

Metodología para la incorporación de información geográfica en Wikidata

Methodology for the incorporation of geographic information in Wikidata

Ángel Obregón-Sierra

angel.obregon@ui1.es  0000-0001-8801-317X

Javier López-Otero

javier.lopez.otero@ui1.es  0000-0002-6543-2926

Antonio Gavira-Narváez

antonio.gavira@ui1.es  0000-0002-5389-8315

Rafael Vega-Pozuelo

rafaelfernando.vega.pozuelo@ui1.es  0000-0003-4982-9285

Universidad Isabel I. Calle Fernán González, 76. 09003 Burgos, España.

INFO ARTÍCULO

Recibido: 02/01/2023
Revisado: 07/02/2023
Aceptado: 08/02/2023

PALABRAS CLAVE

Wikidata
Datos abiertos
SPARQL
GIS
Análisis espacial
Vértices geodésicos

KEYWORDS

Wikidata
Open data
SPARQL
GIS
Spatial analysis
Triangulation station

RESUMEN

Wikidata es una base de conocimiento libre, multilingüe y editable por cualquier usuario, que ha experimentado un crecimiento exponencial desde su creación hasta ser una de las más grandes de internet. Sin embargo, son pocos los estudios sobre la presencia y tratamiento de la información geográfica en dicho repositorio, entre otros motivos, debido a que no existe una metodología explícita para la inclusión de algunos datos.

Así pues, en esta investigación se propone una metodología que incluye 6 fases que permiten la introducción de información geográfica en Wikidata en grandes cantidades, y disponer posteriormente de ella para la realización de análisis espaciales mediante un GIS. Dicha metodología se ha puesto en marcha a través de la introducción de una base de datos de 11.143 vértices geodésicos de España. Representa una contribución a la geografía, ya que permite el uso de Wikidata en un GIS, lo que puede contribuir a mejorar la eficiencia de los modelos que hagan uso de dicha información.

ABSTRACT

Wikidata is a free knowledge base, editable by any user, which has experienced exponential growth since its creation to become one of the largest public knowledge repositories on the Internet at this moment. However, there are few studies on geographic information storage and analysis in Wikidata, among other reasons, because there is no explicit methodology for it.

Thus, this research proposes a method that includes 6 phases that allows the introduction of a huge amount of geographic data in Wikidata, which in turn will be available for spatial analysis using a GIS. This methodology has been implemented through the introduction of a large database of geodesic 11.143 triangulation stations of Spain. It represents a contribution to geography, since it allows the use of large Wikidata databases in a GIS, which can help improve the efficiency of models that make use of such information.



1. INTRODUCCIÓN

Los cambios que ha originado la presente sociedad de la información y, más concretamente el avance de las nuevas tecnologías de la información y la comunicación de la cuarta revolución industrial,¹ ha multiplicado el tráfico de software, documentos, mapas y sistemas de manufactura (Stark et al., 2006). Son varios los factores que lo han provocado, entre ellos cabría destacar el desarrollo de algoritmos de inteligencia artificial (IA) que consumen gran cantidad de información para entrenar sus modelos (González & Evans, 2019), la cual es provista a partir de los macrodatos procedentes de sensores, aplicaciones web, redes sociales y estadísticas públicas.

Del mismo modo, existe un creciente desarrollo de modelizaciones 2D y 3D del territorio y de los objetos que consumen igualmente ingentes cantidades de espacio (Shan & Sun, 2021). En este sentido, la existencia de repositorios de información se ha ido haciendo cada vez más necesaria y en las últimas décadas se ha multiplicado la información y conocimiento almacenado en éstos. Sin embargo, la gestión y organización no ha sido homogénea, ya que mucha información se distribuye en espacios de almacenamiento creados por corporaciones, cuyos contenidos son esencialmente privados y estancos, mientras que otras se distribuyen en una multiplicidad de repositorios de propiedad pública en los que existe abundante información inconexa debido a sus características, así como a los métodos manejados para su obtención y guarda.

Frente a este modelo predominante hasta el momento, cabe destacar el proyecto Wikidata. Este tiene la particularidad de ser una base de conocimiento libre propiedad de la Fundación Wikimedia, una asociación sin ánimo de lucro. Wikidata fue creado para dar soporte al resto de los proyectos de la fundación, como Wikipedia o Wikimedia Commons. Además, al tratarse de un proyecto colaborativo, las utilidades destacadas anteriormente no son las únicas, ya que sirven como base de conocimiento libre a la que puede acceder y editar cualquier usuario con acceso a internet (Obregón, 2022).

