
ESTUDIOS / RESEARCH STUDIES

Inserción de metadatos de las bibliotecas españolas en Wikidata: un modelo de datos abiertos enlazados

Ángel Obregón Sierra

Universidad Isabel I

Correo-e: angel.obregon@ui1.es | ORCID iD: <http://orcid.org/0000-0001-8801-317X>

Recibido: 25-02-21; 2ª versión: 26-05-21; Aceptado: 30-06-21; Publicado: 15/06/2022

Cómo citar este artículo/Citation: Obregón Sierra, A. (2022). Inserción de metadatos de las bibliotecas españolas en Wikidata: un modelo de datos abiertos enlazados. *Revista Española de Documentación Científica*, 45 (3), e330. <https://doi.org/10.3989/redc.2022.3.1870>.

Resumen: El término datos abiertos enlazados se ha hecho muy habitual en el ámbito de la biblioteconomía en los últimos años. Esto es debido al uso que se puede hacer de los metadatos con los que cuentan las bibliotecas. Normalmente se habla de liberar esta información que poseen, pero no sobre dónde se encuentran físicamente estas instituciones u otros datos de la propia institución, para que también puedan utilizarse por cualquier usuario. Esta información podría ser útil para realizar búsquedas de las bibliotecas más cercanas, acceder a los medios de comunicación con los que cuentan o consultar los identificadores que tienen en otras bases de datos. El objetivo de este artículo es el de mostrar el procedimiento utilizado para inserción de todas las bibliotecas de España en una base de datos libre, con la intención de que sean accesibles por todo el mundo. Utilizando diversas herramientas libres se limpiaron los datos y se insertaron 7861 bibliotecas nuevas en Wikidata, para después corregir duplicaciones existentes e insertar nuevos campos. Por último, se presentan diferentes posibilidades de reutilización de los datos en Wikipedia, demostrando que esos datos pueden ser útiles para los usuarios que quieran utilizarlos en el futuro.

Palabras clave: Wikidata; ciencia abierta; bibliotecas de España; metadatos; datos abiertos.

Insertion of metadata from Spanish libraries in Wikidata: a linked open data model

Abstract: The term 'linked open data' has become very common in the field of library science in recent years. This is due to the use that can be made of the metadata held by libraries around the world. There has been much discussion about releasing this information, but not other data such as where these institutions are physically located or details concerning the institutions themselves, so that they can also be used by any user. This information could be useful for searching the nearest libraries, accessing their media or consulting the identifiers they may have in other databases. The objective of this article is to show the procedure used to insert all libraries in Spain in a free database, with the intention that they are accessible all over the world. Using various free tools, the data was cleaned and 7861 new libraries were inserted into Wikidata, to later correct existing duplications and insert new fields. Finally, different possibilities of reusing the data are presented in Wikipedia, showing that these data can be useful for users who want to use it in the future.

Keywords: Wikidata; open science; Spanish libraries; metadata; open data.

Copyright: © 2022 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0).

1. INTRODUCCIÓN

1.1. Datos abiertos enlazados

En los últimos años ha aumentado la preocupación de los profesionales de las bibliotecas respecto a la publicación de datos abiertos. La divulgación de los metadatos y vocabularios se ha extendido siguiendo el modelo de datos abiertos enlazados o *linked open data* en inglés (Alvite, 2017). Estos datos enlazados son un conjunto de buenas prácticas que han sido aprobadas por el World Wide Web Consortium para la publicación de información a través de la web, lo que facilita la comunicación entre quienes la comparten (Ávila, 2020a).

Muchas instituciones públicas y compañías privadas han contratado diversos servicios web para la reutilización de grandes conjuntos de datos, ya que sus bases de conocimiento se estaban convirtiendo en zonas de almacenamiento sin posibilidades de reutilización por otras personas (Martínez-Méndez y otros, 2020). Los datos abiertos, el acceso abierto y la posibilidad de disponer de otro tipo de información alternativa como las *altmetrics* están despertando el interés de los investigadores y las instituciones, ya que constatan una serie de ventajas en su utilización (Uribe y Ochoa, 2018).

A pesar de ello, todavía existe un cierto rechazo debido a la complejidad de su utilización, además de los recursos humanos y económicos que conlleva aprenderlo y ponerlo en marcha (Martínez-Méndez y otros, 2020). El World Wide Web Consortium (2005) nos indica que debemos usar URIs y URIs HTTP, obtener información de los recursos gracias a un modelo de intercambio de datos como RDF, y añadir enlaces a otros URIs, pero, además, se debería utilizar un lenguaje de consulta de tripletes RDF como Sparql, que es el recomendado por el W3C.

Las investigaciones sobre datos abiertos enlazados son muy recientes, y en su mayoría se refieren a casos de uso en ciertas instituciones interesadas en exponer sus catálogos en línea (Fermoso y otros, 2019). También se han publicado artículos sobre el estado en el que se encuentran las instituciones respecto a la implantación de datos abiertos enlazados (Maseda-Seco y otros, 2017), propuestas de publicación de datos abiertos (Marroco y Simionato, 2019) o sobre la visualización gráfica de la información (Ávila, 2020b).

Sin embargo, no se han encontrado investigaciones que se refieran a los datos físicos de las bibliotecas. Para acceder a la información sobre la ubica-

ción de la sede, situación, teléfonos, redes sociales o la página web, únicamente podemos obtener información de las instituciones más importantes, y gracias a consultas en buscadores generalistas y propietarios como Google o Bing.

