

Biblioteca Digital del Patrimonio Iberoamericano: open source technology in the service of a major cooperative project.

José Luis Bueren Gómez-Acebo

Head of the Digital Library Area, National Library of Spain, Madrid, Spain.
biblioteca.digital@bne.es

Elena Sánchez Nogales

Head of the Digital Library Section, National Library of Spain, Madrid, Spain.
biblioteca.digital@bne.es



Copyright © 2013 by **José Luis Bueren Gómez-Acebo, Elena Sánchez Nogales**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

Abstract:

The Digital Library of Ibero-American Heritage was presented at the National Library of Spain in September 2012. This new portal provides access to the digital collections of the Ibero-American National Libraries grouped in ABINIA. It currently provides access to the collections of six National Libraries: Brazil, Chile, Colombia, Panama, Portugal and Spain. In the near future new partners are expected to incorporate to the project.

The development, after some time of study, has been done under open standards, such as XML, Dublin Core or MARC. The technology of the search engine is the open source Lucene/SOLR and the platform is based on Java and HTML.

The portal is already the biggest repository to search Latin-American content and, over time, it could turn into a complementary tool to the most important projects in digital cultural content: Europeana and the Digital Public Library of America.

This paper outlines the history of the Project in the context of a growing Ibero-American focus on digital projects and access to cultural heritage, the current situation, the technological architecture and the impact within end users.

Keywords: Digital libraries, Ibero-America, open source, international cooperation.

1. FOREWORD: WHAT IS BDPI?

The Digital Library of Ibero-American Heritage (BDPI) is a project started by the Association of National Libraries of Ibero-America (ABINIA) with the aim to create a portal that would provide a single source from which to access the digital resources of all the participating libraries.

The portal was made public during the XXIIIth Meeting of ABINIA, hosted by the National Library of Spain (BNE) in September 2012. The goal is to keep growing with the incorporation of as many libraries as possible and thus turning it into a reference site and reinforcing its presence in the web and networks of all the participants.

The portal integrates the bibliographic descriptions of the digital objects hosted by the different contributing libraries. An own tailor-made metadata model (similar to Dublin Core) has been adopted. Participant libraries submit metadata in the most standardized version they are able to provide (mainly Dublin Core or MARC) and they are indexed in a server hosted by the BNE. Searches are performed in this server and the digital objects are viewed in the provider library site. BDPI does not keep digital objects, only metadata. Both the content of the bibliographic descriptions and the management of the digital objects are responsibility of the different participating libraries.

At the moment of writing this paper the available materials in BDPI were:

❖ Drawings, engravings, photographs:	41,853 titles
❖ Books:	62,966
❖ Manuscripts:	8,002
❖ Cartographic material (printed and manuscript):	6,943
❖ Scores (printed and manuscript):	21,641
❖ Newspapers and magazines:	9,232
❖ Electronic resources:	89
❖ Sound recordings (musical and non-musical):	17,033
❖ Web sites:	660

Logo: Biblioteca Digital del Patrimonio Iberoamericano

Welcome Bienvenidos Bem-vindo

Contacto Ayuda

Inicio Acerca de BDPI Colecciones Buscar en BDPI Todos los campos Búsqueda avanzada

Buscar en el texto de los documentos

Manuscritos

Dibujos, grabados y fotografías

Mapas

Biblioteca Nacional de Brasil 19.507 registros

ABINIA BNE BNP

Títulos A-Z | Autores A-Z

2. THE PROJECT: BACKGROUND AND HISTORY

The idea of creating BDPI dates back to 2009 when ABINIA agreed the creation of a portal that provided a single access point to the digital collections of the Latin-American national libraries.

A similar Project had been developed between ABINIA and UNESCO with the creation in 2001 of the digital Library [El Dorado](#). This project was extended until 2003 and its scope was to make available digital works of the Latin-American national libraries, using the z39.50 protocol. Since 2003 the situation around digitisation has significantly varied but it is fair to say that today's achievements are to a great extent the successful outcome of all those previous efforts of cooperation and standardization.

When the creation of BDPI was proposed back in 2009, the BNE, who had just started a systematic digitisation project and launched their new portal "Biblioteca Digital Hispánica" (BDH), offered to coordinate and lead the project.

Initially, the BNE studied the adaptation of the National Libraries Global Prototype developed by the National Library of New Zealand. The intention was to customize that software to the project needs. Such adaptation required the development of certain functionalities as well as some changes in the interface look and feel.

