

EVENTOS

2022

III Jornadas de Gestión del
Patrimonio Bibliográfico



Integración de una base de
datos de procedencias en el
entorno MARC y de datos
enlazados: el caso de la
Biblioteca Nacional de España.



crue

Universidades
Españolas

Red de Bibliotecas
REBIUN

III Jornadas de Gestión del Patrimonio Bibliográfico

Integración de una base de datos de procedencias en el entorno MARC y de datos enlazados: el caso de la Biblioteca Nacional de España.

Integrating a provenance database in MARC and linked data environments: the case of the National Library of Spain

Autores:

Ricardo Santos Muñoz. Biblioteca Nacional de España.
(ricardo.santos@bne.es)

Eva María Molleja López. Biblioteca Nacional de España.
(eva.molleja@bne.es)

Marta Vizcaíno Ruiz. Biblioteca Nacional de España. (marta.vizcaino@bne.es)

Resumen:

En este artículo se presenta, a grandes rasgos, el trabajo llevado a cabo durante los últimos años en la Biblioteca Nacional de España, para normalizar e integrar la información de antiguos poseedores en su catálogo automatizado, adaptando la información de procedencias al formato MARC, y desde ahí, darle salida a través del ecosistema de datos enlazados de la biblioteca, a un entorno más dinámico que permite ofrecer un buscador especializado en el portal datos.bne.es

Palabras clave: Procedencias, datos Enlazados. Linked data

This paper presents the efforts and the current situation of integrating a provenance database into the two National Library of Spain main library metadata workflows: the MARC-based library system and the RDF-based publication pipeline that builds into datos.bne.es, the BNE linked data application.

Keywords: Provenances, Linked Data, Marc 21

Introducción

La necesidad de integrar la ingente cantidad de información existente sobre procedencias de las colecciones patrimoniales de la Biblioteca Nacional de España (en adelante, BNE) en su sistema de gestión bibliotecaria, y ofrecerla a los investigadores a través de un buscador especializado, deriva en 2018 en la convergencia de este proyecto con el de datos.bne.es, desarrollado desde 2010 de forma paralela e independiente con el objetivo de gestionar y publicar los metadatos de su catálogo según las buenas prácticas y los formatos y estándares de la web semántica.

El trabajo en procedencias en la BNE

En 2015, un pequeño grupo de trabajo interdepartamental se propuso elaborar un procedimiento para normalizar, identificar y ampliar la información disponible sobre los antiguos poseedores de los fondos que, por una razón u otra, han acabado formando parte de la colección patrimonial de la BNE.

La información sobre la procedencia de los ejemplares de la BNE era escasa y poco normalizada, lo que dificultaba la investigación y reconstrucción histórica de sus fondos. Tomando como modelo algunas bases de datos ya existentes tanto españolas como internacionales ¹, el trabajo se vertebra en tres ejes:

- **Estudio de antiguos poseedores**, reuniendo datos sobre personas o entidades cuyos fondos pasaron a la BNE, así como de la historia de la colección en su conjunto.
- **Estudio de marcas de procedencias**, reuniendo información sobre ellas, e identificándolas con el objetivo de atribuir correctamente los fondos a sus antiguos propietarios.
- **Estudio de los ejemplares**, consignando de forma normalizada los datos relacionados con su procedencia, lo que permitirá unificar y por tanto reconstruir las bibliotecas y colecciones de los antiguos poseedores.

Para lograr el ambicioso objetivo de reconstruir la historia de los fondos de la BNE, y trabajar en estas tres líneas mencionadas, ha sido necesario desarrollar una serie de procedimientos internos, como son: elaboración de tesauros para la descripción de marcas, tanto heráldicas como pictóricas, digitalización, clasificación y descripción de las marcas, desarrollo de unidades de trabajo compartidas con todos los departamentos, realización de cursos de formación para abordar las distintas tareas mencionadas y finalmente crear un manual de procedimiento para todos los implicados en la catalogación, restauración o adquisición de fondos patrimoniales.