1.2. Wikidata

Wikidata es el proyecto más moderno de la Fundación Wikimedia, lanzado el 29 de octubre de 2012 para servir como base de conocimiento compartido para el resto de proyectos de la citada fundación (Lemus-Rojas y Pintscher, 2018). Proporciona una gran cantidad de datos estructurados de diferentes temáticas, facilitando el acceso, actualización y reutilización de la información que contiene (Candela, 2019).

Desde su fundación se ha observado un aumento constante en la introducción de datos, pero todavía en 2021 podemos encontrar áreas no explotadas, ya que en la mayoría de casos dependen de la inserción de datos de usuarios particulares, que introducen sus propios conjuntos de datos tabulados. Las investigaciones científicas sobre esta base de datos libre están creciendo en los últimos años, pero su integración sigue siendo un desafío para las instituciones (Tharani, 2021). Pocas son las disciplinas que han realizado estudios con esta base de datos, siendo dirigidas principalmente por investigadores europeos (Mora-Cantallos y otros, 2019).

Algunas de estas investigaciones explican cómo han utilizado Wikidata para crear otras herramientas que puedan utilizarse en línea, destacando algunas como WikiGenomes, herramienta especializada en organismos (Putman y otros 2017), Histropedia, herramienta para realizar líneas temporales, Inventaire, clasificador de libros de una colección, o Scholia, que muestra información sobre revistas especializadas e investigadores (Lemus-Rojas y Pintscher, 2018).

La propia Fundación Wikimedia y sus voluntarios buscan vías de colaboración entre las instituciones y sus proyectos para mejorar la información y hacer crecer a su comunidad de usuarios. Habitualmente estas actividades están centradas en la realización de talleres de formación para editar en Wikipedia, fundación de comunidades de usuarios, concursos de creación de artículos y mejora de las competencias de investigación, redacción y referenciación. Pero a la vez que se desarrollan estas actividades, centradas principalmente en Wikipedia, cada vez es más frecuente relacionarlas con los datos abiertos enlazados, y los catálogos y colecciones de las bibliotecas (Ojeda y Tramullas, 2019).

Es frecuente ver en herramientas como Edit-Groups la introducción de miles de datos en Wikidata por parte de usuarios voluntarios (Delpuch, 2020). Algunos de ellos han publicado sus experiencias en artículos científicos o capítulos de libros, explicando sus métodos para la inserción de miles de datos, utilizando herramientas como Zotero, Source MetaData y QuickStatements (Lemus-Rojas y Odell, 2018), Mix'n'match y QuickStatements2 (Neubert, 2017) o programando bots (Waagmeester y otros 2019).

Tal y como se ha comentado anteriormente, se recomienda disponer de un URI para cada elemento que queremos consultar, pero además estos elementos deberían estar vinculados con los identificadores de otras bases de datos o catálogos externos, lo que nos permitiría mejorar los catálogos en línea cerrados para disponer a partir de entonces de catálogos abiertos (Agenjo y Hernández-Carrascal, 2020).

Tras esta revisión bibliográfica sobre la actualidad de los datos abiertos enlazados y Wikidata, el objetivo que se ha buscado en este artículo es mostrar cómo se pueden insertar todas las bibliotecas de España en Wikidata, utilizando herramientas libres con el fin de que cualquier usuario pueda acceder a esta información libremente y pueda replicar el proceso para la introducción de otros conjuntos de datos tabulados.

2. FASES

2.1. Elección de los datos

En primer lugar, se utilizó el servicio de consultas Sparql de la Fundación Wikimedia, Wikidata Query Service (<https://query.wikidata.org/>) para constatar el número de bibliotecas que ya habían sido introducidas por otros usuarios en Wikidata. La consulta devolvió 303 bibliotecas en toda España, una cantidad muy baja que demostraba la falta de información recogida hasta el momento en Wikidata.

A continuación, y siguiendo el flujo de trabajo con datos abiertos enlazados (Wikidata:Linked open data workflow, 2021), se consultaron todas las páginas web de las bibliotecas de las comunidades autónomas y los gobiernos comunitarios para conseguir un listado abierto con información actualizada de las bibliotecas. Se realizó una prueba piloto insertando las bibliotecas del País Vasco gracias a la base de datos del Gobierno Vasco (Bibliotecas Públicas de Euskadi, 2019). Tras la introducción de todas las bibliotecas del País Vasco y tras un tiempo sin constatar ningún problema se procedió a la búsqueda del resto de bibliotecas estatales.

Se consultó la página web del Gobierno de España para conocer el estado de su base de datos de bibliotecas (Directorio de bibliotecas españolas, 2020). Esta contiene varios tipos de formatos estructurados, por lo que fue necesario revisar todos ellos para ver las diferencias que pudieran existir entre cada uno. Por ejemplo, en los listados predefinidos no aparecen las coordenadas de las bibliotecas, pero en el directorio general sí.

Obtuvimos un fichero .xlsx con los datos estructurados de todas las bibliotecas españolas con su código, nombre, país, comunidad autónoma, provincia, municipio, dirección, código postal, tipología, teléfono, correo electrónico, página web, año de fundación, coordenadas, catálogo, titularidad y gestión.

