



Enhancing Digital Services at the National Library of Portugal



Portugal

Helena

Patrício

Abstract

This article presents the National Digital Library as a service for online access to digitised documents of the National Library of Portugal (NLP)¹ collections. An overview of the NLP strategic and tactical goals for the Library digital services is provided. The content available online is briefly characterised and the main functionalities of the digital information system are explained, underlining the case of special search, browse and display features regarding complex objects such as newspapers, dictionaries and encyclopaedias, as well as the role and value of the eBooks on Demand (EOD) service. Examples are presented of integrated services for access to NLP digital content through the Portuguese union catalogue PORBASE, Google, Europeana portal and The European Library (TEL).

Early developments and evolution

In 2002 the National Library of Portugal (NLP) launched the National Digital Library² as a special project for the quality improvement of its services, offering online access to digital content. Like many other digital libraries emerging at the time, the National Digital Library was financially supported by European Union Structural Funds and had a strong focus on IT innovation and experimentation, since there was a lack of software commercially available for imaging, online file display, metadata production, archival or digitisation workflow management.

Five years later, in 2007, deep reforms in the Portuguese public administration led to major changes in the structure of the NLP, namely

the creation of a department in charge of the production and management of digital content, with a strong focus on the full integration of this team with other NLP services, like the special collections services, for content selection and digital content organisation; the reproduction services, for on-demand digitisation; and the preservation and conservation service, for specific document interventions before or after the digitisation process. At the end of the restructuring stage, the National Digital Library was no longer just a special project, managed alongside the traditional organisation; it was starting to become a regular service of the NLP, as it is today.

A strategy towards growth and consolidation (2007-2011)

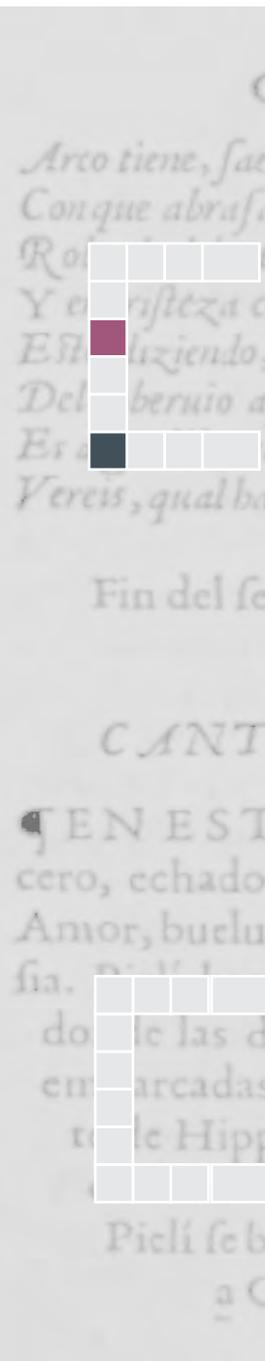
In the last five years, the strategic goals established for the NLP digital services focused on digital collection growth, strengthening of digitisation capabilities, improvement of public access services and development of new services for institutional cooperation. The actions taken to implement the NLP digital strategy and achieve these specific goals are outlined below.

Fostering growth of the digital collection

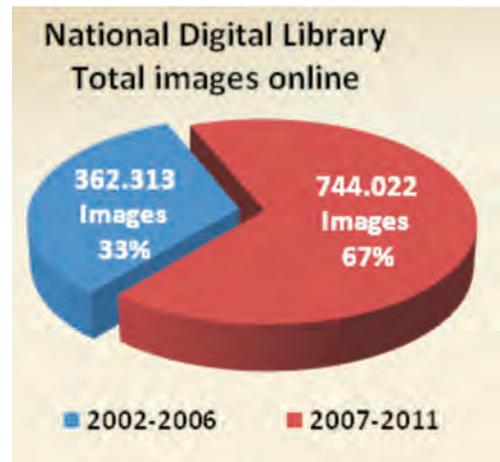
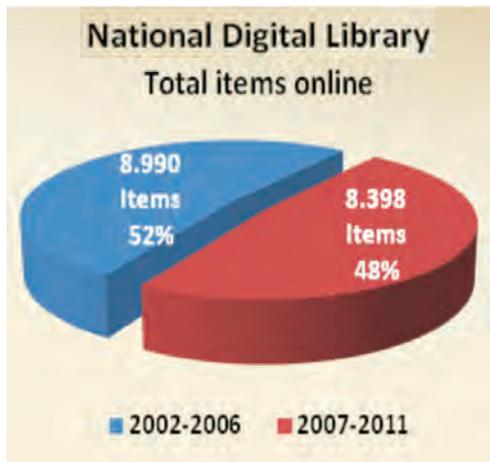
In the first quarter of 2011, the NLP digital collection reached 1 106 335 online images of 17 388 items.

