



**REPORT FOR:  
BIBLIOTECA NACIONAL DE ESPAÑA AND FEDERACIÓN DE GREMIOS DE EDITORES  
DE ESPAÑA**

**COMMENTS AND SUGGESTIONS CONCERNING THE ENCLAVE Editores-BNE PROJECT  
Numilog, Denis Zwirn, June 2009**

- 1. UNDERSTANDING OF THE PROJECT AND SCOPE OF THE CONSULTING MISSION ..... 2**
- 2. THE NEW DIGITAL BOOK CHAIN ..... 2**
  - A. DIGITIZATION DOES NOT LEAD TO DISINTERMEDIATION ..... 3
  - B. DIGITIZATION LEADS TO NEW FUNCTIONS, WHICH ARE THE TRANSPOSITION, IN THE DIGITAL BOOK CHAIN, OF TRADITIONAL FUNCTIONS OF THE PRINT BOOK CHAIN..... 3
  - C. DIFFERENT POSITIONS IN THE DIGITAL BOOK CHAIN DEFINE ONLY FUNCTIONS, NOT PLAYERS ..... 4
  - D. FINALLY, THE FRONTIERS BETWEEN THE PLAYERS AND THEIR TRADITIONAL ROLES MAY EVOLVE..... 5
- 3. TECHNICAL MATTERS: EBOOK FORMATS, PROTECTION, AND METADATA..... 6**
  - A. DOWNLOAD FORMATS ..... 6
  - B. ONLINE READING..... 7
  - C. PROTECTION AND DRM..... 8
  - D. METADATA STANDARDS ..... 9
- 4. THE ARCHITECTURE OF THE ENCLAVE EDITORES-BNE PROJECT ..... 10**
  - A. THE DIFFERENT PARTNERS OF THE PROJECT..... 10
  - B. THE ECONOMIC MODEL AND THE SERVICES OFFERED TO USERS ..... 11
  - C. THE REQUESTED FILES ..... 12
  - D. THE REQUESTED METADATA..... 12
  - E. THE EXCHANGE FLOW OF FILES AND METADATA..... 13
  - F. THE GLOBAL ARCHITECTURE OF THE PROJECT: A SYNTHETIC SCHEME OF THE ROLES OF THE DIFFERENT PLAYERS..... 14
- 5. COMMENTS AND SUGGESTIONS ..... 14**
  - A. GENERAL COMMENTS ..... 15
  - B. TECHNICAL SUGGESTIONS ..... 15
- 6. FURTHER COMMENTS AND SUGGESTIONS ON ECONOMIC MODELS ..... 19**
  - A. THE ECONOMIC MODELS OF EBOOK STORES ..... 19
  - B. THE ECONOMIC MODEL OF ENCLAVE EDITORES-BNE..... 20

## 1. Understanding of the project and scope of the consulting mission

The Biblioteca Nacional de España (BNE) and the Federación de Gremios de Editores de España (FGEE) initiated, in November 2008, a common project, “ODIBNE-Pro”, with partial financial support by the Ministry of Industry. This project aims at creating a stable knowledge center dedicated to the analysis, definition, implementation and diffusion of a model for including copyrighted digital works in digital libraries. Relevant technical, economic, and legal conditions for this inclusion must therefore be identified, as well as criteria for the selection of the works.

On May 2009, and according to the work plans of ODIBNE, a new name for the project was agreed-upon in order to carry out all future dissemination and public communication activities. This name is ENCLAVE Editores-BNE (hereafter “ENCLAVE”).

The first phase of the project aims at adding copyrighted works to the Biblioteca Digital Hispánica (BDH) website, created by the BNE in 2008. The field of digital works targeted in this first phase is defined as follows:

- Digital texts (or “eBooks”)
- Original (rather than critical) cultural or scientific works
- Books existing in both digital and print versions
- Books for which one publisher explicitly holds the digital rights
- eBooks for which a full-text access is possible, according to conditions defined by the publisher

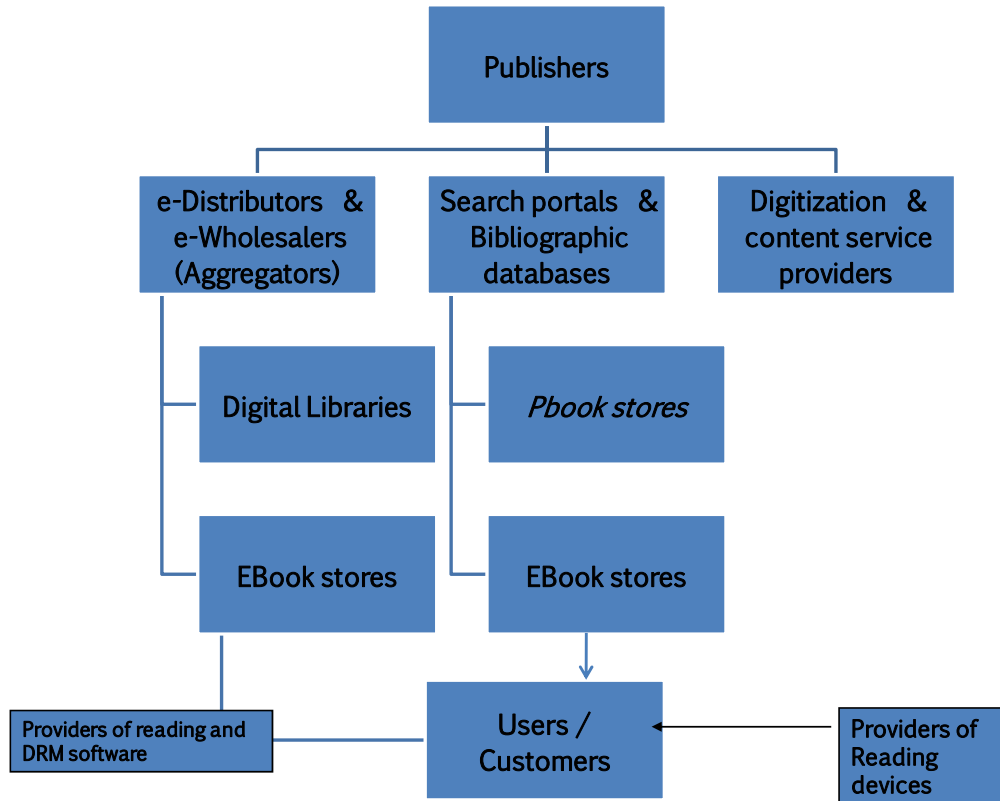
The technical conception and architecture of this phase of the project is handled by Neturity, a Spanish company specialized in information systems and new technologies. The FGEE, in agreement with the BNE and Neturity, wishes to add an “outsider’s view” on the first phase of the project and more specifically on its technical architecture. They wish for this study to be provided by Numilog, a French company specialized in the distribution of eBooks since 1999, and one of the e-distributors partnering since 2008 with Gallica, the website of the Bibliothèque Nationale de France for digital works.