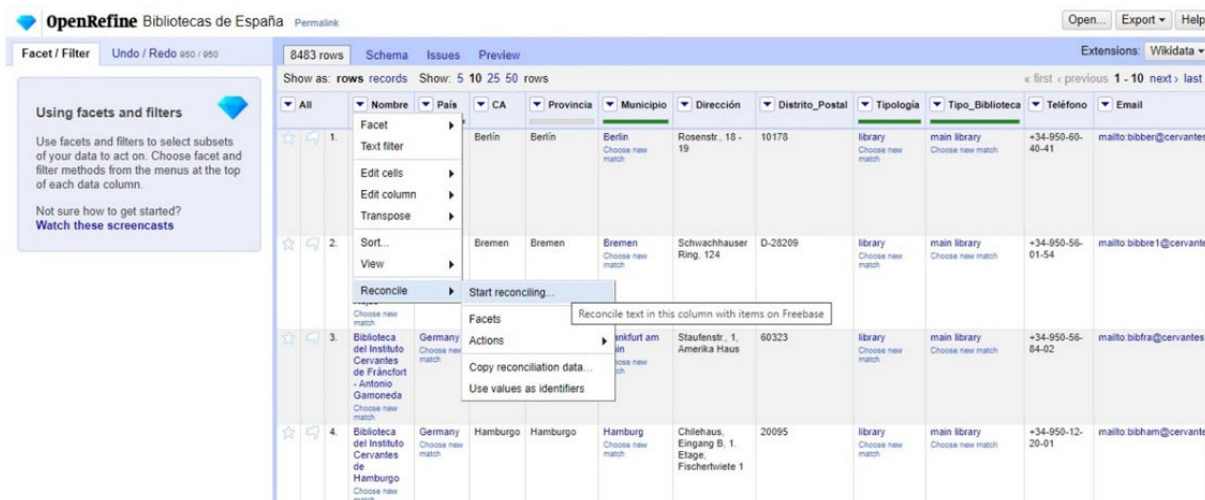
2.2. Proceso de limpieza

Una vez descargado el archivo de datos se procedió a la limpieza de este con LibreOffice Calc, ya que mucha información no es correcta para introducirse tal cual en Wikidata, por lo que hay que comprobar el tipo de dato que contiene cada campo. Por ejemplo, el formato de los números de teléfono, que no siempre tienen la misma estructura y en ocasiones se añaden extensiones al número. También las direcciones de correo electrónico hay que retocarlas, ya que pueden aparecer dos en un mismo campo. Además, para poder ser insertadas en Wikidata necesitan tener delante "mailto:". Por su parte, la coordenada debe contener la latitud y la longitud en la misma celda, separadas por una coma. Para saber qué tipos de datos se introducen habitualmente en las bibliotecas se utilizó una herramienta que está incluida en Wikidata, Recoín.

Una vez que se limpiaron estos casos con LibreOffice Calc, se importó el fichero en OpenRefine (Huynh, 2021). Ciertos campos, como la dirección o el código postal, contienen una cadena de caracteres que se introduce directamente en Wikidata, sin necesidad de realizar cambios. Sin embargo, el país o el tipo de biblioteca, son datos que deben existir previamente en Wikidata, por lo cual hay que reconciliar esta información. Esta acción nos permite comprobar si los elementos que buscamos están en Wikidata, evitando que dupliquemos información. Durante esta fase se encontraron 206 bibliotecas que ya existían en Wikidata, por lo que en esos casos se introducirían los valores junto a los ya existentes. En caso de no existir ese elemento se creó nuevo.

Algunos de los datos que se obtuvieron del fichero fueron descartados, como la comunidad autónoma o la provincia, ya que Wikidata dispone de esa información. Si queremos obtenerla

Figura 1: Opción para comenzar con la reconciliación de los datos en OpenRefine.



se pueden realizar consultas en Wikidata Query Service. Para saber las propiedades donde se iban a introducir los datos nos ayudamos de la herramienta ReCoin, que puede ser habilitada en las preferencias de nuestra cuenta de usuario de Wikidata.

Finalmente, se incluyó la etiqueta de la biblioteca en español, que correspondía con el campo "nombre del archivo", ocho idiomas en la descripción, donde se indicó que era una biblioteca de España, y once propiedades, tal y como muestra la tabla I. Cada propiedad incluía su referencia, la cual indicaba la fecha de acceso, la url donde se puede consultar el fichero original, el idioma de dicha página y su título.

Tabla I: Lista de propiedades de Wikidata para bibliotecas

Nombre de la propiedad	Id de la propiedad
instancia de	P31
fecha de fundación	P571
país	P17
situado en la entidad territorial administrativa	P131
coordenadas	P625
organización matriz	P749
situado en la dirección postal	P6375
código postal	P281
número de teléfono	P1329
correo electrónico	P968
página web oficial	P856

2.3. Añadiendo valores a Wikidata

Una vez que tenemos los datos reconciliados, OpenRefine nos permite crear un esquema de los datos que serán introducidos utilizando un formato muy similar al que veremos después en Wikidata. Una vez terminado el esquema, el propio OpenRefine permite realizar las ediciones en Wikidata sin salir de su programa, pero no podemos rastrear el estado de éstas, ver los errores que se produjeron, ni saber cuánto tiempo tardará. Para realizar un mejor seguimiento de las inserciones se exportaron los datos al formato de QuickStatements (Manske, 2021a).

Tras copiar los datos en QuickStatements, si pulsamos el botón "Run in background" conseguimos que los datos se guarden en el servidor y que puedan ser recuperados en caso de cortar su ejecución. Esa opción es recomendable si el número de ediciones que se van a realizar es alto. Una vez que la herramienta finaliza, podemos verificar los errores que se han producido en ella, y pulsar el botón "Try to reset errors", lo que volverá a ejecutar la instrucción, intentando solucionar los problemas anteriores. A medida que avanza la herramienta en la introducción de los datos, podemos ver en nuestra cuenta cómo los valores son añadidos en Wikidata.

