Universidad Complutense Madrid

Case study 1: Google Books at the Complutense University of Madrid

CERL Annual Seminar 2012

October 30 2012, British Library

José Antonio Magán Wals
Antonio Moreno Cañizares
Manuela Palafox Parejo

Complutense University of Madrid Library

The Complutense University of Madrid and its Library

- The Complutense University of Madrid has 85,000 students and 6,289 scholars
- 34 Libraries
- 3 million books, the first academic library in Spain
- 11,300 seats for reading and 1,500 computers
- 411 librarians
- An important digital collection with more than 600,000 objetcts and millions of scanned pages

Our commitment to collaboration for digitization and dissemination of scientific production and heritage

- Opting for open access dissemination through both the called "green route" as in the "golden path“.

- The library acts as digital publisher in collaboration with other university services.

- Collaboration for the dissemination and digitization:
  - **with external institutions and agencies:** Spanish Government, Federal Government of Madrid, Madrid Academic libraries consortium, Europeana, Hathi Trust, Internet Text Archive
  - **with private institutions:** Google, Santander Universities, Health Sciences Foundation, Editorial Extramuros,
  - **with commercial publishers and distributors:** Springer, Thomson Reuters, Proquest, E-Libro

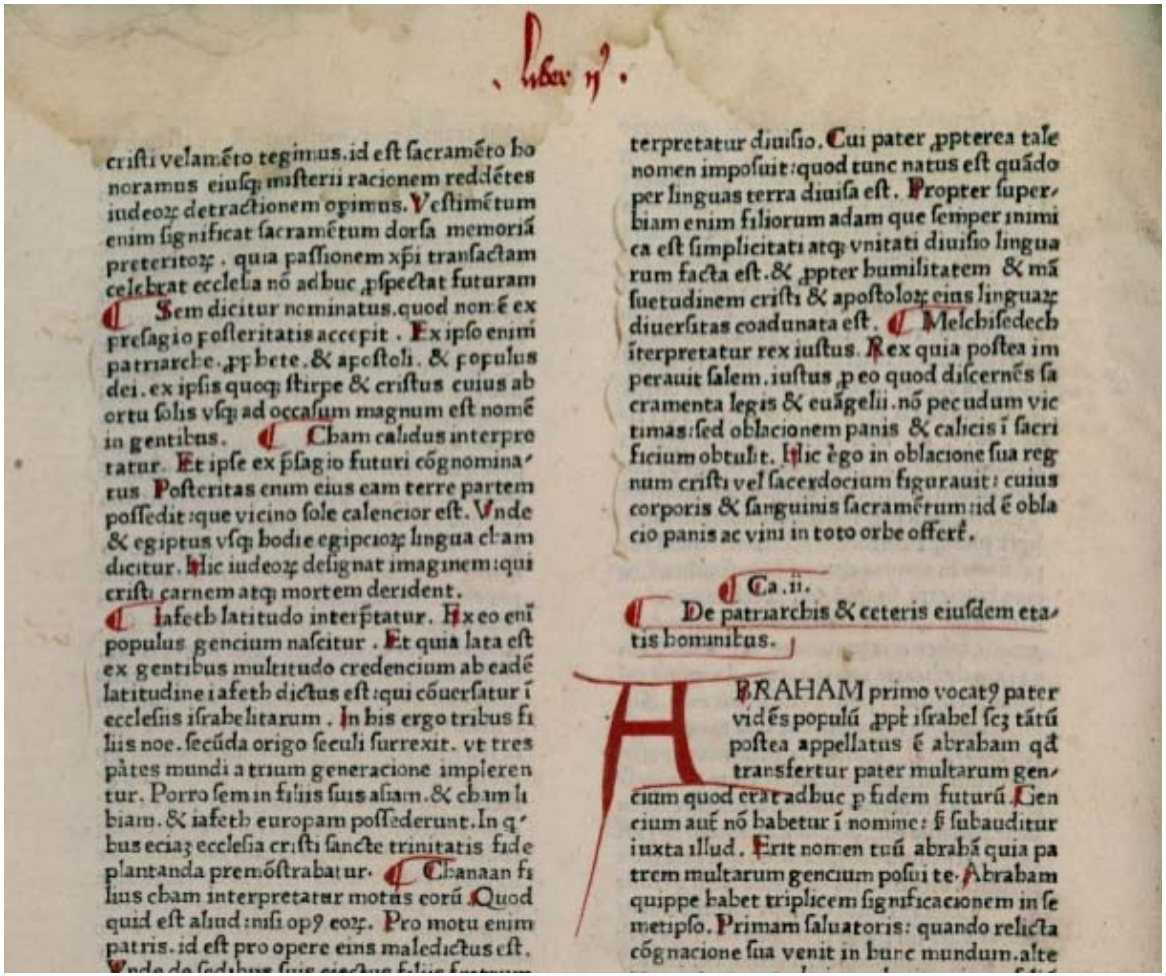
Complutense Digital Colections:
a) Academic works

- 25,000 digital dissertations (5,500 of them in open access)
- 30,000 articles from journals published by our university
- 11,000 e-prins in open access

Complutense Digital Collections: b) Materials for research support

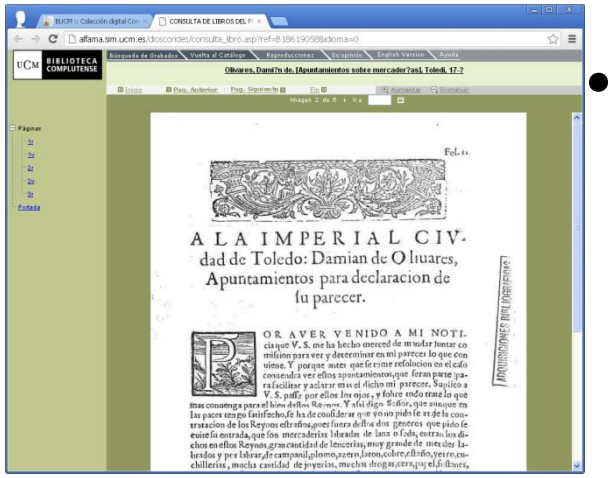
- 400,000 newspapers, photographs from Spanish Civil War, Drawings from the School of Arts…

Complutense University of Madrid Digital Colections :
c) Ancient books and cultural heritage



- 125,000 out of copyright books digitized
- 47,000 Prints
- The greatest digital collection in Spain of Ancient Books

Status of Complutense ancient books digitization in 2006 (*Dioscorides Collection*)



- 2,800 scanned books in 10 years (the largest university collection in open access in Spain).
- At this rate it would take 435 years scanning to digitize the number of works digitized with Google in 3 years.
- Portal deficiencies:
  - No long term digital preservation
  - No multilingual
  - No copyright managed
  - Not adapted to social web

Complutense-Google Agreement

- September 2006. Signing of the cooperation agreement to mass digitization of library collections free of copyright.

- Google Books has scanned more than 20,000,000 books, much of them from libraries:
  - U.S.: Michigan, California, Harvard, Stanford, New York Public Library...
  - Europe: Oxford, Bavarian State Library, Complutense of Madrid, Catalonia, Ghent, Lausanne, Lyon Municipal Library. ..
    Recently joined several national libraries: Florence, Rome, Czech Republic, Austria, Netherlands ...
  - Asia: Library of Keio University, Japan

- Each 6 months more than 90% of the books are visited

A controversial project

- For some people:
  - Violation of the rights of authors and publishers
  - Risk of monopolization of access to the content of books
  - Transfer of public cultural heritage to a commercial company
  - Scanning without enough quality: poor images and OCR.

- For others:
  - Unique opportunity to democratize knowledge by digitizing
  - Creates a free tool that allows you to query the contents of millions of books and download them for reading
  - Stimulates other public and private projects of mass digitization

- The facts:
  - Participating libraries have created with their digital copies important public collections of scanned books.
  - Now when you search Google you find not only infomation in websites but books that can be downloaded

What does Google do?

- Scans documents and is responsible for the costs: Books are scanned twice to avoid errors
- Out of copyright scanned books are freely searchable and downloaded from Google Books.
- Create an exclusive interface for the University and its users, to access to and download digital works of the program.
- Gives to the Complutense a copy of the scanned books.

What does the Complutense Library do?

- We provide the books and experts who oversee the selection of the works to be scanned.
- Update metadata.
- Select and organize the movements of the books to ensure the integrity of collections.
- Preserve and disseminate its digital copy: these copies are used for academic projects

Universidad Complutense Madrid

BEGIN CONSULTATION

http://www.ucm.es/info/dialogycabddh/presentacion.html

HOME
PRESENTATION
TEAM
COLLABORATING INSTITUTIONS
FINANCING
ACKNOWLEDGMENTS
DISSEMINATION
LINKS
CONTACT

Presentation

The goal of **Dialogyca BDDH** (**B**iblioteca **D**igital de **D**iálogo **H**ispánico) is to progressively provide the corpus of all Hispanic literary dialogues written in the different languages of the Peninsula, including Latin-based, Spanish-Hebrew and Spanish-Moorish. During this first phase, we have concentrated especially on the dialogues of the 15th and 16th centuries.
Each data sheet or record corresponds to a single dialogue for which the following information is provided: data on the authorship and literary characteristics, primary bibliography, secondary bibliography and, in addition, the reproduction of the testimony most relevant for the transmission of the text. Thus, when the dialogue on one record is part of a volume along with another type of text or other dialogues, two data sheets have been created. Both of these have internal links to make the consultation easier: one link is for the dialogue and other is for the book in which it is conserved.

