

# EVENTOS

II JORNADAS DE  
GESTIÓN DEL PATRIMONIO BIBLIOGRÁFICO

2019



crue

Universidades  
Españolas

Red de Bibliotecas  
REBIUN

**EL FONDO ANTIGUO DE LA BIBLIOTECA DE LA UNIVERSIDAD DE  
SEVILLA EN INTERNET ARCHIVE**

Javier Villanueva Gonzalo

Eduardo Peñalver Gómez

II Jornadas de Gestión de Patrimonio Bibliográfico  
Santiago de Compostela, 6 y 7 de junio de 2019



Esta obra está sujeta a la [licencia Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional \(CC BY-NC-SA 4.0\)](https://creativecommons.org/licenses/by-nc-sa/4.0/)

**EL FONDO ANTIGUO DE LA BIBLIOTECA DE LA UNIVERSIDAD DE SEVILLA EN INTERNET ARCHIVE**

**SPECIAL COLLECTIONS OF THE LIBRARY OF THE UNIVERSITY OF SEVILLA ON INTERNET ARCHIVE**

EDUARDO PEÑALVER GÓMEZ<sup>1</sup>  
JAVIER VILLANUEVA GONZALO<sup>2</sup>  
Biblioteca Universitaria de Sevilla

**Resumen:** En junio de 2016 la Sección de Fondo Antiguo de la Biblioteca de la Universidad de Sevilla comienza un proyecto de difusión de sus fondos patrimoniales a través de la biblioteca digital Internet Archive, con el fin de mejorar la visibilidad y usabilidad de las imágenes digitalizadas e incrementar la accesibilidad de los textos. También con la esperanza puesta en mejorar la productividad en los procesos de digitalización. Esta comunicación muestra el itinerario de este proyecto a lo largo de estos tres años y los beneficios obtenidos.

**Palabras clave:** Digitalización, patrimonio bibliográfico, difusión

**Abstract:** In June 2016, the Special Collection Section of the Library of the University of Seville began a project to disseminate its bibliographic heritage through the Internet Archive Digital Library with the aim of improving the visibility and usability of the digitised images and increasing the accessibility of the texts. Also with the hope of improving productivity in the digitization processes. This communication shows the itinerary of this project throughout these three years and the benefits obtained.

**Keywords:** Digitization, bibliographic heritage, dissemination

---

<sup>1</sup> [epenalver@us.es](mailto:epenalver@us.es)

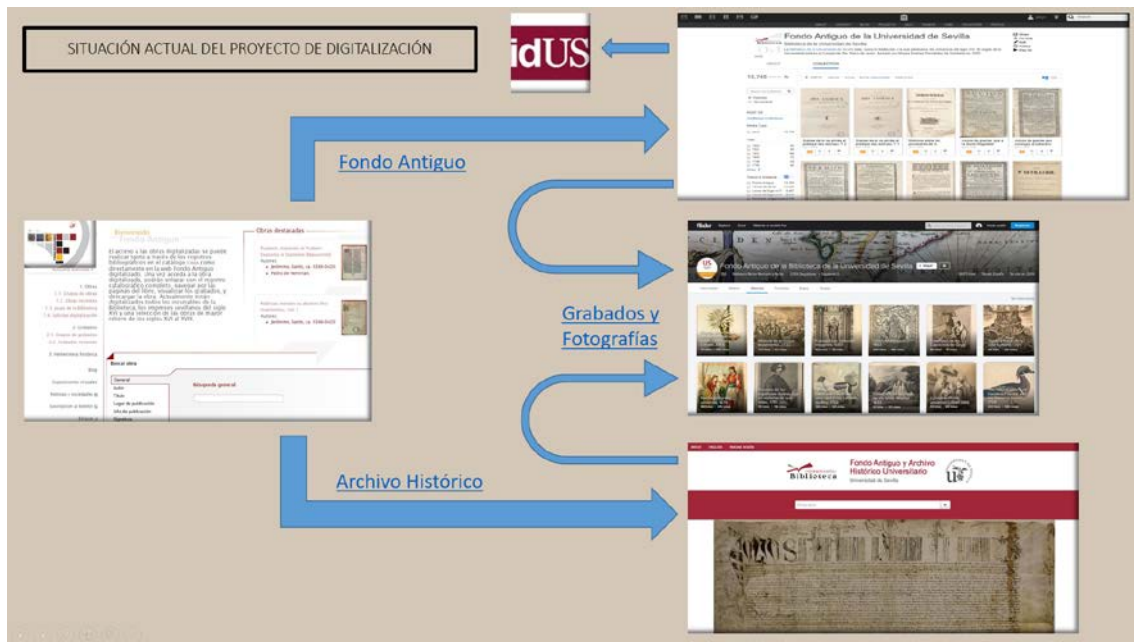
<sup>2</sup> [jabyn@us.es](mailto:jabyn@us.es)