Tras realizar la reconciliación con OpenRefine, se habían encontrado en Wikidata 206 elementos que coincidían con alguna de nuestras bibliotecas, por lo que, al introducir nuestros datos en ellos, en caso de que fuesen datos diferentes, se introducía un segundo valor en dicho campo. Por lo tanto, se realizaron consultas para mostrar el número de valores que existían en la dirección, las coordenadas

Figura 2: Esquema creado en OpenRefine y opción para exportar la información a QuickStatements en la parte superior derecha del programa.

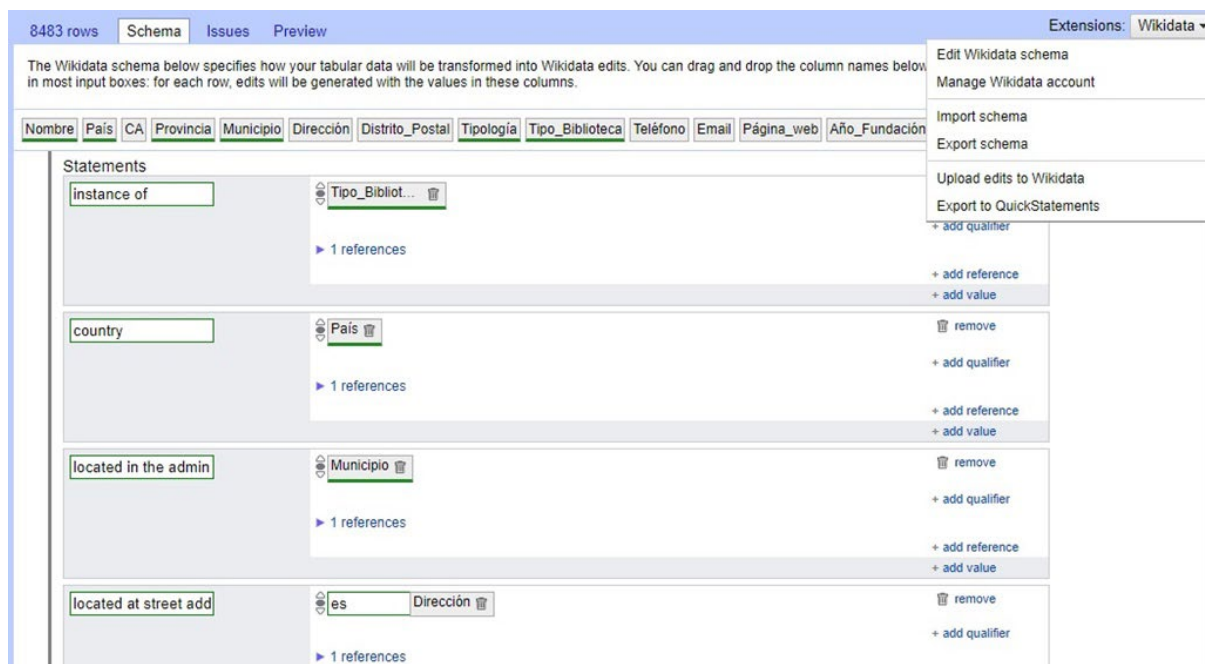


Figura 3: Elemento de Wikidata que muestra parte de los datos que fueron introducidos en un nuevo elemento.