Target audience

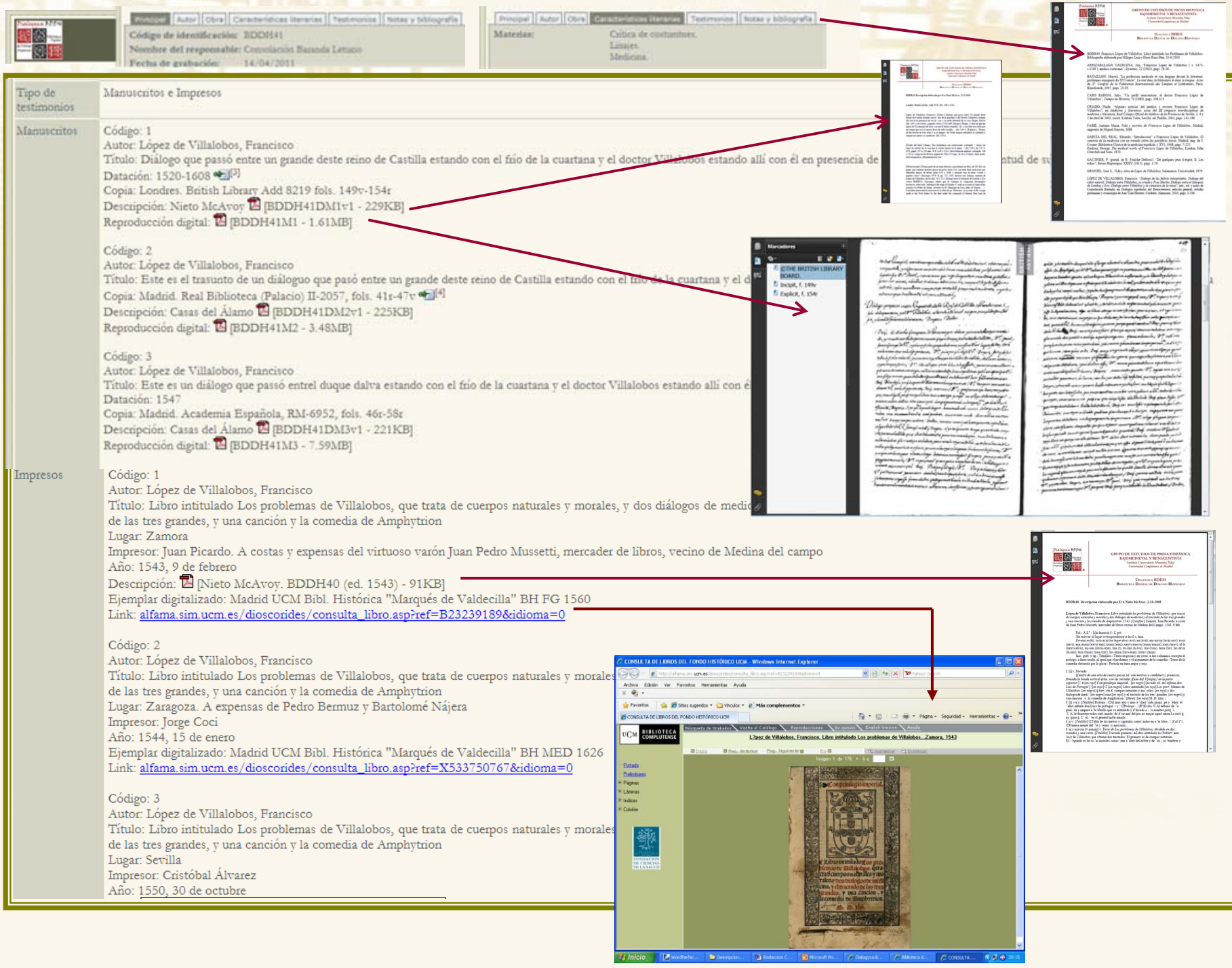
Dialogyca BDDH is aimed at researchers and people interested in a wide variety of specialities, since dialogue is a genre that has served for the transmission of information from all of the branches of knowledge: philosophy, medicine, engineering, spirituality, astronomy, etc. The purpose is to provide the community with comprehensive, coherent, stable and rigorously prepared information created with the tools of philology.

Access

There are two types of access: A) Free: for those wishing to access general information about the work, and B) Restricted: for researchers who would also like to consult more specialised documents (digitised text, bibliographic descriptions and secondary bibliographies). The project is open to all interested researchers. To make collaboration easier, the team has developed a set of users' guidelines or standard operating manual (Plan de Trabajo Normalizado) that can be found at: http://eprints.ucm.es/8638/

REGISTER
CONSULTATION HELP
AUTHORS AND WORKS IN DIALOGYCA BDDH
HOW TO REFERENCE BDDH
BIBLIOGRAPHIC REPERTORIES CITED IN DIALOGYCA BDDH (WITH ABBREVIATED REFERENCE FORMULA)
MANUAL METHOD STANDARD WORK
IU MENÉNDEZ PIDAL

©Dialogyca BDDH. ISBN 978-84-691-8286-4. Ppdad intelectual: M-000333/2010. Registro de marca: 2855599


Tipo de testimonios
Manuscritos e Impresos
Manuscritos
Código: 1
Autor: López de Villalobos, Francisco
Título: Diálogo que passó entre un grande deste reino de Castilla estando con el frío de la cuartana y el doctor Villalobos estando allí con él en presencia de
Datación: 1520-1608 [3]
Copia: Londres. British Library Add 8219 fols. 149v-154r
Descripción: Nieto McAvoy [BDDH41DM1v1 - 229KB]
Reproducción digital: [BDDH41M1 - 1.61MB]
Código: 2
Autor: López de Villalobos, Francisco
Título: Este es el trasunto de un diálogo que pasó entre un grande deste reino de Castilla estando con el frío de la cuartana y el d
Copia: Madrid. Real Biblioteca (Palacio) II-2057, fols. 41r-47v [4]
Descripción: Casas del Álamo [BDDH41DM2v1 - 225KB]
Reproducción digital: [BDDH41M2 - 3.48MB]
Código: 3
Autor: López de Villalobos, Francisco
Título: Este es un diálogo que passó entrel duque dalva estando con el frío de la cuartana y el doctor Villalobos estando allí con é
Datación: 1547
Copia: Madrid. Academia Española, RM-6952, fols. 46r-58r
Descripción: Casas del Álamo [BDDH41DM3v1 - 221KB]
Reproducción digital: [BDDH41M3 - 7.59MB]
Impresos
Código: 1
Autor: López de Villalobos, Francisco
Título: Libro intitulado Los problemas de Villalobos, que trata de cuerpos naturales y morales, y dos diálogos de medic
de las tres grandes, y una canción y la comedia de Amphytrion
Lugar: Zamora
Impresor: Juan Picardo. A costas y expensas del virtuoso varón Juan Pedro Mussetti, mercader de libros, vecino de Medina del campo
Año: 1543, 9 de febrero
Descripción: [Nieto McAvoy. BDDH40 (ed. 1543) - 91KB]
Ejemplar digitalizado: Madrid UCM Bibl. Histórica "Marqués de Valdecilla" BH FG 1560
Link: alfama.sim.ucm.es/dioscorides/consulta_libro.asp?ref=B23239189&idioma=0
Código: 2
Autor: López de Villalobos, Francisco
Título: Libro intitulado Los problemas de Villalobos, que trata de cuerpos naturales y morales
de las tres grandes, y una canción y la comedia de Amphytrion
Lugar: Zaragoza. A expensas de Pedro Bermuz y Bartolomé Nájera
Impresor: Jorge Coci
Año: 1544, 15 de enero
Ejemplar digitalizado: Madrid UCM Bibl. Histórica "Marqués de Valdecilla" BH MED 1626
Link: alfama.sim.ucm.es/dioscorides/consulta_libro.asp?ref=X533750767&idioma=0
Código: 3
Autor: López de Villalobos, Francisco
Título: Libro intitulado Los problemas de Villalobos, que trata de cuerpos naturales y morales
de las tres grandes, y una canción y la comedia de Amphytrion
Lugar: Sevilla
Impresor: Cristóbal Álvarez
Año: 1550, 30 de octubre
Principal
Autor
Obra
Características literarias
Testimonios
Notas y bibliografía
Código de identificación: BDDH41
Nombre del responsable: Consolación Baranda Leturio
Fecha de grabación: 14/04/2011
Materias:
Crítica de costumbres.
Linajes.
Medicina.
THE BRITISH LIBRARY BOARD
Incipit, f. 149v
Explicit, f. 154r
Dialogo que passo entre el Duque de Alua estando con el frio de la cuartana y el doctor Villalobos estando allí con el
Pero, el dicho Duque no se quexaua mucho de la calentura, porque
CONSULTA DE LIBROS DEL FONDO HISTÓRICO UCM - Windows Internet Explorer
UCM BIBLIOTECA COMPLUTENSE
López de Villalobos, Francisco. Libro intitulado Los problemas de Villalobos. Zamora, 1543
Portada
Preliminares
Páginas
Láminas
Índices
Colofón
Con priuilegio imperial.
Inicio

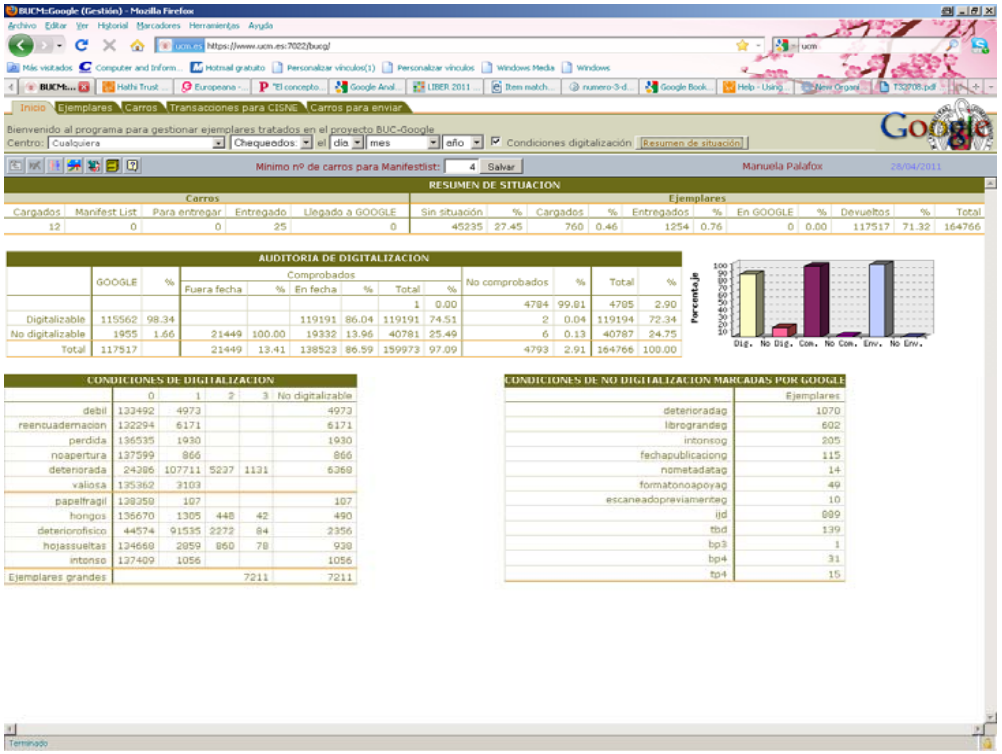
Project Planning and Design: 2007 Actions