Digital images available online since 2007 represent 67% of the collection, corresponding to 48% of all items published in the digital library.

Digital Collection	2002-2006	2007	2008	2009	2010	2011 (Jan-March)	Total
Images	362 313	21 197	98 948	188 231	246 344	189 302	1 106 335
Items	8 990	581	253	778	5 083	1 703	17 388



NLP digital objects can be searched and are available through other services like PORBASE, the online union catalogue of 170 Portuguese libraries, The European Library (TEL), Google, and Europeana.



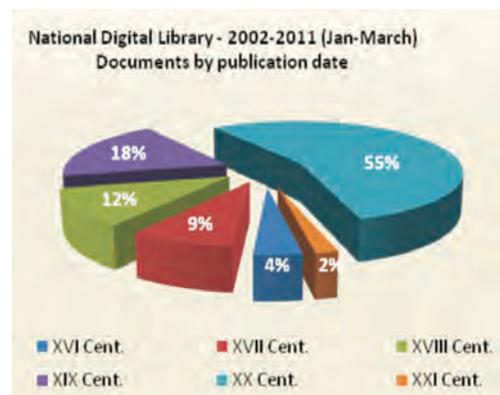
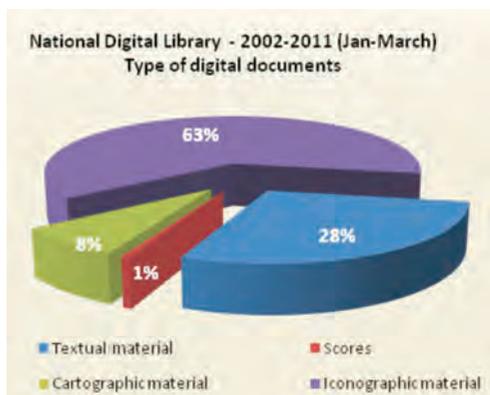
In order to achieve the goal of increasing the digital collections, the NLP has paid special attention to improving its production capabilities, by revising and renewing digitisation procedures and equipment, as well as the storage infrastructure, as it will be described further in this article.

Currently, the NLP digital collection is mainly composed of iconographic documents (63%), which were selected in early digitisation projects for preservation purposes. For the same preservation reasons, cartographic material was also intensively digitised in 2002-2006, amounting to 8% of the total content.

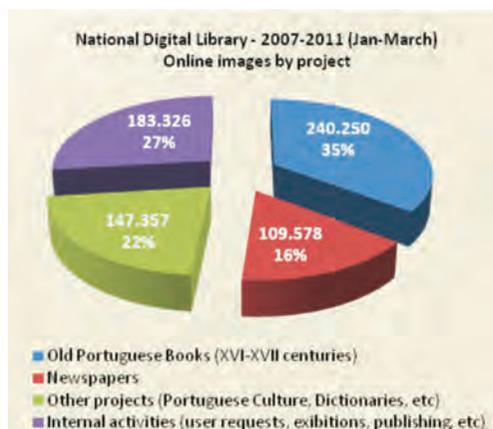
By date, the oldest document online is the 12th

Century handwritten copy of Santo Isidoro de Sevilha *De natura rerum ad Sisebutum*³. Most available documents, however, were published in the 20th Century (55%), including not only iconographic documents (mostly posters), but also born digital resources like NLP editions and academic thesis.

