



1 de enero de 2017 | Vol. 18 | Núm. 1 | ISSN 1607 - 6079

ARTÍCULO

ESBOZOS SOBRE LA EVOLUCIÓN Y EL FUTURO DE UN PIONERO DE LAS HUMANIDADES DIGITALES HISPÁNICAS: EL PROYECTO PHILOBIBLON

(<http://www.revista.unam.mx/vol.18/num1/art02/>)

Óscar Perea Rodríguez

*(Doctor en Historia Medieval, Universidad Complutense,
Madrid)*

Charles B. Faulhaber

(Doctor en Filología Románica, Yale University)

ESBOZOS SOBRE LA EVOLUCIÓN Y EL FUTURO DE UN PIONERO DE LAS HUMANIDADES DIGITALES HISPÁNICAS: EL PROYECTO PHILOBIBLON*

“...los aspectos que hemos procurado cuidar más es el de la presencia del proyecto en las redes sociales.”

Resumen

El presente artículo es un sucinto repaso del devenir histórico y tecnológico de *PhiloBiblon*, uno de los proyectos pioneros en las humanidades digitales aplicadas al estudio de las fuentes primarias de las literaturas hispánicas e ibéricas escritas durante la Edad Media y el Renacimiento. La historia del proyecto tiene como hilo conductor las diferentes bibliografías de la misma: BETA (Bibliografía Española de Textos Antiguos), BITECA (Bibliografía de Textos Antics Catalans, Valencians i Balears), BITAGAP (Bibliografía de Textos Antigos Galegos e Portugueses) y BIPA (Bibliografía de la Poesía Áurea). Se revisan también los diferentes soportes, del papel impreso a la web, pasando por el *cederrón*, en que la base de datos ha sido publicada. Además, se enseña de forma concisa el funcionamiento del programa, sobre todo en modo consulta a través de la web. Por último, se exponen los objetivos deseados para el futuro de *PhiloBiblon*, en especial su transformación en una base de datos con la tecnología de la web semántica y datos abiertos enlazados.

Palabras clave: *PhiloBiblon*, bases de datos, web semántica, datos abiertos enlazados, humanidades digitales, Edad Media, Renacimiento, impresos, manuscritos, España, Portugal, Península Ibérica.

Outline About the Evolution and Future of a Pioneer of Hispanic Digital Humanities: The Philobiblon Project

This article is a succinct review of the historic and technological development of PhiloBiblon, one of the pioneer Digital Humanities projects as applied to the study of the primary sources of the medieval and early modern literatures of the Iberian peninsula. The guiding thread of its history are the four bibliographies that comprise it: BETA (Bibliografía Española de Textos Antiguos), BITECA (Bibliografía de Textos Antics Catalans, Valencians i Balears), BITAGAP (Bibliografía de Textos Antigos Galegos e Portugueses), and BIPA (Bibliografía de la Poesía Áurea). The various media used for publication of the databases, from paper to the web by way of CD-ROM, are also reviewed. The way in which the program functions, especially on the web, is explained concisely; and, finally, PhiloBiblon's long-term goals are set forth, especially its transformation into a semantic web database using Linked Open Data.

Keywords: *PhiloBiblon, databases, Semantic Web, Digital Humanities, Middle Ages, Renaissance, printed books, manuscripts, Spain, Portugal, Iberian Peninsula, Linked Open Data.*

[*] Este trabajo forma parte del proyecto de investigación *De los manuscritos al mundo digital* (FFI2014-55537-C3-1-P), financiado por el Ministerio de Economía y Competitividad de España con aportación de fondos FEDER y dirigido por la Dra. Gemma Avenzoa. Todos los datos contenidos en esta investigación pueden consultarse en línea a través de *PhiloBiblon*, cuyo acceso abierto se hace mediante la siguiente ruta fija: http://bancroft.berkeley.edu/philobiblon/index_es.html [2016-10-20].

ESBOZOS SOBRE LA EVOLUCIÓN Y EL FUTURO DE UN PIONERO DE LAS HUMANIDADES DIGITALES HISPÁNICAS: EL PROYECTO PHILOBIBLON

Introducción

En la España de la dictadura franquista era una quimera hablar de avances tecnológicos y, desde luego, mucho menos de tecnología aplicada al conocimiento de las disciplinas humanísticas. La depauperada economía no daba para más, tal como demuestran las cifras globales de cualquier estudio económico serio sobre esta época, suficientes para explicar una ausencia forzada, no voluntaria, en el devenir de la historia de la tecnología en España. Más allá de las anécdotas sobre las primeras máquinas de tarjetas perforadas que fueron alquiladas por RENFE y otras utilizadas por el embrión del *Ciemat* a finales de los años 60 del siglo pasado (Rodríguez Herrera, p. 95), la paupérrima situación la certificaba de forma oficial una encuesta realizada en 1969 por el entonces Ministerio de Educación y Ciencia, en la cual se censaban los ordenadores existentes en España: 615, muy lejos de los 1.176 de Italia, a considerable distancia de los 4.370 de Alemania y a un trayecto sideral de los 55.606 que poseía Estados Unidos por aquellas mismas fechas (Valero Cortés-Mompín Poblet, p. 9).

Por este motivo, aunque no desconocidas, en España sí eran prácticamente imposibles de ser aplicadas las andanzas del padre Roberto Busa (1974-80 y 1980), el sacerdote tomista reconocido de forma casi unánime como el iniciador de la disciplina que hoy conocemos con la etiqueta académica de humanidades digitales, tal como reconocen Rojas Castro (2013b, p. 11) y Faulhaber (2016, p. 75), aunque en algún momento se prefirió llamar al fenómeno computarización de las Humanidades (Svensson); en España llegó a denominarse en sus inicios como "Industrias de la Lengua", aunque el nombre no obtuvo éxito duradero (Marcos Marín, 1992). En cualquier caso, cuando finalmente el proyecto de Busa saltó de las páginas impresas a los bytes del lenguaje informático (Schmidt, p. 275), comenzó con propiedad la era de la simbiosis entre avances tecnológicos y estudios de humanidades, en el más amplio de los sentidos posible.

En lo que se refiere a España, fueron los años de la agonía del franquismo los que vieron nacer varios intentos que, más que pioneros, deberían de ser considerados verdaderos adelantados a su tiempo, por trabajar ya con una amplitud de miras encomiable y muy por encima de la precariedad tecnológica con la que contaban. Tal fue el resultado de los desvelos de Rafael Lapesa, Manuel Ariza y Francisco Marcos Marín en la disciplina que entonces era llamada lingüística computacional (Marcos Marín, 2013b). Pese a estos esfuerzos heroicos, a finales de los años 70 del siglo XX todavía era predominante entre los investigadores españoles la queja de tener que "intentar un trabajo de computadora, pero sin computadora" (González Cuenca, 1, p. 11).

