

Análisis de enlaces hacia bibliotecas y archivos digitales de patrimonio cultural desde *Wikipedia* en español y catalán

[\[Versió catalana\]](#)

TOMÁS SAORÍN 

Profesor de la Facultad de Comunicación y Documentación
Universidad de Murcia
tsp@um.es

EMILIO JOSÉ RODRÍGUEZ POSADA

Estudiante pre-doctoral. Escuela Superior de Ingeniería
Universidad de Cádiz
emiliojose.rodriguez@uca.es

Opciones



Resumen [\[Abstract\]](#) [\[Resum\]](#)

Objetivo. Describir y evaluar el uso en *Wikipedia* de enlaces a las colecciones digitalizadas en bibliotecas, archivos y otras instituciones culturales.

Metodología. El estudio se realiza sobre la totalidad de los artículos de las ediciones en español y catalán de *Wikipedia*, usando una herramienta de análisis de wikis. Se realiza una selección amplia de 81 colecciones digitales españolas de diferente alcance. También se toman datos de otros proyectos de digitalización para poder comparar los resultados.

Resultados. Se muestra una presencia aún débil de enlaces desde *Wikipedia*, excepto para la *Biblioteca Virtual Miguel de Cervantes*, cuyas magnitudes son sensiblemente diferentes. Algunas colecciones especializadas son más usadas, pero en general se aprecia una falta de atención hacia estas colecciones desde el colectivo de editores de *Wikipedia*, lo cual debería tenerse en cuenta en el desarrollo de los proyectos de digitalización tipo *Europeana*.

1 Introducción

La *Agenda Digital para Europa* ha supuesto la inversión de considerables cantidades de dinero en ampliar la disponibilidad digital del patrimonio cultural europeo. Su mayor exponente es *Europeana*, que agrega recursos de múltiples instituciones culturales. En España han crecido los contenidos y repositorios de carácter patrimonial, al amparo de las convocatorias de ayudas del Ministerio de Cultura, con unos exigentes requisitos de calidad de digitalización, publicación en acceso abierto y cumplimiento de estándares de metadatos. España es uno de los cinco países que más aporta a *Europeana* (el 8,31 % de sus más de 20 millones de objetos).¹

Sin embargo, *Europeana* concede cada vez más importancia en sus estrategias al impacto en el medio digital, y no sólo los datos de ítems digitalizados. La mera puesta a disposición de un gran catálogo no implica que sea altamente visible para el ciudadano, y por lo tanto se intenta orientar el portal hacia los usos y prácticas digitales que conecten con la realidad actual de la red. Las instituciones culturales temen además perder visibilidad en el espacio digital, porque sus registros descriptivos no sean usados como fuente de referencia en relación a los objetos culturales que conservan, y se degrade la calidad de la información cultural (Ferrer-Sapena; Peset, 2012).

Se ha planteado que se deben elaborar nuevas métricas para medir el impacto de sus acciones en

el medio digital, que en lugar de medir simplemente el tráfico en sus propios sitios web, tengan en cuenta "la generación de accesos en sitios ya frecuentados por los usuarios". Contemplan en un lugar prioritario a *Wikipedia*, de la cual resaltan que "Toda institución debería preocuparse de sus recursos en *Wikipedia*. Tendría que formar parte de su plan de negocio habitual" (Verwayen; Arnoldus; Kaufman, 2011).

El interés del sector de archivos, bibliotecas y museos en *Wikipedia* se ha consolidado en un movimiento conocido como [GLAM](#). Desde febrero de 2012 la biblioteca digital *Europeana* incluye un *widget* para "[citar en Wikipedia](#)", que busca facilitar que sus contenidos digitales sean usados como referencia en artículos de cualquier edición de *Wikipedia*. De esta iniciativa cabe deducir que en un futuro cercano se podría usar la presencia en *Wikipedia* como uno de los indicadores de impacto social digital, aplicable a repositorios de la misma naturaleza.

Los aspectos de colecciones, políticas, metadatos y preservación de las colecciones digitales patrimoniales son tratados en otros estudios (Estivill Rius; Gascón García; Sulé Duesa, 2010), pero existe menos información sobre el uso efectivo de sus recursos. Este trabajo se enmarca en lo que diferentes modelos de evaluación de sitios web denominan *visibilidad* o *popularidad* (Codina, 2006; Guallar; Abadal, 2009), pero analizando una única fuente proveedora de enlaces de considerable impacto y uso. Dado que el *pagerank* de Google combina la abundancia y diversidad de enlaces con la visibilidad del emisor del enlace, los resultados de este estudio también pueden tener interés en el campo del posicionamiento web (Codina; Marcos, 2005, p. 86). Dentro de las investigaciones centradas en *Wikipedia*, existe una línea de investigación sobre el análisis del uso y cita de fuentes, que es un elemento básico de la calidad y credibilidad de los artículos de la enciclopedia (Nielsen, 2007; Luyt; Tan, 2010). Otros estudios han abordado el aumento del tráfico originado en *Wikipedia* hacia las colecciones digitales, cuando se han enriquecido artículos e incluido enlaces sobre las temáticas que recogen (Lallye; Dunford, 2007; Elder; Westbrook; Reilly, 2012).

Este trabajo tiene como objetivo realizar una primera aproximación a la presencia de enlaces de fuentes contenidas en bibliotecas digitales españolas del ámbito del patrimonio cultural. Se extraerán datos individuales y globales de cada repositorio, que sirvan para evaluar y comparar en el tiempo su impacto directo en *Wikipedia* en español y en catalán (*Viquipèdia*).

2 Metodología del estudio

Para el estudio se han seleccionado los repositorios y archivos digitales españoles con contenidos del área de patrimonio cultural (patrimonio bibliográfico y documental, patrimonio histórico, patrimonio fotográfico). Los documentos de carácter patrimonial conservados en archivos, bibliotecas, museos e instituciones culturales han sido de difícil acceso hasta su reciente digitalización y difusión a través de portales web.

Tras haberlos identificado, se realiza el análisis de enlaces que apuntan a los repositorios u objetos que contienen desde cualquier artículo de las ediciones en español y catalán de *Wikipedia*.² Los datos de los artículos de *Wikipedia* han sido procesados con una herramienta estadística especialmente enfocada al análisis de wikis llamada [WikiEvidens](#) (ver apéndice 3) A ella hemos añadido las funciones necesarias para parametrizar la lista de fuentes, los patrones de comparación y la edición de *Wikipedia* que se analiza. WikiEvidens es software libre y puede ser descargado gratuitamente.

Se han identificado para cada proyecto todas las variantes posibles de dominios y subdominios que resultan de todas las formas de enlazar (a documentos, fichas de metadatos, secciones, etc.).³

A la hora de establecer el patrón de enlaces directos a los documentos podemos encontrar el problema de las "selecciones ad hoc". Cuando la necesidad de promocionar un contenido supone "sacarlo" del repositorio, y presentarlo a través del sistema de archivos del gestor de contenidos, puede afectar a que el wikipedista escoja un enlace externo al repositorio, y por lo tanto no sea tenido en cuenta al analizar los enlaces al repositorio.⁴ También quedan fuera del estudio todas aquellas digitalizaciones singulares publicadas fuera de repositorios identificables.

Otro elemento que hemos considerado que ofrecía resultados de utilidad era la comprobación de la persistencia de los enlaces, dado que en los tests preliminares detectamos enlaces que habían dejado de funcionar (conocidos como *enlaces rotos*). Para ello hemos utilizado un pequeño *crawler* (disponible entre los módulos de WikiEvidens) para revisar sistemáticamente los enlaces y detectar los que habían dejado de estar accesibles.

Hemos optado por realizar un análisis panorámico, debido a la reducida presencia de enlaces. En una segunda aproximación podría contemplarse la relación entre el tamaño del repositorio y el número de citas recibidas en *Wikipedia*.

También hemos considerado conveniente centrar los resultados en la presencia de enlaces sólo en artículos, eludiendo las páginas de discusión y las de coordinación de la comunidad wikipedista. Tampoco diferenciaremos para cada caso si el enlace aparece como nota, bibliografía o enlaces externos; sólo presentaremos datos globales. Cuando las citas a repositorios patrimoniales sean más abundantes, podrán considerarse estos análisis segmentados, así como otros sobre las categorías de los artículos o el número de enlaces por artículo.

Un aspecto que no hemos desarrollado ha sido el análisis del destino de los enlaces (fichas de metadatos; documento digitalizado, imágenes, categorías, colecciones, búsquedas, exposiciones digitales, etc.), lo cual podría aportar información muy específica de interés para la mejora de la experiencia de usuario y arquitectura de la información de estas bibliotecas digitales. En algunos casos pueden haberse recogido enlaces a registros que no permiten el acceso al documento completo, o enlaces genéricos al portal del repositorio. Si los datos de enlaces fueran más abundantes se estaría en condiciones de afinar más sobre el destino preferido para el enlace.

2.1 Selección de bibliotecas digitales

En este trabajo no se tienen en cuenta los numerosos repositorios científicos en acceso abierto, que contienen sobre todo publicaciones académicas y de investigación.⁵ Se ha optado por aquellos repositorios con documentos históricos y patrimoniales (bibliotecas y archivos digitales).