In parallel to those development tasks, the BNE sent at the end of 2010 some forms to the Latin-American national libraries in order to gather information about the status of their digital libraries and their future integration in the portal.

During 2011 the prototype was evaluated and the conclusion highlighted the value of some functionalities such as OAI-PMH protocol; records search and retrieval; popular records, collection management, facets to filter results; records translation, multilingualism...

However, as a consequence of working with a prototype, there were a number of technical inconveniences:

- No possibility to choose in which index to search.
- Inexistence of an advanced search.
- Impossibility to order the results.
- Loss of some navigation elements.
- Only DC visualisation possible.

These technical inconveniences were repairable but the project faced, furthermore, some strategic problems. On one side, the prototype by that time was not in production in any place, on the other, the developing company gave no service in Spain so any development required a different company to take responsibility of the code and to start almost from scratch.

As a result of this evaluation, it was decided to discard the prototype and use instead the technology under which the *Biblioteca Digital Hispánica* (BDH) had been developed. It is important to take into account that when the BNE started to study the possibility of developing BDPI, the BDH had just been launched. By 2011 more than two years have

passed and in the meantime the BNE's portal had established and shown adaptability and strength to support a big project.

The development allows to incorporate all the functionalities foreseen in the prototype as well as to adapt the interface to the designed look and feel. Furthermore it is possible to include in BDPI all the services implemented in BDH with no added costs. The development is Lucene/SOLR based and has been built on the specifications of the BNE so there's great control of the adaptation and evolution possibilities.

During 2012 different libraries were aggregated to the portal and the development and design works ended. Finally, it was made public in September 2012, within the XXIII Meeting of ABINIA.

3. MORE ABOUT BDPI: TECHNOLOGICAL INFRASTRUCTURE

The development of the BDPI portal is based on open-source standards, extensively widespread in the library domain, such as OAI-PMH and the SOLR search engine. Summing up, its main features are the following:

Search engine: it is designed on the basis of the open-source technology Lucene / Solr. This search engine indexes the content from all the sources providing data to BDPI. Queries are made against this engine.

Programming language: the BDPI search application is essentially a web app complying with J2EE standards. It is Java-based and displayed on an Apache Tomcat application server. All the pages in the search application are implemented in HTML format.

Indexing process in the search engine: the source data (descriptive metadata) from the BDPI libraries are obtained either via OAI-PMH (whenever the option is available, being the preferred way of data exchange) or a static XML file containing the descriptive metadata.

Once these metadata from the different providers are captured, XSLT templates are applied to enable conversion and integration of the original metadata into the fields set-up within the Solr structure. This conversion process generates IDX files, which are new XML files in the format accepted by Solr to perform the indexing process, which will eventually feed the search engine.

Servers hosting BDPI:

BDPI search components are hosted in two servers:

⇒ One of them hosts the search engine (Solr). Features of this server:

- 540 GB Capacity.
- 4 processors Intel (R) Xeon (R) CPU E5450 3.00GHz
- 12 GB RAM.

⇒ A second server hosts the search web application, together with all the indexing processes (XSLT conversion files and indexing scripts) and the database (containing MYSQL tables to keep track of statistics about the application usage). Main server

features:

- 247 GB Capacity.
- 2 processors Intel (R) Xeon (R) CPU E5540 2.53GHz
- 4 GB RAM

OAI-PMH protocol:

The preferred way of data exchange is metadata harvesting via OAI-PMH. However, some of the participant libraries were not able to make records available through OAI, and data submission was completed with alternative means such as sending a physical storage unit or by email.

The data sources for each participant library are so far:

- ⇒ **National Library of Spain:** via OAI-PMH and based on Marc21 format:
<http://bibliotecadigitalhispanica.bne.es/OAI-PUB?verb=ListRecords&set=bne01&metadataPrefix=marc21>
- ⇒ **Newspaper collection of the National Library of Spain:** via OAI-PMH, in Dublin Core-Extended format:
http://hemerotecadigital.bne.es/oai.vm?verb=ListRecords&metadataPrefix=dc_ext
- ⇒ **National Library of Portugal:** via OAI-PMH, in Marcxchange (UNIMARC) format:
<http://oai.bn.pt/servlet/OAIHandler?verb=ListRecords&set=bndlivre&metadataPrefix=marcexchange>
- ⇒ **National Library of Brazil:** records were submitted in a static XML containing metadata (Dublin Core-based format) corresponding to the library's digital resources.
- ⇒ **National Library of Colombia:** via OAI-PMH in Dublin Core format:
http://www.bibliotecanacional.gov.co/tools/open_archive.php?verb=ListRecords&metadataPrefix=oai_dc
- ⇒ **National Library of Chile:** records were submitted in a static XML containing metadata (Dublin Core-based) corresponding to the library's digital resources.
- ⇒ **National Library of Panama:** via static XML, Dublin Core-based.