The purpose of Numilog’s consulting mission is therefore first to check the coherence of the technical architecture of the project and then, if necessary, to make suggestions in order to improve or review certain points. Suggestions about the economic model and on possible further phases of the project will also be proposed, replacing them within the general context of a new “digital book chain”.

## 2. The new digital book chain

The typical organization of the eBooks industry tends to be structured around several functions and players, and can be summarized by the following general scheme:

### The eBook chain : main functional positions



This scheme is associated with the following theses:

#### a. Digitization does not lead to disintermediation

Digitization is often associated with dreams or nightmares (depending on where you stand!) of disintermediation. Indeed, the distribution of digitized content on the Internet leads to new facilities for authors to bypass publishers, for publishers to bypass book stores or even libraries. And these facilities are and will be used by some of them, for certain books. But, at least in the first stage of this new era, this phenomenon should remain marginal. The role of **publishers** in regard to creating, selecting and promoting “good books” will remain essential in the eyes of authors looking for professionals able to help them to realize their books and to push them towards the readers. And so will the role of **eBook stores** and of **digital libraries**, since, when the available contents become more and more massive, it is even more necessary to be able to select and to present them on unique access points where they are organized and marketed, and to be ready to advise the readers. Finally, **bibliographic databases** will always be required, in order to offer exhaustive and neutral catalogs of eBooks with relevant metadata to the whole chain as well as to end-users.

#### b. Digitization leads to new functions, which are the transposition, in the digital book chain, of traditional functions of the print book chain

Even if there is no disintermediation, some new jobs will be required by the digitization of books. These new jobs are in fact transpositions, in the digital sector, of jobs of the traditional book chain:

- **Digitization and Conversion service providers** are required, in order to produce the eBook files, either from paper (scanning and OCR), or from DTP files (pure conversion of format and enrichment of files). This function mirrors the function that printers or DTP studios play in the print chain.
- **E-wholesalers** are required, like wholesalers in the print book chain, in order to negotiate, on behalf of publishers, with eBook stores wishing to sell and promote eBooks selected within large catalogs.
- **E-distributors** are required in order to offer digital warehouse services for the eBook files and metadata, to deliver the metadata to retail channels and to digital libraries, and to deliver the eBook files to end-users, with technical measures, or digital rights management (DRM) systems which warrant the copyright protection. This function is probably the newest and the more technological one of the chain. It mirrors the function of distributors in the print chain, but this time the books are delivered directly from digital warehouses to readers, who order them on eBook stores or borrow them from digital libraries. The development of new kinds of information systems and web technologies is required in order to be able to provide relevant access to the files along various channels, e.g. online, by download, by the phone networks, and according to various economic models, e.g. free browsing, selling of complete books, selling of chapters, renting, lending, and subscription.
- Of course, **technological players**, such as providers of reading software and of DRM technologies, or providers of reading devices, will also play a major role which should ideally remain neutral and transversal to the roles of “content” players.

### c. Different positions in the digital book chain define only functions, not players

One single player may hold several functions in the digital book chain. New functions may be held by traditional players or by new players. Hence, the functions described in the scheme may lead to several different kinds of organization of the industry. For instance:

- E-distributors and e-wholesalers are often the same player, called “aggregator” on the US market. E-distributors often offer digital conversion services. Some aggregators may also sell the eBooks directly on their own eBook store, as well as on the online eBook stores of traditional book stores.
- Publishers or publishing groups may be their own e-distributor and e-wholesalers and may also have a website for direct sales of digital publications. This is often the case for major scientific publishing groups, selling online accesses to academic libraries.
- Traditional printers are offering conversion services for structuring the content (in XML) and for creating eBooks in various formats to publishers or to e-distributors, and some will offer “print on demand” services to the entire chain, using these eBook files.
- Traditional book stores will sell eBooks online along with print books on their websites, and some of the largest ones may be their own aggregators - or wish to be.
- New players are entering the chain, thanks to their major roles in connection to end-users or customers in other fields:
  - Search engines or search portals, such as Google first and foremost, which offer new tools for selecting books with “search inside” technologies, and recently announced opening an eBook store.
  - Pure players of the Internet, such as Amazon first and foremost, but also other kinds of online shops, which develop their eBook stores.
  - Mobile carriers or Internet access providers.

All these new players will progressively try to climb back up the digital book chain, and even to offer transversal technological services or products, like eBook devices and reading software. This is of course an important future challenge for the traditional players.

In the different countries where the eBook market has begun to structure itself, the distribution of the digital functions between the different players varies:

- In the USA, in the U.K. and in Australia, publishers, aggregators and eBook stores are mainly different and independent companies, and choose to contract on the basis of competition

between the services offered, even if the digital warehouse function tends to be more and more concentrated.

- In Germany, publishers and book stores have launched a common initiative for distributing and selling eBooks (Libreka), through their common association (Börsenverein). But this has not yet resulted in a successful sales experiment. Coming to the market with a common commercial strategy might not be in the relevant scope of a professional association, even a very integrated one (this is indeed different from just developing common tools, like a search portal). Besides this initiative some individual propositions of aggregators and eBook stores are developing.
- In France, the organization of the market is still in the making, but it tends to be structured around e-distributors or aggregators owned by or working exclusively for certain publishing groups, just like in the physical distribution industry. Nevertheless the concentration should be greater in the digital market.

Even if certain players hold several functions, the separate identification of these functions is useful in order to understand the workflow of the access to eBooks, and to thereby define a relevant organization of the industry, or the architecture of a project concerning eBooks.

#### **d. Finally, the frontiers between the players and their traditional roles may evolve**

Even though there is no massive disintermediation, and though the players of the print book chain are still players of the digital book chain, the digitization process offers risks or opportunities of modification of their roles and respective frontiers. To quote only the simplest examples among many:

- In the digital book chain, distributors are delivering the books directly to the readers, not to the book stores or the libraries. Hence the temptation of e-distributors to become eBook stores or (commercial) digital libraries themselves;
- Printers, some of which are now implementing industrial processes for creating eBooks from print books or from files, may also offer digital warehouses and become distributors themselves;
- Bibliographic database organizations or online search portals may add new digital services such as full-text search engines and browsing platforms, and could also wish to offer distribution and sales of the eBooks files required for browsing.
- Public or academic libraries which decide to become digital and to lend eBooks to their patrons also have the opportunity to add new functionalities for searching and browsing books, hence to play the role of online search portals or of public bibliographic databases; reciprocally, they could sometimes be exclusively focused on this new role for some books, when no acceptable economic model or financial resources have been found for lending.