Hemos seleccionado los principales proyectos de bibliotecas digitales españolas. Se incluyen también las bibliotecas digitales regionales, y aquellos repositorios digitales de universidades o archivos digitales de otras instituciones culturales que cuentan con un fondo antiguo importante. Se ha de tener en cuenta que existen numerosos proyectos de digitalización monográficos que no se integran en repositorios digitales, así como digitalización de patrimonio publicada electrónicamente de forma dispersa y, por lo tanto, no identificados en este estudio.

También se recogen los proyectos europeos en los que hay una aportación significativa de contenidos españoles. Aunque el alcance es fundamentalmente nacional, se realizan algunas tomas de datos con fuentes de datos de colecciones digitales importantes como la de la Library of Congress, la *Biblioteca Digital Mundial* de la Unesco y *Memoria Chilena* de la Biblioteca Nacional Digital de Chile. A modo de elemento de comparación hemos incluido también las hemerotecas históricas del ABCy *La Vanguardia*, que abarcan todo el siglo xx y son de acceso abierto.

Para la elaboración de la lista de repositorios tomamos como punto de partida la selección realizada por la Biblioteca Nacional de España bajo el título "[Colecciones digitales españolas](#)", vinculadas en su mayor parte a las ayudas de digitalización del Ministerio de Cultura de los últimos años. Este listado se ha complementado con una selección de otros de los 561 proyectos del directorio de colecciones digitales [Hispana](#).

Las colecciones históricas de algunos repositorios universitarios presentaban problemas para su inclusión, ante la imposibilidad de separar estos registros del resto de contenidos científicos e institucionales. Muchos de sus contenidos están disponibles a través de *Hispana*, pero su publicación indiferenciada hace que perdamos estos datos.⁶ En algunos de los análisis hemos aplicado un modulador a los resultados según el porcentaje que suponga la colección histórica en el total del repositorio, indicándolo en nota. Así, por ejemplo, en el repositorio *Roderic* de la Universitat de València, el fondo histórico supone la mitad de los contenidos, mientras que en *Digitum*, de la Universidad de Murcia, es el 5 %.

Se ha prestado atención a la inclusión de colecciones digitales catalanas, aunque fueran de pequeño tamaño, con el fin de captar de la mejor forma el comportamiento específico de la edición en catalán de *Wikipedia*.

También presentan problemas de ajuste los datos que recogemos de *Hispana*, porque se mezclan repositorios de producción científica con colecciones patrimoniales.

Para obtener datos diferenciados según el alcance de cada biblioteca digital, hemos establecido las siguientes categorías: EU (europeas), ES (ámbito nacional), REG (regionales o autonómicas), LOC

(locales o provinciales), UNI (universidades), INS (otras instituciones culturales). En la categoría TEST se incluyen las colecciones digitales usadas para contextualizar las cifras de enlaces recogidos.

3 Resultados

En total se han seleccionado 81 bibliotecas digitales (3 EU; 7 ES; 22 REG; 5 LOC; 28 UNI y 16 INS). Se han incluido 6 colecciones en la categoría TEST. El listado completo y las URL base para localizar los enlaces se presentan en el apéndice 0.

Se han recuperado los enlaces procedentes de 372.286 artículos de *Wikipedia* en catalán (WP-CAT) y 885.935 de *Wikipedia* en español (WP-ES), a partir de sus [dumps](#), con fecha de 19 y 20 de abril de 2012 respectivamente.

Wikipedia en español cuenta con 6.460 enlaces hacia las bibliotecas que hemos seleccionado, lo que representa el 0,21 % del total de enlaces existentes en esta enciclopedia. Estos enlaces están distribuidos por 4.393 artículos, que constituyen solo el 0,50 % del total de artículos disponibles.

Wikipedia en catalán cuenta con 1.666 enlaces distribuidos en 1.075 artículos, lo que supone un 0,11 % del total de enlaces y un 0,29 % del total de artículos.

Edición de la <i>Wikipedia</i>	Enlaces a bibliotecas digitales	Total enlaces	%
WP-ES	6.460	3.104.964	0,21 %
WP-CAT	1.666	1.516.015	0,11 %

El número de enlaces a bibliotecas digitales dentro de un mismo artículo presenta patrones similares en ambas enciclopedias: 1,47 enlaces por artículo en la española y 1,55 en la catalana.

Para poder poner en contexto estas cifras, hemos obtenido las de las hemerotecas del *ABC* y *La Vanguardia*, dos grandes hemerotecas de periódicos españoles del siglo xx. La prensa es una gran proveedora de enlaces a *Wikipedia*: *El País* ocupa el puesto 17 del ranking de enlaces de [Wikipedia en español](#) y la *BBC* el puesto 6 en [Wikipedia en inglés](#). Por ello las magnitudes de este estudio pueden contrastarse con los 4.148 enlaces desde *Wikipedia* en español hacia la Hemeroteca del *ABC* distribuidos en 2.325 artículos y los 1.203 enlaces hacia la Hemeroteca de *La Vanguardia* en 722 artículos. En *Wikipedia* en catalán encontramos 2.217 enlaces desde 942 artículos para la Hemeroteca de *La Vanguardia* y solamente 160 enlaces hacia la Hemeroteca del *ABC*, desde 137 artículos.

También nos será útil saber los datos de otras conocidas colecciones digitales extranjeras: *Google Books*, con las salvedades anteriormente indicadas, recibe 18.497 enlaces en WP-ES y 1.131 en WP-CAT. Las colecciones digitales de la Library of Congress de los Estados Unidos reciben 1.052 enlaces en WP-ES y 276 en WP-CAT. El portal *Memoria de Chile* recibe 1.462 desde WP-ES y 9 desde WP-CAT. Finalmente, la *Biblioteca Digital Mundial* de la Unesco apenas genera 26 enlaces en WP-ES y 3 en WP-CAT.

Las siguientes tres tablas recogen el número de enlaces que apuntan hacia las distintas bibliotecas y repositorios digitales seleccionados, en WP-ES, WP-CAT y de forma conjunta. Aquellos sitios no enlazados no han sido añadidos a las tablas (ver apéndice 2). Los datos de las bibliotecas digitales mixtas (científicas y patrimoniales) marcadas con (***) pueden ser poco fiables para colecciones patrimoniales. Por el contrario, las marcadas con (*) tienen una alta probabilidad de ajustarse al alcance del trabajo.⁷

Biblioteca digital	Enlaces
<i>Biblioteca Virtual Miguel de Cervantes</i>	4.372
<i>Biblioteca Digital Hispánica y Hemeroteca Digital</i> de la Biblioteca Nacional de España	320
<i>Biblioteca Digital de Castilla y León</i>	226
<i>Biblioteca Digital del Real Jardín Botánico</i>	172
<i>Biblioteca Virtual de Derecho Aragonés</i>	152
<i>Dipòsit digital de documents de la UAB</i> (***)	87

<i>Biblioteca Digital Leonesa</i>	86
<i>Biblioteca Virtual de Prensa Histórica</i>	76
<i>Pares (Portal de Archivos Españoles)</i>	68
<i>Biblioteca Digital de Castilla-La Mancha</i>	63
<i>Fondo Documental Histórico de las Cortes de Aragón</i>	55
<i>Biblioteca Saavedra Fajardo de Pensamiento Político Hispánico</i>	54
<i>e_Buah de la Universidad de Alcalá (***)</i>	48
<i>Biblioteca Virtual del Patrimonio Bibliográfico</i>	47
<i>Fondo antiguo de la Universidad de Sevilla</i>	47
<i>Biblioteca Digital Complutense</i>	42
<i>Biblioteca Digital del repositorio Gredos de la Universidad Salamanca (*)</i>	40
<i>Memoria Digital de Canarias</i>	39
<i>Red Digital de Colecciones de Museos de España</i>	38
<i>Memoria Digital Vasca</i>	38
<i>The European Library</i>	30
<i>Memòria Digital de Catalunya</i>	29
<i>Biblioteca Digital del Ateneo Madrid</i>	27
<i>Digitum. Fondo antiguo de la Universidad de Murcia (***)</i>	26
<i>Helvia: Repositorio Institucional de la Universidad de Huelva (**)</i>	22
<i>Arias Montano: repositorio institucional de la Universidad de Huelva (**)</i>	21
<i>Fondo Antiguo de la Universidad de Granada (*)</i>	19
<i>BiNaDI. Biblioteca Digital de Navarra</i>	18
<i>Repositorio da Universidade da Coruña (***)</i>	17
<i>Rodín. Universidad de Cádiz. Patrimonio bibliográfico (**)</i>	15
<i>Biblioteca Digital de Castellón (*)</i>	13
<i>Biblioteca Virtual Ignacio Larramendi</i>	12
<i>Memoria de Madrid</i>	11
<i>Jable: Archivo de Prensa Digital de la Universidad de las Palmas de Gran Canaria</i>	11
<i>BiValdi (Biblioteca Valenciana Digital)</i>	10
<i>Europeana</i>	9
<i>ANC (Arxiu Nacional Catalunya)</i>	8
<i>Biblioteca Digital de la Comunidad Madrid</i>	7
<i>Fototeca del Patrimonio Histórico</i>	6
<i>ARCA (Arxiu de Revistes Catalanes Antiques)</i>	6
<i>Biblioteca Digital de la Región de Murcia</i>	6
<i>Biblioteca Virtual del Principado de Asturias</i>	6
<i>Biblioteca Virtual de Andalucía</i>	6
<i>Fondo Histórico Digital de La Rábida de la Universidad Internacional de Andalucía</i>	6
<i>DUGi Fons Especials: Repositori Digital de la UdG</i>	5
<i>UVaDOC: repositorio documental de la Universidad de Valladolid (*)</i>	5
<i>Recursos de Investigación de la Alhambra</i>	4
<i>Fondo Antiguo Universidad Zaragoza (**)</i>	4
<i>RIUMA. Universidad de Málaga. Patrimonio (***)</i>	4
<i>Biblioteca Digital de la Real Academia de Historia</i>	3
<i>Liburutegi Digitala de la Biblioteca Foral de Bizkaia</i>	3
<i>Biblioteca Virtual de La Rioja</i>	3
<i>Minerva: Repositorio Institucional da Universidade de Santiago de Compostela (**)</i>	3
<i>Hispana (**)</i>	2
<i>Biblioteca Virtual de la Diputación de Zaragoza</i>	2
<i>Somni. Fons històric de la Universitat de València (*)</i>	2