API for external search¹: One of the milestones in the development of BDPI was the launching of a web service which allows launching queries directly from external search interfaces. Similarly to the Europeana API, the aim is to attract visitors from other services, which in return also increase and enhance the functionalities offered to their users.

OCR: It should be mentioned that the portal allows full-text searches for those records with such data included in their descriptive content.

¹ Technical information about the BDPI API is available in:

<http://www.bne.es/webdocs/RedesSociales/Blog/API.pdf>

4. BUILDING BDPI: METADATA AND INGESTION PROCESS

Once data have been submitted by the provider or harvested via OAI, the process of conversion and integration in BDPI portal takes place.

The first phase of this process implies a thorough and in-depth analysis of the metadata provided: amount and quality of data, fields or format descriptive elements used, identifiers, categorization of structural content such as type of material, and actual access to the digital objects. The aim is to extract the highest amount of information, while homogenizing and clustering data together as much as possible.

In connection with this, an important aspect in the development of the entire project, with heavy implications beyond metadata readiness, comes from the wide range of technological development levels among the participating libraries. While some libraries are in a position to offer and share tens of thousands of digital objects, structured and standards-compliant metadata, and long implemented communication protocols, others are still at an early stage of the process, and have a small part of their collection digitally available, with descriptions not fully subject to cataloguing standards and protocols such as OAI-PMH not in place yet.

The BNE has undertaken an exhaustive analysis of data received in order to ensure that indexing becomes as accurate as possible. Data was received in the following models:

- ⇒ BNE: MARC21
- ⇒ BNP: UNIMARC
- ⇒ BNB: Dublin Core
- ⇒ Chile: Dublin Core
- ⇒ Colombia: Dublin Core

Records are indexed according to a data schema which is quite straightforward but essentially allows searching and retrieving information from the basic data fields, to help describe, identify and discover the resource.

1. Author
2. Title
3. Date
4. Subject(s)
5. Edition
6. Identifier
7. Resource_type
8. PID
9. Description
10. Related information
11. Thumbnail
12. Institution/Provider

Obviously, a higher complexity and density in the level of bibliographic description means a more complex mapping, along with an eventually better identified and retrievable resource. Such was, for instance, the case of Portugal.

Special efforts had to be deployed specifically for the fields of Date, Type of resource and Subject, which were heterogeneously formatted and particularly sensitive as long as they are offered and presented as facets that had to keep a unified format, to enable users to filter within a cohesive categorization. The case of Brazil furthermore posed duplication problems, as subjects were given both in Portuguese and English, with no possibility of automatic sorting from the xml. This forced to manually carry out the equivalence and matching between Portuguese/English subjects, in order to present as facets only one of the two versions (Portuguese).

Other important tasks in the metadata analysis and mapping were:

- ⇒ definition of the element which would in each case work as ID for the resource;
- ⇒ selection and presentation of the thumbnail (specifically included in the metadata in the cases of Spain and Portugal; obtained through an automated process which extracts the image from the .pdf – and whenever a .pdf is not available, the library logo),
- ⇒ identification and validation of the URL giving access to the actual resource, be it directly to the digital object (preferred) or to the description of the digital resource from where a link is offered to the object. The latter was the case of Chile - who requested links to be directed to their Memoria Chilena portal - and Portugal, where rights issue might arise from the fact that in some cases two versions (public and internal) of the digital object are offered.

5. SHOWCASING BDPI: INTERFACE, COLLECTIONS AND FUNCTIONALITIES

Functionalities offered in BDPI portal intend, on one side, to facilitate information search and retrieval and, on the other, to provide users with tools that ease navigation in case they don't have specific information needs.

