



Mass-digitization at the Complutense University Library. Access and preservation of its bibliographic heritage

*José Antonio Magán, Manuela Palafox, Eugenio Tardón, Amelia Sanz.
Universidad Complutense de Madrid*

Barcelona, June 29, 2011

Overview

- Bibliographic Heritage Digitization in the UCM
- Preservation of the Complutense digital contents
- Collaborative Annotation of Digitized Literary Texts

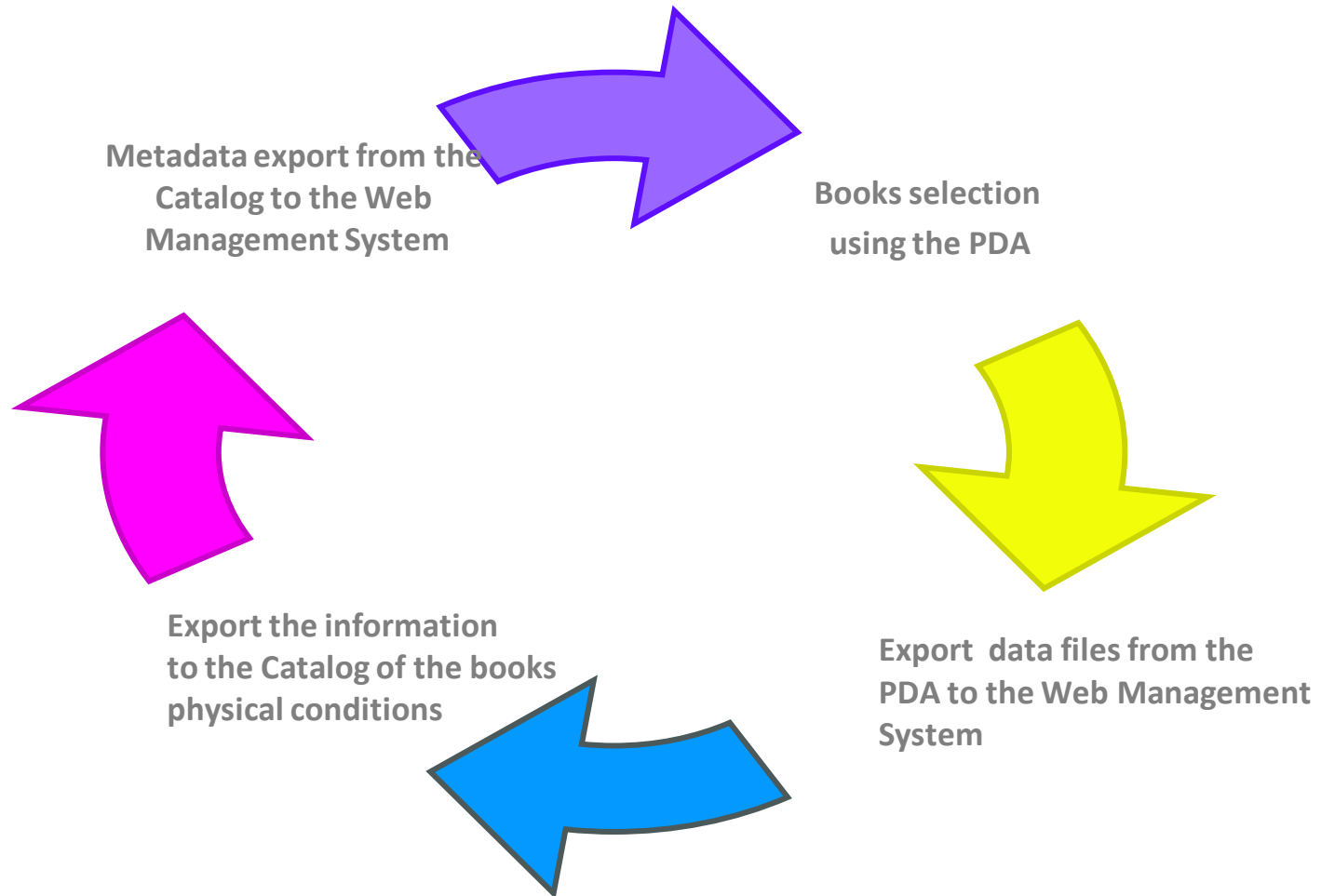
Goals of the digitization Project with Google