La proyección y las características que tiene el proyecto ha propiciado que muchas instituciones, y proyectos, como OpenStreetMap o WikiGenomes (Farda-Sarbas & Müller-Birn, 2019) hayan visto las ventajas de contar con una base de conocimiento abierta que recoja información generalista, ya que pueden hacer consultas federadas (consultas que permiten acceder al contenido de otras bases de datos diferentes a Wikidata) y contar con toda la información que proporciona, sin necesidad de tener que guardar esa información en sus respectivas bases de datos (Lemus-Rojas & Pintscher, 2017). Ello ha multiplicado el uso de Wikidata, de este modo, en 2019 contaba con más de 55 millones de elementos creados (Piscopo & Simperl, 2019), mientras que tres años más tarde ha superado los 100 millones de elementos de temáticas tan diversas como artículos científicos, edificios de interés cultural o personas.

A pesar de ello, no existen muchos artículos científicos sobre esta base de conocimiento, lo cual se explica entre otros motivos por la novedad de este proyecto, que se lanzó en el año 2012. Concretamente a finales de 2022, Google Académico mostraba 822 resultados con el término Wikidata en el título de los artículos, mientras que Scopus devolvía 255 documentos ante la misma búsqueda de información. Como puede observarse por el resultado de la búsqueda, las investigaciones sobre Wikidata se encuentran en las primeras etapas de estudio, publicándose en la mayoría de los casos en actas de conferencias (Mora-Cantalops et al., 2019).

Por otra parte, la puesta en uso de este repositorio ha suscitado la necesidad de elaborar una metodología para trabajar con Wikidata (Obregón, 2022), realizando análisis de la calidad de la información introducida (Shenoy et al., 2022), mostrando experiencias prácticas concretas de introducción de datos (Gurtovoy & Gottschalk, 2022) o desarrollando herramientas para prevenir el vandalismo (Heindorf et al., 2016).

Asimismo, cabe señalar a la biblioteconomía como una de las ramas más interesadas en esta base de conocimiento, ya que identifica a Wikidata como una plataforma global abierta y colaborativa para compartir e intercambiar metadatos entre bibliotecas. Ciertamente, es útil como repositorio para desambiguación y vinculación de datos de autoridad, aportando su conocimiento al bien público (Tharani, 2021).

1. Este concepto hace referencia a la revolución más importante hasta la actualidad, aquella que permite unir tecnologías en prueba o desarrollo que ayudan a difuminar las fronteras existentes entre los espacios físico, biológico y digital.



Otro campo de expansión de Wikidata, sin duda, es el de las ciencias geográficas. La geociencia es una disciplina que trabaja a distintas escalas, desde el ámbito planetario, e incluso extra-planetario, hasta el local. La posibilidad de unificar los distintos repositorios de información dispersos entre un increíble número de instituciones, así como el acceso a información planetaria o local distinta a la del país de residencia y en varios idiomas globales son aspectos que pueden revolucionar una ciencia como la geografía. Esta disciplina consume cantidades ingentes de datos para producir resultados, como es el caso de la mayoría de las ciencias físicas y sobre este respecto ya existen algunos trabajos en los que se ha hecho uso de la plataforma para abordar diferentes estudios relacionados con esta disciplina. Tal es el caso de la climatología (Wu et al., 2022), la biología (Mittermeier, 2021) o la oceanografía (Garijo & Poveda-Villalón, 2020).

La geografía física estudia fenómenos globales, por lo que es lógico que se hayan puesto en marcha algunos trabajos relativos a esta disciplina, que participen de los datos de la plataforma y ayuden a establecer conclusiones relevantes. Sin embargo, aunque la geografía humana trabaja en escalas más diversas y en diferentes regiones del planeta, Wikidata presenta unas características que sin duda pueden ser de utilidad, al ofrecer información de ámbitos locales distintos a los que pueden existir en los países de origen y en varios idiomas globales.

Así pues, en este trabajo se propone una metodología específica para geografía que permita incorporar datos geográficos a Wikidata y que posteriormente puedan ser analizados por la disciplina de geografía, mediante herramientas de análisis espacial, como un GIS. Asimismo, dicha metodología y análisis se aplicará sobre una base de datos de vértices geodésicos de España.

2. WIKIDATA

Tal y como se había adelantado, Wikidata fue creada por la Fundación Wikimedia en 2012 con el propósito de servir de repositorio de datos estructurados para el resto de los proyectos de la fundación. Sin embargo, a pesar del propósito inicial de este repositorio y dada su naturaleza pública y abierta, Wikidata pasó a ser usado progresivamente por los usuarios para un destino más amplio que el de Wikipedia y el resto de los proyectos de la institución. La única condición para la inclusión de nuevos datos es que se introdujesen con una licencia de dominio público, por lo que se trata de una base de conocimiento abierta, libre y de carácter público.