In order to reach this goal, BDPI's portal counts on the following functionalities:

- Simple search allows you to search all record fields quickly and intuitively. By default, the search is restricted to specific bibliographic record fields from the catalogue (namely title, author, subject, description or all of these). Furthermore it is possible to select in which index the search should be done. Finally, it is possible to indicate whether the query should be executed in the full text (OCR) field or not.
- An advanced search enables you to search in more than one field at a time, indicating which indexes should be used to restrict the search results. In addition to the options already included in the simple search, the query can be done on these other fields: "Place of publication", "Publication data", "Date" of publication, "ISBN/ISSN", "Call number", "Universal Decimal Classification (CDU)" and "Geographic location". It is also possible to set limits or filters to the search under the following criteria: Institution, Type of document and Language.

Viewing the results

- If the search retrieves more than one result, a list with abbreviated information (thumbnail, title, and author) will be displayed identifying the records obtained. By default, the records are sorted according to relevance criteria calculated internally by the BDPI search engine. The order can be changed using a dropdown menu; the possible order options are title, author and date.
- Detailed view shows as much information as possible and available about the works described.

Facets

In order to facilitate the results exploration some faceted fields allow refining and limiting navigation. It is possible to filter by Institution, Subject and Type of document.

Collections

The collections offer the user additional means of accessing documents, grouped into cross-linked collections based on common characteristics relating to the type of document, theme, or their particular relevance, interest, appeal or importance within the documents contained in BDPI. They are meant to be an inviting and attractive way of discovering some of the usually 'hidden' treasures of our heritage.

From the navigation bar on the homepage it is already possible to access the collections of *Manuscripts*; *Pictures, engravings and photographs*; and *Maps*. Furthermore there is a "Collections" page with more featured collections and suggestions for accessing common or relevant content from all the collections in the BDPI. At the time the portal was launched, these were: *Geography and travel*, *Music*, *Music scores* and *Literature and literary studies*. More collections recently incorporated were *Magazines and press* and *Tales and legends*.

Highlighted collections



Geography and Travel

Selection of documents, books, maps and pictures, testimonies of the great travels and explorations which built bridges across countries and became a means for Science and knowledge progress since the 16th century.

[View collection](#)



Music

Around 16.600 sound recordings, featuring the wide variety of Latin American music genres, gathered for the first time in this collection.

[View collection](#)



Periodicals

Selection of newspapers and magazines, as direct witnesses for their times which help us to travel through the History of Latin America from different perspectives.

[View collection](#)



Tales and legends

Selection of texts inspired by legends, myths and folk tales, in their original versions or transformed to create new stories and visions of Latin American culture.

[View collection](#)

Libraries



National Library of Brazil
19,507 records



National Library of Chile
21,558 records



National Library of Colombia
7,117 records



National Library of Spain
108,675 records



National Library of Panama
925 records



National Library of Portugal
10,970 records

6. EVALUATING IMPACT AND USE OF BDPI

Usage of the portal since it was launched has been constant and stable with around 400 users per day. Although it is true that Spain encompasses 40% of access, it is interesting to observe high interest in the portal coming from Brazil (13%), México (7%), Colombia, Panamá, Argentina (5% each); Chile, EE.UU., Portugal y France (2-3%).

The study of the statistics data (gathered through Google Analytics) confirms ~~(although this analysis has not been done in depth)~~ that there are two main user profiles for BDPI, different and complementary: on one hand, those who approach BDPI as a research source where they can widen their information search; on the other, those who reach BDPI thanks to some news, recommendation or one-time event (a new collection, for example). This conclusion is reached from the fact that traffic sources are, firstly Facebook and on the third place BNE's *Biblioteca Digital Hispánica*

Considering Facebook own features and nature, it is not surprising to see that access from this channel gets concentrated on specific periods, particularly on the first days after the portal launching and dissemination within social networks, and also following the publication of news or highlighted collections, directing users to the portal. The average duration of these visits is 03:32, slightly below the overall average duration of visits from all sources.

The other important traffic source for BDPI is Biblioteca Digital Hispánica (the digital library of the BNE), since November 2012. The reason behind is the integration in this portal of an API which displays, once a query is made in BDH, the results obtained from the same query in BDPI, which can be then accessed to directly only by clicking on the API link. It is very interesting to note that visits coming from this source often double (or even triple) the average numbers in terms of time spent and pages seen in BDPI. And these numbers let us think that the API does indeed widen the functionalities and services offered by Biblioteca Digital Hispánica, and as such the API tool is proved to be appreciated and used among our users.

The same goes for the Europeana API, also integrated in BDH and offering the same service as a gateway to this pan-European important resource.