¹ Por ejemplo, los proyectos de las universidades de Salamanca, Complutense, de Barcelona, la Real Biblioteca, o internacionales como Nymelyo, el Institut de France, la BnF, la Universidad de Cambridge, BL, Biblioteca Nazionale di Napoli y muchas otras

Tomando como modelo de referencia el de FRBR, de IFLA, se hizo un intensivo trabajo de análisis de los datos bibliográficos y de autoridades, para a través de sucesivos procesos mapeos de los registros MARC, extraer las entidades Persona, Entidad, Obra, Expresión, Manifestación, Ejemplar y Tema, y convertirlos a RDF (Resource Description Framework, el estándar básico de la web semántica), utilizando diferentes vocabularios, fundamentalmente los de IFLA (Vila Suero, 2013).

Tras un paréntesis de dos años, toma forma la idea de desdoblarse el proyecto en dos líneas paralelas pero interconectadas: por un lado, continuar con la transformación de datos a formatos y estructuras naturales de la web semántica, y por otro construir sobre ellos una interfaz o portal para mostrar al usuario final un enfoque distinto sobre los datos del catálogo, basado en las relaciones Obra-Expresión-Manifestación-Item, Personas, Entidades y Conceptos, extraídas de FRBR, y visibilizar conexiones entre los datos de BNE y otros conjuntos externos (Santos, 2015). Así, en noviembre de 2014 se lanza el portal <https://datos.bne.es>, cuyo esquema de datos y de navegación sigue el patrón básico de LRM2 Agente – Obra – Expresión – Manifestación – Ítem.

El análisis y transformación de datos desde MARC21 a RDF se lleva a cabo con la herramienta Marimba, realizada por el equipo del OEG de la Universidad Politécnica de Madrid, y desarrollada posteriormente por la BNE y la empresa Memorándum. Los datos transformados a RDF por Marimba son posteriormente utilizados tanto para su consulta a través de Sparql³ como para su uso en el portal datos.bne.es, a través de un conjunto de aplicaciones de código abierto (Manchado, 2016).

El origen de datos.bne.es, por lo tanto, está en los registros bibliográficos y de autoridad. Son la fuente primaria para construir todas las relaciones, y la calidad de origen de los mismos es la pieza básica para el éxito de todo el sistema (Obregón, 2019). Actualmente, se procesa el conjunto entero de registros bibliográficos, autoridades, materias y fondos.

² *Library Reference Model*: https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712-es.pdf