- Promoting access, dissemination and preservation of the bibliographic heritage
- Adding UCM holdings to those of some major libraries in order to facilitate the access of the global academic community to the scholarly knowledge
- Designing a plan for the conservation and restoration of damaged books
- Providing to UCM scholars a corpus of digital materials which makes possible to develop digital projects

UCM and Google partnership agreement

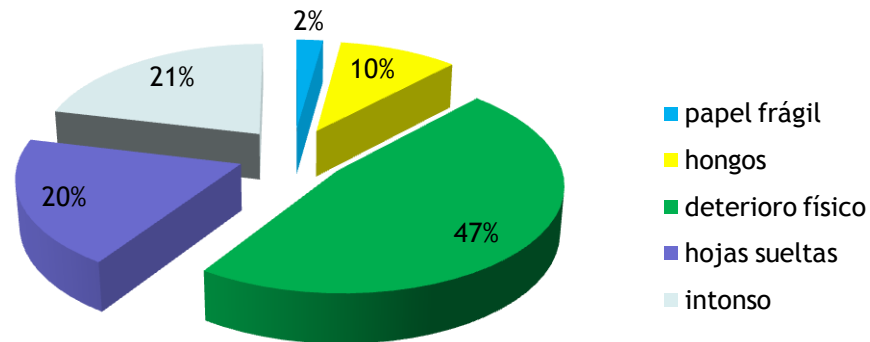
- Cataloging Plan: 220,000 volumes were cataloged, from the 16th to 19th centuries
- Criteria Selection Guidelines:
 - Date of publication: from 16th Century to 1870
 - Book physical conditions: dimensions, text block and binding
- Binding of the 19th century books
- Over 120,000 books digitized
- Dissemination of the bibliographic heritage of the Complutense University Library

Information flow



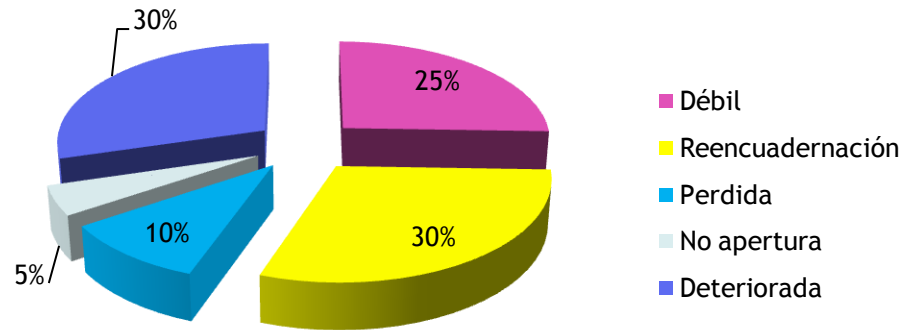
Physical conditions of the digitized books by Google

Non digitized for text-block



Physical conditions of the digitized books by Google

Non-digitized for the binding



Dissemination and access

The screenshot shows the 'Catálogo Cisne' of the Universidad Complutense de Madrid. The search results page displays the following information:

- Autor:** Álvarez, Francisco
- Título:** Breve tratado de relojes solares, y arquitectura / compuesto por Francisco Álvarez ...
- Publicación:** En Madrid : en la oficina de Lorenzo Francisco Mojados, 1727
- Lugar Impr.:** España Madrid

Below the metadata, there is a search bar for the book and a table of holdings:

UBICACIÓN	SIGNATURA	ESTADO	TIPO PRÉSTAMO	NOTA
B.Histórica-Foo. Guerra	BH FG 669	DISPONIBLE	PROTEGIDO ESP.	Olim: 10.03. Ingresó en la ... (CLICK PARA MAS)

At the bottom of the page, contact information for the library is provided: Biblioteca de la Universidad Complutense de Madrid, Ciudad Universitaria - 28040 Madrid, bucweb@buc.ucm.es, © UCM 2010.