The ENCLAVE Editores-BNE project and the relations between its groups of partners have to be recast within this new digital book chain, and its general dynamics. The project may play an important role in stimulating publishers towards an eBook initiative, and in helping to structure the first period of the eBook industry in Spain, much in the way Gallica played this role in France. In its first phase, the project may be analyzed as a “search portal” project (for all books, be they under copyright or not) within a digital library (for public domain books only). This project will of course preserve both the traditional and new players of the book chain and of the rights chain: authors, publishers, e-distributors / wholesalers, (e-)book stores. It will more especially associate publishers with a national bibliographic database institution and with e-distributors, the latter being defined only by functional requirements. Technical flows and a contractual chain have to be defined accordingly.

### 3. Technical matters: eBook formats, protection, and metadata

#### a. Download formats

Many file formats and reading solution have been suggested for downloading eBooks, and there is no unique standard as of yet. However at the time of writing, three main eBook formats are suggested by publishers and used by eBook customers for reading downloaded eBooks. They are described in the following table, with their technical characteristics:

**The main eBooks formats**

| FORMATS           | PDF<br>Adobe  | ePub<br>Adobe  | PRC<br>Amazon                                    |
|-------------------|---|--|--|
| Reading software  | Adobe Digital Edition                                 | Adobe Digital Edition  | Mobipocket Reader                                |
| Specific features | Print possible,<br><b>Conserves lay-out and fonts</b> | Print possible,<br><b>Reflowing</b>                            | <b>Reflowing</b>                                 |
| Reading Devices   | PC, Mac   | PC, Mac<br><b>Sony Reader</b><br>(+ other mobile devices soon) | PC<br>Illiad, Bookeen,<br>Pocket PC,<br>Nokia... |
| DRM technology    | Adobe Content Server 4                                | Adobe Content Server 4   | EBook base<br>(Mobipocket server)                |

Beside these three dominant formats and reading softwares, others exist for the same kind of use, e.g. Microsoft Reader (LIT format), eReader, Zinio Reader, Delivery Reader...or for a more specific use, e.g. Booksurfing or XMDF for reading e-mangas in Japan. Some of these formats and e-reading solutions are also used for offering online reading solutions, e.g. Zinio and Delivery.

Two main reasons explain this diversity:

- The first reason of course is competition between companies which have various strategies for getting revenues from the use of their file formats. Some companies, like Adobe or Microsoft, are traditional software companies which play a transversal role in the eBooks chain and gain revenues from production tools or from the DRM technology associated with their solution. Other companies try to close the market by offering a vertical solution where a proprietary format may be read only on proprietary devices and where all the eBooks in this format are distributed only by one single e-distributor and one single eBook store (the “iTunes temptation”).
- The publishing industry tries to resist what may be considered a risk for them (and for other players of the book chain) by promoting common and open formats and reading solutions. Two formats

satisfying this goal are especially recommended by the International Digital Publishing Forum (IDPF): Adobe PDF and ePub.

- The second reason is more user-oriented: even if standards are highly desirable for the market and for users, certain specific kinds of books (e.g. comics), of devices (e.g. mobile phones), of use (e.g. in school) may require different reading solutions. This justifies for instance the coexistence of PDF and ePub:

- The PDF format is recommended for a perfect rendition of books with complex layouts, e.g. technical, scientific or practical books ; it allows to preserve the reference to the page number of the print book, which is an important feature in an academic context;
- The ePub format allows reflowing, the layout of the book being automatically adapted to the size of the screen and of the font chosen by the reader; it is especially suited when one wants to read a novel on an e-ink reading device, a net PC a Smartphone without having to use scroll buttons.

Interoperability, which means the possibility of reading one same file on various kinds of devices, is also a very important matter for eBook users. At the time being, one can lament that the ePub format's interoperability is still weak, since for instance it is not possible to read a protected ePub file on a Smartphone or even on any other e-ink reader than the Sony ones. But this situation should change in the next few months, since the strategy of the Adobe company is to deliver SDK (software development kits) which can be used for creating protected ePub reading solutions on any device.

The production costs of the eBooks files in these two formats depends on the sources (paper, DTP file, print PDF, XML), on the complexity of the text (layout, notes), on the level of enrichment (hypertext links, bookmarks, multimedia add-on). PDF is generally cheaper than ePub, but it is cheaper to produce PDF files from ePub files than the opposite. Also the same PDF file may be used in different contexts: download, online reading, browsing, and indexation (eventually with minor adaptations).

### **b. Online Reading**

The question of the format of the eBook files is less sensitive and strategic in the context of online reading. The reason is simple: the users do not directly access a file, to be opened with reading software specific to the format of this file, on a device compatible with this software, but instead access the text through a reading solution which is generally compatible with any computer and any Internet browser. The format of the file is transparent for the users, who stream the text page by page and never receive an actual file.

For the users, the interoperability matter is then “nearly” reduced to a question of Internet access. The restriction is due to the fact that some add-in may be requested, for instance the latest version of Flash or Silverlight software. These add-ins are generally very light and simple to install (and often pre-installed), but they may be incompatible with certain reading devices, such as certain Smartphones like the iPhone.

For the publishers, the question of format is only a question of production, not a question of market. Many existing online reading solutions require sources to be PDF or simply image files and include an automatic conversion process from these files to final format displayed..

It is to be noted that, when the book is digitized from paper, it is now standard and highly recommended to use an OCR (optical character recognition) process in order to scan the text in each page of the PDF file, both for download and for online reading. This process is never perfect but allows to perform searches inside the book with more than 98% success.

### c. Protection and DRM

Protection of copyright is a key matter for the future of the digital book chain. And the ENCLAVE Editores-BNE project, which associates a national library and a national federation of publishers, naturally considers that ensuring the conditions of this protection is essential. This general issue raises different precise questions for the different players:

- For publishers: How can one include electronic rights within the author contracts and then make sure that all the right holders agree to distribute digital versions of the original book?
- For digital libraries : how can one find the relevant right holders and economic models for digitizing books in their shelves when the works are out of print or are orphan works, a question which is addressed by an ongoing European project, ARROW.
- For publishers and e-distributors: what kind of technical measures should and could be taken in order to protect the eBooks files?