- Collection Analysis of the works and the libraries involved.
- Progress reports: Data on facilities, access to repositories of books, etc.
- Selection criteria guide : fixed criteria (date of publication) and criteria for the scanning condition.
- 19th Century books bookbinding plan and Recommendations Guide.
- Scanning Program: workflows, schedule and logistics operations (even cleaning books and the book repositories)

Project Planning and Design: 2008-2011 Actions

- Cataloging Plan: 220,000 books cataloged.
- Analysis of the conservation status and selection criteria for scanning in 145,000 books.
- Scanning in the Google Scanning Center of more than 200,000 Books (120,000 from Complutense Library, the rest from Catalonia libraries)
- In june 2011 we finished the operations

Technological Developments: Web application for project management:

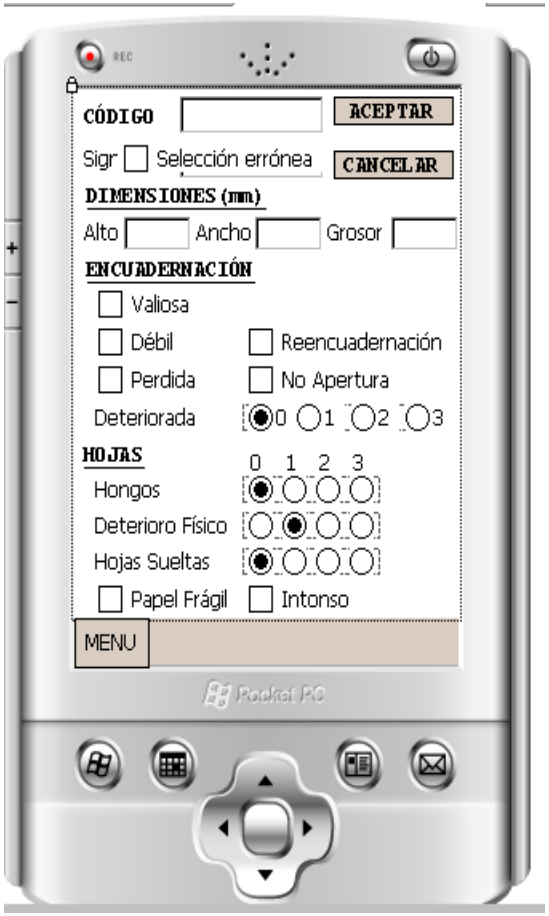


- Offers information online and in real time of all the daily movements of books, Google shipments, returns, preservation status of the books, etc. with statistical data on project operations.

- Stores the metadata of the books included in the digitization project.

Universidad Complutense Madrid

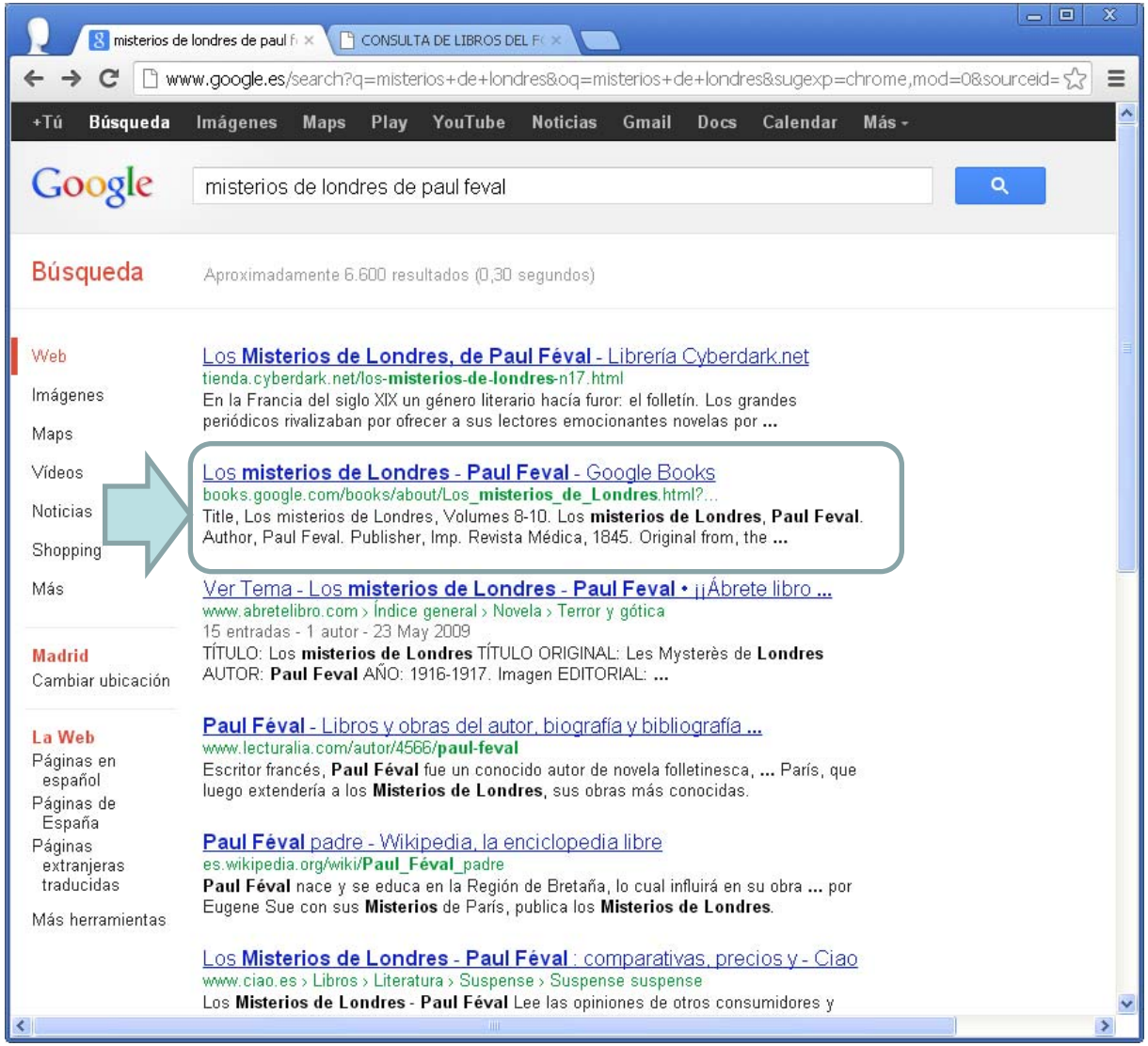
Technological Developments: PDA application



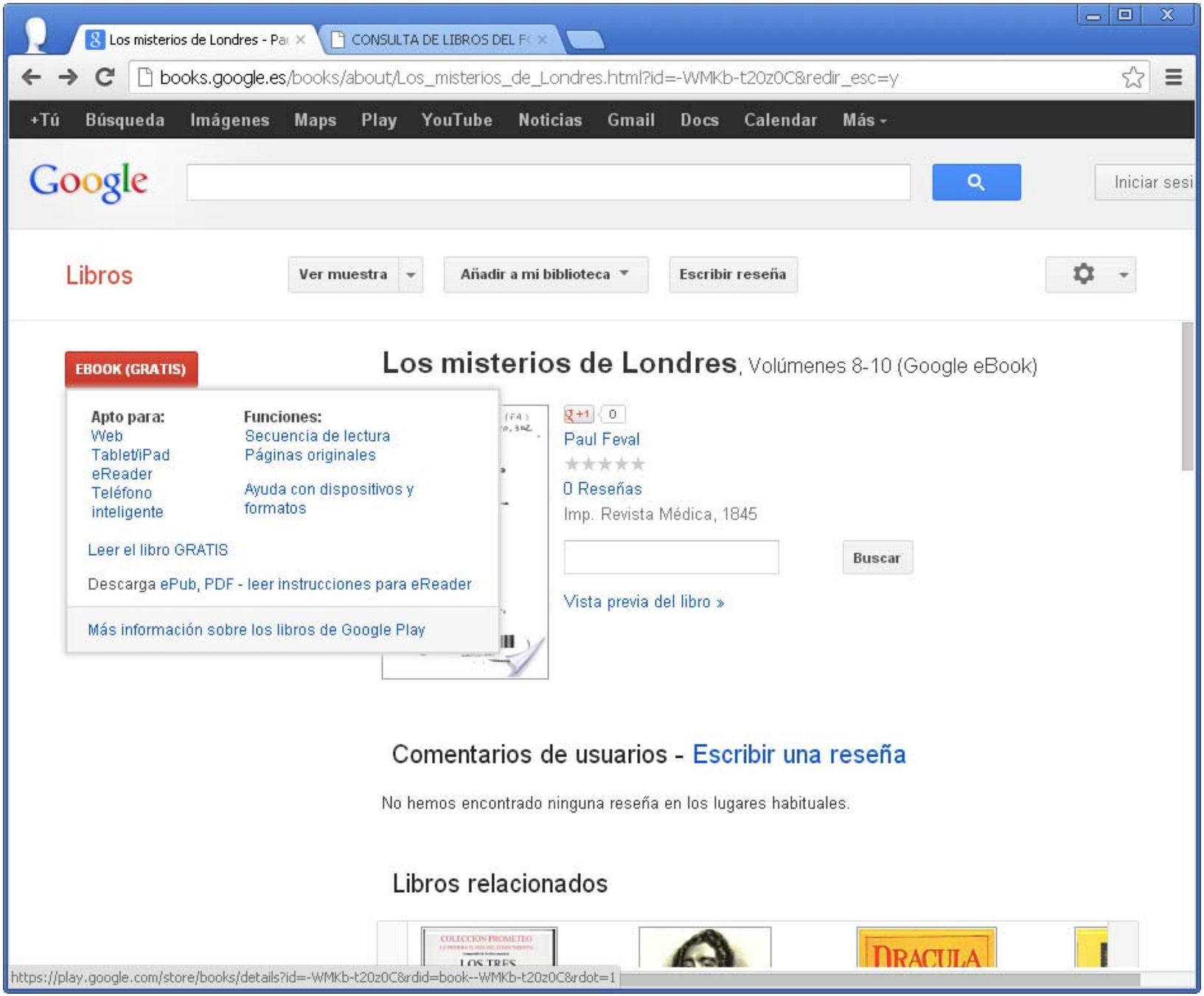
- For selection tasks in storage libraries
- It reads the bar code book.
- We present the characteristics and condition of the book according to the selection criteria guide
  - Dimensions: height, width and thickness.
  - Binding Type: Valuable, weak, lost, rebinding, with opening problems, impaired.
  - Sheets: fungi, physical deterioration, flyers, fragile paper, uncut.
- Such information is exported to Web Management System and then to the Library Catalog