A pesar de ello, algunos años antes, en el ámbito de los estudios hispánicos pero al otro lado del Atlántico –sin olvidar los esfuerzos de Ignacio Soldevila en Canadá que cristalizarían más tarde en el grupo TAUM (Chevalier)–, ya había un pionero que había comenzado a dar sus primeros pasos y del que más tarde se derivaría el proyecto a cuya evolución pasada, a su actualidad del presente y a sus deseos para el futuro queremos dedicar este artículo.

PhiloBiblon: de proyecto "ancillar" a una entidad propia

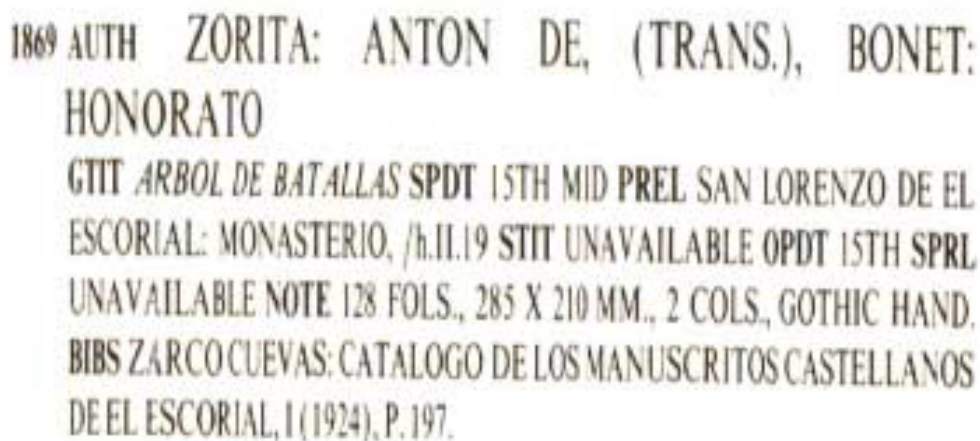
No es el adjetivo "ancillar" uno de los que más se usen para describir un proyecto, tal vez por toda la carga semántica de dependencia que se destila del mismo. Sin embargo, en el caso de *PhiloBiblon*, es obligado reconocer que deriva de BOOST –*Bibliography of Old Spanish Texts*–, más tarde rebautizado como BETA –Bibliografía Española de Textos Antiguos– para la edición del proyecto ADMYTE –Archivo Digital de Manuscritos y Textos Españoles–, al que nos referiremos más adelante. BOOST era el proyecto "ancillar" del DOSL, *Dictionary of the Old Spanish Texts*, dirigido por Lloyd Kasten en la Universidad de Wisconsin, Madison. En sus principios se concibió como un "lexicón citacional", al estilo del *Oxford English Dictionary (OED)*, en el que todas las acepciones de los términos y todas sus formas inflexionales estarían basadas en ejemplos concretos extraídos de manuscritos e impresos anteriores a 1501. Pese a ser la matriz de *PhiloBiblon*, la ingente tarea a realizar demoró la publicación del DOSL hasta 2002 y, además, el corpus quedó restringido a los textos en prosa de Alfonso X el Sabio (Kasten y Nitti). Aunque casi todos los textos trabajados y codificados en esta época están hoy a disposición de los investigadores en la *Digital Library of Old Spanish Texts* (Gago Jover), lo cierto fue que BOOST, en principio tenía como misión catalogar, ubicar y describir las fuentes primarias que se iban a utilizar para la confección del DOSL, acabó por cobrar vida propia debido a que el maestro Lloyd Kasten aceptó el reto de sus jóvenes estudiantes para consolidar un proyecto de forma paralela, como indica Faulhaber (1986-87, p. 158).

A pesar de lo novedoso de su concepción y desarrollo, BOOST llegó a las bibliotecas de forma clásica, en forma de libro impreso en papel que conoció tres ediciones (1975, 1977 y 1984). Sólo en la primera se restringió a la catalogación de manuscritos e impresos anteriores al año 1501, pues todos ellos formaban parte de la esencia del proyecto, tanto en aquellos tiempos como ahora: construir una base de datos bio-bibliográfica que recopila los textos romances literarios escritos en la península ibérica durante la Edad Media y el temprano Renacimiento (BOOST₁, VII; Faulhaber, 1989, p. 214).

BOOST₁, basado exclusivamente en las fichas de la segunda edición de la *Bibliografía de la literatura hispánica* de Simón Díaz (1963-65) y en las del viejo pero excelente catálogo de los manuscritos escurialenses del padre Zarco Cuevas (1924-29), contenía 966 registros; BOOST₂, que ya recogía manuscritos posteriores a 1500, casi duplicó los números con 1.869; la última versión impresa, BOOST₃, alcanzó unas hasta entonces insólitas cifras de 3.738 registros. Todos ellos fueron realizados a través de un sistema de control de base de datos de tipo universal en fichero plano, orientado al trabajo con textos, llamado FAMULUS. El programa estaba diseñado para el control de fichas bibliográficas personales, escritas con un código universal ASCII creado por cualquier editor de textos electrónico, para que luego cada ficha bibliográfica fuera indexada en hilera (Faulhaber y Nitti, p. 288).

BOOST₁ consistió básicamente en presentar todas las fichas ordenadas jerárquicamente por orden alfabético (autor / título / fecha / ubicación actual) y seguidas de los índices de cada uno de sus diez campos, más los índices de los campos relativos a autores, títulos, bibliotecas, y lugar y fecha de copia. Este libro fue impreso mediante fotocomposición mecánica de imágenes, el famoso *photo-offset*, en formato de impresora

de ordenador a línea seguida vertical (Faulhaber y Nitti, p. 291), con todas las letras en mayúscula y sin signos diacríticos. BOOST₂ mantuvo el mismo número de campos y las mismas limitaciones en la indexación con una presentación algo más elegante, en formato horizontal, a doble columna y con una atractiva fuente del tipo *serif* en mayúscula. Dos años antes de la publicación de BOOST₃ se volvió a rescribir el sistema para adaptarlo a un nuevo control de base de datos personalizado que guardase todas las ventajas del antiguo FAMULUS pero sin sus limitaciones en cuanto a número de campos, extensión de los mismos o número de caracteres indexables. Por ello, BOOST₃ contenía catorce campos por cada ficha y permitió imprimir términos de indexación de hasta 159 caracteres.



1869 AUTH ZORITA: ANTON DE, (TRANS.), BONET:
HONORATO
GTIT ARBOL DE BATALLAS SPDT 15TH MID PREL SAN LORENZO DE EL
ESCORIAL: MONASTERIO, /h.II.19 STIT UNAVAILABLE OPDT 15TH SPRL
UNAVAILABLE NOTE 128 FOLS., 285 X 210 MM., 2 COLS., GOTHIC HAND.
BIBS ZARCO CUEVAS: CATALOGO DE LOS MANUSCRITOS CASTELLANOS
DE EL ESCORIAL, I (1924), P. 197.