Terminado

Mass-digitization at the Complutense University Library. Access an preservation of its bibliographic heritage. LIBER 2011. Barcelona 29/07/2011.

Preservation of the UCM digital contents

- For the implementation of a digital preservation program, 2 needs were identified:
 - A large-scale digital content repository
 - A scalable technological and organizational potential
- Conclusion:
 - The digital preservation could only be achieved as a result from the cooperative involvement of academic institutions following the international library community standards.
 - UCM joined HathiTrust on November 2010



HATHI
TRUST

HathiTrust Digital Library

The screenshot shows the HathiTrust Digital Library website in a Mozilla Firefox browser. The browser's address bar displays 'http://www.hathitrust.org/home'. The website header includes the HathiTrust logo and navigation links: Home, About, Public Collections, and My Collections. The main content area is divided into three columns: 'Catalog Search', 'Full-text Search', and 'Collections'. Each column contains a search form with a 'Find' button and a 'Full view only' checkbox. Below the search forms are three sections: 'Recent News and Publications' featuring a 'New!' badge for 'WorldCat Local catalog prototype!', 'Currently Digitized' with statistics on total volumes, book titles, serial titles, pages, and storage, and 'Featured Collection' for 'UM Press' (University of Michigan Press).

Catalog Search
Search information **about** the items.
[Search Box] Find
All Fields [v] Full view only
[Advanced Catalog Search](#) | [Search Tips](#)

Full-text Search
Search words that occur **within** the items.
[Search Box] Find
 Full view only
[Search Tips](#)

Collections
Browse, search, or make HathiTrust collections.
[View Public Collections]
Collections are a way to group items for private or public use. The full-text of items within a collection can be searched independently of the full digital library. [Read more about creating collections.](#)

Recent News and Publications
New! **Try our new WorldCat Local catalog prototype!**
[Tell us what you think](#) or [Read the FAQ](#)

- [2011 Mid-year Review](#)
- [U. of Michigan Library Opens Up Orphan Works](#)
- [Orphan Works Wager: University of Michigan Giving Scholars On-Campus Access](#)
- [University of Michigan to stop worrying about lawsuits, start releasing orphan works](#)
- [U-M Library to share HathiTrust orphan works](#)

[view more news >>](#)

Currently Digitized
8,825,583 total volumes
4,819,092 book titles
214,726 serial titles
3,088,954,050 pages
396 terabytes
104 miles
7,171 tons
2,409,526 volumes (~27% of total) in the public domain
[View visualizations of HathiTrust call numbers, languages, and dates](#)
[statistics information >>](#)

Featured Collection
UM Press
Univ. of Michigan Press's full-text titles available in HathiTrust

Mass-digitization at the Complutense University Library. Access an preservation of its bibliographic heritage. LIBER 2011. Barcelona 29/07/2011.



Books digitized by Google and in-house in HathiTrust

HathiTrust Digital Library - Curso Completo de Anatomía del Cuerpo Humano / por ... 1 - Mucilla Firefox

Home About Public Collections My Collections

Search in this text Find

Classic View Try our new views! Scroll Flip Thumbnails Plain Text

Jump to Section Go

Jump to Go

About this Book

Curso Completo de Anatomía del Cuerpo Humano / por ... 1

View full catalog record

Copyright: Public Domain, Google-digitized.