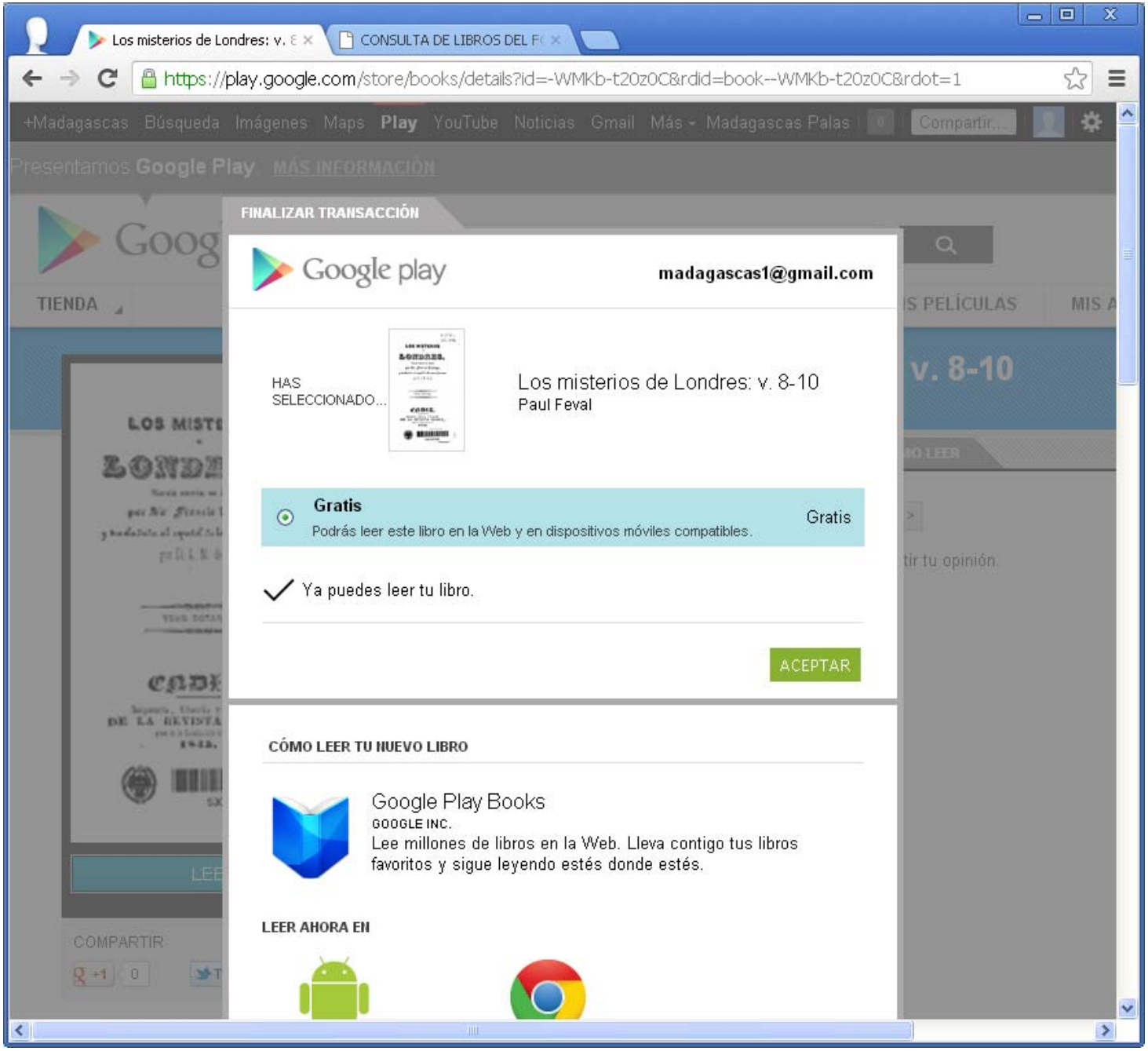
Universidad Complutense Madrid

How do you access the Complutense digitized books? 1. Searching anything in Google (or Google Books or Google Play)



Universidad Complutense Madrid


Los misterios de Londres - Pa
CONSULTA DE LIBROS DEL F
books.google.es/books/about/Los_misterios_de_Londres.html?id=-WMKb-t20z0C&redir_esc=y
+Tú
Búsqueda
Imágenes
Maps
Play
YouTube
Noticias
Gmail
Docs
Calendar
Más
Google
Iniciar sesi
Libros
Ver muestra
Añadir a mi biblioteca
Escribir reseña
EBOOK (GRATIS)
Apto para:
Web
Tablet/iPad
eReader
Teléfono inteligente
Funciones:
Secuencia de lectura
Páginas originales
Ayuda con dispositivos y formatos
Leer el libro GRATIS
Descarga ePub, PDF - leer instrucciones para eReader
Más información sobre los libros de Google Play
Los misterios de Londres, Volúmenes 8-10 (Google eBook)
+1 0
Paul Feval
0 Reseñas
Imp. Revista Médica, 1845
Buscar
Vista previa del libro »
Comentarios de usuarios - Escribir una reseña
No hemos encontrado ninguna reseña en los lugares habituales.
Libros relacionados
DRACULA
https://play.google.com/store/books/details?id=-WMKb-t20z0C&rdid=book--WMKb-t20z0C&rdot=1


Los misterios de Londres: v. 8
CONSULTA DE LIBROS DEL F
https://play.google.com/store/books/details?id=-WMKb-t20z0C&rdid=book--WMKb-t20z0C&rdot=1
+Madagascas Búsqueda Imágenes Maps Play YouTube Noticias Gmail Más Madagascas Palas
Compartir...
Presentamos Google Play MÁS INFORMACIÓN
TIENDA
FINALIZAR TRANSACCIÓN
Google play
madagascas1@gmail.com
HAS SELECCIONADO...
Los misterios de Londres: v. 8-10
Paul Feval
Gratis
Podrás leer este libro en la Web y en dispositivos móviles compatibles.
Gratis
Ya puedes leer tu libro.
ACEPTAR
CÓMO LEER TU NUEVO LIBRO
Google Play Books
GOOGLE INC.
Lee millones de libros en la Web. Lleva contigo tus libros favoritos y sigue leyendo estés donde estés.
LEER AHORA EN
v. 8-10
COMPARTIR

Los misterios de Londres - Pa

CONSULTA DE LIBROS DEL F

books.google.es/books?id=-WMKb-t20z0C&printsec=frontcover&hl=es#v=thumbnail&q&f=false

+Madagascas Búsqueda Imágenes Maps Play YouTube Noticias Gmail Drive Calendar Más