Considering only the latest point, one can make the following observations:

- For the moment, most publishers in the world wish to use technical measures in order to prevent uncontrolled diffusion and uncontrolled use of the eBook files.
- When the eBooks are downloaded, these measures may rely on “digital rights management” (DRM) technologies or on lighter ones, like watermarking.
  - o DRM technologies consist of encrypting the files in such a way that reading the eBook is possible only on one single device or on a fixed number of devices associated with a customer account. The main DRM technology for PDF and ePub files, Adobe Content Server 4, for instance, allows users to read each eBook on six devices activated with the same account. These technologies allow also monitoring some uses of the eBook. For instance, Adobe Content Server 4 allows publishers to set up parameters such as rights of print, of copy and paste, and the reading time.
  - o Watermarking, or “social DRM”, is a lighter protection level which does not technically block the possibility of copying the file and opening it on another computer, but instead adds within the pages of the eBook some information such as the name / e-mail of the customer in order to indicate the legal owner of the file.
- When the eBooks are read online, there is a first natural level of protection since the file is only accessed by streaming, page by page. However, it is possible to add other technical protection measures in order to prevent any possibility of downloading the pages, of copying and pasting and printing. Different technical environments offer different levels of security to prevent the flux sent to the browser from being captured, Silverlight being at the moment the most secure one.

Whatever the protection policy chosen by each publisher and by their e-distributor, it is essential for the coherence of this policy that the protection chain be never broken: the level of protection is in fact always the lowest level of the chain. That is why, in a project of search and browse portal, the browsing platform technology must be controlled by publishers.

#### **d. Metadata standards**

Metadata are data about other data, i.e. all the information that may be used in order to describe books. This information may concern their content, the rights over them, their commercial features and ways to promote them. Metadata is generally public information, whether the content referred to is under copyright or not. They may be of various forms: short texts, long texts, numbers, images, URL...

Considering the various players of the digital book chain and the various contexts in which the description of a same eBook is required, standardization of the list and format of metadata is again a very desirable goal. Several standardization initiatives do exist, but for the moment metadata often remain different within the libraries environment and the publishers / book stores environment. Hence, it is necessary to define correspondences when a digital project associates the two environments.

The main metadata standards used by libraries are Marc and Dublin Core. The recommendation for publishers and bookstores is to use ONIX, which provides a standard XML template for organizing metadata storage and offers a very powerful metadata framework. It is of course possible to define partial correspondence rules between the fields of metadata of these standards, but ONIX includes many more fields than Dublin Core and is better suited to describe eBooks since it is possible for instance to define the formats of the eBooks.

Finally, a very interesting conceptual framework to consider for building a metadata system is “FRBR”, i.e. the Functional Requirements for Bibliographic Records. FRBR is an initiative of librarians and documentation specialists in order to build a theoretical model for representing metadata. However complex, it relies on a very powerful conception of the different levels of description of any “cultural production”, including books. The same term “book” may refer to:

- a “work”, for instance “Don Quijote de la Mancha” written by Miguel de Cervantes ;
- an “expression”, for instance the bilingual text in French and in Spanish translated by Marie Menechal;
- a “manifestation”, for instance the paperback version published in 2007 by the publisher “Langues pour tous” or the PDF version published in 2010 (if any);
- an “item”, for instance this unique paper version you have at home, held between your hands.

A “work” is realized through an “expression” (for instance in different languages, in audio or text form) which is embodied in a “manifestation” (for instance on different kinds of support, with different layouts), which is exemplified by an item.

This conceptual framework can be of special interest for eBooks and for search portals. Indeed:

- In order to distinguish eBooks from print versions, it is highly recommended to give them specific ISBN numbers. A discussion is still opened whether each format of eBook should have its own ISBN (...and we would say yes!). But then it is not possible to offer a general bibliographic solution in order to know that such eBook version is the digital version of such print version (or reciprocally), which can be very useful for many reasons.
- A user may also be interested by all the expressions of a same work (for instance which audio book exists for “Don Quijote”, and some of these expressions may not be found simply by the title (because the title has been translated) or on the contrary a search by title may lead to a lot of noise.

Hence, the differentiation in the metadata of the respective work, expression, manifestation to which refers the “book” would be very useful. A first concrete attempt to do this is the ISTC (International Standard Text Code) initiative, which proposes to build an international registry for recording works.

ISTC numbers could be attributed besides ISBN in order to link different print or digital manifestations of the same work.

#### 4. The architecture of the ENCLAVE Editores-BNE project

In this part, we will describe our understanding of the main lines of the project. Our sources are the documents elaborated by Neturity (“Requisitos técnicos en ENCLAVE Editores-BNE”, 10/05/2009 ; “Fase piloto ENCLAVE Editores-BNE”, 10/05/2009) and the different meetings with the partners of the project to which Numilog was invited to assist.

##### a. The different partners of the project

In the first phase, the ENCLAVE Editores-BNE project will rely on the following partners:

|   |                                   |   |
|---|-----------------------------------|---|
|    | Spanish National Library          | Central player of the project on the Biblioteca Digital Hispanica |
|  | Spanish Federation of Publishers  | Represents publishers in order to define an acceptable scheme     |
|  | Publishers & Rights Holders       | Providers of copyrighted contents and files                       |
|  | Bibliographic information service | Provider of metadata  |
|  | E-distributors                    | Access points for eBooks browsing / Digital warehouses            |
|  | EBooks stores                     | Commercial points for access to full contents                     |

Dilve is the central player for organizing the files and metadata flow between publishers and the Biblioteca Digital Hispánica (BDH). The Federación de Gremios, the publishers participating to the first phase and the Spanish National Library are the decision group. E-distributors and eBook stores will enter the project on the basis of the principles and architecture decided by this group and of the choices of each publisher. Hence it is important to identify which are the goals and constraints of the two parts of the decision group, as far as we understood them:

The Spanish National Library wishes:

- To offer the largest possible access to digitized books - and the best possible information about them - within the scope of their documentary policy

- To offer a very ergonomic solution, maximizing the homogeneity between search tools and browsing services and the fluidity of the user experience
- Not to be assimilated to a book store or to a commercial agent
- To minimize the costs for the reception and displaying of files and metadata
- To use for the project the indexation tool already used by Biblioteca Digital Hispánica, i.e. « DigiTool » of Exlibris. This last point has some precise technical consequences:
  - o Metadata have to be collected in Marc format
  - o Indexation files (for the search engine) have to be in PDF format.

The publishers generally wish:

- To learn from this project in order to begin to structure their digital strategy
- To promote their books through the project
- To control and monetize access to full texts
- To control the browsing of long extracts, not only for copyright protection matters but also for ergonomic and commercial matters: the browsing solution should also be a point from which sales points are indicated to the users
- To minimize the marginal costs of the project for the production of files and metadata required and for the exchange protocols.