Get this Book

Download PDF - left page

Download PDF - right page

Download PDF - whole book

Add to Collection

Login to make your personal collections permanent

Select Collection Add

Share

Permanent link to this book
<http://hdl.handle.net/2027/ucm.532>

Link to this page
<http://hdl.handle.net/2027/ucm.532>

Digitized by Google Original from UNIVERSIDAD COMPLUTENSE DE MADRID

CURSO COMPLETO DE ANATOMÍA DEL CUERPO HUMANO

FOR

EL DOCTOR DON JAIME BONELLS,
Médico de Cámara de la Real Academia de Ciencias Médicas de Madrid, y de las Reales Academias de Ciencias naturales y Artes, y de Medicina práctica de Barcelona, y de la Real Sociedad de Medicina de París,

Y FOR

EL LIC. DON IGNACIO LACABA,
Cirujano de Cámara de S. M. y Catedrático de Anatomía del Real Colegio de Cirujía de San Carlos.

TOMO I.

OSTEOLOGÍA.

CON SUPERIOR FERRIS.
EN MADRID EN LA IMPRENTA DE SANCIANO DE S. SOC. SECT.
de Madrid en la Librería, calle del Lobo.

Digitized by Google Original from UNIVERSIDAD COMPLUTENSE DE MADRID

HathiTrust Digital Library

Home About Public Collections My Collections

Search in this text No text to search in this

Classic View Try our new views! Scroll Flip Thumbnails Plain Text

Jump to 20 Go

About this Book

De consolatione Philosophiae / Anicio Manlio Boecio. De disciplina ...

View full catalog record

Copyright: Public Domain.

Get this Book

Download PDF - this page

Download PDF - whole book

Add to Collection

Login to make your personal collections permanent

Select Collection Add

Share

Permanent link to this book
<http://hdl.handle.net/2027/ucm.531>

Link to this page
<http://hdl.handle.net/2027/ucm.531>

Digitized by Google Original from UNIVERSIDAD COMPLUTENSE DE MADRID

Liber.

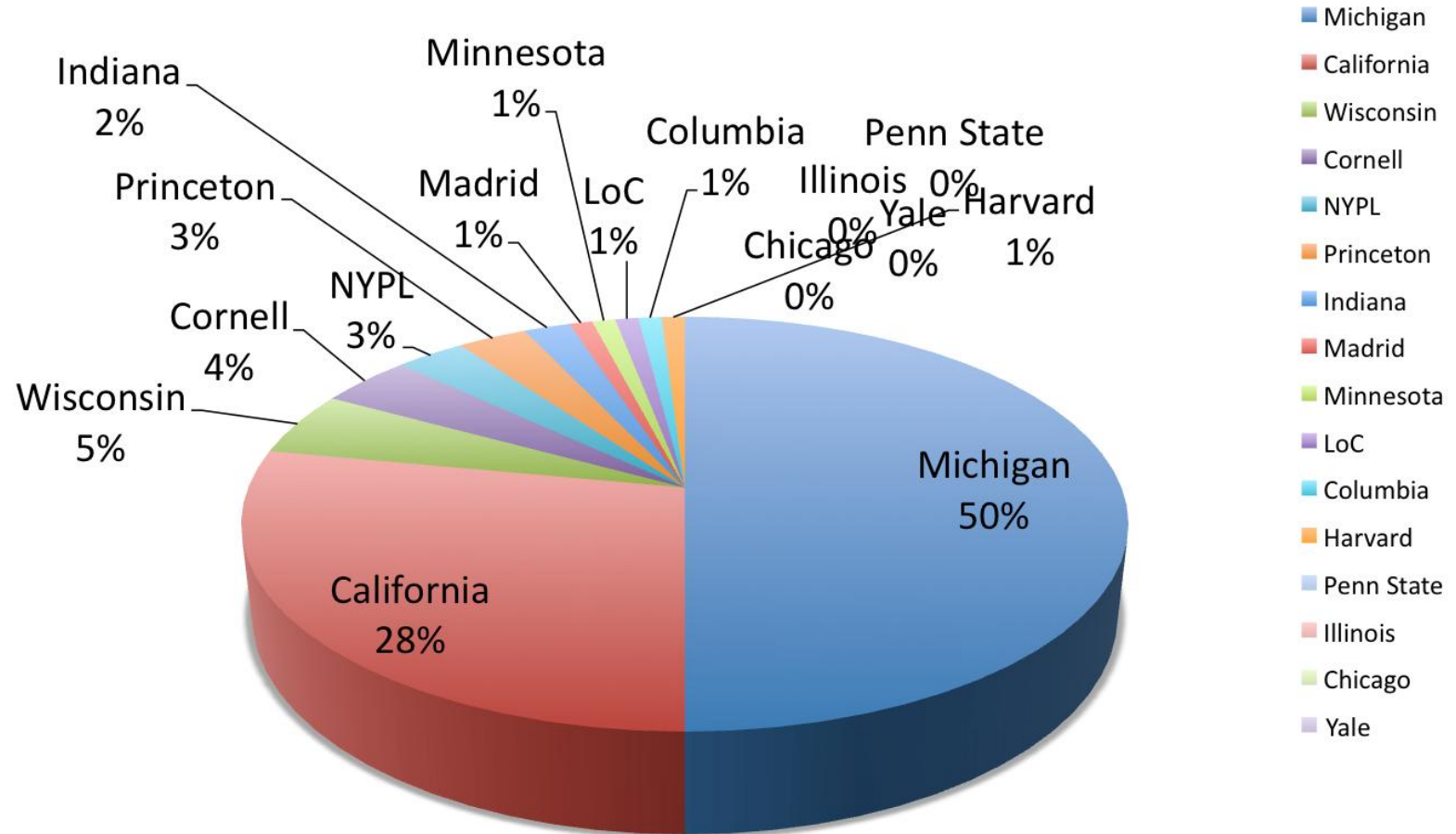
... quibus scientiæ vera fœlicitas et ebullitudo indebo ...
... quibus scientiæ vera fœlicitas et ebullitudo indebo ...
... quibus scientiæ vera fœlicitas et ebullitudo indebo ...

