“Google Libros” y la digitalización masiva: La aportación de la Universidad Complutense de Madrid

José Antonio MAGÁN WALS
Universidad Complutense de Madrid. Biblioteca
magan@ucm.es

Eugenio TARDÓN GONZÁLEZ
Universidad Complutense de Madrid. Biblioteca
tardon@ucm.es

Recibido: Marzo 2014
Aceptado: Mayo 2014

Resumen: Estudio del proyecto de digitalización masiva *Google Libros* que ha permitido escanear más de 20 millones de libros en todo el mundo (el 80% proveniente de las bibliotecas participantes y el resto de más de 50.000 editoriales que participan en el programa) y al que se sumó la *Biblioteca de la Universidad Complutense de Madrid* en 2006 con 120.574 libros que pueden consultarse hoy en *Google*, plataformas propias de la Universidad, *Europeana* y *Hathi Trust.* Ello ha mejorado sensiblemente la difusión y preservación de este patrimonio bibliográfico.

Palabras clave: *Google Libros*; digitalización masiva; *Hathi Trust; Europeana; fair use;* derechos de autor, preservación digital; bibliotecas digitales

“Google Books” and mass digitization: The contribution of the Complutense University of Madrid

Abstract: Study of mass digitization project that has enabled *Google Books* scan more than 20 million books worldwide (80% from participating libraries and the rest from more than 50,000 publishers participating in the program). The *Library of the Complutense University of Madrid* joined in 2006 and 120.574 books were scanned. These are now available in *Google*, UCM's catalogue and databases, *Europeana* and *Hathi Trust*. This has improved the dissemination and preservation of this bibliographic heritage.

Keywords: *Google Books;* mass digitization; *Hathi Trus;*, *Europeana;* copyright; fair use; digital preservation; digital libraries

1 INTRODUCCIÓN

Uno de los grandes anhelos del ser humano ha sido disponer de la totalidad de los libros publicados, poder interrogar el contenido de cada una de sus líneas y relacionarlo con el de otros libros de manera sencilla y rápida. Aunque las bibliotecas y algunas editoriales científicas habían realizado incipientes avances en esta línea, la llegada de *Google Libros* en 2004 acercó este sueño a la realidad de forma notable.

El proyecto *Google Libros* (anteriormente *Google Print*, *Google Book Search* y *Google Books*), en el que se involucraron multitud de editoriales y un grupo de bibliotecas que juntas conservan una parte sustancial del patrimonio bibliográfico mundial, estableció como objetivo digitalizar once millones de libros, permitir buscar en su contenido y su lectura y descarga en caso de estar libres de derechos de autor (Samuelson, 2010). A fecha de hoy esta meta se ha cumplido claramente: *Google* ha digitalizado sobradamente más de 20 millones de libros (la cifra se acerca en algunas fuentes a los treinta millones de libros en la actualidad) muchos de ellos provenientes de las bibliotecas que han participado en el programa (en torno al 80%), y una parte importante de los mismos, por encima del 20%, pueden ya ser leídos y descargados desde las páginas de Google y de las bibliotecas participantes al ser de dominio público.

Muchos vieron como una oportunidad histórica que no debía ser desaprovechada esta posibilidad de digitalización masiva del patrimonio bibliográfico pues permitiría consultar y leer los libros en formato digital. Sin embargo, el proceso ha sido controvertido al haberse digitalizado en Estados Unidos obras sujetas a *copyright* sin contar con el permiso de sus autores o editores, bajo el amparo del concepto jurídico anglosajón de "uso justo". Tras ello hubo un intento de acuerdo entre Google y las asociaciones estadounidenses de editores y autores que permitiría, entre otras cuestiones, la digitalización y acceso a millones de obras huérfanas hoy fuera del dominio público (Frosio, 2011; Durantaye, 2012) y una larga serie de litigios que, pese a la reciente sentencia a favor de Google, aún están por resolverse definitivamente.

Desde la perspectiva de *Google* y las bibliotecas que digitalizaron libros sujetos a *copyright* (Michigan y California), la digitalización se había realizado bajo las excepciones que la legislación americana otorga al *copyright* y que permiten que una obra pueda ser copiada en determinados casos al prevalecer el bien público sobre el derecho privado. Frente a esta opinión, hubo quienes veían en el proyecto un peligro de monopolización del conocimiento albergado en los libros por parte de *Google* y un uso abusivo del "*fair use*".

En 2014 el panorama ha cambiado significativamente: la cifra de libros digitalizados ha duplicado sobradamente el objetivo inicial, la lectura de los libros electrónicos se ha incorporado de lleno al hábito lector y los derechos de autor se han posicionado de forma rotunda en el epicentro del proyecto, cuestionándolo y

reformulándolo mediante sucesivos acuerdos de *Google* con los editores y las asociaciones de derechos de autor. Por otro lado, quienes veían el peligro que este proceso suponía al trasladar el conocimiento preservado en libros celosamente custodiados por las bibliotecas a una compañía privada que tras el proceso de digitalización podría realizar un monopolio del acceso digital a estos libros, constatan ahora cómo las bibliotecas colaboradoras con *Google* han creado con sus copias digitales las mayores colecciones públicas de libros digitalizados que existen en la actualidad (Codina, 2010). De hecho, proyectos que en su momento se crearon con el claro propósito de ofrecer una contrapartida pública (y desde Europa) a la iniciativa de *Google*, como *Europeana*, hoy se nutren mayoritariamente con los contenidos que *Google* digitaliza y, gracias a ello, han incrementado de forma notable su oferta digital inicial.