Of course both parts agree that these goals may be achieved in a first phase by focusing the project on the field defined in §1. :

- Digital texts (or “eBooks”)
- Original (rather than critical) cultural or scientific works
- Books existing in both digital and print versions
- Books for which one publisher explicitly holds the digital rights
- eBooks for which a full-text access is possible, according to conditions defined by the publisher

#### **b. The economic model and the services offered to users**

The project will lead to mix two different economic models and functional roles:

- The BDH will be a **full digital library** for public domain works, offering users the possibility to access the full texts on its website.
- The BDH will be a **search portal** for copyrighted books, offering users the possibility to search inside the texts of the books and in metadata, and to browse them, either through short extracts on the BDH site, or through ones on e-distributors platforms, the access to the complete eBook being possible only through commercial conditions.

This model can be summarized by the following three step table:

### ENCLAVE: 3 steps

|   |   |   |  |
|---|---|---|--|
| <b>STEP 1</b><br>Search Portal          | FREE  | Biblioteca Digital Hispanica (BDH)                        | Full -Text Search<br>Search by subject and key words<br>Display of metadata<br>Short extracts (e.g. 1st chapter) |
| <b>STEP 2</b><br>Browsing sites         | FREE  | BDH (public domain)<br>E-distributors (copyrighted books) | Long extracts (e.g. 20% )<br>Full-text search  |
| <b>STEP 3</b><br>Access to full content | FREE (public domain books)<br>COMMERCIAL (copyrighted books ) | BDH<br>E-Book stores                                      | Various economic models and formats  |

Step 1, on the website of the BDH, essentially offers search functionalities, but also puts forth some browsing possibilities through short extracts of the books, typically the first pages. Step 2 offers more a complete browsing service through longer extracts which may cover the entire book (from page 1 to last page), but with set of pages not accessible for free. The browsing solution of step 2 must be linked to ways of reading the entire books, through various commercial models and formats of eBooks, on various eBook stores (step 3).

#### c. The requested files

The ENCLAVE project documents require that publishers create and send the following files to the relevant players in order to offer all the services described in the previous scheme:

- File of the complete eBook: at least in one of the standard formats of the market (PDF, ePub, PRC) in order to offer along a chosen commercial model at least one digital access to the full text;
- A PDF file for the indexation of the full text with Digitool, which may be the same than the PDF eBook file if any, or prepared specifically for indexation, e.g. without the layout, or without some parts of the books that the publisher does not consider relevant for searching (e.g. index, copyright page, “from the same author” page...);
- A file for the browsing solution on the e-distributor platform, whose format depends of the e-distributor system specifications but is generally a PDF file, and may be the same than the eBook file;
- A PDF file for the short extract (or “partial content”) on the BDH.

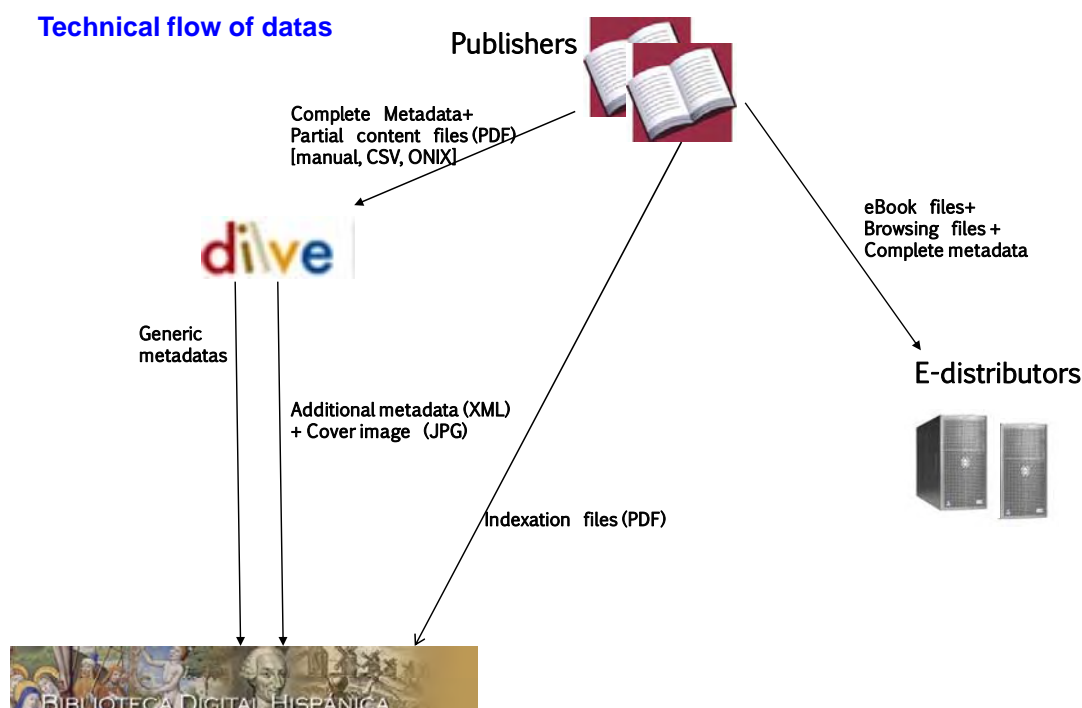
#### d. The requested metadata

The metadata are split by the project in two kinds:

- Book metadata already sent by Dilve and registered by the Biblioteca Digital Hispánica, since each eBook in the project must have a print version ;
- Additional metadata for searching and describing the digital version. These metadata are explicitly limited to the followings :
  - o ISBN of the print version
  - o ISBN of the digital version
  - o Publication date of the digital version
  - o Cover image (JPG format, high quality, 700x700 pixels)
  - o Summary (pure text, < 500 terms or < 2000 characters)
  - o Table of contents (pure text, with “--“ to separate the sections)
  - o E-distributor platform URL for browsing the book
  - o Optional : specific key words

#### e. The exchange flow of files and metadata

The exchange flow of files and metadata between the different players that has been suggested is summarized by the following scheme:



The indexation files with the full text will be sent by the publishers directly to BNE / BDH, where these files will be used for searching but will not be displayed to the public.