... quibus scientiæ vera fœlicitas et ebullitudo indebo ...
... quibus scientiæ vera fœlicitas et ebullitudo indebo ...
... quibus scientiæ vera fœlicitas et ebullitudo indebo ...

Mass-digitization at the Complutense University Library. Access an preservation of its bibliographic heritage. LIBER 2011. Barcelona 29/07/2011.



HathiTrust Digital Library



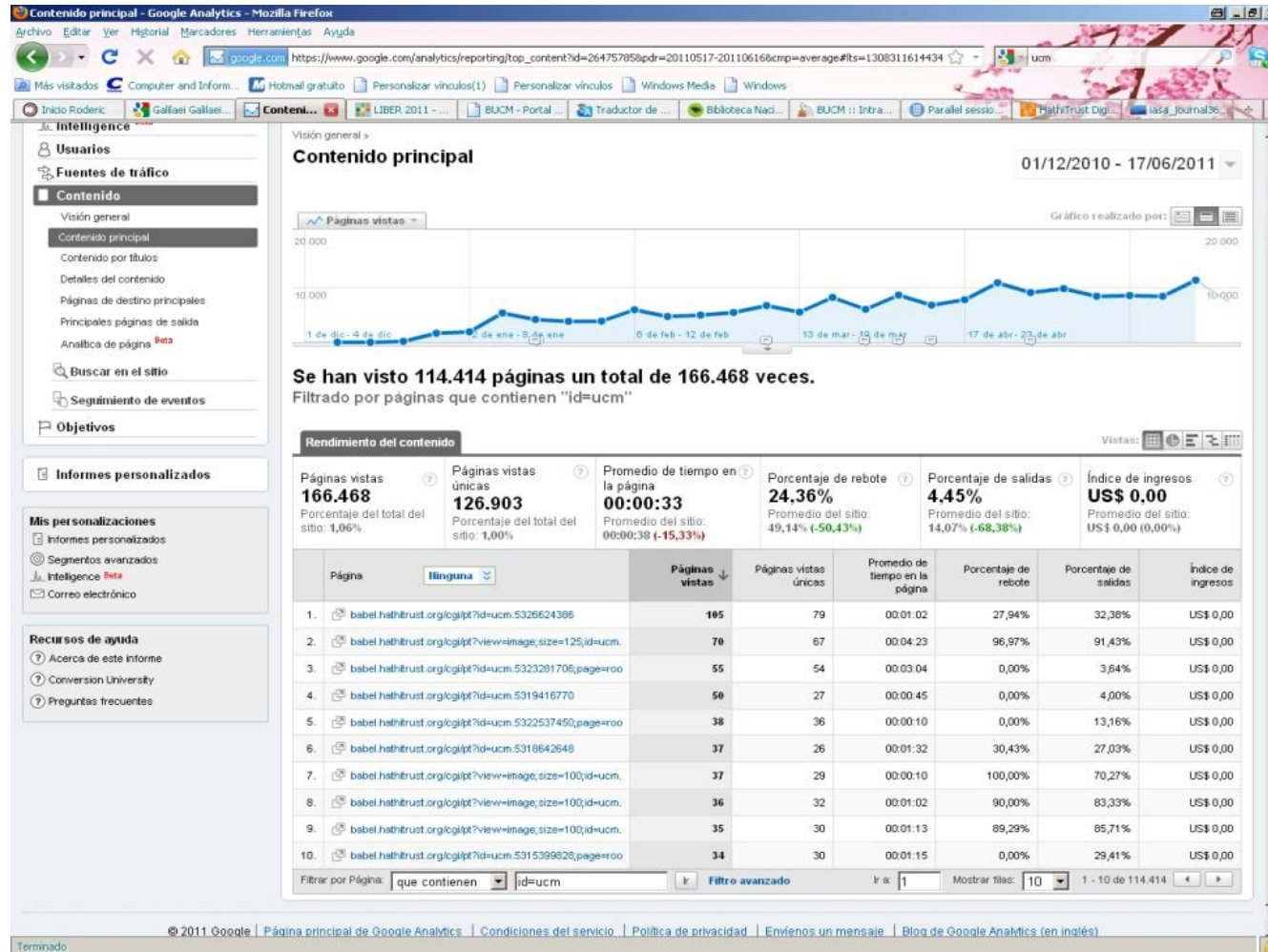
Source: Jeremy York (data as of May 1, 2011)

Mass-digitization at the Complutense University Library. Access an preservation of its bibliographic heritage. LIBER 2011. Barcelona 29/07/2011.

UCM Ingest

Date	Volumes	Total
December 2010	78,256	78,256
January 2011	1,156	79,412
February 2011	6,125	85,537
March	2,774	88,311
April	15,486	103,797
May	2,947	106,744

Visits of the UCM book pages in HathiTrust



Mass-digitization at the Complutense University Library. Access a preservation of its bibliographic heritage. LIBER 2011. Barcelona 29/07/2011.

Europeana Libraries

The screenshot shows a web browser window displaying the Europeana Libraries homepage. The browser's address bar shows the URL: version1.europeana.eu/web/europeana-libraries/home/-/blogs/572040. The page features the Europeana Libraries logo, a navigation menu with links for Home, Project Partners, Workplan, Outcomes, Communications Tools, Newsletter, Contact Us, and Partner Login. The main content area includes a paragraph about the project's goal to digitize 5 million objects from 19 leading European research libraries. Below this, there is a section titled 'The work of Europeana Libraries will:' followed by a bulleted list of three objectives: creating a valuable resource for scholars, building a robust network of national, university and research libraries, and establishing an efficient aggregation model. Further down, it states that four organizations form the bedrock of the project, with logos for Europeana Foundation, CENL (Foundation Conference of European National Libraries), Consortium of European Research Libraries, and LIBER (Association of European Research Libraries). On the right side, there is a 'Blogs' section with a post titled 'A Closer Look: Collection of Historical Notices' dated May 12, 2011. The post text describes an extraordinary collection of historical notices from the 16th century, known as Wickiana, assembled by Swiss pastor Johann Jakob Wick. It mentions that the collection contains reports on political and religious events, including catastrophes, accidents, and crimes. At the bottom of the post, there is an image showing two pages of a historical manuscript. The left page is a broadsheet with a landscape illustration at the bottom, and the right page features anatomical drawings of human figures. The browser's taskbar at the bottom shows a file named '2011-05OCLC-colespe....pdf' and a button to 'Mostrar todas las descargas...'. The browser window also displays various icons in the top bar and a search bar.