³ <https://datos.bne.es/sparql>

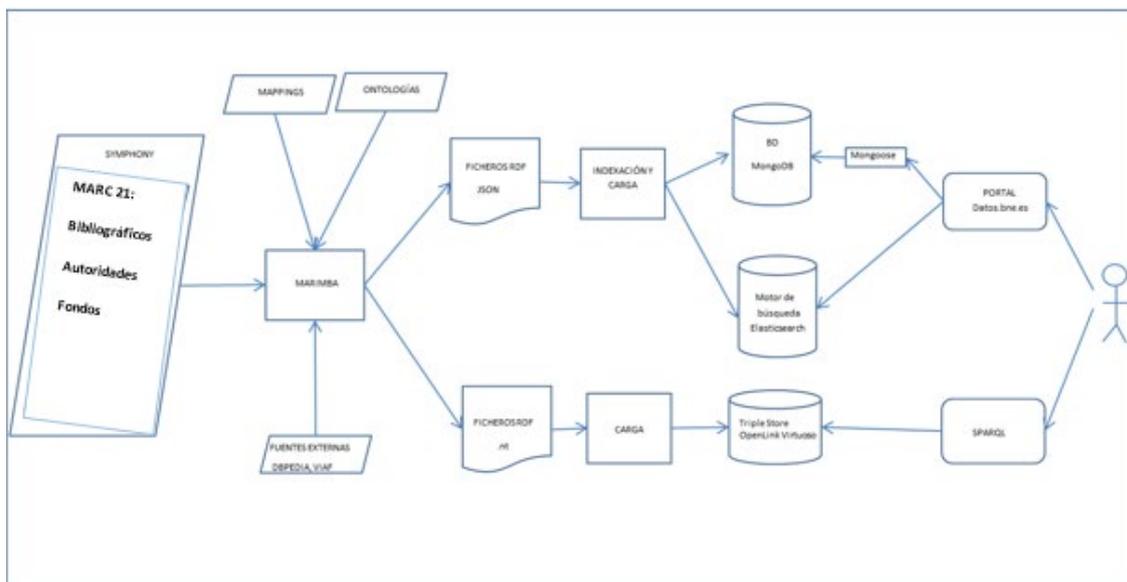


Figura 2. Flujo de datos entre MARC, Marimba y datos.bne

Debido al aumento de la complejidad en los datos y las relaciones entre ellos, se decidió utilizar para anotar los datos una ontología propia⁴, relacionada con otras ontologías de referencia en el sector (FRBR y RDA), aunque se utilizan algunas propiedades de RDA⁵ y Dublin Core para ciertos elementos. Las materias, nombres geográficos, género/forma, están descritas utilizando SKOS.

El portal web explota este árbol de relaciones para, por ejemplo, poder obtener todas las ediciones de una obra, o poder buscar en autores que cumplan ciertos criterios (lugar o fecha de nacimiento, profesión...) (Santos, 2021)

Integración en MARC21

La primera intención del trabajo de procedencias fue construir una base de datos de antiguos poseedores independiente y ajena al catálogo y a los metadatos bibliotecarios. No obstante, por un lado, se plantearon posibles problemas de armonización de estos datos con los generales del catálogo y una segura duplicación del trabajo (el estudio de los poseedores realmente constituye un nuevo registro de autoridad de los mismos). Al mismo tiempo, los desarrollos del nuevo portal de datos.bne.es incluyen la posibilidad de poder buscar autores por atributos como lugar o fecha de nacimiento o profesión⁶. Se decide, por tanto, estudiar la integración de los datos, en primer lugar en registros MARC21 y, desde ahí, en datos.bne.es.

La primera intención del trabajo de procedencias fue construir una base de datos de antiguos poseedores independiente y ajena al catálogo y a los metadatos

⁴ Documentada en: <https://datos.bne.es/def/index-es.html>

⁵ <https://www.rdaregistry.info>

⁶ <https://datos.bne.es/avanzada>

bibliotecarios. No obstante, los posibles problemas de armonización de estos datos con los generales del catálogo, la necesidad de no duplicar esfuerzos, y el desarrollo paralelo del nuevo portal de datos.bne.es, que ya ofrecía la posibilidad de poder buscar autores por atributos como lugar, fecha de nacimiento o profesión⁷, motivaron el abandono de la esta idea inicial, y se pasó a integrar los datos de procedencias en los registros MARC21 y, desde ahí, ofrecerlos en un buscador especializado desde el portal datos.bne.es.

Los datos biográficos recogidos en la tabla anteriormente mencionada encuentran relativamente fácil acomodo gracias a los campos introducidos en el MARC21 de autoridades desde 2014, con la aparición de RDA, sobre todo en el bloque 3XX⁸

Por su parte, los datos relativos a los ejemplares propiamente dichos se registran habitualmente en los campos MARC de los registros de fondos, en los campos 541, 561, 563 y 852 (del que se utiliza el subcampo \$z (nota pública) para describir la marca de propiedad presente).

El mayor problema, no obstante, lo presenta integrar la descripción de las marcas de propiedad en el catálogo, ya que MARC21 no posee ningún campo específico para ello. Para su traslación al formato MARC21, se utilizó un esquema basado en el proyecto de Antics Posseïdors de la Universitat de Barcelona⁹ (Astals, 2010).