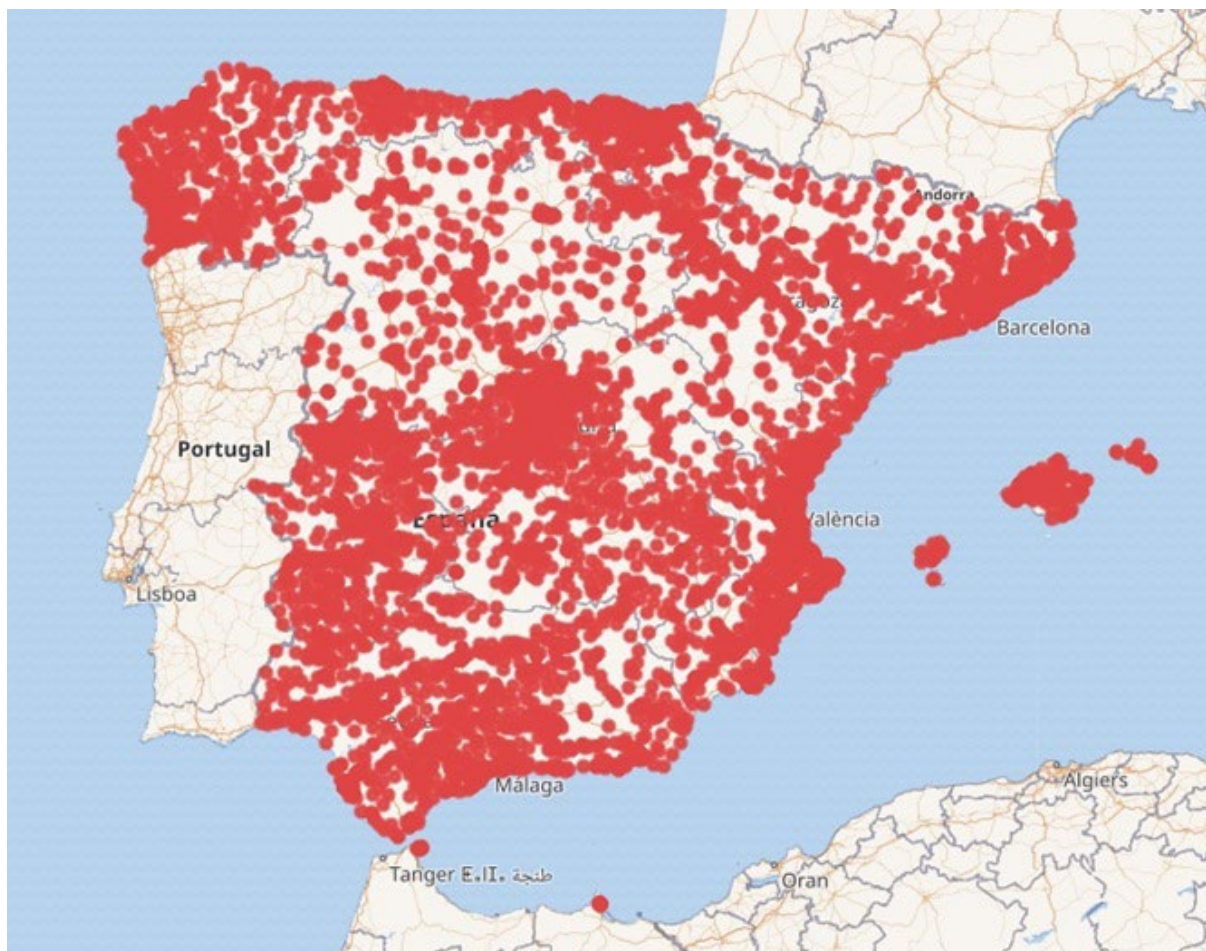


o la página web. En caso de contar con más de un valor, se revisaron uno a uno para borrar la información errónea.

A continuación, se comprobó si podíamos incluir una imagen de las bibliotecas en la propiedad correspondiente de Wikidata (P18). Para ello se utilizó la herramienta WikiData Free Image Search Tool (Manske, 2021b), que nos sugiere posibles fotografías para los elementos de una consulta. Gracias a esta herramienta se añadieron 23 nuevas fotografías y logos. Se incluyeron también, en el caso de bibliotecas sucursales, aquellas de las que formaban parte (P361) y en las bibliotecas principales, aquellas de las que estaban compuestas (P527). En el caso de las bibliotecas universitarias, además, se incluyó "forma parte de (P361)" la Universidad en cuestión.

A pesar de todas las precauciones que tomemos, siempre pueden existir fallos, ya que el fichero origen puede contener información errónea o que la reconciliación realizada en OpenRefine no encontrase todas las coincidencias con los nombres de las bibliotecas. Por lo tanto, el trabajo no termina con la introducción de esta información, sino que hay que seguir rastreando posibles errores. Algunos son muy obvios, como coordenadas en mitad del mar, o en otros países, que son solucionados con la ayuda de Wikidata Query Service, mostrando todas las coordenadas de las bibliotecas de Es-

Figura 4: Todas las bibliotecas de España mostradas con ayuda de Wikidata Query Service



paña con la vista de mapa, tal y como muestra la figura 4.

En ocasiones los errores no son tan obvios, ya que las coordenadas pueden mostrarnos la ubicación en lugares erróneos de la ciudad, bien porque la biblioteca ha cambiado de ubicación, o porque la base de datos no dispone de la información correcta. Para solucionar estos inconvenientes necesitamos la ayuda de otras personas, habitantes de esos lugares que conozcan la ubicación correcta, la dirección y otros datos que se hayan incluido, para verificar que son correctos.

Al final, todo este proceso es un trabajo en equipo, es difícil realizarlo solo, ya que también faltaría incluir la categoría correspondiente de Wikimedia Commons o una imagen ilustrativa de cada biblioteca, en caso de disponer de esa información y que WikiData Free Image Search Tool no la encontrase. Ese trabajo no se puede automatizar, por lo que se realizaron varios comunicados en Twitter mostrando el avance del proceso y pidiendo ayuda para

completar estos datos e incluir además identificadores de otras bases de datos como VIAF, BNE, BnF, GND o libraries.org.

Durante los siguientes días los mensajes fueron retuiteados por varios bibliotecarios, que en ocasiones escribieron mensajes privados para conocer cómo podían colaborar. Wikimedia España publicó en su blog un artículo sobre todo el proceso, explicando cómo se podrían completar los datos de las bibliotecas. Además, el Colegio Oficial de Bibliotecarios y Documentalistas de la Comunidad Valenciana creó para sus socios unas píldoras formativas por videoconferencia, donde se les explicó todo el proceso y cómo podían colaborar. También se realizaron otras videoconferencias de formación en Wikidata y OpenRefine para explicar el proceso de inserción y la importancia de continuar mejorando y actualizando los datos aportados.

En este punto constatamos que disponíamos del Niden INE, el identificador del Instituto Nacional de Estadística, pero este no disponía de propiedad

en Wikidata. Estas deben solicitarse para que sean aprobadas por la comunidad de usuarios y esperar hasta que se consiga el suficiente apoyo para proceder a su creación. Se realizó el proceso y una vez se consiguió su creación se añadió a todas las bibliotecas introducidas.

2.4. Wikipedia

Una vez con todos los datos introducidos y revisados en Wikidata, fuimos a la versión en español de Wikipedia para buscar si existía una ficha específica para las bibliotecas. Estas son plantillas que habitualmente se encuentran en la parte superior

Figura 5: Ficha que se encuentra en el artículo de la Biblioteca de La Casa de Cultura de Santoña, en Wikipedia en español tras recoger la información de Wikidata.

Biblioteca de La Casa de Cultura de Santoña



País

 [España](#)

Tipo

[biblioteca pública](#) y [biblioteca principal](#)

Fundación

[1993](#)

Ubicación

[Santoña](#)

Dirección

Paseo de Camilo José Cela, 1

Coordenadas

 [43°26'33"N 3°27'17'O](#)

Acceso y uso

Miembros

9160

Información adicional

Empleados

4

[Página web oficial](#)

[\[editar datos en Wikidata\]](#)

derecha de los artículos para mostrar los datos más importantes del artículo en cuestión. En el caso de las bibliotecas existe una ficha específica en 63 idiomas diferentes.

