An Approach for a National Digital Library

JOSÉ LUIS BORBINHA

National Library of Portugal – Director for Innovation and Development
Jose.Borbinha@bn.pt

Keywords: National Libraries, Interoperability, Digital Preservation, Digital Deposit, Digitization, Digital Publishing.

Introduction

The National Library of Portugal is a patrimonial library. This presents special motivations to pursue some goals at both national and international levels, related with a wide range of new technical area especially concerning with digital publishing; digitization; preservation of cultural artifacts; metadata creation, processing and exchange; services for resource discovery, access and interoperability; etc. Each of these areas brings new expectations and requirements for new skills, which BN has been identifying and developing during recent years in trials and pilot initiatives. As a result, valuable expertise has been built, making it possible to conceive the actual model for the National Digital Library. This has been conceived as a coherent group of services and resources for users, developed and maintained by professionals, with the involvement of external agents. The technical solutions are based on open and scalable technology, and have been being built and integrated incrementally, with the purpose to be consolidated with the traditional library's services. This paper describes the motivations, requirements and results of that case.

The Problem

As most of the national libraries with similar missions, BN - “Biblioteca Nacional” (National Library of Portugal) has been concerned with its future in a "digital world". BN has been, for a relative a long time, an active player in the areas of technology and digital libraries, since the earlier times of the UNIMARC and bibliographic metadata developments [1], until several of the recent most relevant international projects and initiatives [2][3][4][5][6][7][8][9]. This experience, consolidated with the expertise from its 250 years of history, is what BN is using to define its strategy for its BND - “Biblioteca Nacional Digital” (Nation Digital Library). The actual problem of the BND is not anymore to be experimental, presenting short-term results, but to be stable and trustworthy. That requires thinking seriously in stable models and long-term sustainable strategies for the developments and convergence with the traditional library, because in the end we don't want to have two different libraries, but just one, working coherently. Also, that has to be done in accordance not only with our general mission, but also with the resources that BN expects to be able to dedicate to it, which are not going to be extraordinary.

To address this problem we focused our attention in four major groups of issues:
• Contents: What new genres of contents will national libraries have to deal with in a digital world? How will those be related with the traditional contents? What are the new challenges in this area?
• Actors: Who will build and maintain this new library? Who will use it? Who will work around it? Are we only going to have to rethink the traditional roles, or will we have to consider new ones?
• Services: Do we need to design new services for the digital paradigm? If yes, how these new services will be integrated with the traditional ones? Are we facing really new issues, or just the need to rethink traditional ones?
• Technology: What is the best strategy to deal with our requirements for the digital library and the fast technological obsolescence? Can we afford and trust in the market? Can we expect mature and trustworthy products and technology? We are far away from having the answers to all of these questions, but we feel that we are already in a good position to move on, with strong basis for correct decisions.

Contents

The legal deposit in Portugal is mandatory for all the printing houses, requiring the deposit of the works at BN. In consequence, BN holds a unique collection of printed works, representing an important potential for digitization and also for the creation of new digital resources. Therefore, deposit, digitization and digital publishing are the three main channels to provide contents to the BND.

Digital deposit

The current Portuguese legal deposit law covers only printed materials forcing publishers or printers to deliver copies of the works to the BN. To date we don’t see a strong need to change this law to cover digital works. Rather, we intend to promote a voluntary deposit model, based on agreements with publishers and authors. The main reason is due to the huge complexity in the identification of relevant digital genres, which will need to take in consideration production processes, publishing and business models, and technological specificities.

In the actual moment, a small national library like BN, in a small country like Portugal, can not manage to be able to design a complete model to address responsibly all these issues in a coherent and uniform scenario. To address this problem BN is giving preference to a strategy based on the identification of specific relevant scenarios, so each one can be addressed as is most suitable, according with its specific formal or technical requirements. Scenarios for the deposit of these genres of resources are being considered: grey literature, periodicals (journals and newspapers), digital monographs, and web harvesting.

Grey literature includes mainly thesis, dissertations and other common technical and scientific reports. This scenario takes advantage of the cooperation with the university libraries and scholars, whom usually produce nowadays those works in common word-processing formats, with no complex requirements for presentation neither complex technical formats. The main actions in this area are being promoted in cooperation with the NDLTD initiative [8].

Journals and newspapers represent an important part of the so called “deep web” (contents available in the web, but published by complex systems or protected by
access control models). A challenge in the deposit of these genres is the development of models of cooperation with the publishers. Technically, we need to assure that the deposit of the contents of publishing systems and databases is done in a way that will make it possible its long-term access. Formally, we will need to be also careful with the rules for access to that content, so we will not put in danger business models based, for example, in the free access to new materials and the selling of the old ones (a common case for newspapers). BN will start in early 2004 an action for the deposit of newspapers, while it is under analysis a collaboration with the initiative LOKSS for the deposit of journals [10].

In a third front, BN is accepting for deposit generic digital monographs or similar contents, which are stored and catalogued in the national bibliography as any other item. An important target is to assure the deposit of at least governmental documents, where a large number of them are being published digitally. This will be the purpose of a strong action in 2004 (the technological framework to support this will be shared with the deposit of grey literature).

Finally, concerning harvesting, BN is developing a service for the selective deposit of web sites (harvested based on their registration). On the same time we are following with strong interest the initiatives dealing with the wide harvesting of “national web spaces”, especially those in collaboration with the Internet Archive [11].