2 LOS INICIOS DEL PROCESO

En 1996 Larry Page y Sergey Brin desarrollaron una investigación apoyada por el *Stanford Digital Library Technologies Project* para recolectar, indexar y relacionar el contenido de libros digitales utilizando como criterio para establecer su relevancia el número y calidad de las citas establecidas entre los propios libros. Posteriormente, este algoritmo sería aplicado con éxito a páginas *web* siendo el origen de *Google*. En 2004 los cofundadores presentan *Google Print* en la Feria del Libro de Franfurkt (Brin, 2010), un programa de colaboración entre *Google* y gran número de editoriales relevantes y en diciembre se anuncia un acuerdo para digitalizar más de once millones de libros de cinco bibliotecas (*Biblioteca Pública de Nueva York* junto a las universidades de Harvard, Michigan, Oxford y Stanford).

Google Print estaba integrado por dos proyectos paralelos: El *Proyecto para Bibliotecas* basado en convenios para la digitalización de libros libres de derechos de autor o susceptibles de un *uso justo* "*fair use"* (en Estados Unidos), que representa el 80 % de lo digitalizado, y el *Programa de Afiliación para Editores y Autores* en donde estos deciden el tipo de visualización y distribución de sus obras (*Google*, 2014).

El objetivo inicial era indexar el contenido de los libros para introducirlos en el índice de *Google* y permitir su descarga, si eran de dominio público, o el hojeado y consulta de ciertas páginas o párrafos de los mismos en caso de estar sujetos a *copyright*. No obstante, amparándose en el concepto de "*fair use*", en las universidades de *Michigan* y *California* se digitalizaron libros que estaban aún bajo los derechos de *copyright* y que no habían sido publicados en Estados Unidos.

Pese a las críticas el proyecto siguió adelante. A las cinco primeras instituciones se suman en 2006 *California University* y la *Universidad Complutense de Madrid*, primer socio no anglosajón del proyecto. Tras estas instituciones se unieron otras destacando algunas catalanas (*Biblioteca de Catalunya*, *Ateneo* de Barcelona o el *Monasterio de Montserrat*).

A esta colaboración debemos añadir la asociación con miles de editoriales tanto generalistas (*Planeta*, *McGraw-Hill*, *Hachette*, *Penguin*, *House Mondadori*, *Hyperion*, *Giunti*…) como académicas (*Elsevier*, *Wiley*, *Taylor & Francis*, *Kluwer*, *Springer*, *Thomson* o *Blackwell*).

Desde sus orígenes *Google Books* ofrecía información muy novedosa respecto al libro, sus autores y sobre la obra. Desde pasajes similares en otros libros, las ciudades o lugares mencionados o un mapa conceptual de los términos que más veces aparecían en la obra. La posibilidad de "hojear" el libro, ver su portada e índice, buscar en su interior, descargarlo en formato *PDF*, *epub* o texto y crear "bibliotecas" propias, supuso una verdadera revolución respecto a la información bibliográfica que otros editores y las propias bibliotecas ofrecían en esos momentos (Dhawan, 2013).

3 CONTROVERSIAS ALREDEDOR DEL PROYECTO

Aunque el impacto del anuncio dejó descolocado a una parte importante del sector del libro y las bibliotecas pronto empezaron a sonar voces críticas al proyecto, especialmente desde el ámbito de la industria editorial y las asociaciones de autores, denunciando que obras pertenecientes a su fondo editorial o de sus asociados habían sido digitalizadas sin permiso previo (Esteve, 2010), ante lo cual *Google* y sus socios bibliotecarios argumentaron el *uso justo* de los trabajos (Ji, 2011).

De forma simultánea algunos intelectuales criticaron la falta de calidad de las digitalizaciones realizadas por *Google*, las carencias de los programas de *OCR*, la pobreza de los metadatos utilizados (James; Weiss, 2012) y la inconsistencia de los resultados de las búsquedas (Nunberg, 2009). Además, manifestaban el riesgo de monopolización del acceso al contenido de los libros por parte de una empresa comercial pese al hecho de que ofreciese sus servicios de modo gratuito. También se alertó, desde ciertos países de Europa liderados por la *Biblioteca Nacional de Francia*, del peligro ante una americanización de la oferta digital y la necesidad de que fuesen iniciativas financiadas desde la administración quienes digitalizaran el patrimonio bibliográfico europeo (Jeanneney, 2004).

Frente a estas voces, otras expresaron con contundencia que *Google Búsqueda de Libros* era una herramienta gratuita que permitía fácilmente consultar y descargar los contenidos, representando una oportunidad única para democratizar el conocimiento mediante la digitalización de millones de libros que hasta entonces sólo eran accesibles para investigadores debidamente acreditados en las propias instalaciones de las bibliotecas que los alojaban y que la calidad de los metadatos, el OCR y las imágenes era adecuada.

Independientemente de la controversia que el proyecto ha levantado (Nguyen, 2011**;** Raff, 2011), es incuestionable que sirvió como revulsivo para la aparición de otros programas de digitalización masiva tanto públicos como privados (entre los que debemos destacar *Microsoft Live Search Books* que, tras digitalizar casi

750.000 libros, desaparece en 2.008 aunque sus libros son recuperables desde el *Internet Archive*) y que, gracias a él, el buscador general de *Google* realiza ahora las búsquedas no sólo contra el contenido de millones de páginas web, sino en millones de libros. Por otro lado, otros servicios de *Google*, especialmente su traductor, han mejorado sensiblemente en base al *corpus* de las obras escaneadas.

Finalmente la sentencia dictada tras largos años de incertidumbre da la razón a *Google* y corrobora el *uso justo* de las digitalizaciones de las obras en base a realizar un uso transformativo y no sustitutivo de los libros y la importancia y variedad de los beneficios públicos derivados del proyecto (Garriga, 2014).