Su nombre en español es "Plantilla:Ficha de biblioteca", pero no recogía la información directamente de Wikidata, sino que los datos se tenían que introducir en cada uno de los parámetros manualmente. Dado que las plantillas se van mejorando con el paso del tiempo para que recojan la información automáticamente de Wikidata, programamos ésta para que lo hiciese. Para comprobar el correcto funcionamiento de la ficha se creó un artículo nuevo, recogiendo los datos introducidos en Wikidata, tal y como se muestra en la figura 5:

Pero el uso que se hizo de esta información no fue tan solo para introducir las fichas en las nuevas bibliotecas, sino que también se crearon artículos automáticamente con cierta información que podía mostrarse en tablas. Por ejemplo, en Wikipedia en español se creó el artículo "Bibliotecas Públicas del Estado", compuesto casi exclusivamente por la consulta que se muestra a continuación:

```
{{Wikidata list
```

```
|sparql=SELECT DISTINCT ?item ?itemLabel ?dir
?cod ?entidad ?entidadLabel ?fecha ?coord ?imagen WHERE {
```

```
SERVICE wikibase:label { bd:serviceParam wiki-
base:language "[AUTO_LANGUAGE],en". }
```

```
?item wdt:P361 wd:Q5727891.
```

```
}
```

```
|columns=item:Elemento de Wikidata,label:Etiqueta,P6375:Dirección,P281:Código postal,P131:Ubicación,P571:Fundación,P625:Coordenadas,P18:Imagen
```

```
|min_section=200
```

```
|links=red
```

```
}}
```

Se trata de una consulta introducida en la plantilla "Wikidata list", que proporciona una tabla con los datos que le indicamos. El resultado de la ejecución de este código fue una tabla que se actualiza periódicamente por un bot y que contiene información recogida de Wikidata, tal y como se muestra en la figura 6:

Figura 6: Tabla generada automáticamente por un bot que recoge la información de Wikidata.

Elemento de Wikidata	Etiqueta	Dirección	Código postal	Ubicación	Fundación	Coordenadas	Imagen
Q5727882	Biblioteca Pública del Estado en Palencia	es:c/ Eduardo Dato, 4	34005	Palencia	1897	42.0133, -4.53462	
Q5727898	Biblioteca Pública del Estado en Mérida - Jesús Delgado Valhondo	es:Avda. de la Libertad, s/n	06800	Mérida	1999	38.914685, -6.35741	
Q5727908	Biblioteca Pública del Estado en Santa Cruz de Tenerife	es:c/ Comodoro Rolín, 1	38007	Santa Cruz de Tenerife	1977	28.4643, -16.2665	
Q5727915	Biblioteca Regional de Murcia - Biblioteca Pública del Estado en Murcia	es:Avda. de Juan Carlos I, 17	30008	Murcia	1956	37.998575, -1.137249	
Q6396201	Biblioteca Pública del Estado en Málaga	es:Avenida de Europa, 49	29003	Málaga	1895	36.702686, -4.444817	

Esta tabla muestra las coordenadas de cada una de las instituciones, y gracias a ello se pudo introducir en la parte superior la plantilla “{{mapa lista de coordenadas}}”, la cual inserta un cuadro que nos permite visualizar dichas coordenadas en mapas como OpenStreetMaps y Bing. Además, permite descargar el fichero con las coordenadas en formato KML, GeorSS o GPX.

3. DISCUSIÓN

Se ha mostrado paso a paso el proceso para enriquecer la información sobre bibliotecas en Wikidata, demostrando además otras tareas que pueden mejorar Wikipedia, recogiendo automáticamente el contenido con fichas o haciendo uso de plantillas específicas como “Wikidata list”. La adición de datos en Wikidata es un proceso largo que requiere mucha limpieza de datos, aunque el hecho de ser un usuario habitual en los proyectos de la Fundación Wikimedia evita tener que dedicar tiempo a las interacciones comunitarias y realizar investigaciones sobre las propiedades o requisitos necesarios para publicar, como ya les ocurriera a otros investigadores (Pfunder y otros, 2015).

Este proceso descrito, puede ser replicado con otro tipo de datos, siempre que se disponga de un fichero estructurado, que puede ser obtenido realizando búsquedas en alguna de la gran cantidad de páginas de datos abiertos que están proliferando

en los últimos años (Cerrillo-Martínez, 2018). Las ventajas de utilizar Wikidata para guardar la información común a todos los elementos de un mismo tipo son enormes, ya que los datos están en abierto, pueden ser mejorados por el resto de la comunidad y existen multitud de herramientas y bots para realizar la introducción y el mantenimiento.

Tal y como habíamos comentado, una recomendación del World Wide Web Consortium (2005) para acceder a la información es a través del URI. Todos los elementos y propiedades en Wikidata poseen un URI persistente al que podemos acceder indicando el espacio de nombres seguido del identificador. Por ejemplo, el URI para mostrar la Biblioteca Municipal de Santander es <http://www.wikidata.org/entity/Q5727806>. Se recomienda que este elemento esté vinculado con otras bases de datos y catálogos (Agenjo y Hernández-Carrascal, 2020), por lo que se adjuntó como mínimo el identificador del Instituto Nacional de Estadística.

El World Wide Web Consortium también recomienda utilizar un lenguaje de consulta de tripletes RDF como Sparql, que en Wikidata está disponible gracias al servicio de consultas Sparql denominado Wikidata Query Service, que además de obtener la información de Wikidata y mostrarla en diversos formatos, permite realizar consultas federadas a otros servidores Sparql como Europeana, la British Library o la Biblioteca Nacional de España (Agenjo y Hernández-Carrascal, 2019).