<i>RUIdeRA</i> : repositorio institucional de la UCLM (***)	2
Archivo del Ateneo Madrid	1
<i>Iuris Digital</i> : Biblioteca Virtual de la Real Academia de Jurisprudencia y Legislación	1
<i>Teatro Siglo de Oro</i> de la Biblioteca Nacional de España	1
<i>Biblioteca Dixital de Galicia (Cidade de Cultura Galega)</i>	1
<i>Galiciana: Biblioteca Dixital de Galicia</i>	1
<i>Fons Sol-Torres</i> del Servei de Biblioteca i Documentació de la Universitat de Lleida	1
Biblioteca Digital del CEU (***)	1
16 bibliotecas digitales no tienen enlaces en la edición en español (ver apéndice 2)	

Tabla 1. Enlaces en *Wikipedia* en español (WP-ES)

Biblioteca digital	Enlaces
<i>Biblioteca Virtual Miguel de Cervantes</i>	786
<i>Memòria Digital de Catalunya</i>	186
<i>Biblioteca Virtual de Derecho Aragonés</i>	174
<i>Dipòsit digital de documents de la UAB (***)</i>	147
<i>ARCA (Arxiu de Revistes Catalanes Antiques)</i>	95
<i>Biblioteca Digital del Real Jardín Botánico</i>	35
ANC (Arxiu Nacional Catalunya)	31
<i>Fondo Documental Histórico</i> de las Cortes de Aragón	31
<i>Biblioteca Virtual de Prensa Histórica</i>	24
<i>Biblioteca Digital Hispánica y Hemeroteca Digital</i> de la Biblioteca Nacional de España	18
<i>BiValdi (Biblioteca Valenciana Digital)</i>	13
<i>Biblioteca Saavedra Fajardo de Pensamiento Político Hispánico</i>	12
<i>Biblioteca Digital de Castilla y León</i>	11
Fondo antiguo de la Universidad de Granada (*)	11
<i>Cartoteca Digital</i> de l'Institut Cartogràfic de Catalunya	10
<i>Biblioteca Digital de Castellón (*)</i>	9
<i>Biblioteca Digital Leonesa</i>	8
<i>Red Digital de Colecciones de Museos de España</i>	7
<i>Fons Sol-Torres</i> del Servei de Biblioteca i Documentació de la Universitat de Lleida	6
<i>Biblioteca Virtual del Patrimonio Bibliográfico</i>	5
<i>Pares (Portal de Archivos Españoles)</i>	4
<i>Biblioteca Digital del Ateneo Madrid</i>	4
<i>Memoria Digital de Canarias</i>	4
<i>DUGi Fons Especials: Repositori Digital de la UdG</i>	4
<i>Digitum</i> . Fondo antiguo de la Universidad de Murcia (***)	4
The European Library	3
<i>Biblioteca Virtual Ignacio Larramendi</i>	3
<i>Biblioteca Digital Complutense</i>	3
<i>Europeana</i>	2
<i>Biblioteca Virtual de Andalucía</i>	2
Fondo antiguo de la Universidad Sevilla	2
<i>Somni</i> . Fondo histórico de la Universitat de València (*)	2
<i>Biblioteca Digital</i> del repositorio <i>Gredos</i> de la Universidad Salamanca (*)	2
<i>Almirall: portal de pensament i cultura del segle XIX</i>	1
<i>Liburutegi Digitala</i> de la Biblioteca Foral de Bizkaia	1
<i>Biblioteca Virtual de La Rioja</i>	1
<i>Memoria Digital Vasca</i>	1

<i>Biblioteca Virtual del Principado de Asturias</i>	1
<i>BINaDI. Biblioteca Digital de Navarra</i>	1
<i>Helvia: Repositorio Institucional de la Universidad de Huelva (**)</i>	1
<i>Repositorio da Universidade da Coruña (*)</i>	1
40 bibliotecas digitales no tienen enlaces en la edición catalana (ver apéndice 2)	

Tabla 2. Enlaces en *Wikipedia* en catalán (WP-CAT)

Biblioteca digital	Enllaços
<i>Biblioteca Virtual Miguel de Cervantes</i>	5.158
<i>Biblioteca Digital Hispánica y Hemeroteca Digital de la Biblioteca Nacional de España</i>	338
<i>Biblioteca Virtual de Derecho Aragonés</i>	326
<i>Biblioteca Digital de Castilla y León</i>	237
<i>Dipòsit digital de documents de la UAB (***)</i>	234
<i>Memòria Digital de Catalunya</i>	215
<i>Biblioteca Digital del Real Jardín Botánico</i>	207
<i>ARCA (Arxiu de Revistes Catalanes Antiques)</i>	101
<i>Biblioteca Virtual de Prensa Histórica</i>	100
<i>Biblioteca Digital Leonesa</i>	94
<i>Fondo Documental Histórico de las Cortes de Aragón</i>	86
<i>Pares (Portal de Archivos Españoles)</i>	72
<i>Biblioteca Saavedra Fajardo de Pensamiento Político Hispánico</i>	66
<i>Biblioteca Digital de Castilla-La Mancha</i>	63
<i>Biblioteca Virtual del Patrimonio Bibliográfico</i>	52
<i>Fondo antiguo de la Universidad Sevilla</i>	49
<i>e_Buah de la Universidad de Alcalá (***)</i>	48
<i>Biblioteca Digital Complutense</i>	45
<i>Red Digital de Colecciones de Museos de España</i>	45
<i>Memoria Digital de Canarias</i>	43
<i>Biblioteca Digital del repositorio Gredos de la Universidad Salamanca (*)</i>	42
<i>ANC (Arxiu Nacional Catalunya)</i>	39
<i>Memoria Digital Vasca</i>	39
<i>The European Library</i>	33
<i>Biblioteca Digital del Ateneo Madrid</i>	31
<i>Digitum. Fondo antiguo de la Universidad de Murcia (***)</i>	30
<i>Fondo antiguo de la Universidad de Granada (*)</i>	30
<i>BiValdi (Biblioteca Valenciana Digital)</i>	23
<i>Helvia: Repositorio Institucional de la Universidad de Huelva (**)</i>	23
<i>Biblioteca Digital de Castellón (*)</i>	22
<i>Arias Montano: repositorio institucional de la Universidad de Huelva (**)</i>	21
<i>BINaDI. Biblioteca Digital de Navarra</i>	19
<i>Repositorio da Universidade da Coruña (*)</i>	18
<i>Biblioteca Virtual Ignacio Larramendi</i>	15
<i>Rodín. Universidad de Cádiz. Patrimonio bibliográfico (**)</i>	15
<i>Europeana</i>	11
<i>Jable: Archivo de Prensa Digital de la Universidad de las Palmas de Gran Canaria</i>	11
<i>Memoria de Madrid</i>	11
<i>Cartoteca Digital de l'Institut Cartogràfic de Catalunya</i>	10
<i>DUGi Fons Especials: Repositori Digital de la UdG</i>	9
<i>Biblioteca Virtual de Andalucía</i>	8

<i>Biblioteca Digital de la Comunidad de Madrid</i>	7
<i>Biblioteca Virtual del Principado de Asturias</i>	7
<i>Fons Sol-Torres del Servei de Biblioteca i Documentació de la Universitat de Lleida</i>	7
<i>Biblioteca Digital de la Región de Murcia</i>	6
<i>Fototeca del Patrimonio Histórico</i>	6
<i>Fondo Histórico Digital de La Rábida de la Universidad Internacional de Andalucía</i>	6
<i>UVaDOC: repositorio documental de la Universidad de Valladolid (*)</i>	5
<i>Liburutegi Digitala de la Biblioteca Foral de Bizkaia</i>	4
<i>Biblioteca Virtual de La Rioja</i>	4
<i>Fondo Antiguo Universidad de Zaragoza (**)</i>	4
<i>Recursos de Investigación de la Alhambra</i>	4
<i>RIUMA. Universidad de Málaga. Patrimonio (***)</i>	4
<i>Somni. Fondo histórico de la Universitat de València (*)</i>	4
<i>Biblioteca Digital de la Real Academia de Historia</i>	3
<i>Minerva: Repositorio Institucional da Universidade de Santiago de Compostela (**)</i>	3
<i>Biblioteca Virtual de la Diputación de Zaragoza</i>	2
<i>Hispana</i>	2
<i>RUIdeRA: repositorio institucional de la UCLM (*)</i>	2
<i>Almirall, portal del pensament i cultura del segle XIX</i>	1
<i>Archivo del Ateneo de Madrid</i>	1
<i>Biblioteca Dixital de Galicia (Cidade de Cultura Galega)</i>	1
<i>Biblioteca Digital del CEU (*)</i>	1
<i>Galiciana: Biblioteca Dixital de Galicia</i>	1
<i>Iuris Digital: Biblioteca Virtual de la Real Academia de Jurisprudencia y Legislación</i>	1
<i>Teatro Siglo de Oro de la Biblioteca Nacional de España</i>	1
14 bibliotecas digitales no tienen enlaces en ninguna de las dos Wikipedias (ver apéndice 2)	

Tabla 3. Enlaces en conjuntos en *Wikipedia* en español y en catalán (WP-ES + WP-CAT)

La distribución de resultados muestra una clara disparidad en el caso de la *Biblioteca Virtual Miguel de Cervantes*, que acumula hasta el 67 % de los enlaces en WP-ES y el 47 % en WP-CAT. En las siguientes tablas de estudio de la frecuencia de enlaces no incluiremos los datos de esta biblioteca digital, sobre la que más adelante profundizaremos mediante análisis ad hoc. Eso permite tener una visión más ajustada del uso de los muchos repositorios culturales que han proliferado en los últimos años.