Fig. 1 – Último registro de BOOST₂

Durante los años 70 y primeros 80 del siglo XX, las fuentes primarias recopiladas por *PhiloBiblon* solo incluían las escritas en castellano, en tanto que, recordémoslo una vez más, el proyecto tuvo en sus inicios la vinculación con el Diccionario de Español Medieval. Pero en 1985 se editó la primera edición de la *Bibliography of Old Catalan Texts* (BOOCT=Bibliografía de Antiguos Textos Catalanes), una base de datos sobre fuentes primarias escritas en catalán compuesta e impresa según los mismos parámetros que BOOST₃. A la pionera del proyecto y responsable de la primera edición, Beatriz J. Concheff (1985), se le unieron Vicenç Beltrán y Gemma Avenoza en 1989, y más tarde Lourdes Soriano, configurando estos tres, después de la muerte de Concheff, el núcleo principal del grupo de investigación de BITECA.

Tres años después, en 1988, comenzó a rodar una nueva base de datos, llamada entonces *Bibliography of Old Portuguese Texts* (BOOPT=Bibliografía de Antiguos Textos Portugueses), y dirigida por Arthur L-F. Askins, Harvey L. Sharrer y Aida Fernanda Días, a los que pronto se uniría Martha E. Schaffer. Ambos proyectos, BOOCT y BOOPT, fueron en 1992 también renombrados para el cederrón de ADMYTE como BITECA (*Bibliografía de Textos Antics Catalans, Valencians i Balears*) y BITAGAP (*Bibliografía de Textos Antigos Galegos e Portugueses*).

Todos estos avances corrieron de forma paralela al proceso de migración de BOOST y BOOCT, iniciado en 1985, desde el viejo sistema de la Universidad de Madison a

un nuevo sistema de bases de datos localizado en la Universidad de California en Berkeley: SPIRES (*Stanford Public Information Retrieval System*=Sistema Stanford de Recuperación de Información Pública). La alta capacidad de SPIRES provocó su elección como método tecnológico más sencillo para solventar algunos de los problemas que presentaba FAMULUS (Faulhaber 1991, pp. 92-93). SPIRES era un programa interactivo y orientado a los datos textuales, además de que permitía una considerable extensión y repetición en la información, aunque todavía consistía en un fichero plano. Por ejemplo, operaciones como la ordenación de datos, las búsquedas y la generación de informes se hacían de manera razonablemente simple. No obstante, la entrada de datos y la corrección de éstos una vez introducidos seguían siendo tan compleja como en el sistema antiguo.

Más adelante, en 1987, y gracias a la financiación de IBM, las bibliografías entonces existentes (BETA y BITECA) fueron transportadas al sistema *Revelation* (más tarde bautizado como *Advanced Revelation*, de *Revelation Technologies*), una base de datos relacional de altas prestaciones que operaba bajo el sistema informático predominante de la época, MS-DOS, descrita por primera vez por Faulhaber y Gómez Moreno (1986, p. 291). El sistema *Revelation* presentaba, como novedad más importante, la posibilidad de que campos de extensión variable y repetida pudieran ser enlazados mediante una estructura de base de datos, quedando descartadas para siempre las limitaciones en el número de registros y de campos. Tan sólo existía un mínimo y razonable límite en el tamaño de cada uno de los registros y fichas, que no podía exceder de 64K, limitación que también operaba en lo relativo a la longitud de cada campo individual. El énfasis de este sistema consistía en la relación entre los diferentes registros de la base de datos, lo que hacía posible cumplir el viejo anhelo de poseer un sistema mucho más sencillo de mantener y de actualizar, al permitir que un único cambio realizado en uno de los registros se efectuase también de forma automática en todos aquellos registros ligados al que había sido modificado. Gracias a ello, quedó desterrada de una vez por todas la vieja operación consistente en tener que hacer el mismo cambio de dato en múltiples registros, como recientemente explicaron Faulhaber y Gómez Moreno (2009a, p. 285).

Desde 1987, todos los datos de las cuatro bibliografías han sido distribuidos y manejados por diferentes versiones de esta misma base de datos, realizada y desarrollada año tras año por *Revelation Technologies*. El diseño original de finales de los '80, desarrollado por el trabajo conjunto del profesor Charles Faulhaber y del programador John May, se ha beneficiado para su evolución de un total de nueve ayudas financieras financiación concedidas por NEH (*National Endowment for Humanities*=Fondos Nacionales para las Humanidades) de EE.UU., además de diversos estipendios más pequeños, pero no por ello menos importantes.

***PhiloBiblon* a través de la evolución de sistemas operativos y de almacenamiento**

A principios de la década de los 90 del siglo pasado, cuando la situación económica y tecnológica española mejoró sustancialmente con respecto a épocas anteriores, la colaboración entre *PhiloBiblon* y España pudo ir más allá de los muchos investigadores que, desde los inicios y de forma muy activa, participaron en la recolección de datos. Así, en

1992, gracias a los buenos oficios de Francisco Marcos Marín y de Ángel Gómez Moreno, tanto las bibliografías existentes –BETA, BITAGAP y BITECA– como el sistema de control de la base de datos formaron parte del ya citado ADMYTE. Se trató de uno de los proyectos culturales financiados por la Sociedad Estatal del Quinto Centenario, que tuvo como objetivo principal conmemorar el quingentésimo aniversario del primer viaje de exploración de Cristóbal Colón al Nuevo Mundo. De hecho, el criterio de selección de textos para esta edición –un total de 64–, transcritos conforme a los criterios de presentación textual del *Hispanic Seminary of Medieval Studies* (MacKenzie-Burrus; Martín de Santa Olalla), se intentó plantear como una recopilación de las obras literarias más importantes de la época colombina, es decir, una muestra representativa de los libros que el almirante podría haberse llevado como compañeros de viaje en sus celeberrimas expediciones.

Así pues, además de estos textos, las tres bibliografías fueron publicadas como parte del disco 0 de ADMYTE, editado por la matritense empresa de tecnología Micro-net que dirigía entonces Gerardo Meiro. La novedad fue que la publicación se hizo en *cederrón*, que en aquella época no solo era la forma de almacenamiento más sencilla y más barata, en términos relativos (Pellen 1993 y 1994), sino que todavía no tenía palabra para ser designada en castellano y se continuaba utilizando la forma inglesa, CD-ROM. Una edición revisada y ampliada de ADMYTE se publicó en el disco I, el cual, a pesar de tal numeración, apareció antes que el disco 0, en 1992, mientras que el disco I es de 1993¹. El disco 0, además de las tres bibliografías de *PhiloBiblon*, también contenía otras herramientas digitales, como TACT (*Textual Analysis Computing Tools*=Herramientas para el Análisis Textual Computarizado), creado por John Bradley, de la Universidad de Toronto, y otro conjunto de programas para crítica textual llamado UNITE, que había sido desarrollado en España bajo la dirección de Francisco Marcos Marín, junto con Víctor Morilla Pardial y José Carlos González Cristóbal, integrantes del Grupo de Sistemas Inteligentes de la Escuela de Ingenieros de Telecomunicación de Madrid, como describe Marcos Marín (1994, p. 181).