4 SITUACIÓN ACTUAL

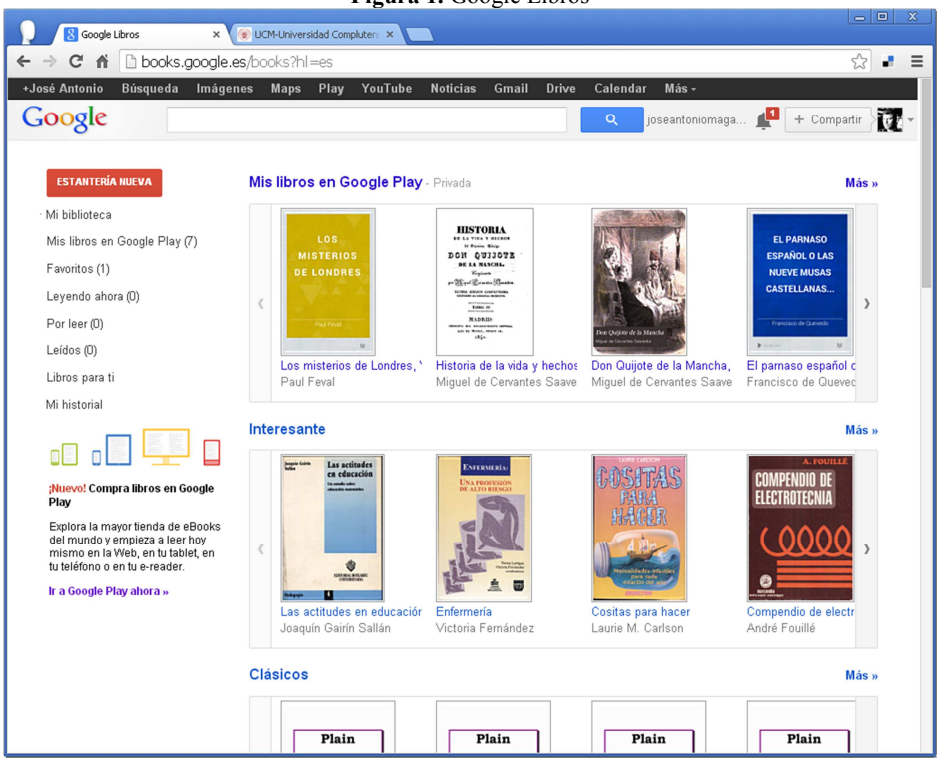
Hoy más de 50.000 editoriales tienen acuerdos de distribución de sus obras con *Google Libros* lo que ha permitido dar una nueva vida a su catálogo editorial, muchas veces inaccesible en las librerías dada la incesante necesidad de cambiar la oferta de libros en venta debido al empuje imparable de las novedades, e incrementar su visibilidad. A ello se une la información estadística sobre las búsquedas y los intereses de los lectores que *Google* les ofrece y que permiten mejorar su oferta de novedades y reediciones.

El *Proyecto para Bibliotecas* cuenta con un número notable provenientes de tres continentes (América, Europa y Asia). Sólo en Europa han sido digitalizadas más de dos millones de obras y, pese a las reticencias iniciales, numerosas bibliotecas nacionales se han ido sumando al proyecto: *Biblioteca Británica*, *Biblioteca Estatal de Baviera* (la institución que alberga el mayor número de incunables a nivel mundial), Holanda, Austria, Hungría, Chequia e Italia. A ellas se unen la *Biblioteca Municipal de Lion* (depositaria de una de las principales colecciones de libros del siglo XIX de Francia) y universitarias como las de Gante o *Sapienza* de Roma.

El proceso de digitalización de millones de libros se ha realizado sin incidentes reseñables y desde la perspectiva de las bibliotecas el éxito es claro: para ellas sería impensable la meta alcanzada de que se visiten al menos una vez cada seis meses más del 90 % de los más de 20 millones de libros escaneados. Sin duda el mayor logro del proyecto es hacer llegar el patrimonio bibliográfico preservado por las bibliotecas durante generaciones al gran público que lo consulta y accede de forma sencilla y cómoda más allá de las instalaciones de las bibliotecas depositarias.

Y hoy las personas que hacen una búsqueda en *Google* desconocen que buscan no sólo en páginas web, sino también en los libros y revistas de algunas de las más prestigiosas bibliotecas y editoriales mundiales, lo que enriquece sustancialmente el resultado de sus consultas. Durante años se alertó de que las búsquedas en Internet carecían de la calidad que los contenidos de las bibliotecas podían aportar. Ahora estas búsquedas aúnan lo mejor de Internet con el saber milenario albergado en los libros que las bibliotecas han preservado para nosotros.

Figura 1. Google Libros



5 LAS COLECCIONES DIGITALES COMPLUTENSES

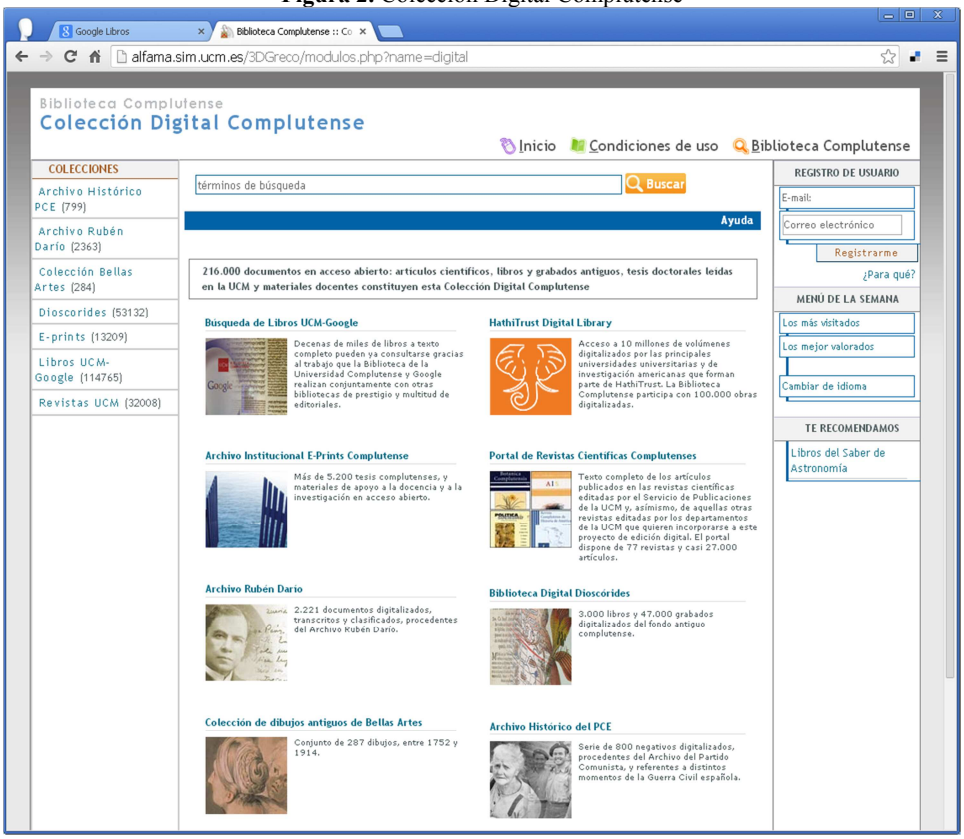
La *Biblioteca de la Universidad Complutense de Madrid* ha creado desde 1995 la mayor colección de libros digitalizados de España con el objetivo de facilitar el acceso y preservar a largo plazo el conocimiento generado por la Universidad y su patrimonio bibliográfico.