Estas consultas también se pueden realizar a través de lenguajes de programación como R o Python, recogiendo los resultados y mostrándolos en nuestras aplicaciones o páginas web, como están realizando diversas instituciones en los últimos años (Candela y otros, 2019). Por ejemplo, si quisiéramos mostrar en nuestra página web todas las bibliotecas de España la consulta sería la siguiente: <https://w.wiki/JmG>.

4. CONCLUSIONES

Gracias a estas consultas, y a otras herramientas como las plantillas anteriormente descritas, podemos introducir nuevos datos en los artículos de Wikipedia. En la actualidad existen más de 300 versiones idiomáticas de Wikipedia, por lo que mantener actualizados los valores, en nuestro caso de las bibliotecas, requeriría una edición por cada modificación y por cada idioma, dificultando su mantenimiento enormemente. Si todas las versiones utilizasen la ficha de biblioteca, y recogiesen los datos de Wikidata, un solo cambio en Wikidata generaría la actualización automática en todos los artículos con dicha ficha, facilitando la actualización de datos susceptibles de modificarse, como la dirección postal o la página web.

Antes de comenzar con la introducción de datos en Wikidata, existían únicamente 303 elementos que se correspondían con bibliotecas situadas en España. Tras introducir todas las bibliotecas recogidas en el fichero del Gobierno de España, se crearon 7861 bibliotecas más, y se mejoraron los datos de 206 de esos elementos ya creados (algunas bibliotecas que existían previamente habían desaparecido y otras eran muy específicas y no aparecían en nuestro archivo). En cuanto al número de GLAMS (galerías, bibliotecas, archivos y museos) en España, existían 2424 elementos creados, siendo el decimotercer país con mayor número, muy por detrás de los 47586 con los que contaba Estados Unidos. Tras la introducción de todas las bibliotecas, España se situó en segundo lugar.

Al ser usuario de varios proyectos de la Fundación Wikimedia desde hace años, no fue necesaria la creación de una cuenta de usuario nueva, aunque es de vital importancia que el resto de la Comunidad tenga conocimiento de nuestro trabajo. En Wikidata es frecuente pertenecer a Wikiproyectos, por lo que es recomendable registrarse en uno en caso de ser un usuario nuevo, ya que nos permite conocer a otros usuarios que trabajan en la misma temática y recibir consejos sobre las propiedades que deben utilizarse en cada momento.

Por último, en este artículo se ha cumplido el objetivo de mostrar de qué forma se han inser-

tado todas las bibliotecas de España en una base de datos libre. El proceso se ha explicado paso a paso con el fin de que pueda ser replicado por otras personas que quieran introducir en Wikidata otros conjuntos de datos tabulados. Sin embargo, es importante señalar la importancia de cada uno de los pasos realizados, ya que de lo contrario podemos cometer errores fácilmente, tales como seleccionar una fuente de datos poco fiable, no limpiar adecuadamente los datos introduciendo valores mal estructurados, insertar datos en propiedades incorrectas, introducir información con menos calidad que la ya existente en Wikidata, con poca vigencia, o incluso añadir valores que se desactualizan con facilidad. Todo ello puede ocasionar que los consumidores de esta información pierdan la confianza en la fuente de los datos por culpa de la desinformación.

5. REFERENCIAS

- Agenjo, X., y Hernández-Carrascal, F. (2019). Agregación de datos bibliográficos por medio de servicios Sparql. *Anuario ThinkEPI*, 13, e13f02. DOI: <https://doi.org/10.3145/thinkepi.2019.e13f02>
- Agenjo, X., y Hernández-Carrascal, F. (2020). Wikipedia, Wikidata y Mix'n'match. *Anuario ThinkEPI*, 14, e14f01. DOI: <https://doi.org/10.3145/thinkepi.2020.e14f01>
- Alvite, M.L. (2017). Colecciones de datos abiertos enlazados: de la búsqueda al descubrimiento de información. En Da Grapa, M.; Borges, S.M.M. (ed.), *Actas del III Congresso ISKO Espanha e Portugal - XIII Congresso ISKO Espanha*, 1049-1059. Universidad de Coimbra. Disponible en: <http://hdl.handle.net/10612/7076> [Fecha de consulta: 02/02/2021].
- Ávila, E. (2020a). *Los datos enlazados y su uso en bibliotecas*. Universidad Nacional Autónoma de México. Disponible en: http://ru.iibi.unam.mx/jspui/handle/IIBI_UNAM/56
- Ávila, E. (2020b). Los datos bibliográficos abiertos enlazados y su comportamiento en la recuperación de información. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, 34(82), 203-227. DOI: <http://dx.doi.org/10.22201/iibi.24488321xe.2020.82.58130>
- Bibliotecas Públicas de Euskadi (2020). *Open Data Euskadi*. Disponible en: <http://opendata.euskadi.eus/catalogo/-/bibliotecas-publicas-de-euskadi/>
- Candela, G. (2019). *Publicación y enriquecimiento semántico de datos abiertos en bibliotecas digitales* [Tesis doctoral, Universidad de Alicante]. Repositorio Institucional RUA. Disponible en: <http://rua.ua.es/dspace/handle/10045/97353>
- Candela, G., Escobar, M.P., Carrasco, R.C., y Marco, M. (2019). A linked open data framework to enhance the discoverability and impact of culture heritage. *Journal of Information Science*, 45(6). DOI: <https://doi.org/10.1177/0165551518812658>
- Cerrillo-Martínez, A. (2018). Datos masivos y datos abiertos para una gobernanza inteligente. *El Profesional de la Información*, 27(5), 1128-1135. DOI: <https://doi.org/10.3145/epi.2018.sep.16>