Seis años después, en 1999, todavía el proyecto dio un paso más adelante con la aparición de ADMYTE II: 290 obras, más del cuádruple que en las ediciones anteriores, en los que por primera vez se conjugaban texto e imágenes, lo que fue un enorme esfuerzo técnico para la época en que todavía hablar de fotografía digital era una quimera (Marcos Marín, 1994, pp. 184-189), como se puede percibir en la siguiente figura.



[1] A pesar de en la página web de ADMYTE se diga otra cosa, la fecha de edición de los cederrones es la que se dice aquí, es decir, 1992 para el Disco 0 y 1993 para el Disco I, tal como indica el director del proyecto, F. Marcos Marín (2013a).

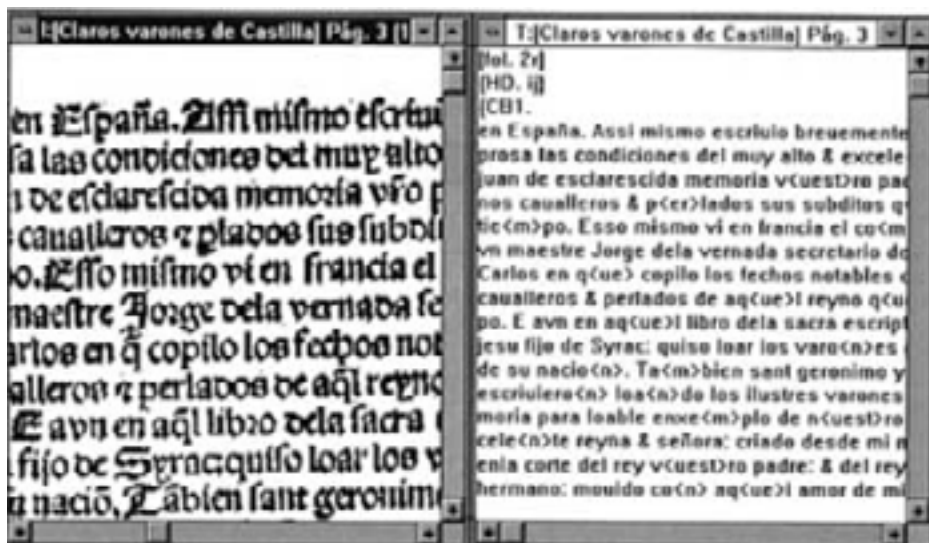


Fig. 2 – Texto e imagen de los *Claros varones de Castilla*, de Hernando del Pulgar

ADMYTE recibió excelentes críticas en la época (Pellen, 1997) e incluso todavía hoy es considerado un magnífico ejemplo de proyecto en humanidades digitales, en opinión de Rojas Castro (2013a, p. 93) y Agenjo (p. 12). Es una lástima que la siempre polémica aplicación de los Derechos de Reproducción Digital (DRM) haya imposibilitado emplear el uso de los discos compactos en cualquier máquina equipada con una versión superior a Windows 95, condenando a los investigadores a acceder a los contenidos de ADMYTE solo mediante el pago de una carísima suscripción a la base de datos en línea.²

Desde la perspectiva de *PhiloBiblon*, el relativo fiasco de los discos compactos de ADMYTE se solventó con la edición en el mismo soporte, durante el año 1999 y gracias al apoyo de la Bancroft Library, de otro disco en el que figuraban las tres bibliografías, BETA, BITECA y BITAGAP, ofreciendo por lo tanto acceso completo a todos los materiales de la base de datos y al resto de funciones de la misma, como la búsqueda y la generación de informes. Las doscientas copias editadas por la Bancroft Library se vendieron al entonces algo elevado precio, pero asequible para universidades, de 100 dólares norteamericanos, agotándose la edición en dos años. Con las ganancias de este temprano ejemplo de financiación colectiva, o *crowdfunding*, el proyecto estaba listo para el siguiente reto: su llegada a la red.

En realidad, el proceso para convertir *PhiloBiblon* en una base de datos de libre acceso en línea ya había comenzado dos años atrás, en 1997, gracias a un complemento de financiación de NEH a las ayudas ya concedidas, así como de un pequeño grupo de ayudas igualmente valiosas: el Programa de Estudios Catalanes Gaspar de Portolà (Universidad de California, Berkeley), el Programa de Estudios Portugueses (Universidad de California, Berkeley) y el Centro de Estudios Portugueses y Gallegos (Universidad de California, Santa Bárbara). Gracias a este esfuerzo, se pudo realizar la página web, posibilitando el acceso en la red a las primeras descripciones detalladas de manuscritos y tempranos impresos en las tres lenguas antes indicadas. Las actualizaciones de la web se hicieron aproximadamente



[2] <<http://www.admyte.com/admyteonline/home.htm>> [2016-10-20].

de forma trimestral, según los distintos equipos de investigación fueran añadiendo datos a sus respectivas bibliografías. También en este mismo año de 1997 se unió al grupo de bibliografías una cuarta, BIPA (*Bibliografía de la Poesía Áurea*), dirigida por José J. Labrador Herraiz y Ralph DiFranco, que extendió el espectro global del proyecto *PhiloBiblon* hasta recopilar textos poéticos de los siglos XVI y XVII. En la actualidad BIPA se encuentra en fase de finalización hasta que Labrador y DiFranco decidan cuándo los contenidos se ajustan al estándar de calidad que desean para su proyecto.

Principales identificadores de *PhiloBiblon*

Hacia el año 2001, la estructura de *PhiloBiblon* consistía en 650 campos de datos individuales. El anteriormente mencionado transporte de datos a la página web comenzó en 1997 se hizo tomando como base las fichas de MS_ED, volcando todo el contenido de ellas en archivos de texto con el código general ASCII ya descrito. Los operarios de la biblioteca de la Universidad de Berkeley diseñaron un código de escritura PERL para convertir los datos de cada manuscrito o temprano impreso de las tres bibliografías de la base de datos (BETA, BITECA y BITAGAP) en lenguaje HTML, el más usado en los archivos de Internet. Estas fichas, y sus correspondientes índices, se pusieron a libre disposición de los investigadores en la red utilizando un sencillo programa de dominio público (SWISH = *Simple Web Indexing System for Humans*). De igual forma, una simple interfaz de protocolo de búsqueda (CGI) permitió realizar en Internet búsquedas por autor, por título o por palabras sueltas en el texto deseado.