**Digitization**

Digitization is one of the most interesting areas of immediate investment for national libraries, who can by this means make available to general access hidden and invaluable treasures. The main target of the digitization program at BN is to build until the end of 2005 a collection of at least one million of images, taken from the most important works in the library, or from important complementary items held in other institutions. These images will be mainly of complete valuable rare works, or works that are at serious risk of deterioration (this will thus not only provide better access but also help to preserve those works). After a few years of trials in several areas, we feel it possible to develop now a wide digitization program based on internal core activities and outsourced contracts.

**Digital publishing**

New interesting digital works, created by the national libraries or by programs promoted by them, can add value to the digitized collections. They can help potential users to better perceive their scientific or cultural relevance, which might represent a new front line in the promotion of the collections. Therefore, the digital publishing initiative at BND will cover the same thematic areas as digitization, with the purpose of augmenting the perception of them and their context. Those comprise, for example, the development of rich biographies of key intellectuals and scientists, or of virtual thematic exhibitions.

Another important objective in the area of digital publishing is the coordination of these activities with the traditional publishing program of the library. This must promote a core publishing activity that will be independent of the medium, where the whole production process (or at least an important part of it) will be common to both online and printed media.
**Actors**

The digital paradigm requires also a new look to the actors of the library. The three classes of actors representing the main interfaces by which the digital library will be operated and perceived are: users, professionals and agents [12].

The end users are expected to be researchers and ordinary citizens. Traditionally, national libraries have been mainly places for research, with unique services and resources. Normal citizens have had in national libraries a guarantee of democratic access to wide sources of information, especially related to the national identity, society, history, culture, and science and technology, but the physical distances have been a barrier to the access to those resources. But those citizens are now linked to the Internet in their homes, demanding virtual access to public services, including the national library. This brings new requirements and the need to develop new services or rethink the traditional ones performed by national libraries, which constitutes a completely new challenge.

Professionals in a “Digital Library” can not be anymore only librarians. Traditional librarians have to share the library now with other classes of specialists on information management, and especially with specialists on technology. To problems in attracting and maintain highly skilled specialists on technology has been an important limitation for the national libraries, where they are competing with the stronger private market. This means that libraries might not be able to embody important areas of knowledge, required to operate in the digital world.

Recognizing this, BN has developed important links with companies, universities and research centers in the country. Also important are the links with other institutions, such as archives and museums. This requires a new mentality for the library's professionals, who have now not to see external actors as producers of goods and services, but as partners!

Finally, agents are external actors comprising a wide range of possible new classes of partners, playing traditional or new roles. For example, it is easier now for a national library to develop direct links or interfaces for publishers, or for the authors themselves, making it easy for them to provide or simply improve the metadata of their resources. In an ideal scenario we would want to apply also standardization and formal interoperability to these interfaces, but the right strategies for that do not seem to be clear yet. Important experience already exists [13], but we are still quite a long way from sustainable solutions, especially in Portugal, where the publishing market is very fragmented and there are no big players.

**Services and Technology**

The main services that the BND is expected to assure are digital resource storage, search, retrieval and preservation. In the area of search and retrieval the problem of interoperability is also very relevant.

The price of the storage technology has been dropping in an amazing rate. However, the prices of complete commercial storage solutions for libraries have been not following this trend in the same speed. Top market companies still price their products too high, especially when we compare them with the prices of the original hardware components and the expected software development. Also, commercial companies have been too slow (too cautious?) in adopting innovative emerging technology for
digital libraries. After realizing this, it was decided to build a new specific framework for BND, based on open technology partnerships, and alliances, and internal developments. The estimated final cost for this action is 25% of the cost of buying a commercial solution, or of the contracting of a new one by using the traditional user's requirements, design and contracting models.

For that purpose, it was chosen as a structural metadata model the new METS format [14], and the FEDORA [15] framework for middleware. For storage, it was developed a 20 TB solution, based on redundant cheap storage servers (at a cost lower than 3.000€ by TByte), running Linux operation system and managed as a GRID network [16]. Finally, for the identification of the resources it was implemented a simple but very efficient PURL service [17].

One interesting aspect of this new networked world is that it is also making it for libraries to cooperate in the sharing and reuse of services and information (metadata). The key word is ‘interoperability’. This means, for a national library in a networked world, the ability to interact and share its resources and services with other cultural heritage partners (libraries, archives, museums, etc.), other public services, and even commercial entities. In practice this means standards! Standards for interoperability are not new for libraries, which have for a long time been recognized as pioneers in the normalization of procedures and models for cooperation. In that sense we decided to use in BND the traditional bibliographic description technology already used for the union catalogue, based on UNIMARC, in what is a perfect marriage between the traditional and the emerging digital library. This means that all the resources of the digital library will be described in the national union catalogue.

To text new areas of interoperability, the records of the BND are made available for harvesting by OAI-PMH [18], while new experiments with Internet search engines are expected to produce new interesting resource discovery services [19].

**Conclusions**

The BND was defined to serve two main purposes. The first and most immediate is the development of a strategic framework by which the BN will address specific problems posed by the management of large collections of digital and digitized items. As a side effect, we expect to reach a minimum stable framework after which we will be able to address the other areas in a complementary approach (especially advanced services and long-term preservation).

It is not easy to present now relevant new conclusions, besides the stressing of the obvious problems already pointed out by the community. There is an emerging general feeling that libraries, and especially national libraries, have spent too much time and efforts discussing the “Digital Library”, with very few developments of real and sustainable cases. One reason for that might have been the reluctance in not acknowledging clearly that, in order to be able to do it properly, they’ll need to incorporate new radical technical skills, which in their turn will imply radical internal reorganizations. That is impossible to assure in a short-term, as also by traditional processes. That is why we in the BND are proceeding in small steps, at the size of our resources, but without any fear of the challenges (and learning by doing it)!
References