- Delpeuch, A. (2020). *EditGroups*. Disponible en: <https://tools.wmflabs.org/editgroups/>
- Directorio de bibliotecas españolas (2020). Disponible en: <http://directoriodibibliotecas.mcu.es/dimbe.cmd?apartado=portada>
- Fermoso, A.M., Manzano, M.I., Armero, A., y Hernández-Hernández, A. (2019). Apertura y publicación de datos bibliográficos con formatos de datos abiertos. Aplicación a un caso práctico. *XV Workshop REBIUN*. Disponible en: <http://hdl.handle.net/20.500.11967/518> [Fecha de consulta: 10/02/2021].
- Huynh, D. (2021). Welcome! OpenRefine. Disponible en: <https://openrefine.org/>
- Lemus-Rojas, M., y Odell, J. (2018). Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project using Wikidata and Scholia. *Journal of Librarianship and Scholarly Communication*, 6(General Issue), eP2272 DOI: <https://doi.org/10.7710/2162-3309.2272>
- Lemus-Rojas, M., y Pintscher, L. (2018). Wikidata and Libraries. En M. Proffitt (Ed.), *Leveraging Wikipedia: Connecting Communities of Knowledge*, 143-158. American Library Association. [Fecha de consulta: 18/01/2021].
- Manske, M. (2021a). *QuickStatements*. Disponible en: <https://quickstatements.toolforge.org/>
- Manske, M. (2021b). *WikiData Free Image Search Tool*. Disponible en: <https://fist.toolforge.org/wdfist/>
- Marroco, D., y Simionato, A. C. (2019). Publicación de datos abiertos en instituciones de patrimonios culturales. *Palabra Clave*, 8(1), e056. Disponible en: http://www.memoria.fahce.unlp.edu.ar/art_revistas/pr.9421/pr.9421.pdf
- Martínez Méndez, F.J., Pastor-Sánchez, J-A., y López Carreño, R. (2020). Linked open data en bibliotecas: estado del arte. *Information Research*, 25(2), paper 862. Disponible en: <http://InformationR.net/ir/25-2/paper862.html>
- Maseda-Seco, D., Bueno-de-la-Fuente, G., y Méndez, E. (2017). Análisis y categorización de los datos abiertos de las bibliotecas municipales españolas: metadatos, interoperabilidad y propuesta para la apertura y reutilización. *El profesional de la información*, 26(3), 392-402. DOI: <https://doi.org/10.3145/epi.2017.may.05>
- Mora-Cantallos, M., Sánchez-Alonso, S., y García-Barriocanal, E. (2019). Una revisión sistemática de la literatura sobre Wikidata. *Tecnologías y aplicaciones de datos*, 53(3), 250-268. DOI: <https://doi.org/10.1108/DTA-12-2018-0110>
- Neubert, J. (2017). Wikidata as a Linking Hub for Knowledge Organization Systems? Integrating an Authority Mapping into Wikidata and Learning Lessons for KOS Mappings. *NKOS@TPDL*, Tesalónica.
- Ojeda, R., y Tramullas, J. (2019). Líneas para el desarrollo de proyectos con Wikipedia y Wikimedia Commons en museos y bibliotecas. *Actas de las IV Jornadas BIMUS*, 111-120). Bimus. Disponible en: <http://www.culturay-deporte.gob.es/dam/jcr:6ebdd83c-70cb-41af-b4d1-cc7ac28c6c63/iv-jornadas-bimus--finales-.pdf> [Fecha de consulta: 13/01/2021].
- Pfundner, A., Schönberg, T., Horn, J., Boyce, R.D., y Samwald, M. (2015). Utilizing the Wikidata System to Improve the Quality of Medical Content in Wikipedia in Diverse Languages: A Pilot Study. *J Med Internet Res*, 17(5): e110. DOI: <http://dx.doi.org/10.2196/jmir.4163>
- Putman, T.E., Lelong, S., Burgstaller-Muehlbacher, S., Waagmeester, A., Diesh, C., Dunn, N., y Good, B.M. (2017). WikiGenomes: an open web application for community consumption and curation of gene annotation data in Wikidata, *Database*, 2017, 2017, DOI: <https://doi.org/10.1093/database/bax025>
- Tharani, K. (2021). Much more than a mere technology: A systematic review of Wikidata in libraries. *The Journal of Academic Librarianship*, 47(2). DOI: <https://doi.org/10.1016/j.acalib.2021.102326>
- Uribe, A., y Ochoa, J. (2018). Ciencia abierta y bibliotecas académicas: una revisión sistemática de la literatura. Universidad Federal de Pernambuco. *Revista do Programa de Pós-Graduacado em Ciencia da informacao*, 4(1), 49-69. Disponible en: <http://hdl.handle.net/10495/12338>
- Waagmeester, A., Schriml, L., y Su, A. (2019). Wikidata as a linked-data hub for Biodiversity data. *Biodiversity Information Science and Standards*, 3, e35206. DOI: <https://doi.org/10.3897/biss.3.35206>
- Wikidata: Linked open data workflow. (2021). En Wikidata. Disponible en: https://www.wikidata.org/w/index.php?title=Wikidata:Linked_open_data_workflow&oldid=1394378153
- World Wide Web Consortium (2005). *Guía Breve de Linked Data*. Disponible en <https://www.w3c.es/Divulgacion/GuiasBreves/LinkedData>