Otra ayuda financiera de NEH, concedida en 2001, hizo posible la migración de datos desde un sistema basado en MS-DOS a un nuevo programa diseñado por *Advanced Revelation* sobre soporte Windows. En el proceso de migración, la estructura de las fichas se hizo de forma más racional, al tiempo que otros campos de datos fueron añadidos hasta llegar al número actual de 1246 campos en diez fichas interrelacionadas.

En aquel momento, sin embargo, surgió un pequeño inconveniente, como la imposibilidad de actualizar la versión en la red de BETA, puesto que el mapa de *software* aplicado a Internet estaba construido sobre la versión en MS-DOS. Sí fue posible, sin embargo, actualizar con regularidad BITAGAP y BITECA, puesto que la migración de ambas –y también de BIPA– al sistema operativo Windows no tuvo lugar hasta el otoño de 2008. Ha sido a través de la web cuando todas las bibliografías han podido reunirse por fin en abierto y con el mismo nivel de desarrollo en todos los sentidos.

A efectos prácticos, para realizar búsquedas en la base de datos y en la web, *PhiloBiblon* utiliza varios identificadores (ID), numeraciones correlativas que el programa asigna automáticamente al crear cada uno de los registros. A continuación, explicaremos los siete más básicos y, por lo tanto, más utilizados, tomando como ejemplo una búsqueda cualquiera en la página web, en este caso las conocidas *Coplas a la muerte de su padre* de Jorge Manrique. Primero se mencionará el identificador (por ejemplo, **Texid**) y se hará una breve descripción del mismo; después, se especificará el nombre de la tabla a la que pertenece (por ejemplo, **Texid** pertenece a UNIFORM TITLE) y se ofrecerá el identificador específico para el caso que hemos tomado como muestra.



Fig. 3 – Búsqueda en la web de *PhiloBiblon* de las *Coplas de Jorge Manrique*

Texid→ Se trata del identificador único de cualquier obra compuesta en la Edad Media, con independencia de que el texto nos haya llegado copiado o impreso en épocas más tardías. En el caso concreto que hemos tomado para ejemplificar, sería:

UNIFORM TITLE → *Coplas a la muerte de su padre* →BETA texid 1623.

Manid→ Asignamos este identificador a cualquier fuente primaria, manuscrita o impresa, que contenga una obra compuesta en la Edad Media. En la imagen capturada se ve uno de los manuscritos que contiene esta obra, conservado en la Real Biblioteca del Monasterio de San Lorenzo de El Escorial, que se representaría de esta forma:

MS_ED → El Escorial, X.II.17 → BETA manid 2165.

Cnum→ Una copia específica de esta obra (**texid**) dentro de una fuente primaria (**manid**). Es decir, para referirse a las *Coplas* manriqueñas contenidas en este manuscrito escorialense se haría de esta manera:

ANALYTIC → BETA Cnum 8144 (es decir, texid 1623 + manid 2165).

Copid→ Se trata de un identificador que se aplica siempre a impresos. La razón es que, a distinción de los manuscritos –que siempre ofrecen variantes de copista–, las ediciones impresas, hasta las más antiguas, suelen guardar un parecido notable entre sí, con variantes muy mínimas y debidas casi siempre a la corrección de erratas durante el mismo proceso de impresión. Así pues, **copid** se utiliza para cualquier otro ejemplar de un texto impreso. Por ejemplo, para desglosar la impresión de las *Coplas* manriqueñas efectuada por el impresor zamorano

Antonio de Centenera en 1483, tendríamos, por supuesto el mismo **texid** (BETA texid 1623), en tanto que es la misma obra, y en este caso un solo **manid** (BETA manid 2004), que se corresponde al ejemplar INC/897 de la Biblioteca Nacional de España, el "ejemplar maestro". Cualquier otro ejemplar de este impreso, conservada en otra biblioteca o en la misma, no tendría un **manid**, sino un **copid**. La elección del "ejemplar maestro" lo hacen los miembros de cada bibliografía siguiendo varios criterios relacionados entre sí: se suele dar prioridad al ejemplar más completo y después a su localización, aunque a veces también se elige a los de más fácil acceso, como los conservados en las grandes bibliotecas de cada lengua, es decir, la Biblioteca Nacional de Madrid (BNE) para BETA, la Biblioteca Nacional de Lisboa (BNP) para BITAGAP y la Biblioteca de Catalunya en Barcelona para BITECA.

BETA texid 1623 → *Coplas por la muerte de su padre* → BETA manid 2004
Zamora: Antonio de Centenera, 1493 [BNE INC/897] → BETA copid 1533 British Library IB.52920.

Bioid→ Cualquier individuo relacionado con obra, fuente primaria o copia específica, de cualquiera de las formas: autor, impresor o copista –si los hubiere–, dedicatario de la obra, etc.

Autor de BETA texid 1623 → Jorge Manrique → BETA bioid 1074.

Bibid→ Con este identificador localizamos bibliografía secundaria que aporte información sobre cualquiera de los otros identificadores que estamos trabajando. De forma general, el Bibid se halla insertado en la tabla de identificador con el que esté más estrictamente relacionado.

Si fuese un artículo o libro que describiera en términos codicológicos la fuente primaria (esc. X.II.17, BETA manid 2165), entonces BETA bibid 1003 (Zarco Cuevas, por ejemplo) se encontraría en la tabla de MS_ED, pues el libro está más relacionado con el manuscrito. Por el contrario, si estuviéramos introduciendo un registro de un trabajo que ahondase en la biografía de Jorge Manrique (BETA bioid 1074), la ficha resultante BETA bibid 4419 (Perea Rodríguez 2001) estaría en la tabla de BIOGRAPHY.

Libid→ Identificador para la biblioteca que contenga el manid que estamos analizando.

BETA manid 2165 → El Escorial, X.II.17 → Real Biblioteca del Monasterio de San Lorenzo de El Escorial = Libid 225.