The most relevant content available online is from textual documents (28%), encompassing old Portuguese books from the 16th and 17th Centuries, important for its uniqueness and heritage value, and Portuguese newspapers of the 19th Century, because of their information richness and fragile nature of the items.



Since 2007, the NLP has followed four priorities/production flows for the digitisation programme: 1) old books printed in Portugal in the 16th and 17th Centuries; 2) 19th century Portuguese newspapers; 3) smaller projects with a specific opportunity, such as the "Portuguese Culture" project⁴ sponsored by the Luso-American Development Foundation; and 4) internal support to NLP activities like publishing or book exhibitions in the context of special events and celebrations.



Strengthening capabilities for digitisation projects

To reinforce NLP resources allocated to digital content production and management, specific measures were implemented in what concerns the revision of internal processes, procedures and workflows and the renewal of NLP production and storage infrastructure.

Since 2006, the financial cutbacks in Portuguese public administration budgets and the lack of EU funds led to the optimisation of internal resources. In this context, in 2007 digitisation processes were redefined, with the corresponding review of procedures and workflows. Guidelines, models and technical requirements for image capture, metadata and online display were also formalised in the last years.



The revised guidelines and technical requirements concern a variety of aspects: selection policies for digitisation; digital file formats (TIFF masters and JPEG/PDF access copies); image resolution (usually 300 dpi for masters and 150 dpi for copies); access files compression (80% quality JPEG and LURA-tech compression of PDF); colour depth (24 bpp) and mode (RGB); and file and directory name conventions. Regarding metadata, MARC XML is used for descriptive metadata; TIFF 6.0 specification is applied to master files tags; local schema are used for technical and rights metadata and METS Schema for declaring documents physical and logical structure. Different metadata sets and images are encoded in METS. The structure of each digital object encoded in METS packages which are used by the NLP PURL system to register digital objects and their items.

Future developments include the simplification of the NLP metadata profile; extracting image file metadata to MIX elements and adopting the JPEG2000 format for the master files of certain high volume, digital objects.

In order to increase the production and improve sustainability of the NLP digital collection, five new book scanners were bought, the network capacity was upgraded to 1 Gb on the digitisation workflow premises, and a new digital content archiving system was acquired. This new system includes NAS storage equipment for public access copies of the digital objects (11 TB), a Content Addressed System (CAS) for master digital objects (51 TB), digital content management software, and a backup system.

Enhancing public access to digital content

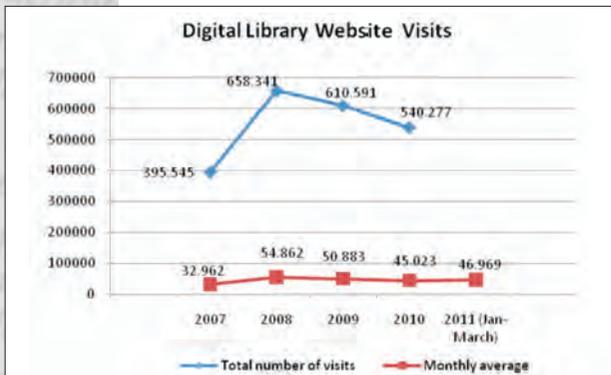
Public services to access digital content have been continuously enhanced in several ways, including the redesign of the National Digital Library website, the development of new in-



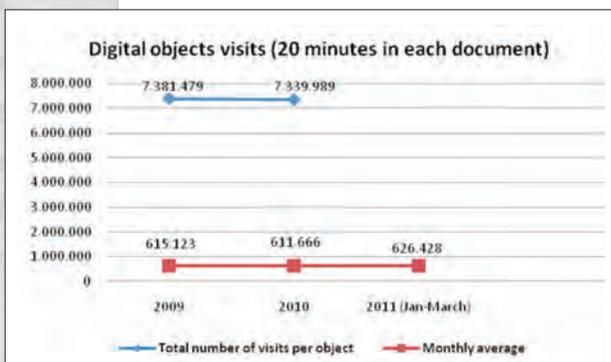
terface features to access and browse online documents, and the provision of new services such as EOD (eBooks on Demand) and POD (Print on Demand).