La distribución de enlaces por categorías de bibliotecas digitales, teniendo en cuenta que la muestra no es exhaustiva para colecciones de carácter local e instituciones culturales, es la siguiente. Se toman los datos conjuntos de WP-ES y WP-CAT.

Tipos de biblioteca digital	Enlaces	%
Regionales	899	35 %
Instituciones	666	26 %
Estatales	615	24 %
Universidades ⁸	264	10 %
Locales	111	4 %
Europeas	44	2 %

Tabla 4. Distribución de enlaces por tipo de Biblioteca Digital (WP-ES + WP-CAT)

En WP-ES, 6 de las 80 bibliotecas digitales estudiadas concentran el 50 % de los enlaces recogidos: *Biblioteca Digital Hispánica* y *Hemeroteca Digital* de la Biblioteca Nacional de España; *Biblioteca Digital de Castilla y León*; *Biblioteca Digital del Real Jardín Botánico*; *Biblioteca Virtual de Derecho Aragonés*; *Dispòsit digital de documents de la UAB (***)*; *Biblioteca Digital Leonesa*.

El 76 % se alcanza con sólo 17 repositorios: *Biblioteca Virtual de Prensa Histórica*; *Pares (Portal de Archivos Españoles)*; *Biblioteca Digital de Castilla-La Mancha*; *Fondo Documental Histórico de las Cortes de Aragón*; *Biblioteca Saavedra Fajardo de Pensamiento Político Hispánico*; *Fondo antiguo de la Universidad de Alcalá (***)*; *Biblioteca Virtual del Patrimonio Bibliográfico*; *Fondo Antiguo y Archivo Histórico de la Universidad Sevilla*; *Colección Digital Complutense*; *Biblioteca Digital del repositorio Gredos de la Universidad Salamanca (*)*.

Es decir, en WP-ES un 21 % de las bibliotecas digitales reciben el 76 % de los enlaces, lo que supone valores cercanos a la distribución de Pareto. Un 80 % de las bibliotecas digitales tiene escasa presencia.

Para WP-CAT la concentración de enlaces es mayor, 6 repositorios acumulan el 76 % de los enlaces: *Memòria Digital de Catalunya*; *Biblioteca Virtual de Derecho Aragonés*; *Dispòsit digital de documents de la UAB (***)*; *ARCA (Arxiu de Revistes Catalanes Antiques)*; *Biblioteca Digital del Real Jardín Botánico*; *Fondo Documental Histórico de las Cortes de Aragón*.

Para WP-CAT encontramos que muchas de las 10 primeras bibliotecas digitales (excluyendo la *Biblioteca Virtual Miguel de Cervantes* y *Dipòsit digital de documents de la UAB*) son de poco uso en WP-ES, especialmente la *Memòria Digital de Catalunya* y *ARCA (Arxiu de Revistes Catalanes Antiques)*, con un 87 % y un 79 % de enlaces originados desde la enciclopedia en catalán.

Biblioteca digital	% enlaces desde WP-CAT
<i>Memòria Digital de Catalunya</i>	87
<i>Biblioteca Virtual de Derecho Aragonés</i>	53
<i>ARCA (Arxiu de Revistes Catalanes Antiques)</i>	94
<i>Biblioteca Digital del Real Jardín Botánico</i>	17
ANC (Arxiu Nacional Catalunya)	79
<i>Fondo Documental Histórico de las Cortes de Aragón</i>	36
<i>Biblioteca Virtual de Prensa Histórica</i>	24
<i>Biblioteca Digital Hispánica y Hemeroteca Digital de la Biblioteca Nacional de España</i>	5
<i>BiValdi (Biblioteca Valenciana Digital)</i>	57
<i>Biblioteca Saavedra Fajardo de Pensamiento Político Hispánico</i>	18

Tabla 5. Distribución de enlaces entre WP-CAT y WP-ES para el *top ten* de enlaces en WP-CAT

Encontramos algunas características en el formato de los enlaces. A primera vista destaca la abundancia de ejemplos de mala praxis para URL limpias y Cool URIs (abundancia de variables, controladores internos, etc.).⁹ La extensión media de las cadenas de la URL, sin contar el dominio, es de 56 caracteres en WP-ES y 55 en WP-CAT. Apenas hay uso en *Wikipedia* de la URL recomendada en los repositorios conforme al sistema [Handle](#). Las URL Handle tienen 11 caracteres (una extensión similar a la usada por un servicio masivo como Youtube). Algunos repositorios como la *Biblioteca Digital Hispánica* doblan el valor medio (111 caracteres) aunque su recomendación de enlace permanente es de 25. Muy pocos casos hacen uso de URL amigables.

En cuanto al número de enlaces rotos, hemos encontrado que representan en torno al 4 % en *Wikipedia* en español y el 12 % en *Wikipedia* en catalán, siendo los errores más frecuentes el de "página no encontrada" (error 404) y el de "tiempo expirado" (consideramos 25 segundos como tiempo máximo de espera). Los datos completos se presentan en el apéndice 4. La estabilidad de los enlaces es algo fundamental en la red —y pocas veces se cumple (la [vida media de una página web en Internet](#) es de 77 día— e ineludible en los repositorios. Un 96 % de enlaces accesibles en el caso de la española, confirma la estabilidad, pero debería estudiarse el origen de los errores.

Los enlaces pueden incluirse en los artículos bien como nota final (referencias), bien como enlaces en sección independiente (ya sea en "Enlaces externos" o "Bibliografía"). Para realizar el cálculo de enlaces hemos utilizado dos *scripts* diferentes. Aunque finalmente el más eficiente de ellos no permitía detectar la ubicación del enlace dentro de cada artículo, los cálculos realizados con el otro indicaban que un poco más de la mitad de los enlaces aparecen dentro de notas (con la etiqueta `<ref></ref>`), referenciando datos específicos del artículo (en *Wikipedia* en catalán un 64,47 % y en *Wikipedia* en español un 56,28 %).

No podemos dejar de destacar los altos resultados obtenidos por las dos fuentes que contienen recursos de hispanoamérica, *Biblioteca Virtual Miguel de Cervantes* y *Memoria de Chile*. Por otro lado, los datos de los dos grandes agregadores analizados, *Europeana* e *Hispana*, son muy bajos (11 y 2 enlaces, respectivamente). Pese a acumular gran cantidad de registros, son poco usados directamente para enlazar.

3.1 El caso de la *Biblioteca Virtual Miguel de Cervantes* (BVMdC)

Con los 4.372 en WP-ES, y 786 en WP-CAT, esta biblioteca se sitúa entre los sitios webs con más enlaces en *Wikipedia*, en la posición 69. La enorme diferencia en las cifras de enlaces, como ya hemos comentado anteriormente, nos permite avanzar en un análisis más detallado sobre las prácticas de enlazado de los editores de *Wikipedia*. La BVMdC es, además, una de las pioneras en digitalización a gran escala de textos completos en español.

Los resultados relativos al total de artículos para WP-ES son más del doble que para WP-CAT (0,49 % frente a 0,21 %). Usaremos sólo los enlaces de la edición en español para describir las pautas de enlazado.

La estabilidad de las URL es importante en los repositorios, orientados a ofrecer un acceso permanente y único a los recursos. Sin embargo es frecuente que los objetos digitales puedan ser enlazados de varias maneras diferentes o que incluso las URL visibles para el usuario no sean adecuadas para el enlazado.

La mayoría de los enlaces de la BVMdC usan el dominio estándar de la web "www.cervantesvirtual.com". Pero un significativo 15 % usa la forma "descargas.cervantesvirtual.com" y un 4 % usa el subdominio "bib.cervantesvirtual.com".

Solo 22 enlaces son genéricos a la página principal de la BVMdC. El resto son vínculos directos a documentos y por lo tanto suponemos una mayor precisión para apoyar la fiabilidad del contenido de los artículos.

Se detectan enlaces antiguos que, no obstante, funcionan correctamente al ser transformados por el servidor conforme a la nomenclatura actual de URL del sitio.¹⁰ Se percibe la evolución hacia URL amigables.