Google

Mad... 0

Libros

Añadir a mi biblioteca Escribir reseña

Cubierta delantera

LEER EBOOK

Conseguir este libro impreso ▼

+1 0

0 Reseñas

Escribir reseña

Los misterios de Londres, Volúmenes 8-10

Escrito por Paul Feval

Buscar en este libro Ir

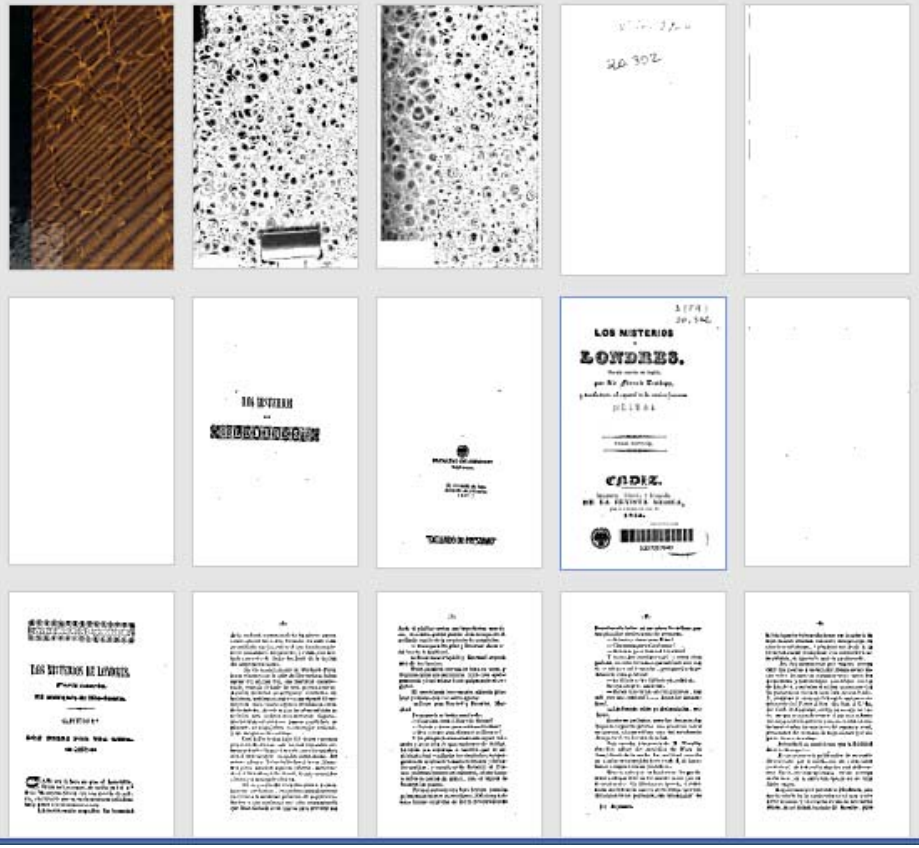
Acerca de este libro

▸ Mi biblioteca

▸ Mi historial

Libros en Google Play

Condiciones de servicio



Los misterios de Londres - Pa | CONSULTA DE LIBROS DEL F

books.google.es/books?id=-WMKb-t20z0C&hl=es&source=gbs_navlinks_s&redir_esc=y

+Madagascas Búsqueda Imágenes Maps Play YouTube Noticias Gmail Drive Calendar Más

Google

Mad... 0

Libros

Añadir a mi biblioteca Escribir reseña

LEER EBOOK

Conseguir este libro impreso

- Mi biblioteca
- Mi historial

Libros en Google Play

Los misterios de Londres, Volúmenes 8-10 (Google eBook)

+1 0

Paul Feval

★★★★★

0 Reseñas

Imp. Revista Médica, 1845

Buscar

Vista previa del libro »

Comentarios de usuarios - Escribir una reseña

No hemos encontrado ninguna reseña en los lugares habituales.

Libros relacionados

Los misterios de Londres - Pa

CONSULTA DE LIBROS DEL F

books.google.es/books?id=-WMKb-t20z0C&hl=es&source=gbs_navlinks_s&redir_esc=y

+Madagascas Búsqueda Imágenes Maps Play YouTube Noticias Gmail Drive Calendar Más

Google

Mad... 0

Libros

Añadir a mi biblioteca Escribir reseña

LEER EBOOK

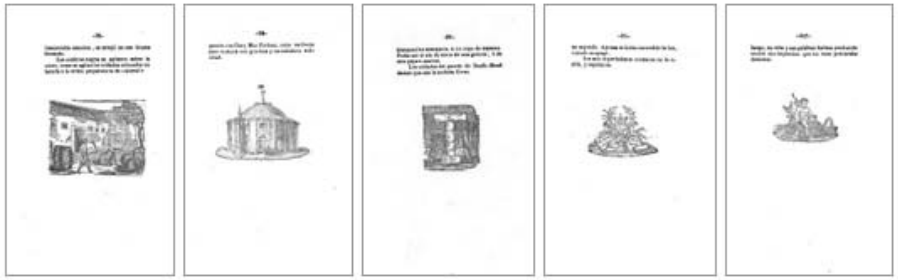
Conseguir este libro impreso

Mi biblioteca

Mi historial

Libros en Google Play

Páginas seleccionadas



Página 76 Página 39 Página 60 Página 21 Página 127

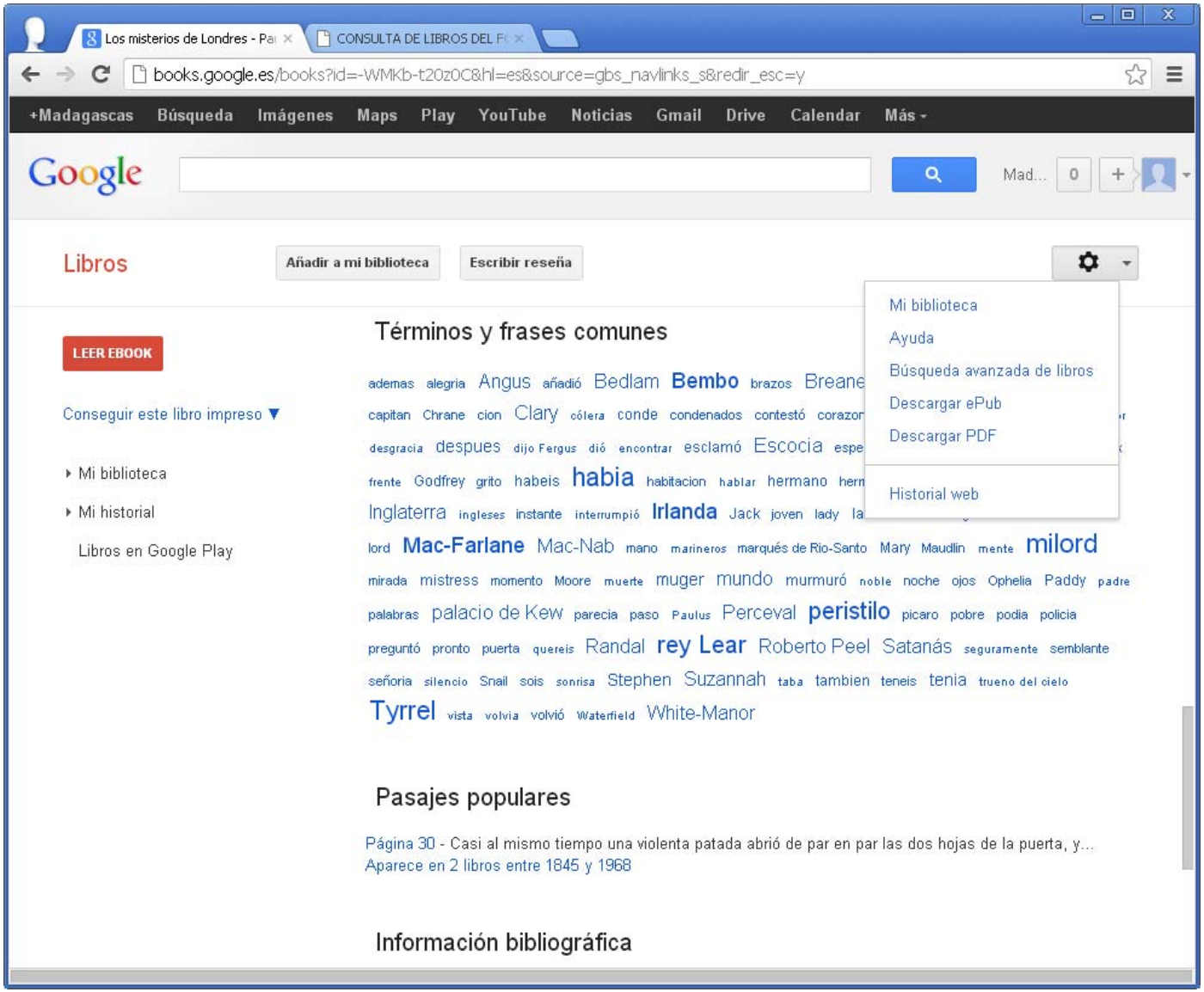
Otras ediciones - Ver todo

1848

1848

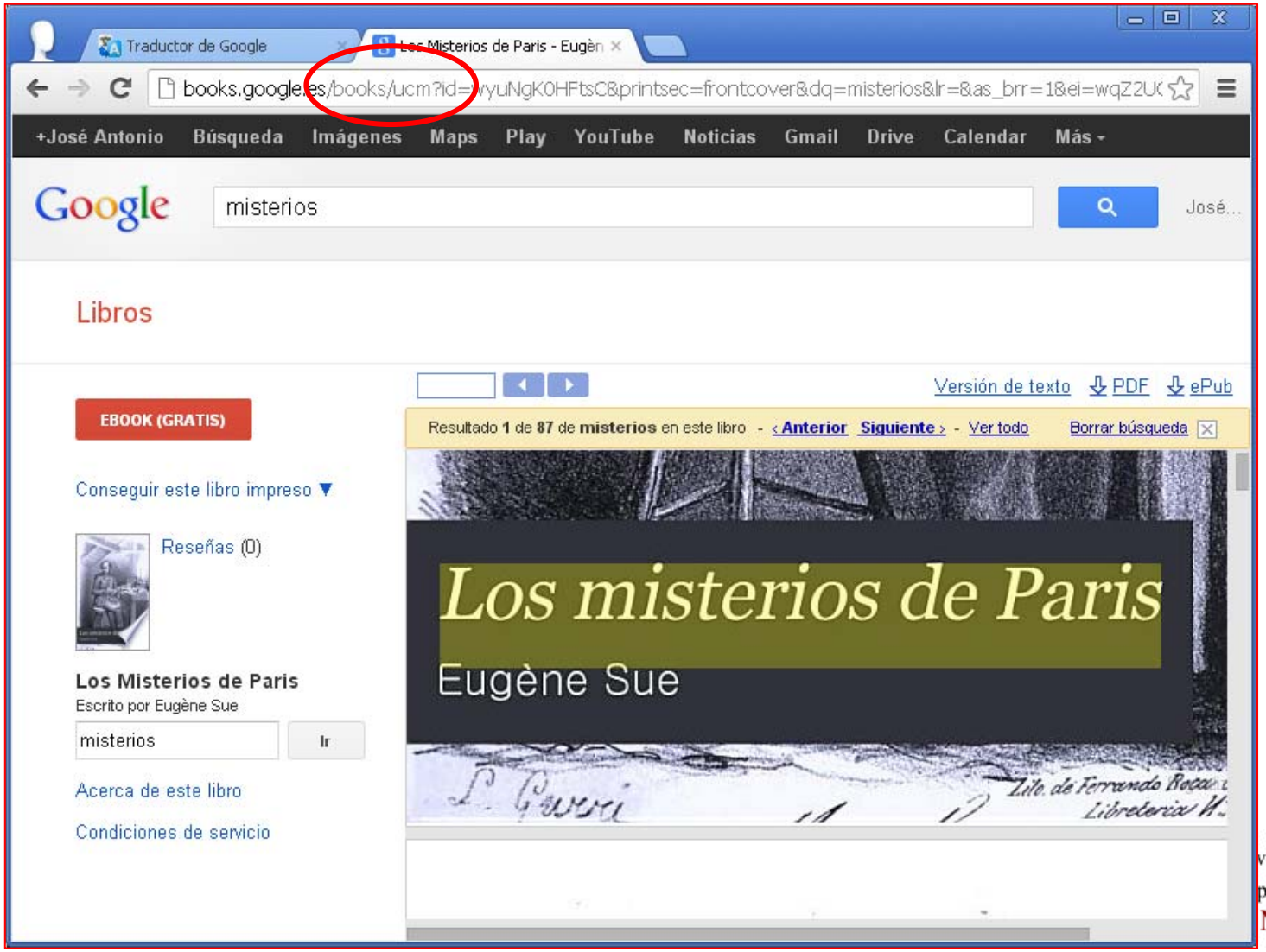
1845

1844



Remember: Every 6 months more than 90% of the 20.000.000 Google Books are visited