Before elaborating on the measures taken to accomplish this goal, it is worth looking at the evolution of figures on statistics of access both to the website and the digital documents themselves.

The table below shows data collected from Google Analytics reports on visits to the web pages of the National Digital Library. In 2008, visits increased to approximately half a million per year, remaining stable since then.



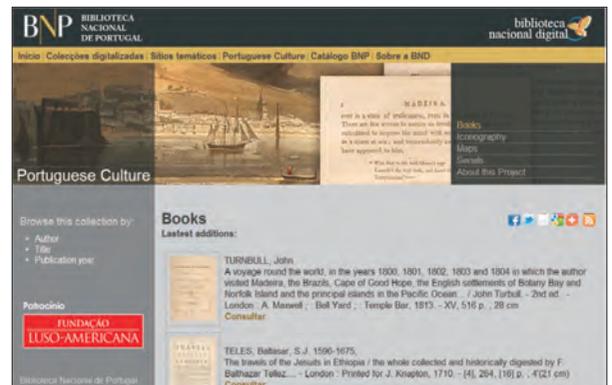
In 2009, the NLP implemented a tool to count the number of accesses to digital documents, because this type of access is not reported by Google Analytics whose tracking code is not inserted in the digital object.



The number of visits to digital objects (circa 7 million per year) is much higher than that of digital library web pages, because most frequently users land directly on the digital object through Google searches, NLP catalogue searches or other services linking directly to digital objects such as the Europeana portal.

National Digital Library website redesign

In 2009, the National Digital Library website was completely redesigned. A new graphic design was developed and the content was reorganized by type of resources, making it easier to browse by author, title, date of publication or latest additions to the digital collection. A deeper integration with the NLP catalogue was also achieved, as well as full compliance with web accessibility requirements.



National Digital Library website 2005-2008

National Digital Library website 2009-2011

SHORT

ARTICLES

FROM

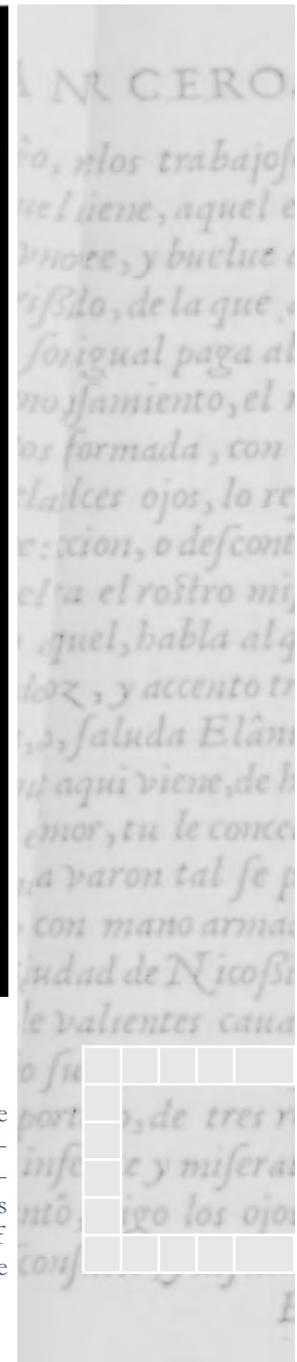
CLOSED

DOORS

TO OPEN

GATES

Portuguese manuscripts from the 12th Century (www.purl.pt/15013); and 15th Century (www.purl.pt/15004) and 1612 (www.purl.pt/14454)



In 2011, social web functionalities were added to both the website pages and the landing page of each digital object, making web sharing of the resources easier and providing the NLP with devices to easily collect more information about the documents most tagged or shared on the Web, as well as the type of users and social media that are being most used.

Special interfaces for complex digital objects

Due to special features of documents like newspapers and reference works, new functionalities for browsing and accessing these resources have been provided since 2009. It is possible to browse and search term entries of dictionaries and encyclopaedias, and to browse newspapers issues by calendar.