El término enlace no es exclusivamente usado como enlace al documento, y además adopta diferentes formas. El 4 % identifica en el enlace las *bibliotecas de autor*, manteniendo en la URL la referencia a la forma de organización de la colección por autores.¹¹ El 6 % de los enlaces son a información de autores, en lugar de a obras. Un 2 % son a contenidos de las secciones específicas de historia.¹² Un 26 % enlaza directamente a documentos PDF y no a páginas web convencionales —sean fichas informativas o textos fuente publicados en HTML.

La BVMdC se organiza en portales temáticos e institucionales. Sin embargo la gran mayoría de enlaces usados en *Wikipedia* no permiten identificar si el documento está englobado en uno de ellos. Al enlazar se pierde parte de esa portalización que resalta el papel de las instituciones que colaboran con sus fondos. Sólo el 6 % de los enlaces hacen referencia a "portal". Los más frecuentes son el portal 0 (37 %), portal 33 (16 %) y portal de Chile (6 %).¹³

Se detectan hasta 20 formas diferentes de construcción de URL. Puede deberse tanto a cambios en la tecnología del servidor o a necesidades de enlace específicas de los usuarios (elección entre obras, autores, materias, páginas informativas, enlaces directos al PDF).

```
/autor/  
bib_autor  
/buscador/  
Buscar.html  
/descargaPdf/  
/escaparate/  
FichaAutor.html  
FichaClasificacionMaterias  
FichaMateria  
FichaObra
```

/obra/ /obra-visor /obras-autor/ Hemeroteca Historia /portal/ BVMC_OBRAS /servlet/Ficha servlet/MuestraForo servlet/SirveObra
--

Figura 1. Principales cadenas en los enlaces a la *Biblioteca Virtual Miguel de Cervantes*

4 Discusión de los resultados

Aunque es arriesgado establecer patrones explicativos para los resultados obtenidos, parece que el valor de los documentos pueda tener tanta importancia como la cantidad de ítems de cada repositorio. Hay algunas cuestiones que merecerían una mayor profundización:

1. ¿Por qué un repositorio muy especializado como el del Real Jardín Botánico tiene tan buenos resultados? La variable coherencia temática de la colección podría estar influyendo en el incremento de su uso y visibilidad.
2. ¿Qué patrones de conducta serían esperables en los editores de artículos de *Wikipedia* para la cita de documentos tan especializados como los de las bibliotecas digitales? Si vemos *Wikipedia* como una enciclopedia generalista, los documentos de las bibliotecas digitales resultan mucho menos adecuados que manuales y obras de síntesis. Si la vemos también como una enciclopedia especializada que abarca cada vez temas más delimitados, estos documentos tienen un mayor valor potencial, y por lo tanto sería razonable esperar encontrarlos citados.
3. ¿Existe relación entre diferentes arquitecturas de publicación digital de contenidos y la frecuencia de citas en *Wikipedia*? Es difícil captar en una investigación cómo una publicación digital con una mejor experiencia de usuario afectaría a su uso desde *Wikipedia*.
4. ¿Qué explicación podemos encontrar a los bajos resultados obtenidos por agregadores como *Hispana* o *Europeana*? Esto puede hacernos reflexionar sobre el nivel de madurez o lo adecuado de la estrategia informativa de los proyectos.

Si aceptamos que la cita en *Wikipedia* es otro indicador de uso de los contenidos de una biblioteca digital, sería útil poder conocer mejor qué factores externos al contenido mismo influyen en el uso. Sin duda aspectos como la calidad de la experiencia de usuario, la organización mediante narrativas digitales atractivas, o la optimización para buscadores, han de tener su influencia en el uso de los repositorios digitales.

La espectacular diferencia de resultados de la *Biblioteca Virtual Miguel de Cervantes* merece una consideración aparte. Es quizá el proyecto con más antigüedad en nuestro país, además de estar ampliamente vinculado con la comunidad iberoamericana, lo cual amplía su alcance y posibilidades. El factor hispanoamericano merecería atención, dado que el proyecto *Memoria Chilena* presenta también resultados muy superiores a los valores medios de las bibliotecas digitales españolas. Además de un volumen considerable de contenidos, ha sido valorada como "un ejercicio permanente de curaduría digital" (Abbagliati, 2011). Ambas han enfocado su organización de contenidos con arquitecturas de información orientadas a portales temáticos y no exclusivamente al repositorio.

Las bibliotecas digitales regionales en conjunto tienen un buen comportamiento, lo que podría indicar cierta preferencia por una agrupación de contenidos locales frente a los grandes proyectos nacionales. También da qué pensar el escaso uso de la *Biblioteca Virtual de Prensa Histórica*, cuando los archivos retrospectivos del *ABC* y *La Vanguardia* son usados frecuentemente. Por otro lado, no se aprecian diferencias a favor de *Viquipèdia*, aunque ha realizado regularmente actividades de concienciación y colaboración GLAM.

5 Conclusiones

Los datos de uso de enlaces a repositorios digitales de patrimonio cultural son, desde cualquier punto de vista, muy bajos. Cabría preguntarse si se trata de una limitación propia de la fuente (documentos históricos) que reduce su potencial de uso en la escritura de artículos enciclopédicos. Quizá las fuentes primarias documentales y bibliográficas sean de interés para la investigación, pero menos para la elaboración de contenidos informativos como los de *Wikipedia*. Sin embargo muchos artículos de *Wikipedia* son homologables a los de una enciclopedia especializada —incluso algunos artículos tratan sobre alguno de los documentos presentes en bibliotecas digitales— donde sí sería esperable el enlace a fuentes primarias.

Estos datos pueden tomarse como un indicador de falta de visibilidad real de los contenidos patrimoniales, y de escasa conexión con una de las comunidades digitales más activas en la difusión y organización de conocimiento. Los editores de *Wikipedia* muestran un interés por encima de la media en fuentes de calidad para mejorar sus artículos, y sin embargo parece que conocen o manejan poco las numerosas bibliotecas digitales que, a priori, serían muy útiles para enriquecer los artículos de ámbito histórico-cultural.

Podría haber muchas y diferentes acciones para estimular el uso de bibliotecas digitales en *Wikipedia*. Algunas son relativas al propio sitio web de la biblioteca digital y otras se refieren a la promoción de estas bibliotecas digitales entre los wikipedistas. Desde el punto de vista de las bibliotecas digitales podría ser útil:

1. Mejorar la construcción de URL, facilitando la identificación de *cool URIs* o *permalinks*, estables en el tiempo, sencillas y cortas.
2. Incorporar un *widjet* "citar en *Wikipedia*" a semejanza de *Europeana*.
3. Mejorar los sistemas de promoción de contenidos destacados, de forma que no sea necesario extraer los documentos del repositorio para difundirlos en los portales temáticos.
4. Mejorar la experiencia de usuario y usabilidad del sitio.
5. Mejorar el posicionamiento en buscadores, especialmente en Google.
6. Integrar los repositorios en narrativas digitales con mayor potencial de conexión con los usuarios. Por ejemplo, realizando exposiciones virtuales conectadas con los repositorios.

Este último punto está presente en las iniciativas puestas en marcha en *Europeana* para integrar la información aportada por usuarios (*user generated content*) y nuevas narrativas digitales (*digital storytelling*). El objetivo es conectar con los usuarios digitales e impulsar el uso creativo de sus recursos. La creación de exposiciones digitales utilizando los objetos contenidos en las bibliotecas aprovecha estrategias de corte museográfico, didáctico o de comunicación social (Saorín, 2011).

La combinación de todas estas acciones sugiere nuevas líneas de investigación para establecer los mejores modelos para publicar bibliotecas digitales que, además de los requisitos técnicos internos, incorporen la mejora de la experiencia de usuario a través de la funcionalidad, accesibilidad, *findability*, utilidad, estética, credibilidad y usabilidad (Hassan Montero, 2006).

Desde el punto de vista de la comunidad de *Wikipedia*, sería interesante que el sector de los museos, archivos y bibliotecas (GLAM) adoptara un papel más activo, promoviendo el uso de las fuentes digitales en artículos y preocupándose por los artículos y wikiproyectos más relacionados con sus temas de interés. La figura del wikipedista en residencia puede ser complementada incluyendo *Wikipedia* en las rutinas de trabajo, como plataforma adecuada para cumplir su misión de difusión del patrimonio.

En definitiva, este estudio cuenta con frentes muy interesantes que pueden ser abordados en el futuro. En primer lugar habría que ampliar la lista de repositorios con los del resto de países de la comunidad lingüística española, e incluso algunos otros internacionales. Por otro lado es importante incluir *Wikimedia Commons* y *Wikisource* para este tipo de análisis, puesto que en ellos se recogen fuentes gráficas y documentales. La naturaleza de muchos de los recursos de estas bibliotecas digitales es dual: sirven como enlace a la fuente y para ilustrar el artículo. Documentos históricos, mapas, pinturas, objetos culturales, etc., son frecuentemente usados desde el Banco de imágenes *Wikimedia Commons*. Queremos resaltar que las limitaciones de licencia de uso libre que tienen en su mayoría estos repositorios —además de frecuentes marcas de agua— excluyen radicalmente su inclusión en Commons. Este aspecto podría estar minando las opciones de ser enlazados, al no responder a una necesidad muy determinada de los editores de *Wikipedia*.

Finalmente, consideramos que los datos de este estudio pueden servir para evaluar de forma periódica los resultados de las acciones que se hayan emprendido para mejorar la difusión y estimular el uso de las bibliotecas digitales.