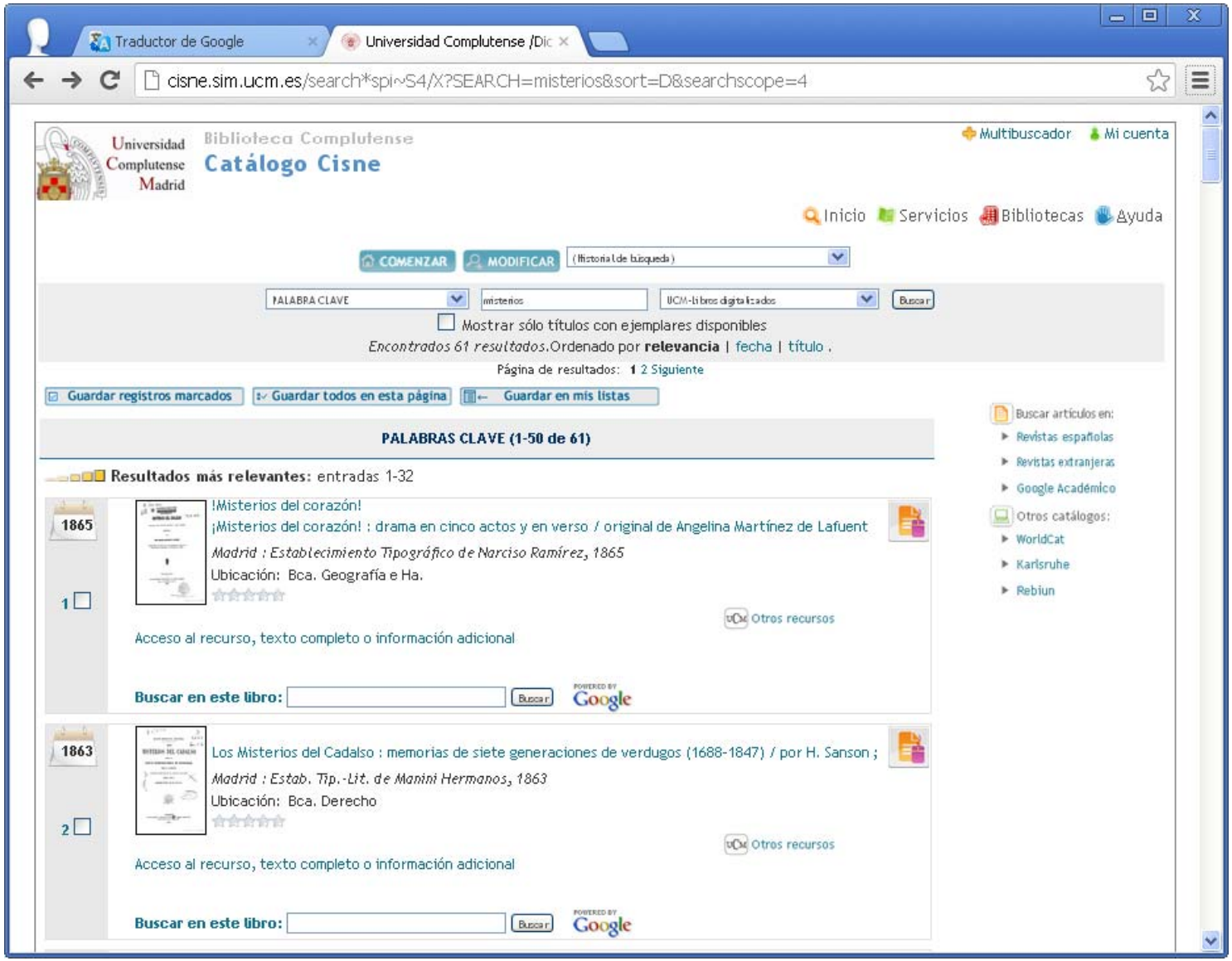
How do you access the Complutense digitized books?
2. Exclusive Google search interface for searching Complutense books .



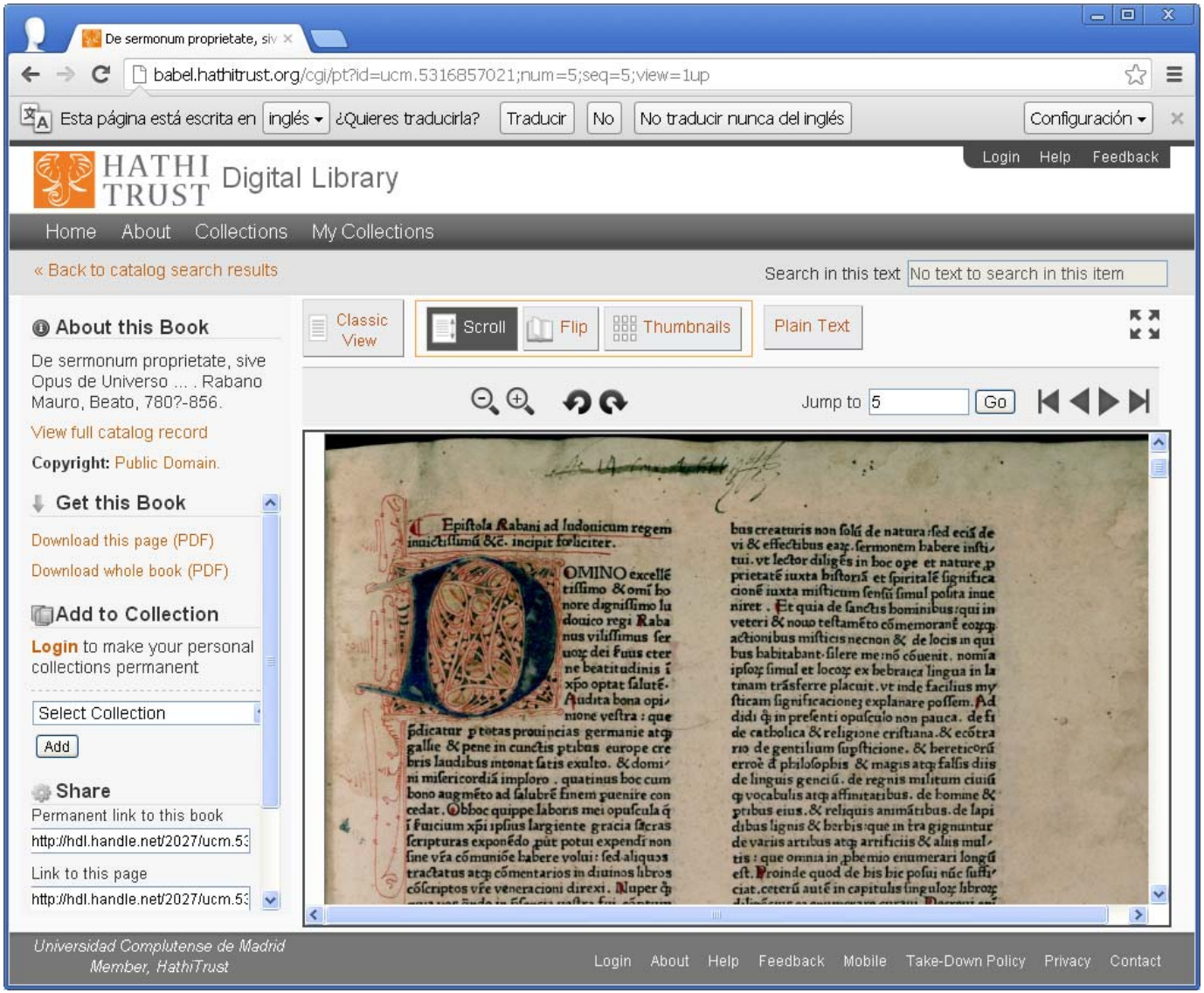
versidad
plutense
Madrid

How do you access the Complutense digitized books?