We can therefore see these two different ways of accessing the content as being both relevant and complementary: Facebook, on one side, providing a high number of one-off visits, as a result of some outstanding content or particular news; BDH's API, on the other, as a much more sophisticated tool used by more specialized users.

These facts lead to the conclusion that BDPI has a great potential to provide services to, summarising, the two main kind of users that a library (national, particularly) can have: general citizens attracted by curiosity or recommendation; and specialists who find here a single and relevant access point that enhances the results of their research.

Dissemination through social networks and the API implementation seem key actions to increase the visibility of the portal. One more way to achieve this objective could be trying to place it in web pages, blogs, educational forums, etc. in which people look for this kind of resources. The second traffic source to BDPI is a Brazilian² web page dedicated to inform about cultural activities and, among them, sites to access digital content.

² catracalivre.folha.uol.com.br

Overall figures along the first seven months:

- ⇒ Unique visitors: 63.222.
- ⇒ Visits: 80.714.
- ⇒ Page views: 374.101.
- ⇒ Average visit time: 03:42.
- ⇒ New visits rate: 78,21%
- ⇒ Pages per visit: 4,63.
- ⇒ Rebound rate: 59,22%

5. ENVISAGING BDPI: THE FUTURE

Quoting a very famous Spanish poem, “the path is made by walking”. And the way we have already went over the recent years brings closer and closer the old dream (and apologize here for the commonplace) of a true global digital library.

Projects like Europeana (2008) and the DPLA (2013) make it real to be able to access, from one single search, the digital content of thousand of European and North American cultural institutions.

It may seem presumptuous to talk about BDPI in connection with Europeana and the DPLA. The level of awareness, the amount of content and the budget it manages is not comparable to the other two projects. However, BDPI is indeed similar in its philosophy and objectives.

Offering a single access point to digital collections in Latin-American libraries is, saving quantitative distances, the same that is offered in Europeana and DPLA. No matter these distances, there is no technical or strategic obstacle that prevents BDPI from being the seed of a global project to access Latin-American content.

Latin-American countries share, as well as the European and North American, history, culture and languages. The presence of their collections under the same context enriches the particular collections of the participating institutions.

The discovery of the New World (at least from the perspective of the Spaniards who reached the Caribbean coasts), the impact of the European presence in the American continent, wars, evangelization, slavery, independency process, literature, indigenous languages... there is a multiplicity of themes in which the Latin-American perspectives enrich the information available in the different institutions.

BDPI is now a solid project with the participation of important libraries, and provides access to around 170.000 works. The reception among users is being very positive and so far it has had an excellent behaviour from a technical point of view.

It is possible to set two concrete objectives that should be possible to reach in the short and medium term.

- ⇒ **Incorporating as many libraries as possible to the portal:** the first step to consolidate the project should be the effective incorporation of all the libraries included in ABINIA. Afterwards, the possibility of offering other big libraries and other cultural institutions the opportunity to aggregate their collections to the portal should be considered.
- ⇒ **Increase the portal's visibility:** Usage statistics since BDPI's launch show that it is interesting for those users who get to know it. However, until now, no big efforts have been done to improve the awareness of the portal among citizens. To reach this goal it is key to count on the cooperation of participating libraries, who may - through their social networks, or with the API implementation - contribute to disseminate the project in their respective countries.
- ⇒ Enriching the portal with blogs, creating an specific Facebook profile, cooperate with Latin-American studies Institutes (or cultural in general), create a distribution list, enhance the portal services with specific tools for the different kind of users, the creation of curated content... are actions that will also contribute to improve the portals visibility.

Moreover these operational objectives it has to be mentioned that BDPI doesn't have until now a real organizational model. It is necessary to agree and establish the way in which it is going to be managed throughout time and to clarify key aspects such as:

- ⇒ Operational responsibilities.
- ⇒ Staff dedicated to the project.
- ⇒ Funding.
- ⇒ Participation requirements.

The creation of BDPI is, so far, the history of a success. No Latin-American cooperative project had ever joined the amount of data and had had a comparable utilization level. In this sense, it is remarkable how technology, particularly the use of standards and open source tools, facilitate the fulfilment of these projects.

The mere existence of this portal should contribute to foster digitization, standardization and open technologies use. Moreover its political significance, BDPI's success will confirm cultural policy makers the importance that digitization and digital information has for citizens. Free access to quality cultural content promotes education, research and leisure.

Latin-America as a cultural reality deserves a self space in the universe of cultural institutions and BDPI is an excellent example of how cooperation in this field is viable and has a very positive impact.