En la actualidad se han digitalizado más de 160.000 libros (entre fondo antiguo y tesis) y, además del programa de digitalización propio en colaboración con *Santander Universidades*, existen otros dos grandes proyectos: con *Google* para digitalizar en torno a 20.000 libros libres de derechos de autor posteriores a 1870 y con *ProQuest* para digitalizar 3.000 tesis doctorales de la Universidad. A esta se unen otras colecciones digitales: la de revistas académicas publicadas por la Universidad, con más de 38.000 artículos, la de Prensa Digital de la Facultad de Ciencias de la Información, con 450.000 periódicos escaneados (en acceso restringido desde la propia biblioteca para fines de investigación), 49.742 grabados pertenecientes a la *Biblioteca Digital Dioscórides*, 13.178 documentos depositados en el *Archivo Institucional E-Prints UCM*, parte del archivo fotográfico del

Partido Comunista de España, el *Archivo Rubén Darío*, la colección de grabados japoneses o los *Dibujos de Academia* de la Facultad de Bellas Artes.

Estas digitalizaciones han sido posibles gracias a la colaboración con distintas instituciones tanto públicas (*Comunidad de Madrid* y *Ministerio de Cultura*) como privadas (*Google*, *Santander Universidades*, *Fundación de Ciencias de la Salud*, *Editorial Extramuros*, *ProQuest*). A ello se une una apuesta decidida para difundir el patrimonio bibliográfico de la UCM e incrementar la visibilidad de su producción científica y académica mediante políticas de acceso abierto, la participación en multitud de proyectos cooperativos digitales y la colaboración con algunas editoriales científicas de reconocido prestigio para la publicación en sus plataformas comerciales de las revistas editadas por la Universidad y de otros materiales académicos (*Gale*, *ProQuest, E-Libro* y *Océano*).

Figura 2. Colección Digital Complutense



6 EL ACUERDO DE LA *UNIVERSIDAD COMPLUTENSE DE MADRID* CON *GOOGLE* PARA DIGITALIZAR OBRAS DE DOMINIO PÚBLICO

En 2006, fecha de la firma del acuerdo de colaboración con *Google* para digitalizar las obras en dominio público de la UCM, la *Biblioteca Complutense* tenía la mayor colección española de libros antiguos digitalizados, la *Biblioteca Digital Dioscórides*, gracias a la colaboración con la *Fundación de Ciencias de la Salud*. Esta colección contaba con casi 2.800 libros y 47.000 grabados digitalizados desde 1995.

No obstante la Universidad estaba muy lejos de digitalizar en un tiempo razonable su colección de fondo antiguo. A este ritmo de digitalización se hubiesen necesitado 435 años para escanear las obras posteriormente digitalizadas con *Google* en tan sólo tres años. A ello se unían deficiencias en la plataforma empleada para difundir esta colección: inexistencia de una interfaz multilingüe, gestión inadecuada de los derechos de autor, falta de adaptación a las nuevas necesidades y demandas de los usuarios (especialmente las relativas a la *web* social) y carencia de herramientas adecuadas para la preservación digital a largo plazo.

Por todo ello, cuando *Google Books* se hizo público en 2004, la Biblioteca Complutense mostró interés en sumarse al proyecto y, gracias al decidido apoyo del rector y presidente de REBIUN Carlos Berzosa y de todo el ámbito universitario, firmó en 2006 un acuerdo para digitalizar y difundir las obras libres de derechos de autor de la *UCM*. *Google* se comprometía a escanear los documentos asumiendo los costes del traslado al centro de escaneado, su digitalización, proceso, difusión y preservación. Los libros se digitalizaron dos veces para reducir errores en un centro de escaneado ubicado en Madrid. Este centro era de pequeño tamaño para la escala de digitalización de *Google* que llegó a escanear más de 4.000 libros diarios en los grandes centros de digitalización de Estados Unidos.

Además de incluir estos libros en el índice general de *Google*, en *Google Libros* y en *Google Play*, la empresa creó una interfaz exclusiva para la Complutense en donde se pueden realizar consultas contra todos los libros del proyecto o sólo los de la *UCM* (http://biblioteca.ucm.es/atencion/25403.php). Esta interfaz de acceso libre y sin coste para la Universidad ha permitido contar con una pasarela multilingüe, adaptada a los desarrollos tecnológicos y que controla debidamente los derechos de autor a nivel internacional.

De cada libro escaneado *Google* crea dos copias: una para *Google* y otra para la *UCM*. El acuerdo incluye expresamente que una parte de los libros digitalizados pueden ser empleados en proyectos conjuntos con la *Biblioteca Nacional de España*, el consorcio de bibliotecas *Madroño* y el *Catálogo Colectivo del Patrimonio Bibliográfico Español* y, aunque existe una limitación para emplear con intenciones comerciales los libros por parte de la Universidad sin conocimiento previo de *Google*, la colección completa está disponible en *Europeana* y *Hathi Trust*.