Cada artículo publicado en Wikipedia tiene su correspondiente elemento en Wikidata, que es compartido por todas las versiones idiomáticas y el resto de los proyectos, como Wikimedia Commons. Desde su fundación, Wikipedia no ha parado de crecer, llegando a acumular en 2022 más de 60 millones de artículos en 318 idiomas y este crecimiento se ha reflejado igualmente en Wikidata, que ha crecido hasta superar los 100 millones de elementos en 2022 (Wikidata Mainpage, 2022).

Otra particularidad de Wikidata es que este repositorio puede ser editado por humanos o por máquinas (Wikidata Mainpage, 2022), de hecho, el 90% de la introducción de datos procede de bots desarrollados por contribuidores (Vrandečić & Krötzsch, 2014). Es por ello que el valor de este repositorio está creciendo de un modo exponencial y es en la actualidad uno de los repositorios públicos de mayores dimensiones del mundo.

Por otra parte, Wikidata muestra algunas singularidades respecto a otras bases de datos convencionales, concretamente es una base de conocimiento que es accesible en múltiples idiomas, lo cual contribuye a su difusión y preeminencia global (Vrandečić & Krötzsch, 2014). Gracias a este rasgo es posible que existan múltiples aportaciones de Wikidata en varios idiomas, por lo que se multiplicarán el número de aportaciones que en última instancia beneficiará a toda la plataforma, ya que el valor de esta depende de su tamaño.

Esta base de conocimiento tiene dos componentes esenciales: los elementos (comienzan por una Q seguida de un número) y las propiedades (comienzan por una P seguida por un número). Los primeros representan entidades concretas o abstractas, por ejemplo, "Miguel de Cervantes" (Q5682). Cada uno de estos elementos está formado por "declaraciones" (o afirmaciones), que nos ayudan a explicar qué es el elemento en cuestión. Cada declaración está formada por una propiedad, que se refiere a los rasgos caracterizadores



de ese elemento, y el valor que toma esta propiedad. Por ejemplo, “lugar de nacimiento” (P19) - “Alcalá de Henares” (Q46940). Cada elemento está formado por varias declaraciones, ya que estas nos ayudan a explicar mejor al elemento en cuestión.

Dado que Wikidata es una base de conocimiento abierta y reutilizable, capaz de combinar los metadatos locales con los globales, muchas bibliotecas han comenzado a trabajar con ella para incrementar su acceso global (Tharani, 2021). De este modo, son varios los desarrollos que ha experimentado Wikidata en este campo, entre ellos cabe destacar la gran difusión de uso que ha tenido en la catalogación bibliotecaria, donde los identificadores de Wikidata han permitido a las bibliotecas acceder a información de Wikidata en relación con los fondos que poseen (Van Veen, 2019).

Desde el punto de vista de otras disciplinas como la medicina, la biomedicina, la biología, la economía o la geografía, la capacidad evolutiva de Wikidata resulta especialmente interesante, ya que la continua aportación de datos permite comparar la eficiencia de los modelos, lo que en última instancia representa una notable mejora en la eficiencia de dichos modelos (Andra et al., 2020; Turki et al. 2019).

Finalmente, también es significativo el conjunto de estudios que han analizado la calidad de los datos introducidos en Wikidata, que a diferencia de otros repositorios similares como DBpedia o YAGO, incluye el concepto de “provenance” según el cual debe referenciarse claramente la fuente de la que proceden los datos (Piscopo & Simperl, 2019). Este rasgo de Wikidata facilita la reutilización de los datos y contribuye a evitar información duplicada. En cualquier caso, son varios los trabajos que consideran que es necesario profundizar en el análisis de la calidad de la información, especialmente porque Wikidata, a diferencia de otros repositorios, es editable por cualquier usuario (Obregón, 2022), lo que multiplica la información disponible, por una parte, pero también la posibilidad de haber más errores, por otra (Piscopo & Simperl, 2019; Shenoy et al., 2022).

Wikidata permite la edición de elementos a direcciones IP, esto es, a personas que no se han registrado. Sin embargo, estas no pueden crear elementos. Para ello hay que registrarse y, así, obtener privilegios como la creación y seguimiento de elementos, la subida de archivos o la creación de su página de usuario (Registered user, 2023). En la parte superior derecha de Wikidata se habilita la opción de crear una cuenta, concretamente en el enlace “Create account” o “Crear una cuenta” si se ha seleccionado la versión española de la página, tal y como puede verse en la figura 1.