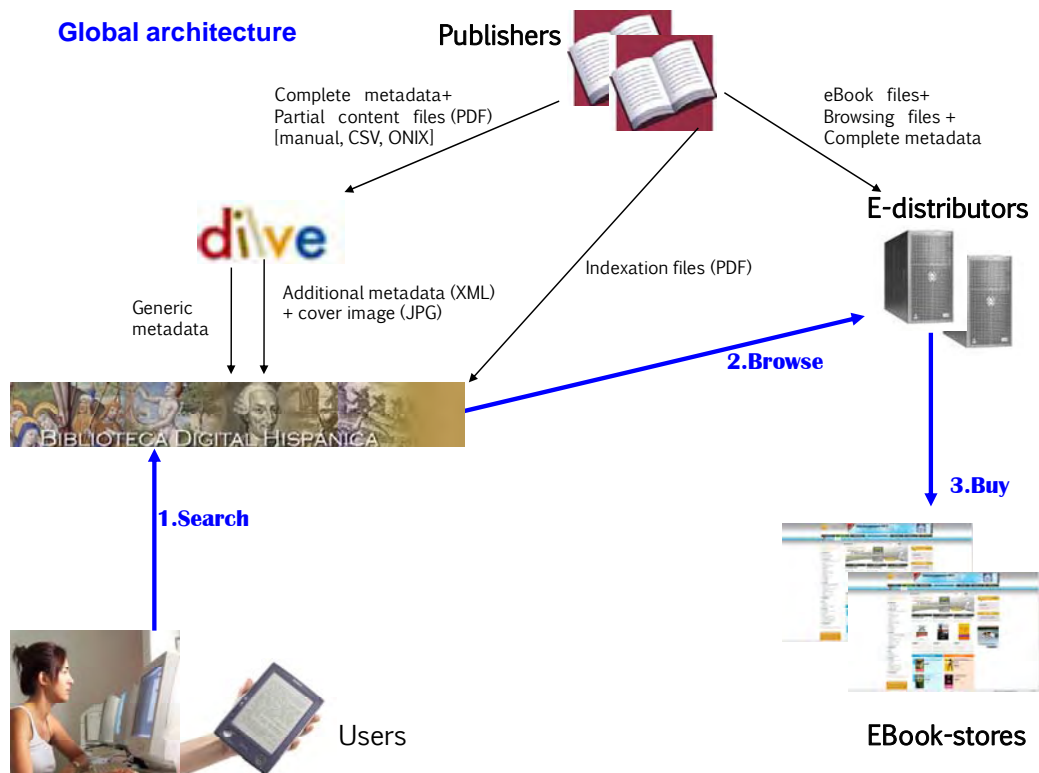
The partial content files will be sent by publishers directly to Dilve. An open alternative is to send them directly to BNE, if the publisher does not want to make these files public on Dilve (all the data registered by Dilve is always public).

The final format of metadata for BNE will be Marc 21. However, this is not a requirement for publishers, since they will enter the Dilve database, following which Dilve will send them to BNE in XML and the conversion in Marc 21 will be achieved by BNE along some correspondence rules.

Finally, publishers must send to e-distributors the complete eBooks and browsing files (generally the same if the eBooks are sent in PDF), and of course the complete set of metadata, including additional informations required for sales, e.g. prices and DRM parameters. As has already been explained, “e-distributor” stands here for a function, which may be held by the publisher himself.

**f. The global architecture of the project: a synthetic scheme of the roles of the different players**

Finally, the economic model and the technical architecture proposed for the project may be summarized by the following scheme:



The technical flow of data allows performing the 3 steps economic model for services. The last players added into the scheme are eBook stores, where the accesses to eBooks are sold. Again, “eBook stores” stands only for a functional, which may be held by traditional online book stores, by e-distributors themselves, or by publishers. In any case, the eBook files will be delivered to the user directly by the e-distributors, which also send the relevant DRM or technical measures of protection after validation of orders by the eBook stores.

**5. Comments and suggestions**

The economic model and technical architecture of the project, as described in the previous part, lead to the following comments and suggestions:

#### **a. General comments**

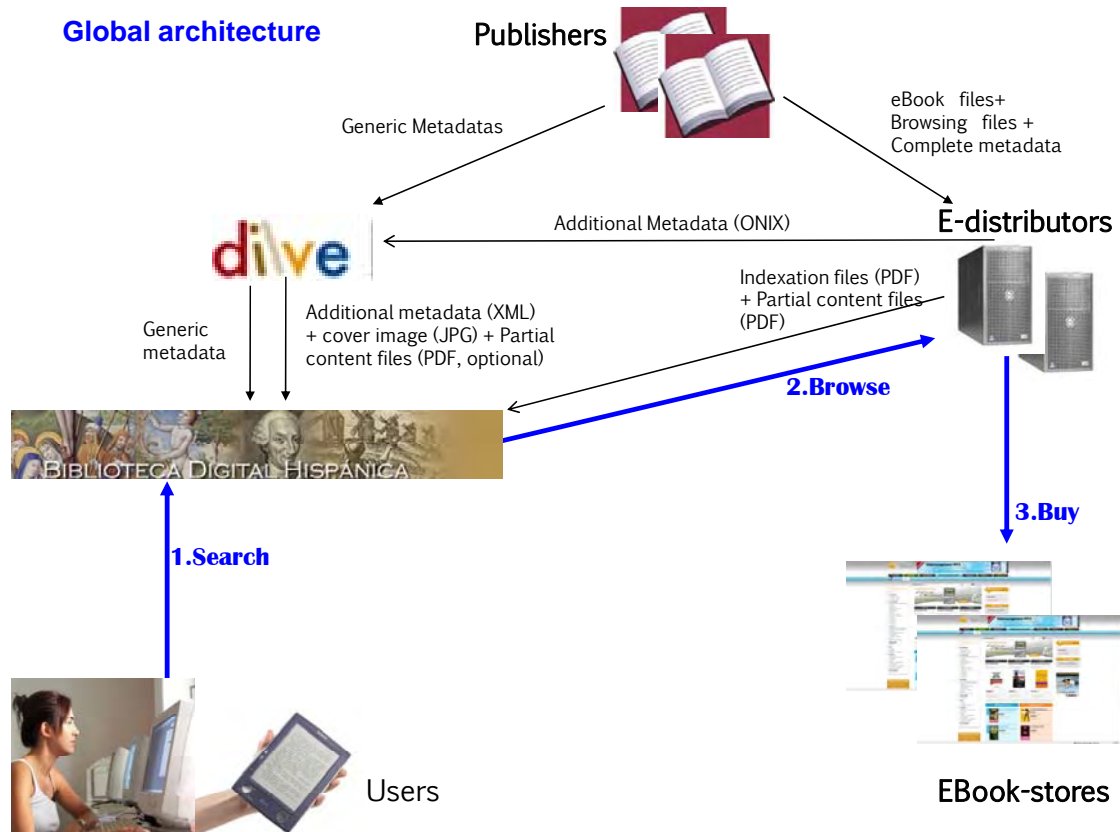
- The global architecture of the project is relevant to its goals, and is economically and technically coherent ;
- The project is faithful to the Report of Marie-Hélène Descamps, “i2010: Vers une bibliothèque numérique européenne“, adopted by the European Parliament in 2007 in order to suggest principles for the future European Digital Library (now called “Europeana”). Hence, it should be easy to consider in a second phase a connection with Europeana.
- The project relies on the same general principles as the main other existing national library project, Gallica, but has its own originality and added value: it offers a possibility for browsing books directly within the national library website; the synergies with existing print metadata and with the national books database organization, Dilve, are emphasized; the requirements of homogeneity between the browsing platforms of e-distributors will be strict.
- The first phases of the project may lead to further phases, enlarging its technical richness, for instance concerning metadata, and the user services on the BDH website. Some of these possibilities will be described below and in the part 6 of this document.