Bibliografía

- Abbagliati, Enzo (2011). "Portal Memoria Chilena: un ejemplo de curaduría de contenidos digitales". *Hlb: revista de historia Iberoamericana*, vol. 4, nº 2. <http://revistahistoria.universia.cl/pdfs_revistas/articulo_166_1325114205359.pdf>. [Consulta: 24/05/2012].
- Codina, Lluís (2006). "Metodología de análisis y evaluación de recursos digitales en línea: Parte II: Parámetros e Indicadores (v. 2006)". *Análisis y métodos en ciencias de la documentación*. <<http://www.lluiscodina.com/metodos.htm>>. [Consulta: 24/05/2012].
- Codina, Lluís; Marcos, Mari Carmen (2005). "Posicionamiento web: conceptos y herramientas". *El profesional de la información*, vol. 14, nº 2 (marzo-abril), p. 84-99, <<http://www.elprofesionaldeinformacion.com/contenidos/2005/marzo/1.pdf>>. [Consulta: 24/05/2012].
- Elder, Danielle; Westbrook, R. Niccole; Reilly, Michele (2012). "Wikipedia lover, not a hater: harnessing Wikipedia to increase the discoverability of library resources". *Journal of web librarianship*, vol. 6, issue 1, p.32-44.
- Estivill Rius, Assumpció; Gascón García, Jesús; Sulé Duesa, Andreu (2010). "Las colecciones digitales patrimoniales españolas: políticas de colección y presentación de la colección". *BiD: textos universitaris de biblioteconomia i documentació*, núm. 25 (juny). <<http://www.ub.edu/bid/25/estivill2.htm>>. [Consulta: 24/05/2012].
- Ferrer-Sapena, Antonia; Peset, Fernanda (2012). "La reutilización de datos culturales". *Anuario ThinkEPI*, vol. 6, p. 193-196.
- Guallar, Javier; Abadal, Ernest (2009). "Evaluación de hemerotecas de prensa digital: indicadores y ejemplos de buenas prácticas". *El profesional de la información*, vol. 18, nº 3 (mayo-junio), p. 255-269. <<http://hdl.handle.net/10760/13048>>. [Consulta: 24/05/2012].
- Hassan Montero, Yusef (2006). "Factores del diseño web orientado a la satisfacción y no-frustración de uso". *Revista española de documentación científica*, vol. 29, nº 2, p. 239-257. <<http://redc.revistas.csic.es/index.php/redc/article/view/291/353>>. [Consulta: 24/05/2012].
- Huvila, Isto (2010). "Where does the information come from? Information source use patterns in Wikipedia". *Information research*, vol.15, no. 3 (September). <<http://informationr.net/ir/15-3/paper433.html>>. [Consulta: 24/05/2012].
- Lallye, Anne M.; Dunford, Carolyn E. (2007). "Using Wikipedia to extend digital collections". *D-Lib magazine*, vol. 13, no. 5-6 (May-June). <<http://www.dlib.org/dlib/may07/lallye05lallye.html>>. [Consulta: 24/05/2012].
- Luyt, Brendan; Tan, Daniel (2010). "Improving Wikipedia's credibility: references and citations in a sample of history articles". *Journal of the American Society for Information Science and Technology*, vol. 61, no. 4, p. 715-722. <<http://onlinelibrary.wiley.com/doi/10.1002/asi.21304/pdf>>. [Consulta: 24/05/2012].
- Nielsen, Finn Årup (2007). "Scientific citations in Wikipedia". *First Monday*, vol. 12, no. 8 (August). <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1997/1872>>. [Consulta: 24/05/2012].
- Saorín, Tomás (2011). "Exposiciones digitales y reutilización: aplicación del software libre Omeka para la publicación estructurada". *Métodos de información*, vol. 2, nº 2, p.. 29-46. <<http://www.metodosdeinformacion.es/mei/index.php/mei/article/viewFile/IIMEI2-N2-029046/740>>. [Consulta: 24/05/2012].
- Smith-Yoshimura, Karen; Rose Holley (2012). *Social metadata for libraries, archives, and museums. Part 3: Recommendations and readings*. Dublin, Ohio: OCLC Research. <<http://www.oclc.org/research/publications/library/2012/2012-01.pdf>>. [Consulta: 24/05/2012].
- Verwayen, Harry; Arnoldus, Martijn; Kaufman, Peter B. (2011). *The problem of the yellow milkmaid: a business model perspective on open metadata*. *Europeana White Paper*, no. 2 (November).

<<http://pro.europeana.eu/documents/858566/2cbf1f78-e036-4088-af25-94684ff90dc5>>. [Consulta: 24/05/2012].

Fecha de recepción: 15/05/2012. Fecha de aceptación: 28/05/2012.

Notas

¹ *Europeana Facts & Figures*, April 2012: <<http://pro.europeana.eu/web/guest/about/facts-figures>>. [Consulta: 24/05/2012].

² *Dump* de *Wikipedia* en español de 20 de abril de 2012 (más de 885.000 artículos). *Dump* de *Wikipedia* en catalán de 19 de abril de 2012 (más de 372.000 artículos).

³ Nos ha sido difícil identificar si se han producido cambios en el formato de las URL durante los últimos años, por lo que de haber sucedido, podría haber enlaces no identificados en nuestro estudio.

⁴ Un caso paradigmático podrían ser las Capitulaciones de Santa Fe, que el Ministerio de Cultura destaca en "Novedades Pares" al ser incluidas por la Unesco como "[Memoria del mundo](#)". Una cita a este documento no contabilizaría en nuestra recogida de datos.

⁵ Que son la mayoría de los recogidos en [BuscaRepositorios](#) y recolectados en [Recolecta](#).

⁶ Por ejemplo, la colección de 341 obras de la *Biblioteca Digital Floridablanca* dentro del repositorio [Digitum](#) de la Universidad de Murcia, o la colección Somni del repositorio [Roderic](#) de la Universitat de València.

⁷ Los repositorios mixtos vienen marcados con un asterisco (*) cuando el porcentaje de la colección patrimonial es muy significativo (más del 50 %), doble asterisco (**) cuando es mayor del 25 % y triple cuando es menor. Como las cifras son bajas, usamos el dato total de enlaces. En el apéndice 1 ofrecemos los valores resultantes de aplicar la ponderación calculada para cada repositorio.

⁸ Usamos la estimación de enlaces consecuencia de los criterios de ponderación de colecciones patrimoniales (ver apéndice 1).

⁹ Tomando las recomendaciones de Cool URIs don't change: <<http://www.w3.org/Provider/Style/URI>>. [Consulta: 24/05/2012].

¹⁰ Por ejemplo: <<http://www.cervantesvirtual.com/FichaObra.html?portal=0&Ref=8183&video=1se>> se transforma en una URL más descriptiva; <<http://www.cervantesvirtual.com/obra/conde-de-gondomar-en-la-real-biblioteca--0/>>. [Consultas: 24/05/2012].

¹¹ Ejemplo: <http://bib.cervantesvirtual.com/bib_autor/alfonsoelsabio/>. [Consulta: 24/05/2012].

¹² Ejemplo: <<http://www.cervantesvirtual.com/historia/textos/medieval/>>. [Consulta: 24/05/2012].

¹³ Ejemplo: <<http://www.cervantesvirtual.com/FichaObra.html?Ref=1102&portal=157>>. [Consulta: 24/05/2012].

Apéndice 0. Bibliotecas digitales seleccionadas

- Ámbito europeo (EU):
 1. **Europeana**
<http://www.europeana.eu>
<http://exhibitions.europeana.eu>
<http://www.europeanaregia.eu>
 2. **Portal Europeo de Archivos**
<http://www.archivesportaleurope.eu>
 3. **The European Library**
<http://www.theeuropeanlibrary.org>

- **Ámbito estatal español (ES):**
 1. **Biblioteca Digital Hispánica y Hemeroteca Digital de la Biblioteca Nacional de España**
<http://bdh.bne.es>
<http://bibliotecadigitalhispanica.bne.es>
<http://hemerotecadigital.bne.es>
 2. **Biblioteca Virtual de Prensa Histórica**
<http://prensahistorica.mcu.es>
 3. **Biblioteca Virtual del Patrimonio Bibliográfico**
<http://bvpb.mcu.es>
 4. **Red Digital de Colecciones de Museos de España**
<http://ceres.mcu.es>
 5. **Pares (Portal de Archivos Españoles)**
<http://pares.mcu.es>
 6. **Fototeca del Patrimonio Histórico**
http://www.mcu.es/fototeca_patrimonio/
<http://ipce.mcu.es/documentacion/fototeca/fondos>
 7. **Hispana**
<http://roai.mcu.es>