Por su parte, la *Universidad Complutense* además de ofrecer su fondo anterior a 1870 creó los equipos de bibliotecarios que garantizaron la selección y préstamo de las obras a digitalizar, de acuerdo con los criterios de preservación y características para el préstamo establecidos por el equipo de su *Biblioteca Histórica Marqués de Valdecilla*, dirigido por Marta Torres. Entre 2008 y 2011 fueron analizados para corroborar si su tamaño, rareza o condiciones de preservación permitían la digitalización 164.875 ejemplares y 120.574 fueron finalmente escaneados (Magán, Palafox, Tardón y Sanz, 2011). En junio de 2011 finalizaron las operaciones de escaneado en Madrid y en 2013 se han retomado las operaciones para fondos posteriores a 1870 libres de derechos de autor, estando previsto digitalizar otros 20.000 libros hasta octubre de 2014.

7 IMPLICACIONES DEL PROYECTO PARA LA GESTIÓN Y DIFUSIÓN DEL PATRIMONIO DE LA *UCM*

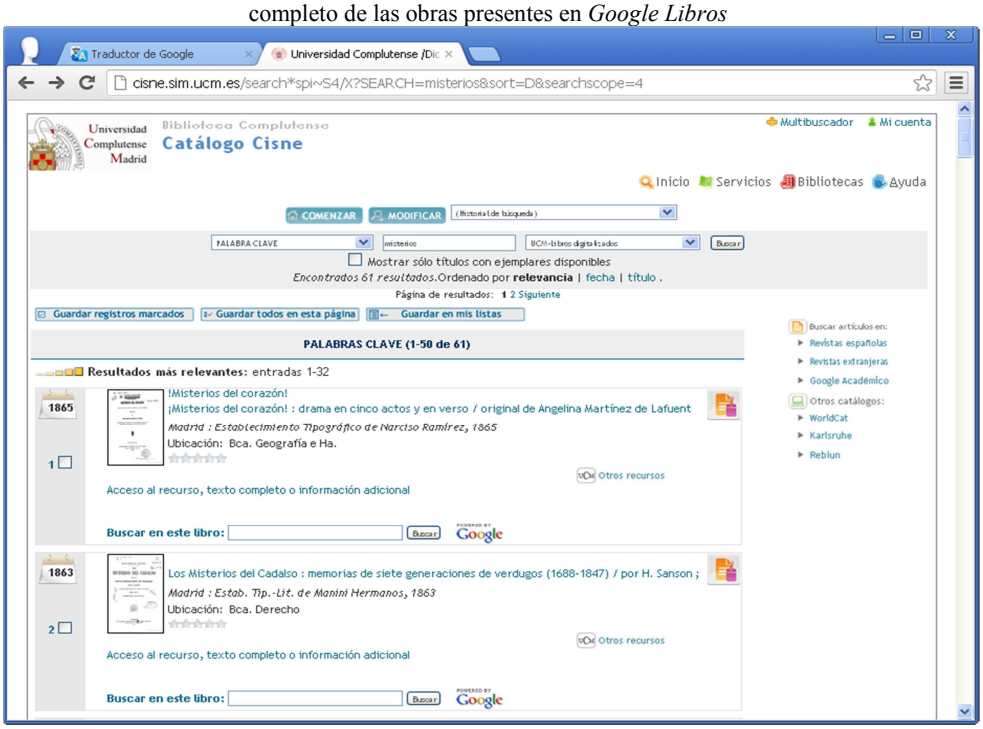
Además de su carga tecnológica, el acuerdo supuso reforzar significativamente los recursos dedicados a la gestión del patrimonio bibliográfico y paliar algunas de sus carencias más acuciantes. Se realizó un análisis detallado de la colección y se establecieron unas políticas de encuadernación y selección de los materiales a escanear, además de recomendaciones relativas a su tratamiento y manuales de procedimiento de los flujos de trabajo y procesos a realizar.

También se realizaron labores de acondicionamiento y limpieza de los depósitos, las estanterías y las obras mismas. Se creó una base de datos en donde se recogió información relativa al estado de preservación de cada ejemplar que ha permitido contar con una información preciosa de los ejemplares valiosos y de aquellos que necesitaban urgentemente actuaciones para evitar su deterioro. Las páginas de cientos de libros intonsos fueron abiertas para permitir su digitalización y se realizaron múltiples restauraciones en la encuadernación y el cuerpo de libros que, sin el aliciente de la digitalización, no se hubiesen producido. Pero la labor más destacable fue la inclusión en el catálogo automatizado de la biblioteca de una parte importantísima del fondo antiguo que aún estaba por introducir (220.000 ejemplares anteriores al siglo XX).

Para este proyecto el equipo de informática de la Biblioteca a cargo de Zacarías Martín Maté desarrolló dos aplicaciones: una para la selección de las obras mediante *PDAs* en los depósitos y otra para la gestión en tiempo real de los envíos a *Google* que permitió introducir en el catálogo metadatos relativos a las condiciones de conservación y encuadernación de cada ejemplar. Este programa contrasta diariamente la información de nuestro catálogo con el *Google Return Interface* (*GRIN*) e incorpora automáticamente, en caso de que un libro de la *UCM* esté en *Google Libros*, una etiqueta 856 que enlaza a la copia digital del libro. Además, se utiliza una *API* de *Google* que permite introducir en la descripción del

libro del catálogo una caja de búsqueda para consultar en el texto completo de cada obra específica.