La difusión del patrimonio bibliográfico es uno de los mayores retos tecnológicos a que se enfrentan las bibliotecas con colecciones patrimoniales.

La Biblioteca de la Universidad de Sevilla comenzó su andadura en este terreno en el año 2004 con la creación del Portal de Fondos Digitalizados en la página <http://fondosdigitales.us.es>.

En todos estos años, el Portal ha respondido con bastante solvencia a las expectativas, pero se han planteado recientemente algunas complicaciones derivadas de la dependencia de un sistema informático propietario. A partir de cierto momento el mantenimiento del Portal se revela insostenible, la aplicación va quedando obsoleta en muchos aspectos y el continuo parcheo de errores de la plataforma que íbamos reportando a la empresa, no solucionaba sus problemas reales – usabilidad de la interfaz, metadatos mal implementados, problemas en la generación del reconocimiento óptico de caracteres, imposibilidad de búsqueda textual tanto en buscadores como en la propia plataforma ni en el documento, falta de optimización para dispositivos móviles.etc., también desde el punto de vista económico se estaba convirtiendo en algo prohibitivo y sin una respuesta acorde al dinero gastado.

En Junio de 2016 se estuvieron barajando varias alternativas, entre ellas la de un repositorio para fondo Antiguo en Dspace, que descartamos porque consideramos muy insuficiente la oferta de descarga y visualización que ofrecía. Por contra, Internet Archive ofrecía esta posibilidad, una bella interfaz para el libro en “streaming”, además de múltiples formatos de descarga, incluido el pdf con OCR. Después de realizar una prueba de carga y flujo de trabajo, presentamos a la dirección de la Biblioteca la posibilidad de migración de parte del Portal de Fondos Digitales a Internet Archive.



El proyecto se desdobló según el tipo de material:

- Los libros y la hemeroteca histórica, se vuelcan en nuestra colección en Internet Archive:

<https://archive.org/details/bibliotecauniversitariadesevilla>

El volcado comenzó en Junio de 2016 y concluyó en julio de 2017. En este momento la colección en Internet Archive aloja 17.000 items (más de 2 millones de páginas). El Portal de fondos digitalizados se cerró al público en marzo de 2019 pero dejó de actualizarse en diciembre de 2017.

Es importante señalar que el pdf con OCR incrustado que genera esta plataforma se incluirá en una colección de nueva creación en IDUS, nuestro repositorio institucional, específica para Fondo Antigo, en el que se incluirán los metadatos necesarios para la interoperabilidad con Hispana, Europea y demás recolectores.

- Las colecciones de grabados de la antigua plataforma se vuelcan en nuestro banco de imágenes en Flickr que contiene en este momento más de 70.000 grabados y arroja una estadística de más de 23 millones de visitas acumuladas desde su creación.
- El Archivo Histórico pasa a una nueva página institucional en la dirección <https://ahus.us.es> montada sobre el software ICA-AtoM y contiene en este momento más de 250.00 imágenes.

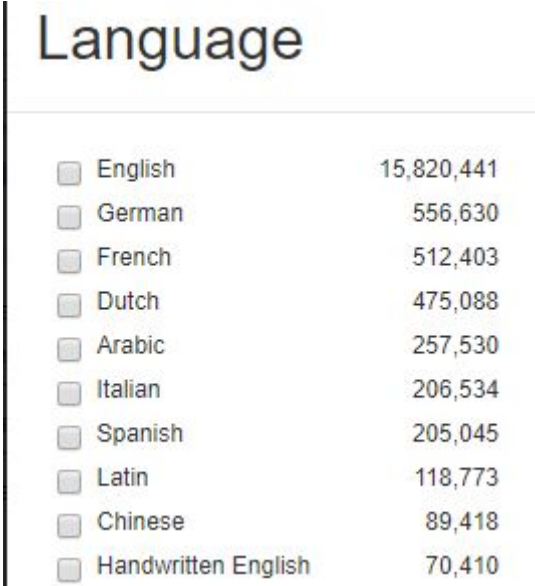
### ¿Qué es Internet Archive?