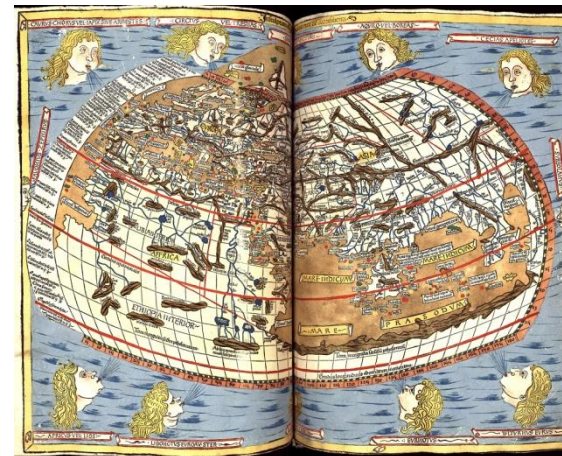
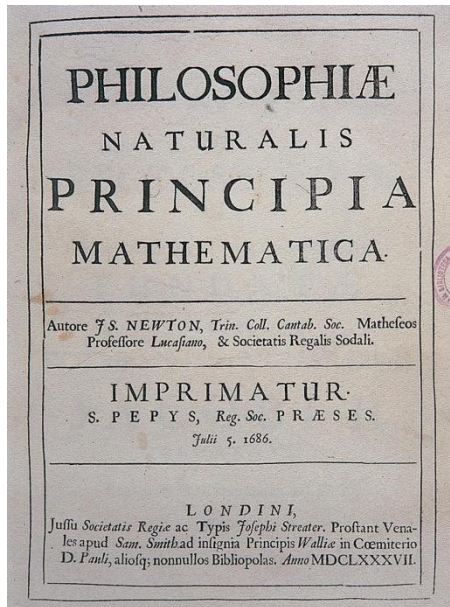
Mass-digitization at the Complutense University Library. Access a preservation of its bibliographic heritage. LIBER 2011. Barcelona 29/07/2011.

Europeana Libraries

- 1,200 film and video clips
- 850,000 images
- 4.3 million texts (books, journal articles, theses, letters)
- Complutense University Library will aggregate:
 - Complutense Rare Books from 16th Century to 1870: Metadata of more of 120,000 books digitized by Google.
 - Engravings of Dioscorides Collection: 50,000 engravings approximately.
 - Complutense University Theses: 6,000 theses.
 - Journal articles published by Complutense University Press: 31,000 articles
 - Fine Art Old Drawings from the Fine Arts Faculty (between 1752 and 1914): 287 drawings.
 - Photos of the Spanish Civil War from the Historical Archive of the Communist Party of Spain: 3,200 photographs approximately.
 - Personal Archive of Ruben Dario: 5,000 documents.
 - Complutense Manuscripts & Incunabula
 - Cartographic material, maps and city views



Europeana Data Model



Mass-digitization at the Complutense University Library. Access an preservation of its bibliographic heritage. LIBER 2011. Barcelona 29/07/2011.



Collaborative Annotation of Digitalized Literary Texts Project



ILSA

Grupo de Investigación en Ingeniería de Lenguajes Software y Aplicaciones

● Investigador responsable:

José Luis Sierra Rodríguez (jlsierra@fdi.ucm.es)

● Universidad:

Universidad Complutense de Madrid



● Líneas de Trabajo:

- Herramientas y Lenguajes Informáticos de Dominio Específico
- Humanidades Digitales
- Informática Educativa



Google research

WEDNESDAY, JULY 14, 2010

Our commitment to the digital humanities



Posted by Jon Orwant, Engineering Manager for Google Books, Magazines and Patents

(Cross-posted from the *Official Google Blog*)