Figura 3. Caja de búsqueda en el *Catálogo Cisne* de la UCM que permite la búsqueda en el texto completo de las obras presentes en *Google Libros*



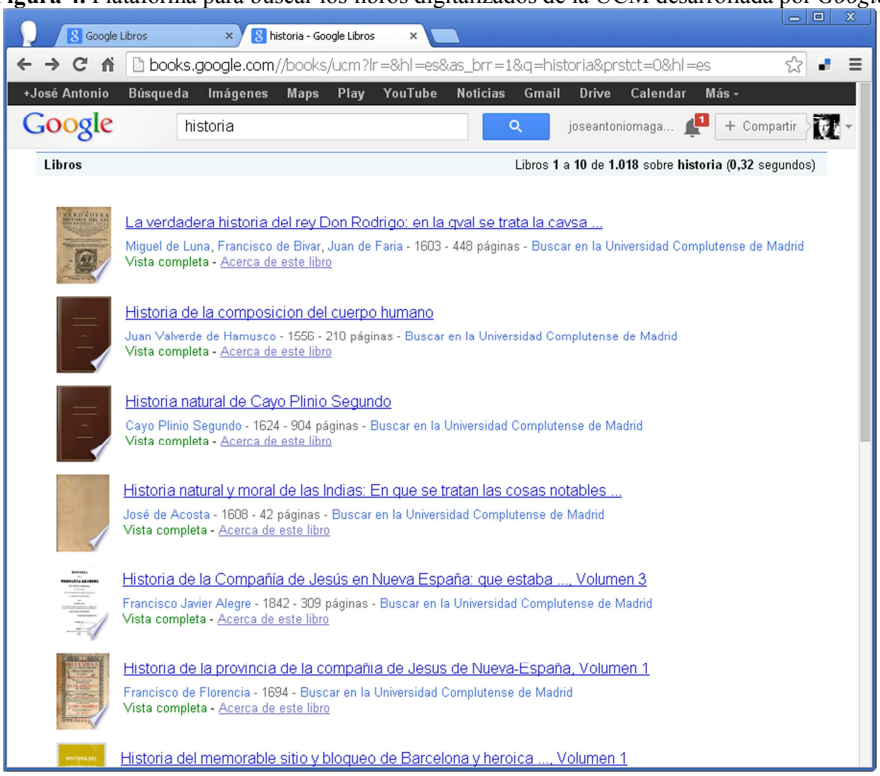
En definitiva, el proyecto de digitalización masiva con Google supuso una reactivación de labores que hasta entonces no habían podido ser acometidas adecuadamente. Hoy, no sólo se ha digitalizado el 76.14 % de los ejemplares anteriores a 1870, sino que sus condiciones de conservación han mejorado sustancialmente y se cuenta en el catálogo con datos relativos al estado de conservación y características físicas de cada ejemplar que permiten una gestión de la colección del fondo antiguo más adecuada.

8 GARANTÍA DE ACCESO PÚBLICO Y PRESERVACIÓN A LARGO PLAZO DE LAS OBRAS DIGITALIZADAS POR *GOOGLE*

Una de las polémicas del proceso de digitalización masiva de *Google* ha sido la duda respecto a la garantía de acceso público a las obras digitalizadas. Hoy los libros complutenses son accesibles libremente desde diferentes pasarelas. La *UCM* garantiza el acceso a las obras en sus catálogos, la aplicación propia "Colección

Digital Complutense" y la plataforma creada por *Google* para la *UCM*. También están presentes en *Hathi Trust*, la principal biblioteca pública digital mundial, y en *Europeana*, el mayor portal de objetos digitales europeo donde, pese a las reticencias que algunos mantenían, los libros de *Google* constituyen una parte importantísima del fondo antiguo incluido. Además varios miles están en el *Internet Text Archive*, lo que convierte a la UCM en la institución española con mayor presencia en este archivo abierto.

Figura 4. Plataforma para buscar los libros digitalizados de la UCM desarrollada por *Google*



No obstante el mayor medio de difusión son los productos de *Google*: *Google Libros*, *Google Play, Google Académico* y el propio buscador general. Este último es especialmente relevante por ser multilingüe, estar en constante evolución tecnológica y, sobre todo, haber introducido el contenido de los libros en la herramienta de búsqueda de información más generalizada entre los investigadores y el público en general.

Para el alojamiento físico de los objetos digitales, su preservación a largo plazo y difusión, la *UCM* ha optado por *Hathi Trust*, siendo el único miembro no norteamericano. Frente a otras "bibliotecas digitales" que son simples recolectores de

metadatos, *Hathi Trust* aloja y gestiona físicamente los objetos digitales, contando con 3.898 millones de páginas pertenecientes a 5.802.345 libros y 290.952 publicaciones periódicas, de las que el 34% están en el dominio público.

Hathi Trust está integrado por más de 80 instituciones de gran prestigio en el ámbito de la digitalización en Estados Unidos más la Universidad Complutense. Entre sus miembros se encuentran la *Library of Congress*, la Biblioteca Pública de Nueva York, la *California Digital Library* y universidades como Columbia, Cornell, Harvard, MIT, Princeton, Stanford, California, Chicago, Michigan o Yale. *Hathi Trust* mantiene un doble papel: como repositorio seguro para la preservación a largo plazo y como servicio público que garantiza el acceso a los objetos depositados empleando para ello los estándares internacionales más exigentes.

La *UCM* se sumó a *Hathi Trust* al no existir en España ni Europa un consorcio similar que permitiera compartir conocimientos en el ámbito de las bibliotecas digitales y garantizase la preservación y difusión de nuestra colección digital a unos costes razonables.

9 CONCLUSIONES

La digitalización masiva por parte de *Google* ha cambiado de forma significativa el panorama del libro y su sector. La intención de digitalizar *todos* los libros sigue avanzando aunque con un ritmo inferior al de hace tres años. Esta apuesta ha coincidido con el cambio del hábito lector y su entorno más importante de los últimos siglos: del objeto físico se ha pasado al digital en diferentes dispositivos de lectura (lectores de libros electrónicos, tabletas, ordenadores, teléfonos…); de la propiedad al uso; del original a la piratería; de no poder consultar su contenido a la facilidad de consulta desde el propio teléfono; del esencial papel del editor tradicional a la autoedición; de las tiradas costosas y reducidas a la edición bajo demanda y la difusión masiva y sin costes que ofrecen los nuevos agentes en la distribución del libro electrónico.

