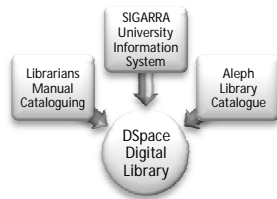
## **3 Integrating metadata in DSpace**

DSpace software [2] was conceived mainly to accommodate institutional repositories of scientific production and uses the Dublin Core format to register metadata that can be imported using an XML format. Using DSpace for our thematic digital libraries has required attention to three issues: the integration of the digital libraries in the university infrastructures; the reuse of cataloguing materials and the adaptation of DSpace features to the thematic digital libraries purposes.

The first problem required a clear view of the existing resources and some agreement with the university services. The digital libraries are sharing existing library and IT staff and this has led to the choice of DSpace, which is already supported by the university.

The second problem has led to the development of two metadata workflows to gather existing object descriptions in the catalogues and in the university information

system. The diagram in Figure 1 illustrates the metadata creation for the digital library based on the new and existing descriptions.



**Fig.1.** Metadata sources for the digital libraries

After analyzing the existing resources within the University we decided to establish methods for exporting record metadata, and the associated digital objects when available, from the Aleph catalogue to DSpace and from the SIGARRA system to DSpace, in order to reuse bibliographic records.

To guarantee that all required record data were exported from Aleph to DSpace we have designed a coding schema for the collections and used it on the UNIMARC records in Aleph, where CCL commands then allow us to select records for export. The resulting XML document with chosen UNIMARC fields is mapped to Dublin Core format [5] using a custom-designed XSL. The export process is controlled by publication date on the bibliographic catalogue.

The export procedure from SIGARRA, the university information system, to DSpace, though not tested yet, will be supported by the existing procedures developed for the university repository.

The third problem mentioned above, regarding the design of a digital library interface in DSpace, is still being addressed and will require several development and test steps using the intended target audience.

Since the digital libraries will accommodate mostly reserved and ancient collections we decided to publish documents in 600x400 format with image resolution of 200 dpi in JPEG format, keeping restricted access to files using 600 dpi TIFF format to guarantee collection preservation. Depending on their condition, the documents will be either digitized or photographed. For text documents, two files will be generated, one using TIFF format for collection preservation and a PDF extracted text for full-text indexing.

The export procedure from SIGARRA, the university information system, to DSpace, though not tested yet, will be supported by the existing procedures developed for the university repository.

## 4 Prototype Digital Libraries

DSpace organizes repositories by communities that can be used to represent the units in an organization- departments, libraries, museums. Collections are organized for a community according to themes or document types. The diagram in Figure 2 illustrates the community and collection structure of the two case studies in DSpace.

Since the Fine Arts collection has documents of different sources we had to consider organizing the communities by services, and collections by document type. The archival cataloguing rules typically use a hierarchical structure capturing document provenance [6]. We created two sub-communities, archive and library, and three sub-communities within the archive community (Portuense Academy of Fine Arts, School of the Arts and Portuense Museum). For the library community there was no need to create sub-communities.

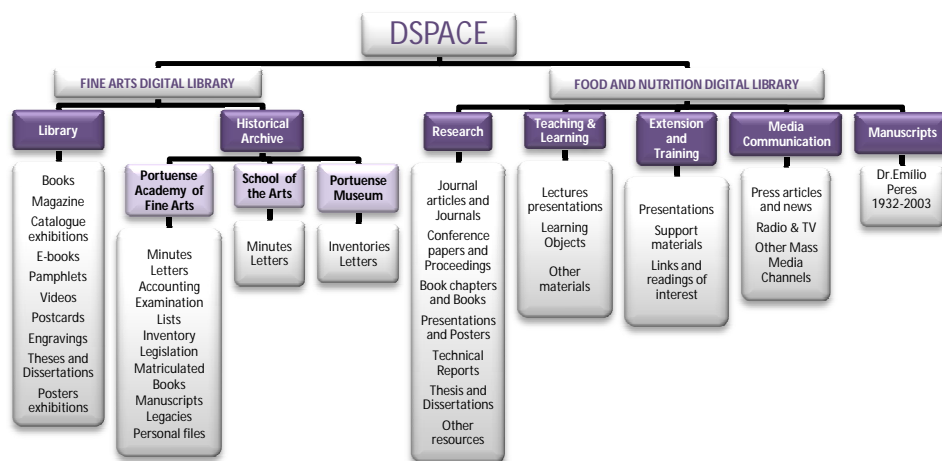


Fig. 2. Communities and collections for the two case studies

The nature of the prototype collection for the Food and Nutrition Digital Library, which is composed by different types of work, from scientific articles and books to newspaper and magazine articles, lecture presentations and didactic material, justifies a functional organization of the digital library in different areas featured as communities of action for faculty members: Research; Teaching and Learning; Extension and Training; Media Communication. These main action fields are functionally viewed as dynamic because some documents types and materials can be of interest to different audiences and used in different contexts by the professionals. Generically, the collection organization tries to respond to the uneven goal of disseminating science for the public interest: from original scientific materials to the local newspaper.

## References

1. Ribeiro, LÍgia M., David, Gabriel, Azevedo, Ana, Santos, J. C. Marques dos. Developing an information system at the Engineering Faculty of Porto University. FEUP, Porto (1997). Available at: <http://repositorio.up.pt/aberto/handle/10216/606>
2. Smith, M., Bass, M., McClellan, G., Tansley R., Barton M., Branschofsky, M., Stuve, D., Walker, J. H.: DSpace: An Open Source Dynamic Digital Repository. D-Lib Magazine. 9 (1) (2003).
3. Pyrounakis, G., Saidis, K., Nikolaidou, M., Lourdi, I. Designing an integrated digital library framework to support multiple heterogeneous collections. 8th European Conference, ECDL 2004. LNCS. Vol.3232, p 26-37, (2004).
4. Wise, M., Spiro, L., Henry, G., Byrd, S. Expanding roles for the institutional repository. OCLC Systems & Services, v 23, n 2, p 216-23, (2007).
5. Permanent UNIMARC Committee. UNIMARC/Dublin Core Mapping. Available at: <http://www.unimarc.net/dubin-core-map.html>.
6. ISAD(G): General International Standard Archival Description, Second edition. <http://www.ica.org/en/node/30000>