La creación de nuevos elementos implica el cumplimiento de uno de los tres criterios de relevancia: contener al menos un enlace a otro proyecto de la Fundación Wikimedia, que el elemento a crear se refiera a una entidad conceptual claramente identificable o que cumpla alguna necesidad para con otros elementos (Wikidata, 2023).

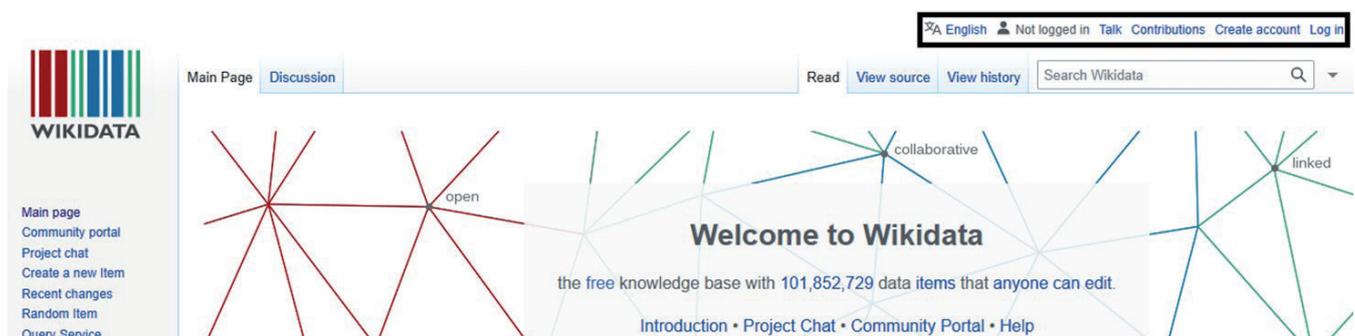


Figura 1. Plataforma Wikidata y área de acceso de editores. Fuente: Wikidata (2023).



3. DATOS USADOS: VÉRTICES GEODÉSICOS

A continuación, se realizará una presentación de los vértices geodésicos, habida cuenta de que es el elemento sobre el que se aplicará la metodología propuesta. Concretamente los vértices geodésicos son una red de postes fijos, estables y visibles a gran distancia, que están ubicados correctamente en un plano y permiten identificar la localización y altura, (respecto al nivel del mar), de cualquier punto en el mapa. Todo ello será posible, siempre que dicho punto esté comprendido entre tres vértices geodésicos, lo que permitirá identificar la mencionada ubicación mediante triangulación (Barranco, 2011).

La Red Geodésica Fundamental española se inició en 1852 y, en 1870, ya se había completado una red con vértices que permitían conformar triángulos de entre 30 y 70 kilómetros (Martín, 1999), este grupo de vértices constituirían una red geodésica denominada “de primer orden” (fig. 2).

La red geodésica “de segundo orden”, se apoyaría en la primera y los lados de sus triángulos variaban entre los 10 y los 25 kilómetros, mientras que la red “de tercer orden, presentaba lados de 5 a 10 kilómetros. Todos los vértices de las redes más grandes eran, a su vez, vértices de las más pequeñas. Todas las bases secundarias se unieron a la triangulación mediante redes auxiliares y se fueron ultimando las observaciones de toda la red que se completó definitivamente en 1915 (Urteaga & Nadal, 2001). Faltaba unir las bases secundarias a la triangulación mediante redes auxiliares y terminar las observaciones en algunos cuadriláteros de los márgenes de la red como el de Cáceres o el de Huesca (Urteaga & Nadal, 2001).