Desde el punto de vista bibliográfico, una parte importante de los libros y su contenido, de los que hasta la aparición de *Google Libros* sólo se disponía de metadatos que, pese a ser detallados en lo relativo a la descripción física del documento, eran fragmentarios en cuanto a su alcance y pobres respecto a la descripción del contenido y sus relaciones semánticas con otros documentos en la web, cuenta ahora con una herramienta global y sin coste que permite la búsqueda en el texto completo de las obras con unas funcionalidades superiores a las herramientas bibliográficas tradicionales. Y, aunque estas nuevas funciones se han ido trasladando a los catálogos de las bibliotecas gracias a las herramientas de enriquecimiento, siguen estando lejos de la gestión global de derechos de autor, aprovechamiento efectivo de la *minería de datos* e innovación que *Google Libros* posee.

Se tienen ahora nuevos datos respecto al libro y el proyecto que ha permitido digitalizar obras en más de cuatrocientas lenguas. Es el caso de la controvertida

estimación realizada por uno de los técnicos de *Google* respecto a la existencia de 130 millones de libros publicados en el mundo (Taycher, 2010) lo que implica que los materiales pendientes de digitalización serían aún numerosos. O que el 16% de los libros estén en dominio público, el 9% estén sujetos a *copyright* y sean accesibles por estar aún en el circuito comercial y que el 75% restante pese a estar sujetos a *copyright* estén fuera del circuito de distribución (Lessig, 2006). Esta información nos muestra cómo la mayor parte de los libros son de muy difícil acceso para el ciudadano medio al ser consultables sólo en bibliotecas de investigación en donde su consulta está sometida a restricciones importantes por motivos de preservación. Estas obras *huérfanas* en el sentido de que nadie gestiona sus derechos ni las publica y, por lo tanto, no pueden ser consultadas suponen una pérdida de saber en potencia de una magnitud colosal (el 75% del conocimiento albergado en libros) que la sociedad del conocimiento no puede permitirse el lujo de perder. Por ello, se impone la necesidad de un cambio legislativo que permita publicar sin lucro estas obras y compensar los derechos legítimos de los autores en caso de que se soliciten de forma individual.

Otro dato para la reflexión es que debido a las diferentes leyes nacionales de protección de los derechos de autor, en la actualidad hay una considerable diferencia entre el número de libros digitalizados por *Google* que pueden leer y descargar los ciudadanos de distintos países. Es el caso de libros de autores españoles guardados en bibliotecas de Estados Unidos y publicados hasta la segunda década del siglo XX que son accesibles en texto completo en América y otros países pero no en España pues nuestras leyes son más restrictivas e imposibilitan en la práctica el pase al dominio público de libros posteriores a los años setenta del siglo XIX. Esto genera un perjuicio claro para los ciudadanos europeos frente a los norteamericanos al no poder acceder en la práctica a una parte importante de su producción literaria y científica que sí puede consultarse en Estados Unidos y otros países con una legislación más garantista del dominio público. Ello es grave pues al no poderse escanear y difundir estos libros en la mayor parte de los países europeos y sí en América se está creando una brecha creciente entre las posibilidades para acceder al conocimiento por parte de los ciudadanos de uno y otro continente.

Un aspecto a destacar es que gracias a este proyecto se han producido innovaciones significativas en las técnicas de escaneo: *Google* ha perfeccionado la logística de las operaciones de digitalización, diseñado estaciones de escaneado que han permitido la digitalización masiva y difusión de los libros electrónicos y desarrollado mecanismos de mejora de los OCR impensables sin su capacidad de empuje. Un ejemplo de ello es reCAPTCHA, un servicio antirobot gratuito empleado por *Google* para corregir los errores en los OCRs empleados para el análisis de los textos de los libros digitalizados en *Google Libros* y que permite a millones de personas que diariamente validan accesos a servicios en la web

reinterpretar cada día caracteres no legibles por los OCRs en más de cien millones de palabras dudosas y mejorar así la digitalización (Vercelli, 2010).

Respecto a los autores y editores, *Google Libros* y *Google Play* les ofrece una pasarela para la difusión de sus obras inimaginable hasta hace poco. El proyecto, en su apartado dirigido a editores y autores ha tenido gran éxito pues permite a las más de 50.000 editoriales participantes y a un número creciente de autores individuales contar con una herramienta muy útil para la difusión de sus obras, analizar estadísticas de acceso y uso de las mismas y, en caso de estar interesados, de su venta en *Google Play* (y anteriormente en *Google Editions*). Autores cuyas obras estaban relegadas al olvido por la falta de interés económico en su reedición vuelven a ser leídas al contar con una difusión y mercado potencial impensables en un contexto anterior a la red. Y muchos otros pueden publicar sus obras de forma autónoma gracias a las facilidades que ofrecen para la autoedición *Google Libros* y las otras grandes pasarelas internacionales de libros electrónicos (*Amazon*, *Barnes & Noble*…). Sin embargo debemos reseñar cómo las asociaciones de autores y editores han visto con preocupación que los preacuerdos firmados entre *Google* y la *Asociación de Editores Americanos* y la *Asociación de Autores* fueran anulados por el juez en base al *uso justo* lo que implica la paralización de las ventajas para sus asociados incluidas en dichos acuerdos.

Sin embargo, los mayores beneficiarios del proyecto han sido los lectores. Debemos tener en cuenta que hasta su aparición no existía una herramienta tan completa (y gratuita) para el control bibliográfico del libro y su difusión entre el público en general. El hecho de que esta herramienta permita, además, la descarga de los libros en caso de estar en el dominio público supone un avance enorme para la difusión de la cultura y la ciencia pues ha permitido que ciudadanos e investigadores de países en donde el acceso a este tipo de libros era imposible puedan hoy, si tienen un dispositivo conectado a Internet, acceder a la mayor biblioteca digital mundial sin coste. Además, el análisis del *corpus* de las obras digitalizadas está permitiendo el desarrollo de estudios notables por parte de los investigadores y la mejora de otras herramientas de *Google*: es el caso de su *Traductor* que gracias al análisis de la información contenida en los libros ha tenido un avance espectacular en los últimos años.