- **Ámbito regional (REG):**
 1. **Biblioteca Digital de la Región de Murcia**
<http://bibliotecadigital.carm.es>
 2. **Biblioteca Virtual de Andalucía**
<http://www.bibliotecavirtualdeandalucia.es>
 3. **Biblioteca Virtual de Aragón**
<http://bibliotecavirtual.aragon.es>
 4. **Fondo Documental Histórico de las Cortes de Aragón**
<http://www.cortesaragon.es/fileadmin/ESTATICAS/fondoAntiguo.htm>
 5. **Biblioteca Digital de Castilla y León**
<http://bibliotecadigital.jcyl.es>
 6. **Archivo de la imagen de Castilla-La Mancha**
http://clip.jccm.es/archivo_de_la_imagen
 7. **Biblioteca Digital de Castilla-La Mancha**
<http://clip.jccm.es/bidicam/es/>
 8. **Biblioteca Digital de la Comunidad Madrid**
<http://www.bibliotecavirtualmadrid.org>
 9. **Memòria Digital de Catalunya**
<http://mdc.cbuc.cat>
 10. **ARCA (Arxiu de Revistes Catalanes Antiques)**
<http://mdc2.cbuc.cat>
 11. **BiValdi (Biblioteca Valenciana Digital)**
<http://bv2.gva.es>
 12. **BINaDI. Biblioteca Digital de Navarra**
<https://administracionelectronica.navarra.es/binadi>
http://www.navarra.es/appsext/bnd/GN_Ficheros_PDF_Binadi.aspx
 13. **Liburuklik: Biblioteca Digital Vasca**
<http://www.liburuklik.euskadi.net>
 14. **Memoria Digital Vasca**
<http://www.memoriadigitalvasca.es>
 15. **Galiciana: Biblioteca Dixital de Galicia**
<http://galiciana.bibliotecadegalicia.xunta.es>
 16. **Biblioteca Dixital de Galicia (Cidade de Cultura Galega)**
<http://www.csborg.org/bibliotecadixital>
<http://www.csborg.org/prensagalega>
<http://www.cmg.xunta.es/visor/prensa/>
<http://www.cmg.xunta.es/mediateca/hemeroteca/>
http://bvg.centromultimedia.net/search*gag/fcmapa*/
<http://www.cmg.xunta.es/mediateca/cartografia/>
 17. **Memoria Digital de Canarias**
<http://mdc.ulpgc.es>
 18. **Archivos Históricos de la Región de Murcia. Proyecto Carmesí**
<http://www.regmurcia.com/servlet/s.SI?METHOD=FRMSENCILLA2&sit=c.373.m.139.serv.Carmesi>

19. **Biblioteca Virtual del Principado de Asturias**
<http://www.bibliotecavirtual.asturias.es>
20. **Biblioteca Virtual de La Rioja**
<http://bibliotecavirtual.larioja.org>
21. **ANC (Arxiu Nacional Catalunya)**
<http://cultura.gencat.net/anc/>
22. **Documentos y Archivos de Aragón**
<http://www.sipca.es/dara/>

Sin ánimo de ser exhaustivos, hemos recogido unas pocas bibliotecas digitales de carácter local o provincial (LOC):

1. **Biblioteca Virtual de la Diputación de Zaragoza**
<http://www.bivizar.es>
2. **Biblioteca Digital de la ciudad de Murcia**
<http://www.murcia.es/bibliotecadigitaldemurcia/>
3. **Biblioteca Digital Leonesa**
<http://www.saber.es>
4. **Memoria de Madrid**
<http://www.memoriademadrid.es>
5. **Liburutegi Digitala de la Biblioteca Foral de Bizkaia**
http://www.bizkaia.net/kultura/foru_liburutegia/bibdigital_spi.asp?Tem_Codigo=2542&Idioma=EU

Hemos seleccionado alguna de las colecciones de fondo antiguo más relevantes de universidades españolas (UNI):

1. **Fondo Antiguo Universidad de Zaragoza (31 %)**
<http://zagan.unizar.es>
2. **Digitum. Fondo antiguo de la Universidad de Murcia (5 %)**
<http://digitum.um.es>
<http://hdl.handle.net/10201/>
3. **Somni. Fons històric de la Universitat de València (51 %)**
<http://roderic.uv.es>
<http://hdl.handle.net/10550/>
4. **Fondo antiguo de la Universidad de Granada (62 %)**
<http://digibug.ugr.es>
<http://hdl.handle.net/10481/>
5. **Biblioteca Digital del repositorio Gredos de la Universidad de Salamanca (61 %)**
<http://gredos.usal.es>
<http://hdl.handle.net/10366/>
6. **Fondo antiguo de la Universidad Sevilla**
<http://fondosdigitales.us.es/fondos/>
7. **Col·leccions digitals de la Universitat de Barcelona**
<http://www.bib.ub.edu/recursos-informacio/colleccions/colleccions-digitals/>
8. **Biblioteca Digital Complutense**
<http://alfama.sim.ucm.es>
9. **Fondo Fotográfico de la Universidad de Navarra**
<http://coleccionfff.unav.es/bvunav>
10. **Dipòsit digital de documents de la UAB (19 %)**
<http://ddd.uab.cat>
11. **Biblioteca Digital de la UIMP**
<http://www.bduimp.es/archivo>
12. **DUGi Fons Especials: Repositori Digital de la UdG**
<http://dugifonsespecials.udg.edu>
<http://hdl.handle.net/10256.2/>
13. **e_Buah de la Universidad de Alcalá (3 %)**
<http://dspace.uah.es/>
<http://hdl.handle.net/10017/>
14. **Helvia: Repositorio Institucional de la Universidad de Huelva (32 %)**
<http://helvia.uco.es>
<http://hdl.handle.net/10396/>
15. **Jable: Archivo de Prensa Digital de la Universidad de las Palmas de Gran Canaria**
<http://jable.ulpgc.es>
16. **Biblioteca Digital de Castellón (69 %)**

- <http://repositori.uji.es>
<http://hdl.handle.net/10234/>
17. **Repositori Obert UdL. Fons Especials**
<http://repositori.udl.cat/handle/10459.2/>
<http://hdl.handle.net/10459.2/>
 18. **Minerva: Repositorio Institucional da Universidade de Santiago de Compostela (34 %)**
<http://dspace.usc.es>
<http://hdl.handle.net/10347/>
 19. **Fondo Histórico Digital de La Rábida de la Universidad Internacional de Andalucía (73 %)**
<http://dspace.unia.es>
<http://dspace.unia.es/handle/10334/105>
 20. **RIUMA. Universidad de Málaga. Patrimonio (14 %)**
<http://riuma.uma.es>
<http://hdl.handle.net/10630/>
 21. **Rodin. Universidad de Cádiz. Patrimonio bibliográfico (42 %)**
<http://rodin.uca.es>
<http://hdl.handle.net/10498/>
 22. **Repositorio da Universidade da Coruña (12 %)**
<http://ruc.udc.es>
<http://hdl.handle.net/2183/>
 23. **RUIdeRA: repositorio institucional de la UCLM (12 %)**
<https://ruidera.uclm.es>
<http://hdl.handle.net/10578/>
 24. **UVaDOC: repositorio documental de la Universidad de Valladolid (74 %)**
<http://uvadoc.uva.es>
 25. **Arias Montano: repositorio institucional de la Universidad de Huelva (25 %)**
<http://rabida.uhu.es/dspace/>
<http://hdl.handle.net/10272/>
 26. **Biblioteca Digital del CEU (8 %)**
<http://dspace.ceu.es>
<http://hdl.handle.net/10637/>
 27. **Fons Sol-Torres del Servei de Biblioteca i Documentació de la Universitat de Lleida**
<http://soltorres.udl.cat>
<http://hdl.handle.net/10459/>
 28. **DigiBUO: Biblioteca Digital de la Universidad de Oviedo**
<http://digibuo.sheol.uniovi.es/>

Hemos seleccionado un conjunto diverso de colecciones digitalizadas de instituciones culturales de diferente naturaleza y alcance (INS):

1. **Biblioteca Virtual de Derecho Aragonés**
<http://www.derechoaragones.es/es/estaticos/contenido.cmd?pagina=estaticos/indice>
2. **Teatro Siglo de Oro de la Biblioteca Nacional de España**
<http://teatrosiglodeoro.bne.es/es/RecursosBibliograficos/indexEstudios.html>
3. **Biblioteca Virtual Miguel de Cervantes**
<http://www.cervantesvirtual.com>
4. **Biblioteca Digital Ateneo de Madrid**
http://www.ateneodemadrid.com/biblioteca_digital/
5. **Biblioteca Digital del Real Jardín Botánico**
<http://bibdigital.rjb.csic.es>
6. **Archivo del Ateneo de Madrid**
<http://archivo.ateneodemadrid.es>
7. **Biblioteca Virtual Ignacio Larramendi**
<http://www.larramendi.es/i18n/estaticos/contenido.cmd?pagina=estaticos/bibliotecall>
8. **Iuris Digital: Biblioteca Virtual de la Real Academia de Jurisprudencia y Legislación**
<http://bvrajyl.insde.es>
9. **Cartoteca Digital de l'Institut Cartogràfic de Catalunya**
<http://cartotecadigital.icc.cat>
10. **Almirall: portal de pensament i cultura del segle XIX**
<http://194.224.194.201:9080/Almirall/>
11. **Recursos de Investigación de la Alhambra**
<http://www.alhambra-patronato.es/ria/>
<http://hdl.handle.net/10514/>
12. **Repositorio institucional Arjona y Cubas de la Real Academia de Córdoba de Ciencias,**

- Bellas Letras y Nobles Artes**
<http://repositorio.racordoba.es>
<http://hdl.handle.net/10853/>
13. **Biblioteca Virtual Sierra Pambley**
<http://bibliotecavirtualesierrapambley.org>
 14. **Biblioteca Saavedra Fajardo de Pensamiento Político Hispánico**
<http://saavedrafajardo.um.es>
 15. **Biblioteca Virtual de la Real Academia Nacional de Farmacia. España**
<http://bibliotecavirtual.ranf.com>
 16. **Biblioteca Digital de la Real Academia de Historia**
<http://bibliotecadigital.rah.es>

Por último, en la categoría TEST hemos considerado las siguientes colecciones:

1. **Hemeroteca delABC**
<http://hemeroteca.abc.es>
2. **Hemeroteca de La Vanguardia**
<http://hemeroteca.lavanguardia.es>
3. **Memoria de Chile**
<http://www.memoriachilena.cl>
4. **Biblioteca Digital Mundial de la Unesco**
<http://www.wdl.org/es>
5. **Colecciones digitales de la Library of Congress**
<http://chroniclingamerica.loc.gov>
<http://memory.loc.gov>
<http://lcweb2.loc.gov/diglib>
<http://www.loc.gov/pictures/collection>
<http://international.loc.gov>
6. **Google Books**
<http://books.google.com>

Apéndice 1. Datos ponderados para repositorios mixtos

Datos conjuntos para WP-ES y WP-CAT. Junto al nombre del repositorio se indica la ponderación aplicada, calculada a partir del porcentaje de las colecciones patrimoniales sobre el total del repositorio, y el número de enlaces estimados. Los datos de las bibliotecas digitales marcadas con (***) pueden ser poco fiables para colecciones patrimoniales. Por el contrario, las marcadas con (*) tienen una alta probabilidad de ser válidos.