PhiloBiblon y el futuro

Casi parece un ingrediente obligatorio en los trabajos de humanidades digitales finalizar el recorrido jugando, aunque sea de forma somera, a pronosticar qué nos deparará el futuro venidero. Acabar así es bastante más problemático de lo que parece si, como es el caso de *PhiloBiblon*, estamos ante un proyecto que ha conocido más de cuatro décadas en la punta de lanza de la disciplina académica (Faulhaber 2014, p. 18) y, pese a ello,

todavía ha recibido el honor de ser mencionado en 2016 entre los más destacados por la Asociación Europea de Humanidades Digitales (EAHD).³

En la actualidad, parece lógico suponer que el sistema que ya se está imponiendo para bases de datos como la de *PhiloBiblon* es el de la Web semántica (Hitzler *et al.*), y a adoptar sus estándares y métodos se encaminan los deseos de todos los equipos de investigación que forman el proyecto, como describe Faulhaber (2016, pp. 88-90). Sin embargo, aún se trata de una fase de estudio para evaluar pros y contras de una decisión que, si bien conlleva cierto riesgo, como todas, parece ser la más apropiada para mantener el proyecto en el lugar que siempre ha ocupado dentro de la disciplina. En este sentido, el presente más inmediato del proyecto se ha beneficiado de la concesión en 2014 de una ayuda a Giulia Hill, concedida por la *Library de Berkeley* como complemento a la ayuda NEH de 2013, con el objetivo de crear un sitio espejo en Internet gracias a un acuerdo de colaboración con la Universidad Pompeu Fabra de Barcelona.⁴ Con el apoyo de Mercè Cabo Rigol, María Morrás, Joan Trench y Marc Esteve, desde la primavera de 2016 por primera vez disponemos de dos páginas web paralelas, una en Berkeley y otra en Barcelona.

En realidad, uno de los rasgos que más ha beneficiado al proyecto es que, en cierto sentido, bastantes de las fórmulas de trabajo y financiación que hoy día son habituales en las humanidades digitales ya fueron utilizadas en el pasado por los equipos integrantes. Así, desde sus mismos inicios, el modelo de proyecto destacó porque en él la cooperación de usuarios, colegas y estudiantes era absolutamente fundamental. Ya en BOOST₁, Kasten y Nitti tuvieron la genial idea de añadir al final del volumen una docena de formularios de actualización de datos (*Entry Update Forms*) con el objetivo de que otros investigadores pudiesen informar de errores contenidos en el volumen, efectuar precisiones a los registros o, desde luego, añadir nuevos registros a la base de datos, como indica Faulhaber (2016, p. 78). Como puede comprobarse, al igual que *PhiloBiblon* ya estaba imbuido del espíritu de las humanidades digitales antes de que el término se popularizase, también fue pionero en fomentar la colaboración masiva, o *crowdsourcing*.



[3] <<http://eadh.org/projects/philobiblon>> [2016-10-20].

[4] <<http://philobiblon.upf.edu/philobiblon>> [2010-10-26].

Conclusión

En la actualidad, esta vía continúa siendo la manera más efectiva de colaborar con el proyecto, pues la web del proyecto ya dispone de unos formularios para su descarga por parte de colegas y estudiantes, con el objetivo de facilitar la mutua colaboración: al tiempo que se benefician de los datos de que disponemos, los usuarios nos pueden ofrecer su pericia en la consulta de las fuentes primarias de la literatura en lenguas romances de la península ibérica, recibiendo, como es lógico, el crédito correspondiente como responsables de los datos que figurarán en la base y en la web. Los formularios son de cuatro tipos distintos: para manuscritos, para impresos, para fuentes secundarias y para biografías.

Al mismo tiempo, reconocemos la dificultad que entraña describir con minuciosidad un código manuscrito o un libro impreso, pues se necesitan unos conocimientos mínimos de varias materias que van desde la Paleografía a la Codicología, pasando por la Bibliografía Material y la Historia del Libro, y cuyo corolario deseado es la aplicación de todas ellas a las humanidades digitales. Para tal efecto, gracias al apoyo de la Fundación San Millán, que coordina Almudena Martínez, y al Instituto de Literatura y Traducción del Cilengua, dirigido por Carlos Alvar, hemos podido organizar el Seminario Internacional Cilengua PhiloBiblon.

El objetivo primordial de este Seminario es el de entrenar a dos decenas de usuarios —estudiantes de maestría o doctorado, recientes doctores y jóvenes colegas— en todo lo que conlleva el "ecosistema PhiloBiblon". Durante cinco días, la última semana de junio de cada año, un nutrido cuerpo docente formado por diferentes miembros de los equipos de *PhiloBiblon* enseña a la veintena de asistentes los fundamentos esenciales de todas las materias necesarias para hacer buen uso de la base de datos, tanto para beneficio de sus investigaciones como preparándolos para colaborar con nuestro proyecto. El Seminario tiene lugar en las pintorescas instalaciones del Cilengua, en el monasterio benedictino de San Millán de la Cogolla, y las dos ediciones que hasta ahora hemos organizado, en 2015 y 2016,⁵ han contado con inmejorables críticas por parte de los estudiantes (Izquierdo; Sagar García), revelándose como uno de los acontecimientos más esperados por quienes aspiran a estar entrenados en las humanidades digitales aplicadas al medioevo ibérico.

Otro de los aspectos que hemos procurado cuidar más es el de la presencia del proyecto en las redes sociales. Estrenamos perfiles en Facebook y en Twitter en el año 2012, mientras que en 2015 hemos inaugurado nuestro canal de Youtube. Cada una de estas redes se ha mostrado útil para propósitos distintos, pero todos ellos de gran valor en nuestra globalizada sociedad. Facebook es un inmejorable vehículo de promoción de eventos y nos sirve para estar en contacto no solo con miembros de la Academia, sino también con un público muchísimo más amplio. El canal de Youtube es inmensamente provechoso a la hora de comunicar instrucciones a usuarios de la base de datos y a los asistentes al Seminario Internacional. Por último, Twitter ha sorprendido por su capacidad para encontrar incluso nuevos registros, pues, al fin y al cabo, todas las bibliotecas tuitean noticias sobre sus fondos, incluidos los antiguos, de manera que disponer de tal información al instante y contrastarla con la de la base de datos es un sueño para aquellos investigadores que, como nos hacen saber los más veteranos miembros del proyecto, comenzaron a trabajar cuando no se contaba con tales métodos de comunicación académica.



[3] I Seminario: <<http://www.cilengua.es/convocatorias/i-seminario-internacional-philobiblon>> [2016-10-20]. Twitter hashtag: #CilenguaPhiloBiblon.
II Seminario: <<http://www.cilengua.es/convocatorias/ii-seminario-internacional-philobiblon>> [2016-10-20]. Twitter hashtag: #CilenguaPhiloBiblon2.

En conclusión, nos gusta pensar que PhiloBiblon afronta el reto de los nuevos tiempos con la madurez de un veterano y con la energía de las nuevas generaciones que se han incorporado al proyecto, tanto los miembros de los equipos de investigación como aquellos que participan bien con la elaboración de formularios, o bien nos hacen llegar sus publicaciones, o a veces simplemente nos envían un detallado email con correcciones, alteraciones o adiciones. Por nuestra parte, pondremos todo el esfuerzo posible para seguir haciendo lo más difícil, que como es de común conocimiento, no es llegar, sino mantenerse.