140

Old books printed in Portugal in the 16th Century:
 1501 (www.purl.pt/6931); 1530 (www.purl.pt/14549) and 1567 (www.purl.pt/14798)



In the second half of 2011, the experience of accessing digital objects will be enhanced by a flash page flip interface, including functionalities such as page turn, multiple page view, auto-play, full text search, zoom, Google analytics code inside the object, full integration with the NLP catalogue, and sharing and printing facilities. This flash interface will not replace PDF and JPEG object versions, which will continue to be available.

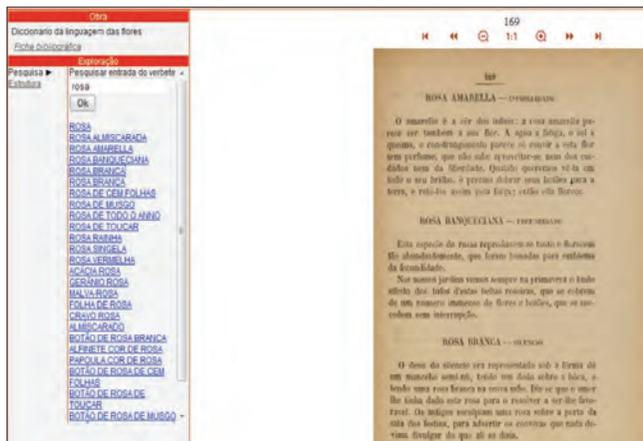
On demand services: EOD and POD

Since 2008, the NLP provides a new service for on demand digitisation, eBooks on Demand (EOD)⁵. It is a full transactional online service that allows users to order eBooks directly from the NLP catalogue, pay online and receive a link to download a full text PDF. This service is provided in the context of the EOD Network, which includes 30 libraries from 12 European countries.

In 2011, the NLP will offer a new EOD Network service, Print on Demand (POD), which will deliver paper copies of digitised books. These reprints are real trade paper books and will be ordered directly from the catalogue of the NLP or through the digital library. Once ordered, each reprint will be available to purchase at Amazon.

Diversifying access channels and creating new services for institutional cooperation

The integration with other, external, access services has always been a goal to make the most of digital assets of the NLP. NLP digital objects can be searched and are available



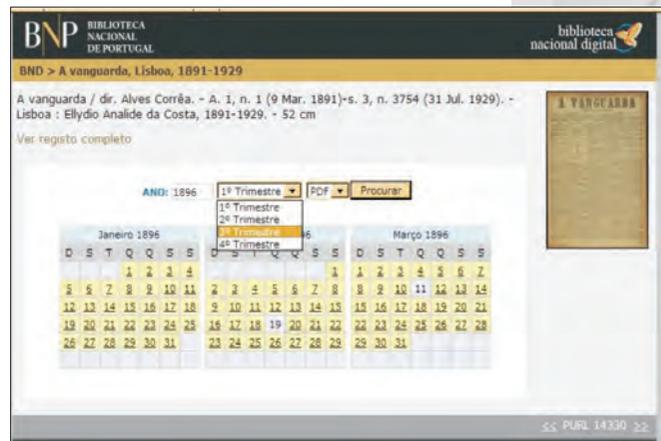
Example of searching and browsing by term entries (www.purl.pt/13929)

through other services like PORBASE⁶, the online union catalogue of 170 Portuguese libraries; The European Library (TEL)⁷; Google and Europeana⁸.

According to Google Analytics reports, since January 2010 access to the National Digital Library was driven by 3 446 traffic sources: the first one is Google, the 13th is Europeana; PORBASE was the 16th and the 32nd one is TEL.



In May 2011, the NLP will launch RNOD - National Registry for Digital Objects⁹, a meta-data repository of bibliographical digital documents, digital collection or information about digitisation projects, provided by Portuguese organisations. The system will serve three major goals: the first is to function as a central point of access to a multitude of scattered sources, providing at the same time the means to share information indispensable for avoiding duplication of digitisation efforts, especially important in what concerns printed material; the second is to act as an aggregator to Europeana, thus providing common means to redistribute data that will increase access to digital content; the third goal, no less im-