Internet Archive es una biblioteca digital sin fines de lucro fundada por Brewster Kahle en 1996 con la misión de proporcionar "acceso universal a todos los conocimientos".

La organización busca preservar el patrimonio cultural del mundo y proporcionar acceso abierto a nuestro conocimiento compartido en el ámbito digital.

Las colecciones digitales de Internet Archive incluyen:

- 349 Billones de páginas web.
- 20 millones de libros.
- 4,9 millones de videos.
- 5.1 millones de audio, etc.
- 30 Petabytes (backups Alejandría, Amsterdam).
- Colabora con más de 1500 bibliotecas en todo el mundo.
- El español aparece como la octava lengua en contenidos.

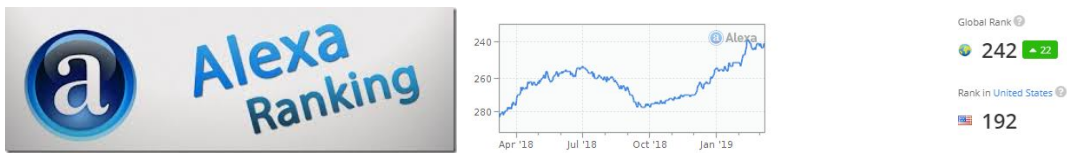


Language	
<input type="checkbox"/> English	15,820,441
<input type="checkbox"/> German	556,630
<input type="checkbox"/> French	512,403
<input type="checkbox"/> Dutch	475,088
<input type="checkbox"/> Arabic	257,530
<input type="checkbox"/> Italian	206,534
<input type="checkbox"/> Spanish	205,045
<input type="checkbox"/> Latin	118,773
<input type="checkbox"/> Chinese	89,418
<input type="checkbox"/> Handwritten English	70,410

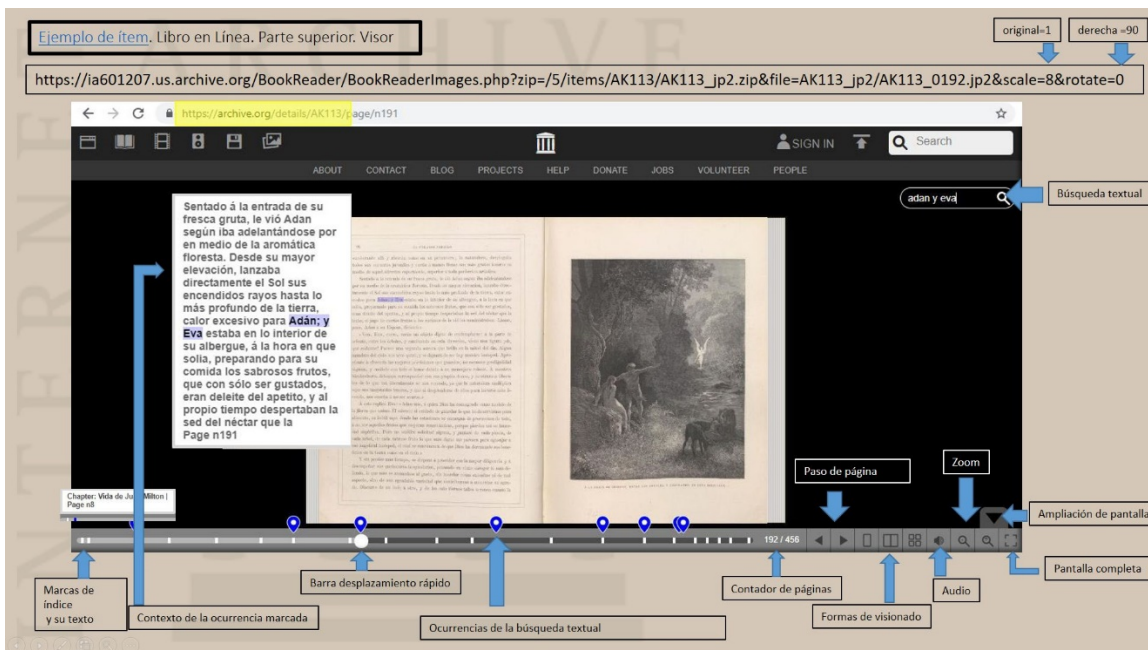
Las ventajas que nos decidieron a apostar por Internet Archive fueron:

- Proyecto de prestigio internacional en una plataforma pública.

- Ahorro de costes en infraestructura informática, frente a la situación con la anterior plataforma.
- Existencia de página individualizada para la colección. Con registro estadístico propio y página de descripción del proyecto:  
<https://archive.org/details/bibliotecauniversitariadesevilla&tab=about>
- Visibilidad internacional de la plataforma. Siempre se encuentran por encima del top 300 en el ranking mundial de Alexa.



- Usabilidad de la plataforma. Mejoras significativas en la visualización y navegación dentro del ítem, entre otras características a reseñar: visionado por defecto a doble página, lo que ofrece una apariencia de libro real en pantalla, generación del libro en “streaming” lo que aporta una url propia para cada página muy interesante para citar y para incrustar el lector de libros en blogs, etc. por la página deseada.



- Disponibilidad del ocr de la obra en pdf, texto plano y en el libro en línea.
- Posibilidad de búsqueda de la información textual a tres niveles. Dentro del libro, en todo nuestro fondo o en toda la plataforma.
- Las búsquedas se pueden restringir con multitud de filtros o facetas: año, materia, idioma, etc.
- 3 opciones de visualización para los resultados de una búsqueda: listado, miniaturas y descripción que a su vez pueden ordenarse, por autor, título, fecha de subida, fecha de publicación, número de vistas, reseñas.
- Presencia en buscadores del texto completo de las obras.
- Funcionalidades de tipo social (Compartir en plataformas, añadir a tu propia colección de favoritos si tienes cuenta en la plataforma, también puedes escribir comentarios en cualquier ítem si tienes cuenta de usuario)
- Cada ítem tiene su registro estadístico que incluye visitas, favoritos y reseñas.
- RSS.
- Optimizado para tablets y móviles.
- Dispone también de otras funcionalidades en este momento no muy útiles para fondo antiguo, como la generación automática de audio en línea y la posibilidad de descargárselo en formato Daisy, y también la posibilidad de descargar las obras como epub o Kindle.
- No menos interesante es la presencia de un fichero torrent entre los derivados que genera la plataforma, asegurando su carácter público si quedaba alguna duda de ello.
- Posibilidad de aplicar diversos niveles de licencia Creative Commons individualmente en cada registro. Nosotros marcamos todo el fondo antiguo como Dominio Público, siguiendo además el ejemplo de otras bibliotecas, tanto nacionales como extranjeras, nos parece la licencia más acertada para este tipo de contenido.

Como desventajas podríamos señalar:

- Interfaz de visualización no disponible en español (solamente en inglés).
- Imposibilidad de intervenir en el diseño gráfico.

### Mejoras en la productividad

-Desaparece el trabajo de edición unido al sistema de gestión antiguo. Había que definir páginas por tipo, crear índices, describir grabados, etc.

- Asociado a lo anterior, mejoras en el flujo de trabajo en digitalización. Solamente tenemos que preparar la carpeta de imágenes y subirla en zip; el sistema automatizado de Internet Archive crea el libro, el ocr, y los derivados.

### Mejoras en la difusión

El cambio de plataforma ha supuesto una mejora considerable en la visibilidad y accesibilidad de nuestro fondo antiguo. Tomando como referencia las estadísticas referidas al Portal de Fondos Digitales de nuestra Memoria del año 2016, observamos una mejora significativa si comparamos con los años siguientes en Internet Archive. Este mes alcanzaremos el millón de visitas.

Fondos Digitales 2016	13.212
Internet Archive 2017	280.284
Internet Archive 2018	479.778

VIEWS — About the New Statistics

Total Views 977,237