3. Catalogue of the Library of the Complutense University.



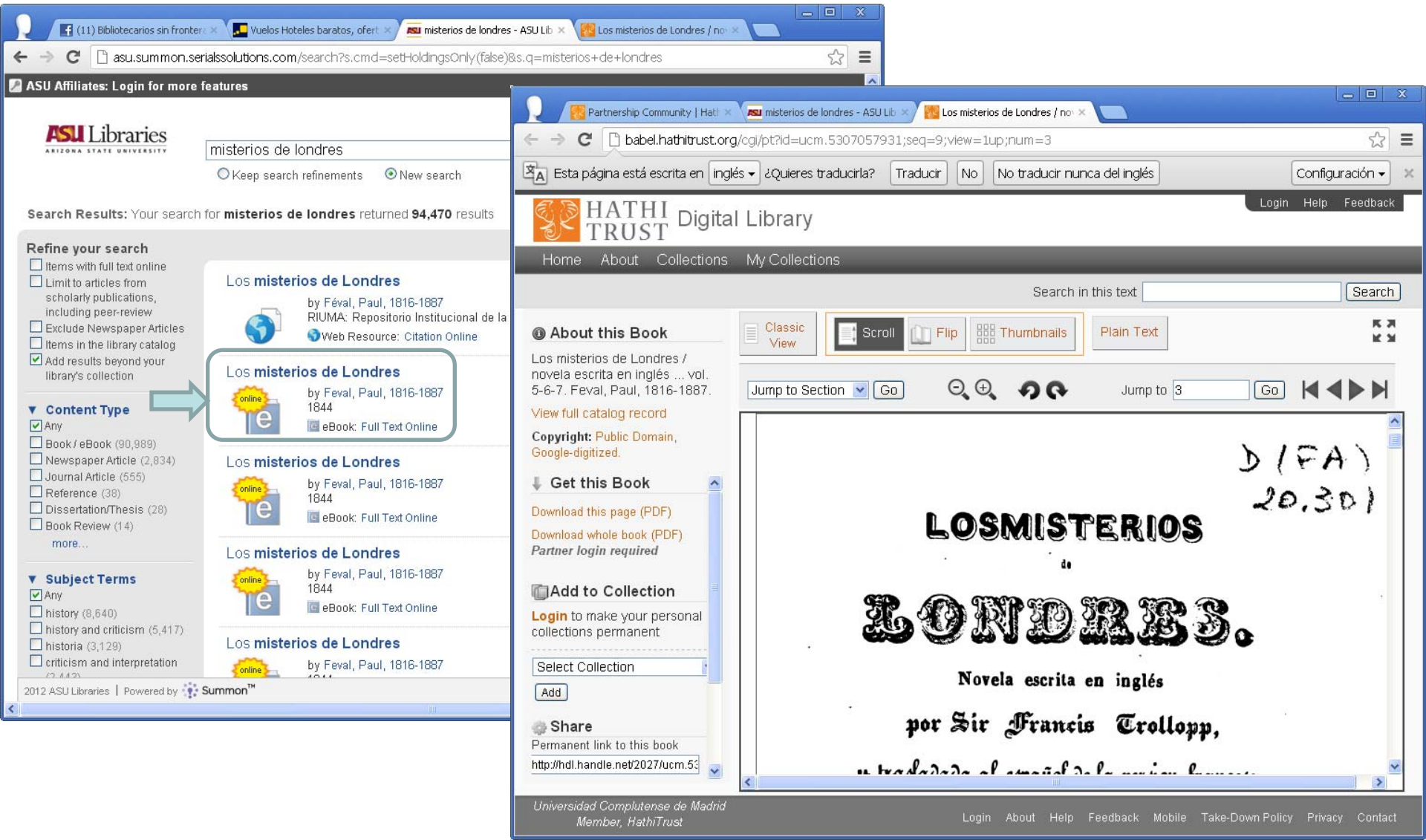
How do you access the Complutense digitized books?
4. HathiTrust Digital Library.



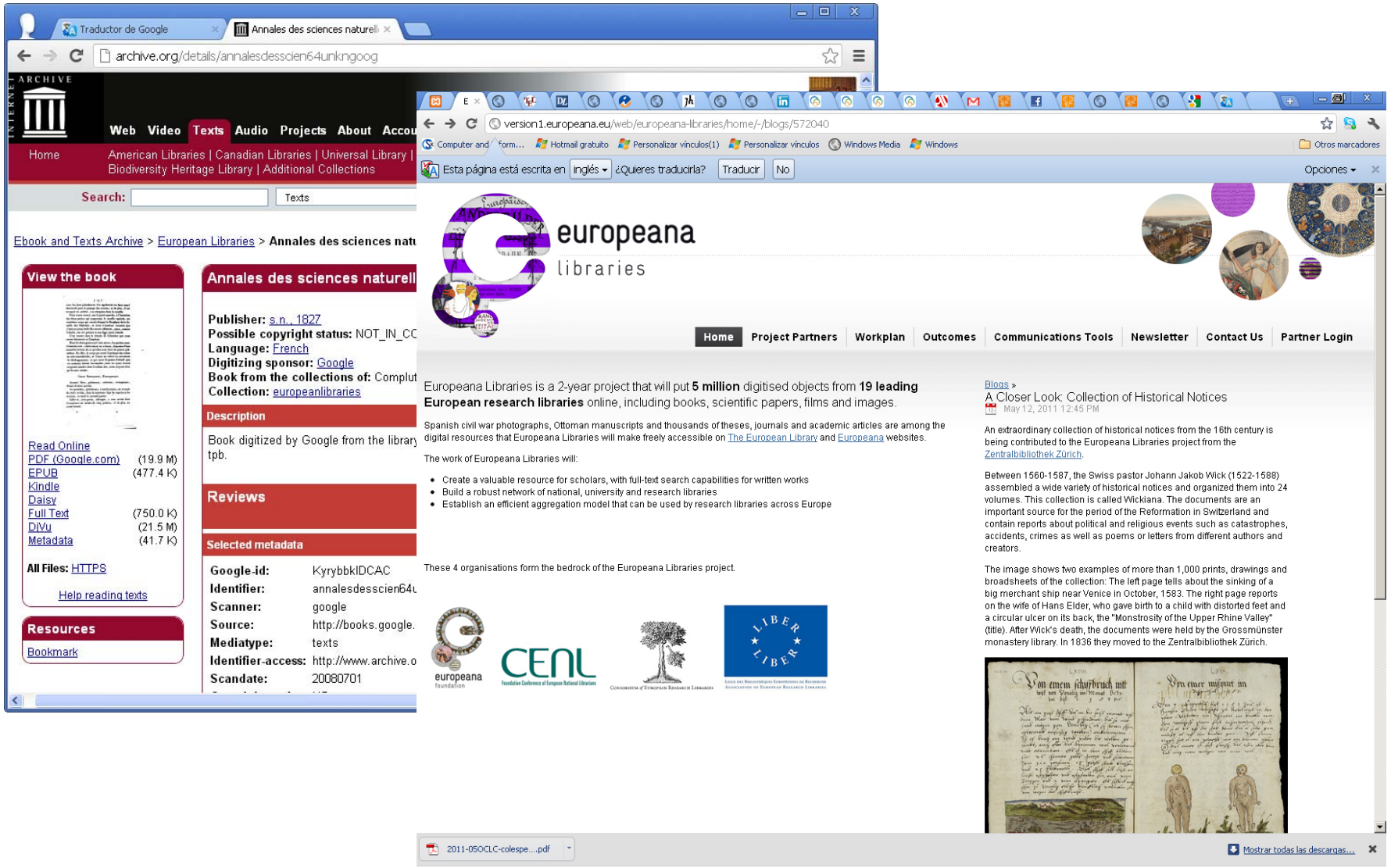
What's Hathi Trust?

- A library consortium to ensure that the cultural record is preserved and accessible long into the future.
- 10,557,655 total volumes digitized: 5,556,767 book titles and 274,642 serial titles
- 31% of total in the public domain
- Complutense joined HT in 2010. The only non-American partner
- Partners: Library of Congress, New York Public Library, California Digital Library and some academic libraries: Columbia, Cornell, Harvard, MIT, Princeton, Stanford, California, Chicago, Michigan, Yale…

How do you access the Complutense digitized books?
5. In your own catalogue if you have a discovery tool as Summon (only one click)



How do you access the Complutense digitized books?
6. More: Internet Archive, Europeana…



Europeana Libraries Project

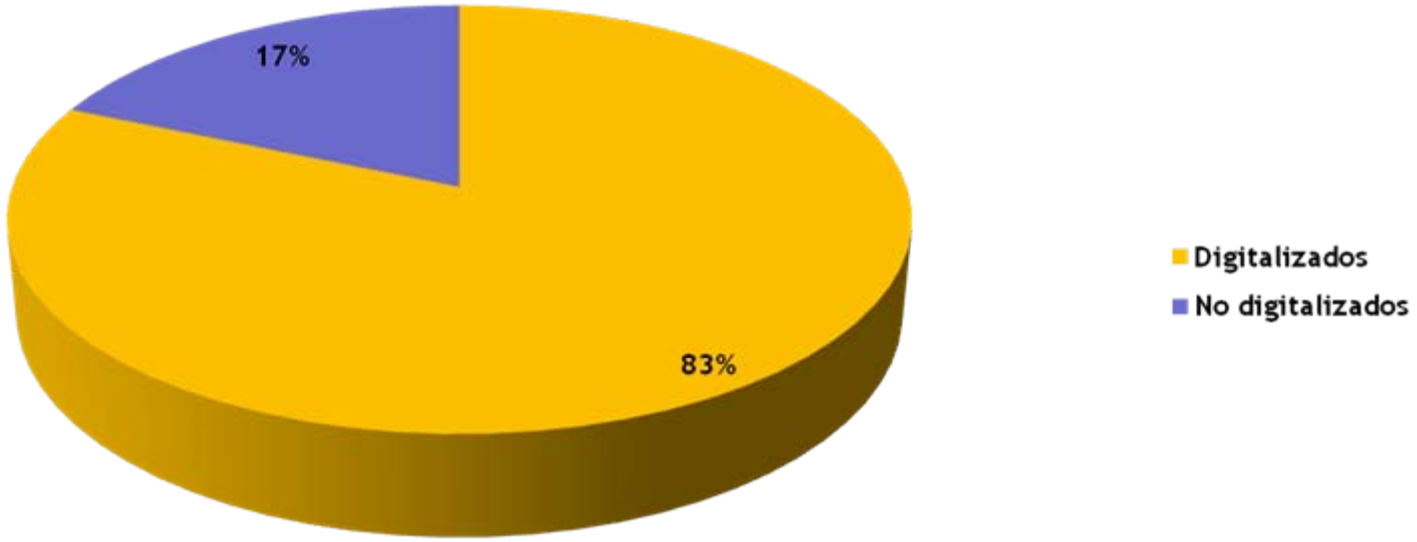
- Complutense collaborates with 18 research libraries from 14 countries: Bavarian State Library, Oxford University, Welcome Library, University College London, Ghent, Trinity College, etc.