Así, para la descripción de la marca de propiedad se creó un campo local de MARC denominado 859. En este campo se definieron, a su vez, varios subcampos:

\$a (NR) Marca (valor: Marca de procedencia, para distinguir de un uso futuro de “Marca de impresor”)

\$b (NR) Tipo de marca de propiedad, según la lista anterior

\$z (NR) Descripción de la marca y/o texto del lema o divisa

\$d (R) Datación. Expresada en siglos

\$e (NR) descriptores, según una lista normalizada de referencia.

\$x (NR) Signatura del libro con la marca tomada como base para la descripción.

\$0 (NR) Identificador único de la marca descrita.

\$y (NR) Texto para el enlace, con el literal “Enlace a la marca”

⁷ <https://datos.bne.es/avanzada>

⁸ <https://www.loc.gov/marc/authority/ad1xx3xx.html>

⁹ <https://marques.crai.ub.edu>

\$u (NR) URL de la imagen digitalizada en la marca.

Además de en datos.bne, en el fichero online de autoridades de la BNE se pueden ver estos registros, y un enlace tanto a la imagen digitalizada de las marcas, como a los registros bibliográficos donde hay un ejemplar asociado al poseedor. Por ejemplo:

http://catalogo.bne.es/uhtbin/authoritybrowse.cgi?action=display&authority_id=XX1175400

El registro de autoridad de un poseedor se complementa con dos campos locales:

- 901, que instruye al catalogador sobre cómo citar al poseedor en la nota de procedencia del registro de Fondos (campo 561)
- 909, es un campo que utilizará datos.bne para enlazar el autor con sus ejemplares.

Integración en datos.bne

El modelo de datos básico que utiliza datos.bne está basado, como se ha dicho, en FRBR (ahora, LRM). No obstante, su representación final viene determinada en buena parte por el origen de los datos y la estructura MARC.

El modelo básico de clases y relaciones de datos.bne es éste:

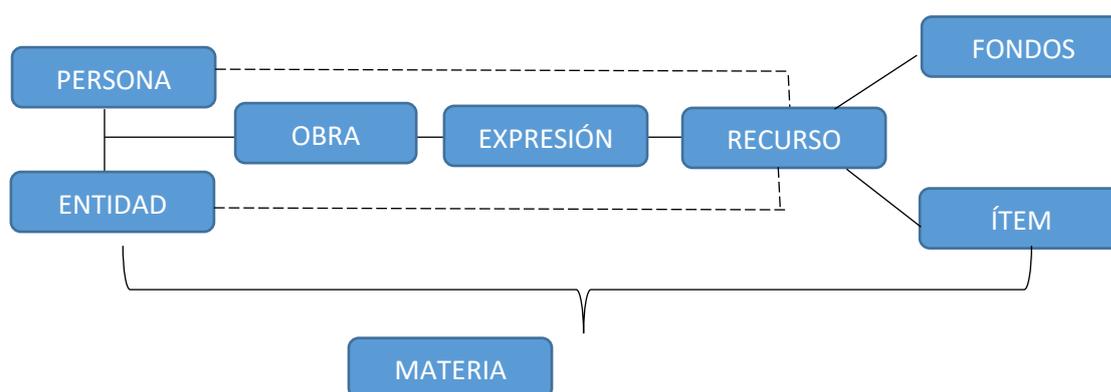


Figura 3. Modelo de datos básico de datos.bne.es

No se ha creado ninguna clase nueva para distinguir a los poseedores, sino que estos son meras instancias de la clase Persona (C1005) o Entidad (C1006), relacionados por una *Object Property* con los ejemplares de los que fueron poseedores, a través de la clase Fondos (C1007).