Repositorios mixtos		Total enlaces	% Pond.	Pond. enlaces
UVaDOC: repositorio documental de la Universidad de Valladolid	*	5	74	4
<i>Biblioteca Digital de Castellón</i>	*	22	69	15
Fondo antiguo de la Universidad de Granada	*	30	62	19
<i>Biblioteca Digital del repositorio Gredos de la Universidad Salamanca</i>	*	42	61	26
<i>Somni. Fons històric de la Universitat de València</i>	*	4	51	2
<i>Rodin. Universidad de Cádiz. Patrimonio bibliográfico</i>	**	15	42	6
<i>Minerva: Repositorio Institucional da Universidade de Santiago de Compostela</i>	**	3	34	1
<i>Helvia: Repositorio Institucional de la Universidad de Huelva</i>	**	23	32	7
Fondo Antiguo de la Universidad de Zaragoza	**	4	31	1
<i>Arias Montano: repositorio institucional de la Universidad de Huelva</i>	**	21	25	5
<i>Dipòsit digital de documents de la UAB</i>	***	234	19	44
<i>RIUMA. Universidad de Málaga. Patrimonio</i>	***	4	14	1
<i>Repositorio da Universidade da Coruña</i>	***	18	12	2
<i>RUIdeRA: repositorio institucional de la UCLM</i>	***	2	12	0

Biblioteca Digital del CEU	***	1	8	0
<i>Digitum</i> . Fondo antiguo de la Universidad de Murcia	***	30	5	2
e_Buah de la Universidad de Alcalá	***	48	3	1

Para el recolector *Hispana* hemos realizado una estimación del 30 % y tan sólo recibe 2 enlaces.

Apéndice 2. Bibliotecas digitales sin enlaces desde *Wikipedia*

Un total de 14 bibliotecas digitales no tienen enlaces en ninguna de las dos ediciones (WP-ES y WP-CAT): *Biblioteca Virtual Sierra Pambley*; *Biblioteca Virtual de la Real Academia Nacional de Farmacia*. España; *Repositorio institucional Arjona y Cubas* de la Real Academia de Córdoba de Ciencias, Bellas Letras y Nobles Artes; *Biblioteca Digital de la ciudad de Murcia*; *Archivo de la imagen de Castilla-La Mancha*; *Liburuklik: Biblioteca Digital Vasca*; *Biblioteca Virtual de Aragón*; Documentos y Archivos de Aragón; *Archivos Históricos de la Región de Murcia*. Proyecto Carmesí; *Biblioteca Digital de la UIMP*; *DigiBUO*: Biblioteca Digital de la Universidad de Oviedo; Col·leccions digitals de la Universitat de Barcelona; Fondo Fotográfico de la Universidad de Navarra; *Repositori Obert UdL*. Fons Especials.

16 casos no tienen enlaces en WP-ES: *Portal Europeo de Archivos*; *Almirall: portal de pensament i cultura del segle XIX*; *Repositorio institucional Arjona y Cubas* de la Real Academia de Córdoba de Ciencias, Bellas Letras y Nobles Artes; *Biblioteca Virtual de la Real Academia Nacional de Farmacia*. España; *Biblioteca Virtual Sierra Pambley*; *Cartoteca Digital* de l'Institut Cartogràfic de Catalunya; *Biblioteca Digital de la ciudad de Murcia*; *Documentos y Archivos de Aragón*; *Biblioteca Virtual de Aragón*; *Archivo de la imagen de Castilla-La Mancha*; *Archivos Históricos de la Región de Murcia*. Proyecto Carmesí; *Liburuklik: Biblioteca Digital Vasca*; *DigiBUO*: Biblioteca Digital de la Universidad de Oviedo; Col·leccions digitals de la Universitat de Barcelona; *Repositori Obert UdL*. Fons Especials; Fondo Fotográfico de la Universidad de Navarra y *Biblioteca Digital de la UIMP*.

40 bibliotecas digitales no tienen enlaces en WP-CAT: *Hispana*; *Fototeca del Patrimonio Histórico*; *Portal Europeo de Archivos*; Archivo del Ateneo de Madrid; *Iuris Digital*: Biblioteca Virtual de la Real Academia de Jurisprudencia y Legislación; *Recursos de Investigación de la Alhambra*; *Repositorio institucional Arjona y Cubas* de la Real Academia de Córdoba de Ciencias, Bellas Letras y Nobles Artes; *Teatro Siglo de Oro* de la Biblioteca Nacional de España; *Biblioteca Virtual de la Real Academia Nacional de Farmacia*. España; *Biblioteca Virtual Sierra Pambley*; *Biblioteca Digital* de la Real Academia de Historia; *Memoria de Madrid*; *Biblioteca Virtual de la Diputación de Zaragoza*; *Biblioteca Digital de la ciudad de Murcia*; *Biblioteca Digital de la Comunidad Madrid*; Documentos y Archivos de Aragón; *Biblioteca Virtual de Aragón*; *Archivo de la imagen de Castilla-La Mancha*; *Biblioteca Dixital de Galicia (Cidade de Cultura Galega)*; *Biblioteca Digital de Castilla-La Mancha* *Archivos Históricos de la Región de Murcia*. Proyecto Carmesí; *Biblioteca Digital de la Región de Murcia*; *Liburuklik: Biblioteca Digital Vasca*; *Galiciana: Biblioteca Dixital de Galicia*; *RUIdeRA*: repositorio institucional de la UCLM; *Fondo Histórico Digital de La Rábida* de la Universidad Internacional de Andalucía; Biblioteca Digital del CEU; *Fondo Antiguo Universidad de Zaragoza*; *Minerva: Repositorio Institucional da Universidade de Santiago de Compostela*; *Rodin*. Universidad de Cádiz. Patrimonio bibliográfico; *DigiBUO*: Biblioteca Digital de la Universidad de Oviedo; Col·leccions digitals de la Universitat de Barcelona; e_Buah de la Universidad de Alcalá; *Repositori Obert UdL*. Fons Especials; Fondo Fotográfico de la Universidad de Navarra; *UVaDOC*: repositorio documental de la Universidad de Valladolid; *Biblioteca Digital de la UIMP*; *Arias Montano*: repositorio institucional de la Universidad de Huelva; *RIUMA*. Universidad de Málaga. Patrimonio; *Jable: Archivo de Prensa Digital* de la Universidad de las Palmas de Gran Canaria.

Apéndice 3. Detalles del proceso de selección de enlaces

Abundando en el aspecto técnico del análisis, hemos utilizado los [dumps](#) de enlaces (*-[externallinks.sql.gz](#)) que proporciona la Fundación Wikimedia en su web. Este fichero incluye todos los enlaces a sitios web externos de *Wikipedia*, insertados en cualquier parte de los artículos. Inicialmente consideramos analizar directamente el texto bruto de los artículos y extraerlos de ahí, pero el análisis con el *dump* de enlaces es más rápido y exacto, si bien es cierto que perdemos información sobre cuándo fue insertado cada enlace.

En el análisis sólo se han tenido en cuenta aquellos enlaces que se encuentran en artículos de la

enciclopedia (conocido como "espacio de nombres" principal), descartando por completo aquellos situados en sus páginas de discusión u otras de mantenimiento.

Los resultados se han exportado a ficheros CSV generando un [dataset](#) con licencia libre, con el ánimo de que nuestros resultados puedan ser analizados y reutilizados por todo aquel que lo desee.

Apéndice 4. Enlaces rotos en WP-ES y WP-CAT

Estado del enlace	Wikipedia en español	Wikipedia en catalán
Correcto (OK)	6.173	1.452
Página no encontrada (error 404)	119	40
error 500	69	37
Tiempo expirado (25 segundos sin respuesta)	68	108
Enlaces mal escritos	18	2
error 504	9	4
error 400	4	1
error 503	0	22
Total	Accesibles: 6.173 (95,56 %) Inaccesibles: 287 (4,44 %)	Accesibles: 1.452 (87,15 %) Inaccesibles: 214 (12,85 %)