Pero todas estas ventajas son a costa de que la empresa más especializada en el control de la información y su uso, gracias a su experiencia en la minería de datos, tenga detallada información sobre qué intereses lectores tenemos, sus conexiones y cómo estos se interrelacionan con nuestros gustos al navegar y comprar no sólo en la red, sino en el mundo real gracias a la información sobre la ubicación y otros usos que ofrecen nuestros terminales. Del análisis de estos datos surgen servicios maravillosos cada vez más personalizados que recuerdan nuestras preferencias y movimientos virtuales y físicos, pero también surge la duda y el recelo fundado respecto al uso que de esta información privada e íntima realizan unas empresas

que, llegado el caso, trasladan esta información a agencias gubernamentales cuyo interés va mucho más allá de lo comercial.

Finalmente, desde un punto de vista bibliotecario, los acuerdos con *Google* han permitido que la misión de preservación y difusión del patrimonio bibliográfico de las bibliotecas participantes se haya mejorado significativamente al digitalizar sin coste una parte sustancial de su patrimonio, liberando recursos para acometer proyectos de digitalización propios más especializados. Dado que cada seis meses más del 90 % de estos libros son visitados en *Google* y están, también, presentes en las plataformas de las bibliotecas y en otras iniciativas públicas como *Hathi Trust* o *Europeana*, creemos que se ha dado un paso de gigante para facilitar al público la búsqueda y consulta del conocimiento contenido en los libros. Nuestra función, nuestra pasión.

10 REFERENCIAS BIBLIOGRÁFICAS

BRIN, Sergey (2009). "A library to last forever". *New York Times*. 8 de octubre. <http://www.nytimes.com/2009/10/09/opinion/09brin.html>. [Consulta: 19/02/2014].

CODINA, Lluís (2010). "Anatomía de *Google Books*: un proyecto de biblioteca digital en la encrucijada". *Bid: Textos universitaris de biblioteconomía i documentació*. n. 24.

DHAWAN, Amrita (2013). "Searching Mindfully: Are Libraries up to the challenge of competing with Google Books?". *Library Philosophy and Practice.*
<http://digitalcommons.unl.edu/libphilprac/907/>. [Consulta: 19/02/2014].

DURANTAYE, Katharina de la (2010). "Finding a Home for Orphans: Google Book Search and Orphan Works Law in the United States and Europe". *Fordham intellectual property media and entertainment law journal*. vol. 21, pp. 229-291.

ESTEVE, Asunción (2010)." Análisis legal del proyecto *Google Books* desde la perspectiva de los derechos de la propiedad intelectual". *Bid. Textos universitarios de Biblioteconomía y Documentación*. n. 24. [Consulta: 19/02/2014].

FROSIO, Giancarlo F. (2011)**.** "Google Books rejected: taking the orphans to the digital public library of Alexandria". *Santa Clara Computer and High - Technology Law Journal*. nº 1, vol. 28. pp. 81-141.

GARRIGA, Abel. "La sentencia Google Books o la importancia de la ratio legis" 24 de enero de 2014.
<http://www.holtropblog.com/es/index.php/blog-uk/it-ip/437-sentencia-google-books>. [Consulta: 19/02/2014].

Google Books (2014). "History of Google Books"
<https://www.google.com/googlebooks/about/history.html>. [Consulta: 21/02/2014].

JAMES, Ryan; WEISS, Andrew (2012). "An Assessment of Google Books' Metadata". *Journal of Library Metadata*. nº 1, vol. 12, pp. 15-22.

JEANNENEY, Jean-Noël (2010). *Quand Google défie l'Europe: plaidoyer pour un sursaut*. París: Mille et une Nuits.

JI , Yuan (2011). "Why the Google Book Search Settlement should be approved: A response to antitrust concers and suggestions for regulation". *Albany Law Journal of Science and Technology*. vol. 231, pp. 231-278.

LESSIG, L. (2006). Is Google Book Search "Fair Use"? <http://www.youtube .com/user/lessig#p/u/20/TmU2i1hQiN0> . [Consulta: 07/04/2014].

MAGÁN WALS, José Antonio; PALAFOX PAREJO, Manuela; TARDÓN GONZÁLEZ, Eugenio; SANZ CABRERIZO, Amelia (2011.) "Mass Digitization at the Complutense Univer-sity Library: Access to and Preservation of its Cultural Heritage". *Liber Quarterly.* nº 1, vol. 21, pp. 48-68.

NGUYEN, Courtney (2011). "A Modern Library Class Action: The Google Book Settlement and the Future of Digital Books". *Hastings Commnication and entertainment law journal*. nº 2, vol. 33, pp. 249-274.

NUNBERG, Geoffrey (2009) "Google's Book Search: A Disaster for Scholars". *The Chronical Review*. <http://chronicle.com/article/Googles-Book-Search-A/48245/>. [Consulta: 21/02/2014].

RAFF, Daniel (2011). "The inmaterial text: digital technology, the Google Book Shettlement, and the distribution of print culture in the United States". *Entreprises et Histoire*. nº 64, pp. 146-168.

SAMUELSON, Pamela (2010). "Google Book Search and the Future of Books in Cyberspace". *Minnesota Law Review*. nº 5, vol. 94, pp. 1308-1374.

TAYCHER, Leonid (2010). Books of the world, stand up and be counted! All 129,864,880 of you. Post en *Google Book Search* blog. Agosto, 05. <http://booksearch.blogspot.com.es/2010/08/books-of-world-stand-up-and-be-counted.html>. [Consulta: 07/04/2014].

VERCELLI, Ariel. Google Books y los cambios en las industrias editoriales. http://www.arielvercelli.org/gbylcelie.pdf. [Consulta: 07/04/2014].