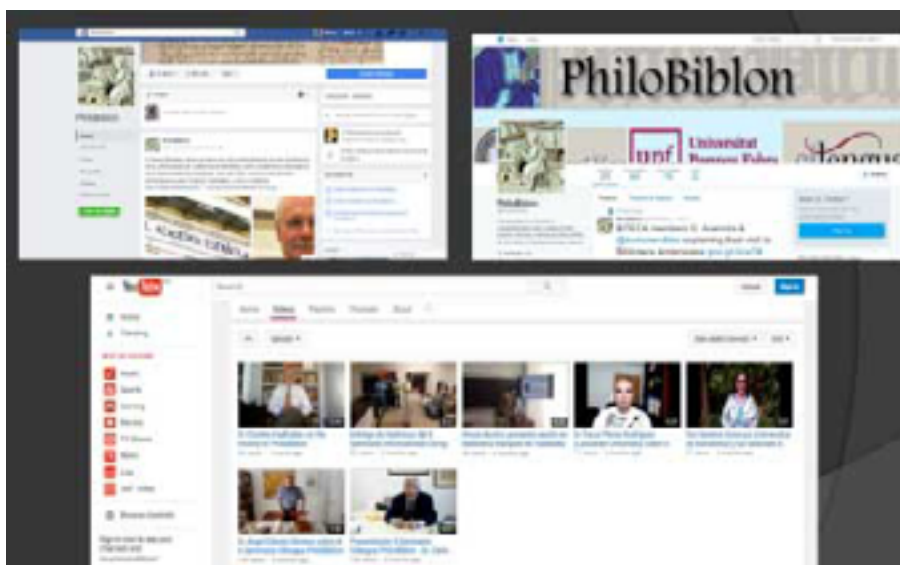


Fig. 4 – Presencia de *PhiloBiblon* en las redes sociales

Bibliografía

- [1] ADMYTE: *Archivo Digital de Manuscritos y Textos Españoles*. Marcos Marín et al. (ed.). Madrid: Quinto Centenario-Biblioteca Nacional-Micronet, 1993.
- [2]_____. *Archivo Digital de Manuscritos y Textos Españoles*. Marcos Marín et al. (ed.). Madrid: Quinto Centenario-Biblioteca Nacional-Micronet, 1992.
- [3]_____. *Archivo Digital de Manuscritos y Textos Españoles*. Marcos Marín et al. (ed.). Madrid: Sociedad Estatal de Conmemoraciones Culturales Centenario-Biblioteca Nacional-Micronet, 1999.
- [4] AGENJO, Xavier. Las bibliotecas virtuales españolas y el tratamiento textual de los recursos bibliográficos. *Ínsula*. [en línea]. 2015, núm. 822 p. 12-15. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=5139045>. ISSN: 0020-4536
- [5] BELTRÁN, Vicenc., AVENOZA, Gemma., SORIANO, Lourdes. *BITECA: Bibliografia de textos antics catalans, valencians i balears*. Valencia: Acadèmia Valenciana de la Lengua, 2013. 428 p. ISBN 978-84-482-5800-9
- [6] BOOST: *Bibliography of Old Spanish Texts*. Anthony Cárdenas., John. Nitti. Jean Gilkison. (ed.). Madison: Hispanic Seminary of Medieval Studies, 1975
- [7]_____. *Bibliography of Old Spanish Texts*. Anthony Cárdenas., John. Nitti. Jean Gilkison. (ed.). Madison: Hispanic Seminary of Medieval Studies, 1977.
- [8]_____. *Bibliography of Old Spanish Texts*. Charles Faulhaber et al. (ed.) Madison: Hispanic Seminary of Medieval Studies, 1984.
- [9] BUSA, Roberto. *Index Thomisticus: Sancti Thomae Aquinatis operum omnium indices et concordantiae in quibus verborum omnium et singulorum formae et lemmata cum suis frequentiis et contextibus variis modis referuntur*. Stuttgart-Bad Cannstatt : Frommann-Holzboog, 1979.
- [10]_____. "The Annals of Humanities Computing: The Index Thomisticus". *Computers and the Humanities*. 1980 , vol. 14, núm. 2, pp. 83-90.
- [11] CHEVALIER, Monique, DANSEREAU, Jules, POULIN, Guy. *TAUM-MÉTÉO: description du système*. Montréal: TAUM-Université de Montréal, 1978. ISBN: 426953449
- [12] CONCHEFF, B. J. *Bibliography of Old Catalan Texts*. Madison: Hispanic Seminary of Medieval Studies, 1985.

- [13] FAULHABER, Charle. "PhiloBiblon and the Semantic Web. Notes for a Future History". En: *La literatura medieval hispánica en la imprenta (1475-1600)*. M. J. Lacarra y N. Aranda García. (ed.). Valencia: Universitat de Valencia, 2016. p. 75-93.
- [14]_____. "PhiloBiblon, Information Technology, and Medieval Spanish Literature: A Balance Sheet". En: *Humanitats a la xarxa: Món medieval. Humanities on the Web: The Medieval World*. (ed.). Lourdes Soriano et al. Bern: Peter Lang, 2014. p. 15-43.
- [15]_____. "Desiderata para el estudio de las literaturas hispánicas medievales". En: *Medioevo y literatura. Actas del V Congreso de la Asociación Hispánica de Literatura Medieval (Granada, 27 septiembre - 1 octubre 1993)*. Juan Paredes. (ed.). Granada: Universidad de Granada 1994. p. 93-107.
- [16]_____. "PhiloBiblon: Problems and Solutions in a Relational Data Base of Medieval Texts". *Linguistic & Literary Computing*, 1991, núm. 2 . p. 89-96.
- [17]_____. "Bibliography of Old Spanish Texts: Evolution of a Data Base". En: *Databases in the Humanities and Social Sciences. Proceedings of the International Conference on Databases in the Humanities and Social Sciences held at Auburn University at Montgomery, July, 1987*. Lawrence J. McCrank. Medford. (ed.). NJ: Learned Information, 1989. p. 213-221.
- [18]_____. "Hispanismo e informática". *Incipit*, 1987, vol. 6, p.157-184.
- [19]_____. ASKINS, Arthur., SHARRER, Harvey. *PhiloBiblon. Electronic Bibliographies of Medieval Catalan, Castilian, Galician, and Portuguese Literature*. Berkeley: The Bancroft Library, 1999.
- [20]_____. GÓMEZ MORENO, Ángel. *La Corónica*. 1986, vol.14, p. 291-292.
- [21]_____. "De BOOST a BETA: de Madison a Berkeley". En: *Los códices literarios de la Edad Media. Interpretación, historia, técnicas y catalogación*. Cátedra, P. M., Carro Carvajal, E.B., y Durán Barceló (eds.). San Millán de la Cogolla: Cilengua-Instituto de Historia del Libro y de la Lectura, 2009. p. 283-292.
- [22]_____. "PhiloBiblon: Pasado y futuro", *Incipit*. 2009, vol. 29 , p.191-200.
- [23]_____. ADMYTE: *Archivo Digital de Manuscritos y Textos Españoles*. Marcos Marín et al. (ed.). La Corónica. 1990, vol.18, p. 131-145.
- [24]_____. NITTI, Jean. "BOOST, Debits and Credits", *La Corónica*. 1983, vol. 11, p. 286-292.