- The objective is to incorporate 5 million digital objects: manuscripts, films and texts belonging to bibliographic and scientific heritage of the participating libraries to Europeana.

Universidad Complutense Madrid

Scanning process total data

Checked Books	Scanned Books	Not scanned Books	%
143,000	**120,000**	**23,000**	**17%**



Universidad Complutense Madrid

Access to Complutense Books in Google
Most visited books (one week)

- Every week more than 60 % of the scanned books are visited
- The most visited book of the Google european partners is from Complutense University.

Accesos	Título	Autor	Año	Centro
12.490	Diccionario etimológico de la lengua castellana (ensayo)	Pedro Felipe Monlau	1856	FLL-DER
12.008	Diccionario geográfico-estadístico de España y sus posesiones de ultramar	Pascual Madoz	1830	VET
8.637	La Ilíada	Homero	1788	FOA
8.275	Vida y viajes de Cristóbal Colón	Washington Irving	1852	GHI
7.520	Enciclopedia moderna	Francisco de Paula Mellado	1851	DER
7.027	Los tres reinos de la naturaleza o museo pintoresco de historia naturaleza: Botánica. Mineralogía	Georges-Luis Leclerc Buffon	1857-1858	MED-FOA
6.468	Diccionario de la lengua castellana	Real Academia Española	1852	FLL
4.450	Mitología universal	Juan Bautista Carrasco	1864	DER
4.205	Linajes nobles de España	Juan José Vilar Psayla	1867	FLL
4.179	Diccionario de agricultura práctica y economía rural	Agustín Esteban Collantes, Agustín Alfaro	1855	MED
4.035	Anatomie descriptive	Jean Cruveilhier	1837	MED
3.984	Anatomia do corpo humano	Bernardo Santucci	1739	FOA
3.671	Diccionario universal latino-español	Manuel de Valbuena	1808	FOA

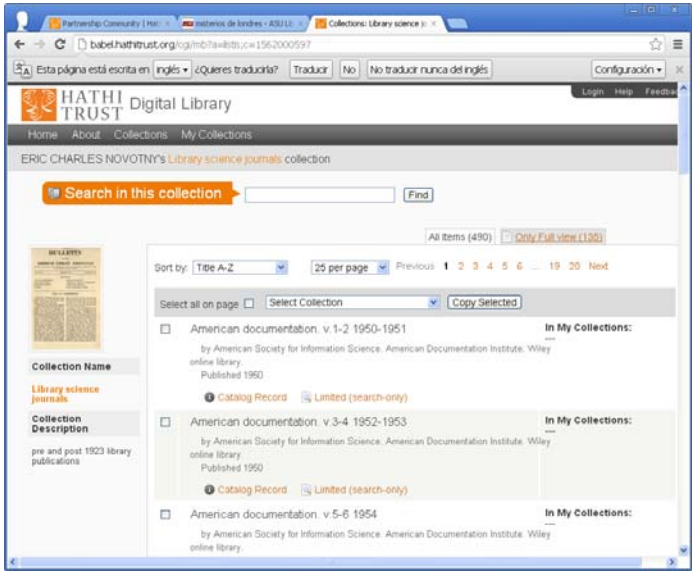
How do we preserve our digitized books? Hathi Trust

- Hathi Trust: Long term preservation (and dissemination).

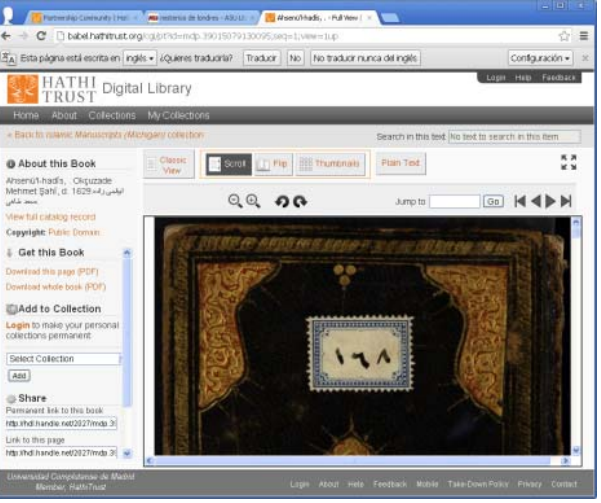
- For us, digital preservation could only be achieved as a result of cooperative involvement with other academic institutions following the standards of the international library community.

- What's Hathi Trust?
  - A repository for storing high quality
  - A scalable technological and organizational potential
  - A portal to access scanned books and journals

HathiTrust Characteristics:



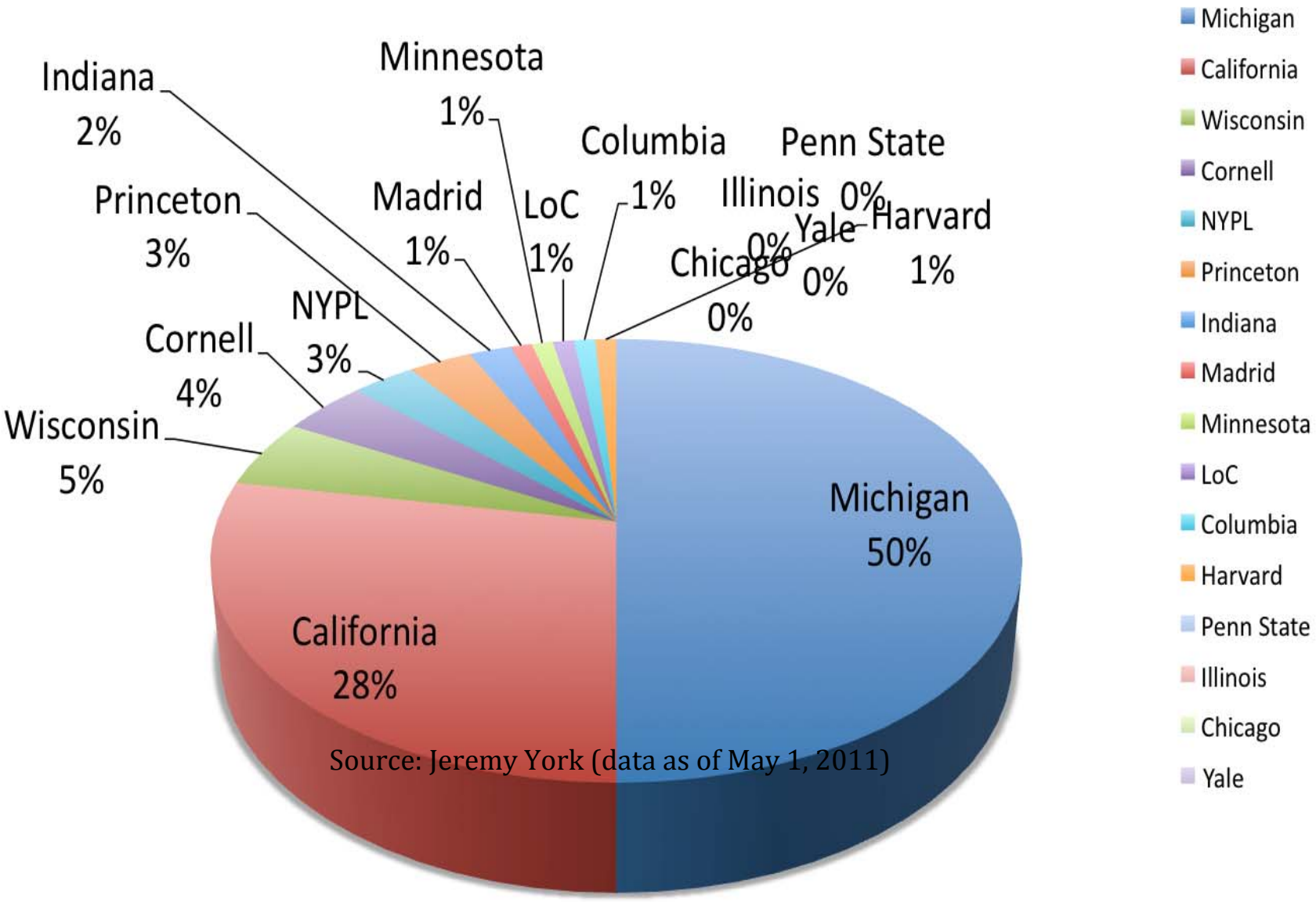
- Bibliographic and full-text search.
- Shibboleth authentication system.
- Bibliographic metadata are managed in a Library Management System (Aleph).
- Access to disabled users
- Offers access to bibliographic data via API for your catalog
- Bibliographic data (and access to scanned books) is included in discovery tools as *Summon*.
- You can download the books in the public domain (pdf, epub).
- Additionally, you can make collections, make them public and share with others.



Characteristics of HathiTrust

- Content Preservation: Image and text representation (if possible)
- Open code technologies: PERL,Linux, MySQL
- International Standards:
  - Trustworthy Repositories Audit & Certification (TRAC)
  - Open Archival Information System (OAIS) Reference Model
  - Preservation Metadata Implementation Strategies (PREMIS)
  - Image formats: TIFF, JPEG 2000
  - Permanent URL

HathiTrust Digital Library Collection



Source: Jeremy York (data as of May 1, 2011)

Conclusions: Objectives achieved in the project

- Scanning an important number of our ancient books (83%)
- Increasing the use of the collection by the general public.
- Supporting researchers offering digitized materials for text analysis.
- Increasing the visibility and long term preservation of our collections.
- Bringing on the library catalog all the books before the twentieth century (much of them in full text).
- Knowing the exact preservation condition of each book
- Establishing a plan of conservation and restoration of damaged books.

Universidad Complutense Madrid

Thank you for your attention!

José Antonio Magán Wals

Antonio Moreno Cañizares

Manuela Palafox Parejo

Complutense University of Madrid Library