50397 readers
BY FEEDBURNER

OTHER GOOGLE BLOGS

We've given awards to 12 projects led by 23 researchers at 15 universities:

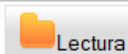
- Steven Abney and Terry Szymanski, University of Michigan. *Automatic Identification and Extraction of Structured Linguistic Passages in Texts.*
- Elton Barker, The Open University, Eric C. Kansa, University of California-Berkeley, Leif Isaksen, University of Southampton, United Kingdom. *Google Ancient Places (GAP): Discovering historic geographical entities in the Google Books corpus.*
- Dan Cohen and Fred Gibbs, George Mason University. *Reframing the Victorians.*
- Gregory R. Crane, Tufts University. *Classics in Google Books.*
- Miles Efron, Graduate School of Library and Information Science, University of Illinois. *Meeting the Challenge of Language Change in Text Retrieval with Machine Translation Techniques.*
- Brian Geiger, University of California-Riverside, Benjamin Pauley, Eastern Connecticut State University. *Early Modern Books Metadata in Google Books.*
- David Mimno and David Blei, Princeton University. *The Open Encyclopedia of Classical Sites.*
- Alfonso Moreno, Magdalen College, University of Oxford. *Bibliotheca Academica Translationum: link to Google Books.*
- Todd Presner, David Shepard, Chris Johanson, James Lee, University of California-Los Angeles. *Hypercities Geo-Scribe.*
- Amelia del Rosario Sanz-Cabrerizo and José Luis Sierra-Rodríguez, Universidad Complutense de Madrid. *Collaborative Annotation of Digitalized Literary Texts.*



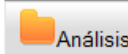
Universidad
Complutense
Madrid



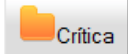
ALL



Lectura



Análisis



Crítica

All

Filter

Select All

Select None

ForgottenBooks.org

not able to light on any map or work giving the exact locality of the Dracula, as there are no maps of this country as yet to compare with the Ordnance Survey Maps; but I found that Bistritz, the post town by Count Dracula, is a fairly well-known place. I shall enter here from my notes, as they may refresh my memory when I talk over my notes with Mina.

In the population of Transylvania there are four distinct nationalities: the Wallachs in the South, and mixed with them the Wallachs, who are the remnants of the Dacians; Magyars in the West, and Szekelys in the East and North. I am going among the latter, who claim to be descended from Attila and the Huns. This may be so, for when the Magyars conquered the country in the eleventh century they found the Huns settled in it.

I read that every known superstition in the world is gathered into the horseshoe of the Carpathians, as if it were the centre of some sort of imaginative whirlpool; if so my stay may be very interesting. (Mem., I must ask the Count all about them.)

I did not sleep well, though my bed was comfortable enough, for I had all sorts of queer dreams. There was a dog howling all night under my window, which may have had something to do with it; or it may have been the paprika, for I had to drink up all the water in my carafe, and was still thirsty. Towards morning I slept and was wakened by the continuous knocking at my door, so I guess I must have been sleeping soundly then.

I had for breakfast more paprika, and a sort of porridge of maize flour which they said was "mamaliga", and egg-plant stuffed with forcemeat, a very excellent dish, which they call "impletata". (Mem., get recipe for this also.)

I had to hurry breakfast, for the train started a little before eight, or rather it ought to have done so, for after rushing to the station at 7:30 I had to sit in the carriage for more than an hour before we began to move.

It seems to me that the further east you go the more unpunctual are the trains. What ought they to be in China?

Más información s

Un párrafo introdu

Un párrafo introdu

Comentario satíric

Conclusions

- The large-scale digitization with Google will allow the Complutense University Library to face successfully to the new challenges are posing to research libraries
- Complutense digital collection of over 120,000 books is in HathiTrust, the most innovative digital preservation project for the library community.
- Complutense Library is collaborating with Europeana Libraries to increase the dissemination of our digitized heritage and participate in an innovative research area related to the semantic web and linked data.

Thank you very much!
Questions?

mpalafox@buc.ucm.es