#### **b. Technical suggestions**

Considering the previous comments, the following technical suggestions will focus only certain details which may optimize the work flow, clarify the contractual relations and open the way to possible future phases:

- The main suggestion concerns the files and metadata flow: in order to secure and to simplify it for the publishers, the role of the e-distributors should be more central. Indeed, it is necessary to ensure that the activation of a book on BDH is posterior to its availability on the browsing platform and as an eBook. And yet this information is under the e-distributor’s exclusive control. If the publishers send the additional metadata and the indexation files to the BNE before the e-distributors have uploaded the files onto the platform, there is a risk of failure when the users want to browse the titles found. Furthermore, the syntax of the browsing URL is built by e-distributors and may evolve. Hence it should be their responsibility to indicate this URL and availability of the title for browsing. Of course, coordination between publishers and e-distributors on these two points is possible, but it would add complexity and work for both parties. Finally, publishers will in any case have to send complete eBook files and metadata to e-distributors. The latter will have the possibility to easily create short extracts and indexation files from these files, following publisher’s specifications. If e-distributors send the indexation files, partial content files and additional metadata to Dilve and to BNE, this will save time and money for publishers, and will facilitate the participation of medium or small companies to the ENCLAVE project.

Hence, the recommended technical architecture is the following:



The additional metadata could be sent by e-distributors to Dilve in ONIX format, enabling a secure automation process for transferring them. Image covers and short extracts could be included within the ONIX files, for instance through links to online covers and extracts.

- Updating protocols should be defined for at least two cases :
  - o The eBook has to be withdrawn from sale and from browsing, for instance because of disagreement with right holders;
  - o A new version of the book and a new digital version are issued: it is in general not relevant to keep two versions of the same “work” or “expression” for online searching and access purposes (except for certain ancient books having a value for book lovers).
  
- If the goal is to minimize production costs, the simplest way is to create one single PDF file that will be used for selling the eBook, for indexation, for browsing (with “holes” created by the browsing platform) and for partial content (it is very simple to cut a PDF file). But it is recommended that publishers not “over-simplify” the production process and rather try to adapt the digital versions to the following cases :
  - o if the book has been digitized from paper, the requirement is to create not only a PDF with the images of the book pages but also to extract the text by an OCR process, in order to enable the reader to search terms inside ;

- PDF eBooks files should be optimized for screen reading and should include minimal enrichments such as bookmarks or hypertext links. The cut marks of the print PDF, if any, should be removed, the digital weight should be optimized, the preferred display mode of the book (one single page, two pages, full screen...) should be chosen, minimal metadata should be added within the PDF ;
  - For literature, it is recommended to offer the possibility to reflow the eBook on eBook devices, in ePub format ;
  - Finally, for indexation, it may be interesting to remove the layout (for security reasons), and to remove certain pages (for relevance reasons, e.g. index, copyright pages).
- The ontology of metadata records should be clarified. Indeed, the basic entity of registration may be a “manifestation”, i.e. a publication with a precise format and identified by an ISBN number, or an “expression”, i.e. a title whose first version has been published at a certain date by a certain publisher, with several further manifestations. In the proposed metadata scheme, there is only one ISBN for each digital version. But there may be several digital manifestations of the same title (“expression”), for instance online version, PDF version and ePub version. If each of them is registered by a different ISBN, this could lead to registering the same title several times and to giving several answers (and book presentation pages) for this title after a search on the BDH site. An alternative would be to allow for several ISBN numbers in the metadata and to link them to the files format, i.e. to register “expressions” rather than “manifestations”. The first phase of the project could ignore this question by requesting publishers to register only one of the digital versions available. But this is surely a question to deal with in further phases.
- In further phases, the list of metadata should be augmented. Here is a non-exhaustive list of relevant candidate-information :
- Formats of the digital versions available (if not included in the first phase as proposed above)
  - Size of the corresponding eBook files
  - E-distributor name
  - “Collection”, in a marketing meaning, as defined by the publisher (if not already registered by the generic metadata over the print version)
  - ISTC numbers, when publishers will create them
  - Information on out-of-print titles
  - Add-in of the digital versions: videos, interactive tools
  - Links to online resources in relation with the work
- On the contrary, it should be possible in further phases to deal with purely digital books. Hence, the requirement of indicating print ISBN in the digital metadata should not be a “hard” constraint of the information system for future evolution.
- The list of generic metadata already registered by Dilve and used by BNE for print books should be explicitly indicated in the final document project.
- The contractual agreements may link BNE to publishers, and publishers to e-distributors (which is the implicit contractual chain in the project documents) or BNE to e-distributors and e-distributors to publishers. Indeed, if one follows the suggestion to ask e-distributors, rather than publishers, to send the additional metadata and files requested, the second contractual chain would be natural.

- Whatever the choice between the two possible contractual chains, it should be requested that the publishers indicate one single link for browsing one single book. This condition is in fact included in the document's project since the additional metadata include one single browsing URL. If the additional metadata are sent by e-distributors, as suggested in the present document, publishers should include exclusivity articles in their agreements with e-publishers in order to warrant this unity. This exclusivity may be limited of course to a list of titles and to the partnership with the ENCLAVE project. And it may be limited to a certain eBook format (a publisher can choose different e-distributors for different formats). But the answer on this last point has to be correlated to the answer to the above question concerning the ontology of metadata: if BDH registers "expressions", then the publisher has to choose only one single e-distributor for this expression, whatever the format of its digitized versions.
- Whatever the choice between the two possible contractual chains, the contractual agreements with publishers and e-distributors should be precise on the following points in order to ensure a minimal homogeneity for the users :
  - o The minimal length or percentage of text of the browsing extracts (not limited to a set of first pages)
  - o The possibility to search terms inside the whole text, even if certain sets of pages are not accessible for free
  - o The deliveries of periodic usage statistics, in order to know how many accesses to the browsing platform have been performed, for which titles, and if possible the direct impact on sales.
- Dilve, which will play a central role for organizing the metadata flow, could be in charge of controlling the conformity of the titles proposed by publishers and e-distributors to the documentary policy of the project.