No obstante, los datos relativos a las marcas de procedencias propiamente dichas (es decir, el contenido del campo 859, anteriormente mencionado), no tiene cabida en ninguna de las clases del modelo, por lo que se decide ampliar el modelo con una nueva clase (C1008) para incorporar los datos sobre marcas de procedencia. Esta clase está declarada (<http://datos.bne.es/def/C1008>) como una subclase de la clase *Mark* del modelo CIDOC-CRM (https://cidoc-crm.org/html/cidoc_crm_v7.1.1.html#E37)¹⁰

Así, el modelo queda completado de esta manera:

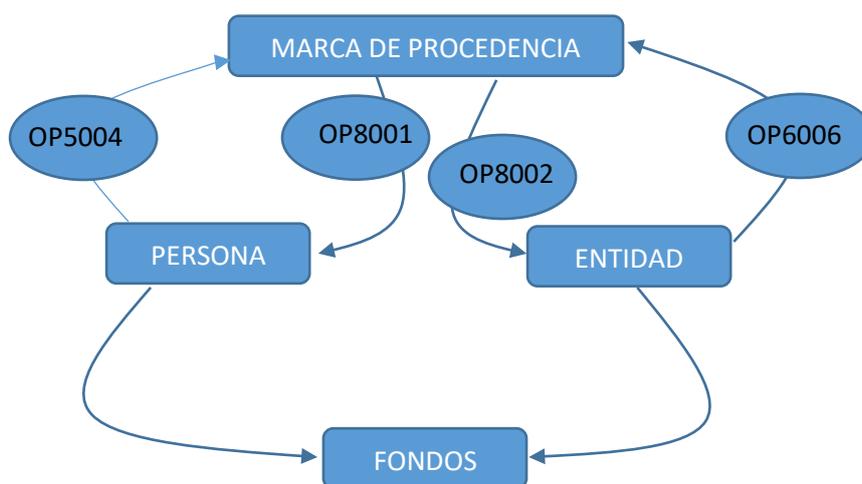


Figura 4. Modelo de datos básico de procedencias en datos.bne.es

Las relaciones entre personas y entidades y sus marcas se definen por sendas propiedades bidireccionales de la ontología. En cambio, la relación entre persona/entidad y sus fondos es dinámica y se construye, en el portal, mediante una consulta del nombre del poseedor (campo 909 mencionado anteriormente), y la propiedad *Nota de procedencia*, de la clase Fondos. (extraído del subcampo 561 \$a).

¹⁰ Definida en CIDOC como “This class comprises symbols, signs, signatures or short texts applied to instances of E24 Physical Human-Made Thing by arbitrary techniques, often in order to indicate such things as creator, owner, dedications, purpose or to communicate information generally. Instances of E37 Mark do not represent the actual image of a mark, but the abstract ideal (or archetype) as used for codification in reference documents forming cultural documentation.

This class specifically excludes features that have no semantic significance, such as scratches or tool marks. These should be documented as instances of E25 Human-Made Feature.”

Buscador de procedencias

En el portal, se ha diseñado un buscador específico para este subconjunto de datos: <https://datos.bne.es/marcas>. Este buscador permite localizar poseedores (personas y entidades) o marcas.

Para visualizar correctamente todas las características se recomienda el uso de **Mozilla Firefox**.

El buscador de Marcas permite buscar por su Descripción, que incluye tanto la descripción libre como por descriptores (859 \$e, \$z), el tipo de marca, campo de selección única (859 \$b) o su datación (selección múltiple, 859 \$d)

El buscador de personas es similar a los buscadores avanzados de autores (<https://datos.bne.es/avanzada>) y permite recuperar autores por características como lugar de procedencia, profesión, títulos o cargos, en el caso de autores, o tipo de entidad, lugar de procedencia, en el caso de entidad.