- [25] GAGO JOVER, Francisco. *Digital Library of Old Spanish Texts* [en línea]. New York: Hispanic Seminary of Medieval Studies, 2011-2016. Disponible en: <http://nubr.co/wYkUY6>. [Consulta 20 octubre 2016].
- [26]_____ "Reseña de PhiloBiblon", *Digital Philology*. 2012. vol.1, núm. 2, p. 323-27.
- [27] SAN ISIDORO. *Las Etimologías de San Isidoro romanceadas*. Joaquín González Cuenca (ed.). Salamanca: Ediciones de la Universidad de Salamanca, 1983.
- [28] HITZLER, Pascal, KRÖTZSCH, Markus, RUDOLPH, Sebastian. *Foundations of Semantic Web Technologies*. Boca Raton, Florida: CRC Press, 2010. 456 p. ISBN: 978142009050
- [29] HOCKEY, Susan. "The History of Humanities Computing". En: *A Companion to Digital Humanities*. Schreibman, Susan., Siemens, Ray., Unsworth, John (ed.). Malden: Blackwell. 2004. p. 3-19.
- [30] IZQUIERDO, Almudena. "Reseña del I Seminario Internacional Cilengua PhiloBiblon (San Millán de la Cogolla, 15-19 de junio de 2015)" [en línea]. *PhiloBlog. PhiloBiblon's Project Blog*. 31 julio 2015. Disponible en: <http://nubr.co/9OakrT> [consulta 20 de Octubre del 2016].
- [31] KASTEN, Lloyd ., NITTI, Jean. *Diccionario de la prosa castellana del Rey Alfonso XI*. New York: Hispanic Seminary of Medieval Studies, 2002.
- [32] MACKENZIE, David., BURRUS, Victoria. *A Manual of Manuscript Transcription for the Dictionary of the Old Spanish Language*. Madison: The Hispanic Seminary of Medieval Studies, 1986.
- [33] MARCOS MARÍN, Francisco. "Industrias de la lengua". En *Idiomas. Todo sobre los idiomas*. 1992, vol. 13, p. 6-11.
- [34]_____ *Informática y humanidades*. Madrid: Gredos, 1994.
- [35]_____ "ADMYTE®, el Archivo Digital de Manuscritos y Textos Españoles" [en línea]. *Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid*. 13 febrero 2013. Disponible en: <http://nubr.co/mVjjJW> [Consulta: 20 octubre 2016].
- [36]_____ "Las humanidades digitales" [en línea]. *Blog del Profesor Francisco Marcos Marín*. 3 diciembre 2013. Disponible en: <http://nubr.co/3oSs1B> [Consulta: 20 octubre 2016].
- [37] *Manual de transcripción para el Diccionario del Español Antiguo*. Martín de Santa Olalla, A. (trad.) . Madison: Hispanic Seminary of Medieval Studies, 1992.

- [38] PELLEN, René. "Les CD-ROM pour médiévistes: premières éléments d'une discographie". *Le Médiéviste et l'Ordinateur* n. 28. 1993, p. 13-18.
- [39]_____. "Les CD-ROM pour médiévistes: première mise à jour". *Le Médiéviste et l'Ordinateur* n. 30.1994, p. 40-42.
- [40]_____. "Le CD-ROM: un nouvel âge pour la recherche? Étude d'«ADMYTE1», base de textes espagnols médiévaux". *Revue de Linguistique Romane* n. 61, 1997, p. 89-131.
- [41] PEREA RODRÍGUEZ, Óscar. "La corte literaria de Alfonso el Inocente (1465-1468) según las *Coplas a una partida de Guevara, poeta del Cancionero general*". En *Medievalismo. Boletín de la Sociedad Española de Estudios Medievales*. 2001, vol. 11, p. 33-57.
- [42] RABEN, Joseph. "Humanities Computing 25 Years Later". *Computers and the Humanities*. 1991, vol. 25, núm. 6, p.341-350.
- [43] RODRÍGUEZ HERRERA, Daniel. Ceros y unos. *La increíble historia de la informática, internet y los videojuegos*. Madrid: Ciudadela Libros, 2011. 220 p. ISBN: 9788496836808
- [44] ROJAS CASTRO, Antonio. "Las Humanidades Digitales: principios, valores y prácticas". *Janus. Estudios sobre el Siglo de Oro*. 2013, vol. 2, p. 74-99. ISSN: 2254-7290
- [45]_____. "El mapa y el territorio. Una aproximación histórico-bibliográfica a la emergencia de las Humanidades Digitales en España". *Caracteres*. 2013, vol.3, p. 10-53.
- [46] SAGUAR GARCÍA, Amaranta. "Breve crónica del II Seminario Internacional Cienlengua PhiloBiblon (27 de junio a 1 de julio de 2016, Madrid-San Millán de la Cogolla" [en línea]. *PhiloBlog. PhiloBiblon's Project Blog*. 06 septiembre 2016. Disponible en: https://www.humboldt-foundation.de/pls/web/pub_hn_query.bibliographia_details?p_externe_id=7000277577&p_lang=de [Consulta: 20 octubre 2016].
- [47] SCHMIDT, R. W. "An Historic Research Instrument: The Index Thomisticus". *The New Scholasticism*. 1976, vol. 50, núm. 2, p. 237-249.
- [48] SIMÓN DÍAZ, José. *Bibliografía de la literatura española*. Tomo III, vols. I-II. Madrid: Consejo Superior de Investigaciones Científicas. Instituto. Miguel de Cervantes de Filología Hispánica, 1963-1965.

- [49] SVENSSON, Patrik. "Humanities Computing as Digital Humanities" [en línea]. *Digital Humanities Quarterly*. 2009, vol. 3, núm. 3. Disponible en: <http://nubr.co/RQRDfr> . [Consulta: 20 octubre 2016].
- [50] VALERO CORTÉS, Mateo., y Mompín Poblet, J. *Ciencia y tecnología. España, siglo XXI*. Madrid: Instituto de España-Editorial Biblioteca Nueva, 2012 [en línea]. Disponible en: <http://nubr.co/eQaAYq> [Consulta: 20 octubre 2016].
- [51] ZARCO CUEVAS, Julián. *Catálogo de los manuscritos castellanos de la Real Biblioteca de El Escorial*. vols. I-II Madrid: Imprenta Helénica, 1924-1926; vol. III San Lorenzo del Escorial: Imprenta del Real Monasterio de El Escorial, 1929.