## 6. Further comments and suggestions on economic models

### a. The economic models of eBook stores

Observing the existing market, one can suggest a (non-exhaustive) typology of the main economic models used, summarized by the following table:

#### Typical eBook stores economic models

|            | Perennial rights             |               | Temporary rights            |                      |
|------------|------------------------------|---------------|-----------------------------|----------------------|
|            | Title by Title               | Packages      | Title by Title (short time) | Packages (long time) |
| Full books | L1<br>Sales of single copies | L2<br>Bundles | L3<br>Renting               | L4<br>Subscriptions  |
| Chapters   | L5<br>Sales of chapters      | L6            | L7                          | L8                   |

The main choices in order to build an economic model for the final customer are indeed to:

- offer, like with a print book, a perennial right to read the title or a temporary right during a period;
- sell each item separately or to offer a price for access to packages, or sets of titles (personal digital libraries)
- sell a complete book or parts, typically chapters (it could also be pages, or articles for a magazine).

The choice of the relevant economic models belongs to publishers, in relation with e-distributors and eBook stores. However, in the context of the ENCLAVE project, our suggestion is to ask publishers to link the browsing service only to economic models of the kind L1 or L3. Indeed, offers per chapter do not propose direct access to full texts. And offers per package are not correlated only to the title found and chosen on the BDH website; hence, they do not propose a homogeneous flow from search on BDH to the access to full texts relevant for that precise search.

In a further phase, the project could also consider ways of associating online print book stores, in order to indicate to users the websites where they can buy the print version.

### b. The economic model of ENCLAVE Editores-BNE

The economic model of the ENCLAVE project in its first phase has been summarized by the “3 steps” model (§4.). This model is the most relevant in order to offer possibilities for any Internet user searching books on the BDH to access to copyrighted texts under two major constraints:

- The respect of copyright and of the commercial goals of right holders
- The difficulty to give any acceptable monetary value to an unlimited access to a book for all Internet users. Indeed, suppose BDH offers to anyone in the world the possibility to read digital version of some recent book for free; it would in this case be required to offer compensation to right holders... and the relevant value of this compensation could be the value of the expected turnover of this book in print version. This would not allow building any realistic deal between public institutions and right holders if important quantities of recent titles are concerned.

Of course, this economic model is associated with the following limit: the functional role of BDH is restricted to a search portal for copyrighted texts. The BDH will not lend the eBooks to their patrons, like a traditional library does. However, in a second phase of the project, one may perhaps go beyond this limitation for certain titles and certain users by offering digital lending services to these users.

In order to study how to build such an alternative model, one can consider again a typology of the main economic models proposed in the market, but this time for digital lending services:

### Digital lending economic models

|  | Purchasing<br>(perennial right of lending) |           | Subscription<br>(annual right of lending) |           |
|--|--|-----------|---|-----------|
|  | Title by Title                             | Packages  | Title by Title                            | Packages  |
| <b>Fixed simultaneous accesses</b><br>(prices depend of number of accesses)  | <b>B1</b>                                  | <b>B2</b> | <b>B3</b>                                 | <b>B4</b> |
| <b>Unlimited simultaneous accesses</b><br>(prices depend of number of users) | <b>B5</b>                                  | <b>B6</b> | <b>B7</b>                                 | <b>B8</b> |

The two main sets of typical business models offered for digital libraries are:

- **Fixed simultaneous accesses**, more often on a title by title basis, where the library may purchase a perennial right to lend the eBooks (B1) or subscribe to an annual right to lend them (B3);

- **Unlimited access to packages of eBooks**, generally on the basis of annual subscription fees (or license agreements), depending for instance of the number of potential end-users (B8).

Hundreds of **local** academic and public libraries in the whole world now offer selections of eBooks along one of those economic models. They are implemented by e-distributors using well installed technologies allowing to:

- Set up check-out periods within the eBooks file (in the DRM parameters)
- Define online reading sessions
- Monitor quantities of simultaneous accesses
- Measure numbers of pages read online for each title
- Control accesses through IP-addresses, log-in and password systems and various kinds of advanced authentication systems such as LDAP directories.

These models may be used for negotiating certain agreements between **national** libraries and publishers, under certain specific conditions:

- The National Library of Norway and Kopinor, an association representing Norwegian right holders associations, negotiated an agreement in order to digitize and lend all the Norwegian titles under copyright up to the 1990s (the project is named Bokhylla.no). These books will not be prepared for print or download, but only for online reading. They will be made available only to Norwegian IP-addresses. Beyond the specificities of Norwegian culture and context, one can observe that the more recent titles are not covered by the agreement and that, besides the fact that access is limited to Norwegian IP-addresses, Norwegian, unlike Spanish, is not an international language.
- In a further stage of the ENCLAVE project, the BDH could offer digital lending services for certain categories of books, especially those interesting searchers and booklovers, not recent trade books. This service could be limited to books published before a certain date (1990s like in Norway?). It could be restricted to registered patrons of the BNE, or to certain categories of these patrons (searchers, teachers, students preparing a thesis...). The access to full texts could be limited to the computers within the premises of BNE, which may be identified by their IP address. But remote accesses could also be proposed for these patrons, identified with log-in and passwords, through secure access control processes. Online accesses should be favored over download. Agreements with publishers could be based on any of the above economic models, especially B1, B3 or B8. This should be discussed on a case by case basis, with each publisher (unlike the Norwegian model where the agreement is global).

On the technical side, these models should be handled again by e-distributors, in order to warrant to publisher's protection over eBooks, control of the reading experience, authentication processes and master of the economic models. Besides external links for browsing, the book presentation pages on the BDH site could add external links for borrowing. Or one may keep one single external link, the possibility to read the full text being offered within the browsing platforms but accessible only after specific authentication of the user and/or IP-address.

This additional economic model and digital lending service would be an original solution proposed by ENCLAVE Editores-BNE. It creates a combination of the search portal model and of the digital library model. Hence it will emphasize the role of BNE as a national library offering accesses to knowledge and cultural texts to those who work with these texts, while ensuring full respect of copyright and of the players of the book chain.