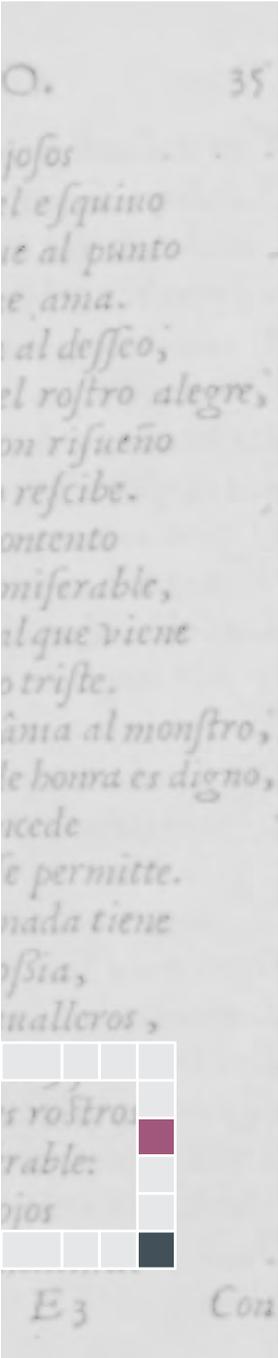


Example of calendar newspaper browsing (www.purl.pt/14330)

portant, is to create opportunities for community building and sharing of expertise and experience regarding technical and managerial aspects of digitisation and digital preservation.

Conclusion

The NLP strategy for digital content and services has been guided by the need to create sustainable and integrated technical and organisational structures capable of continuously

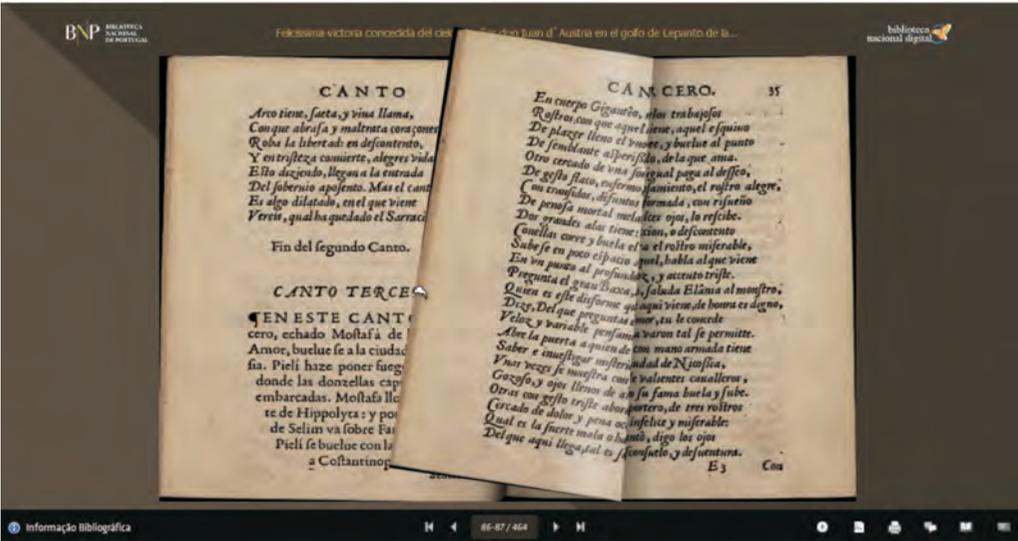


feeding a critical mass of cultural resources online that can serve, at the same time, heritage preservation and research needs, as well as contributing to shaping cooperation by integration services both at the national and at the international levels.

Within the next years, efforts will continue to focus on delivering digital content reuse services for end users and other digitisation projects, establishing cooperation projects with the publishing industry, enhancing digital content to allow new usages and looking for opportunities to create structures for mass digitisation activities.



1. www.bnportugal.pt/
2. <http://purl.pt/index/geral/PT/index.html>
3. [www.purl.pt/15013](http://purl.pt/15013)
4. [www.purl.pt/index/porCulture/EN/index.html](http://purl.pt/index/porCulture/EN/index.html)
5. www.books2ebooks.eu/
6. <http://porbase.bnportugal.pt/>
7. www.theeuropeanlibrary.org
8. www.europeana.eu
9. <http://rnod.bnportugal.pt/rnod/>



New document display (to be implemented in 2011)