La página del poseedor es una página de autor a la que se añaden una descripción de sus marcas, junto con una imagen de muestra (que, por el momento, solo se visualiza si se usa el navegador mencionado), y una relación, denominada “Su biblioteca”, que conduce a los fondos identificados de su colección

Marcas

Descripción: Monograma estampado en la cubierta anterior con las letras del apellido "Pimentel" entrelazadas dentro de una corona de laurel

Tipo: Supralibros
Datación: S. XIX.S. XVIII

Descripción: Etiqueta en papel adherida a la hoja de guarda con letra C. y un número

Tipo: Signaturas
Datación: S. XIX.S. XVIII

En la BNE

Autor de 4 Obras

Su biblioteca 433 Obras

Figura 5. Detalle de la página de la Condesa-Duquesa de Benavente: <https://datos.bne.es/persona/XX1118611>

Como en el resto de la aplicación, todos los recursos de datosbne ofrecen URIs estables, tanto la de cada uno de los diferentes objetos, como los propios resultados de las búsquedas.

Conclusión y próximos pasos

Además de los lógicos ajustes o mejoras necesarias, tanto de contenido, navegación, diseño o modelo de datos, los próximos pasos se dirigirán sobre todo a la creación de más contenido para ser explotado desde esta aplicación, esto es, el paulatino incremento de nuevas colecciones, nuevos poseedores y nuevas descripciones de marcas. Con ese objetivo, se incorporarán a este flujo de trabajo nuevos tipos documentales, es decir, otro tipo de objetos de colecciones más allá de los libros: partituras, mapas, dibujos, grabados y todo el inmenso material que conserva la BNE.

La decisión de integrar la base de datos de procedencias en el sistema general de metadatos de la biblioteca, si bien limita en alguna medida las posibilidades de explotación, evita la duplicidad del trabajo y facilita su gestión y mantenimiento, poniéndolo en conexión con otros datos y elementos de la biblioteca.

Bibliografía citada

Astal, Isabel.: Ruiz, Marina; Verger, Neus. La base de dades Antics posseïdors de la Biblioteca de Reserva de la Universitat de Barcelona. En: *BiD: textos universitaris de biblioteconomia i documentació*, juny, 2010, núm. 24.

Disponible en:

http://diposit.ub.edu/dspace/bitstream/2445/13685/1/astals_bid_24.pdf

Manchado, Ana; Santos, Ricardo (2015). *Datos.bne y otras políticas abiertas de datos bibliográficos en BNE*. En: XIV Jornadas Españolas de Documentación (FESABID), Gijón, 28-30 mayo de 2015, p. 128-145. Disponible en: (<https://www.fesabid.org/wp-content/uploads/2021/07/2015-Gijon.pdf>)

Obregón, Eva; Gómez, Roberto; Santos, Ricardo (2019). Uso de la web semántica como implementación de RDA y FRBR: el caso de datos.bne. En: *I Simposio Internacional: Tendencias en la Organización de la Información y del Conocimiento. Universidad de Costa Rica, 30 octubre – 1 de noviembre 2019*.

Disponible en http://www.ebci.ucr.ac.cr/simposio-2019/docs/05_Web%20semantica%20RDA%20y%20FRBR.pdf

Santos, Ricardo; Manchado, Ana, Vila Suero, Daniel (2015). Datos.bne.es: a LOD service and a FRBR-modelled access into the library collection. En: *World Library and Information Conference, Ciudad del Cabo, 15-21 agosto 2015*. Disponible en: <http://library.ifla.org/1085/1/207-santos-en.pdf>

Santos, Ricardo. Adaptando RDA con perspectiva de datos enlazados. Algunos pasos dados en la Biblioteca Nacional de España. En: *XI Encuentro de Catalogación y Metadatos (virtual), 11-13 de agosto de 2021*

Vila Suero, Daniel.; Villalón-Terrazas, Boris; Gómez-Pérez, Asunción. (2013) Datos.bne.es: A library linked dataset. En: *Semantic Web*, vol. 4, nº 3 (2013), p. 307–313