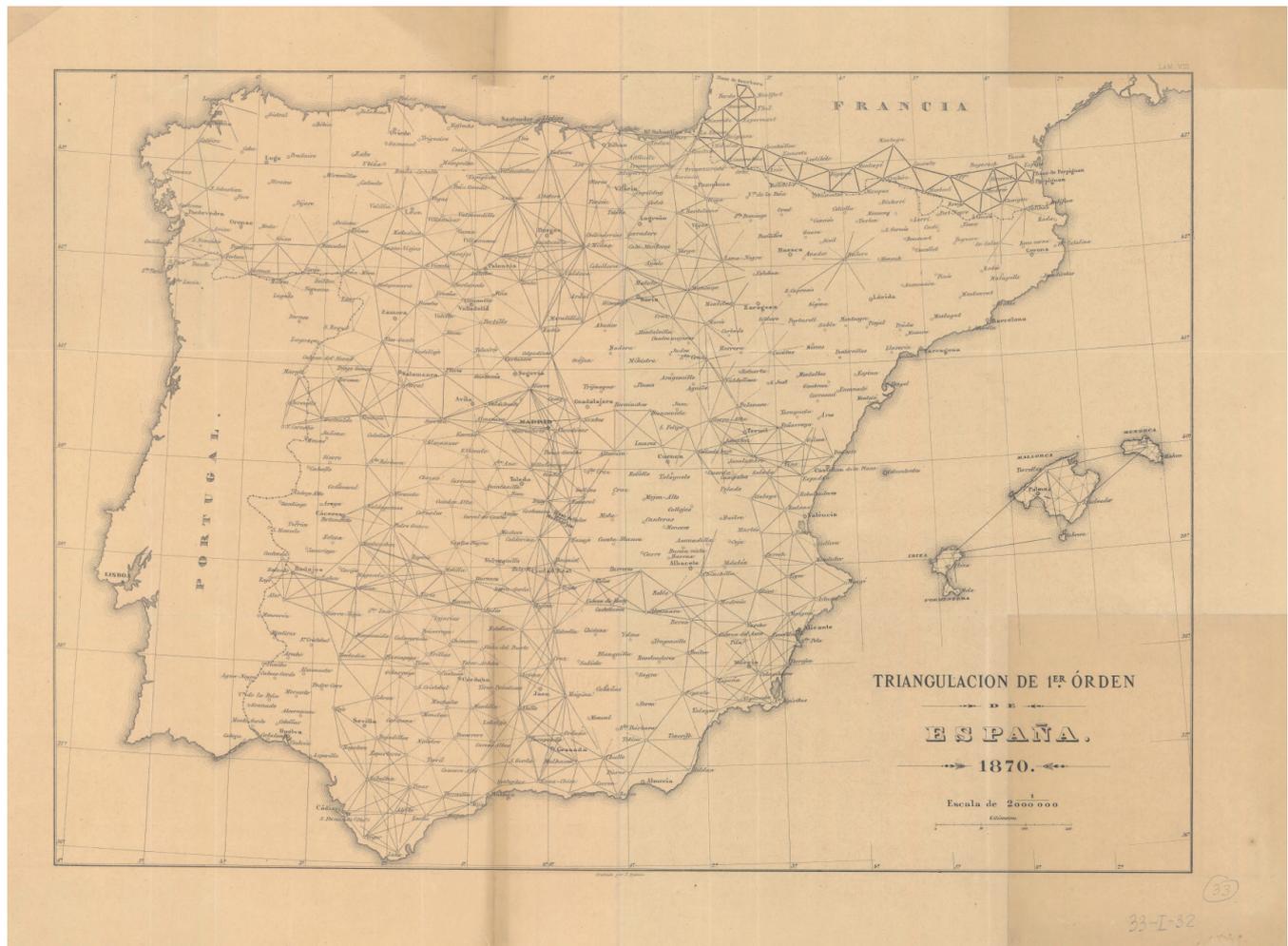


Figura 2. Mapa de España. Triangulación geodésica, escala 1:2.000.000 (IGN, 1870).



A mediados de los años 70 del siglo XX, las redes geodésicas españolas cambiaron su estructura para pasar a estar formada esta vez por dos órdenes, primero e inferior (BOE, 1975). La Red de Primer Orden estaba formada por vértices de la antigua red de primer orden con un total de unos 680 vértices; y la Red de Orden Inferior (ROI), lo conformaban las antiguas redes de segundo y tercer orden; los vértices de la red de primer orden formaban parte, a su vez, de la ROI. A su vez, desde 1975 éstas se encuentran protegidas por la Ley sobre Señales Geodésicas y Geofísicas (BOE, 1975).

Posteriormente, la aparición de la observación geodésica por técnicas espaciales, en concreto el GPS, materializó una nueva red apoyada en los 1.200 vértices de la ROI y uniformemente distribuida por toda España.

A principios del siglo XXI la necesidad de establecer una cartografía europea unificada hizo indispensable la conversión de las coordenadas de los marcos de los Sistemas Nacionales al Marco ETRS89. Tal determinación requirió el conocimiento de ambas clases de coordenadas en un número muy elevado de estaciones uniformemente distribuidas, debiendo ser este número mayor cuando más irregularidades presentase el Marco local (Regidor et al., 2000). Desde entonces esta red pasó a llamarse Red Geodésica Nacional por Técnicas Espaciales (REGENTE).

En la actualidad el Instituto Geográfico Nacional ha desplegado una nueva red de observación continua tanto GPS como GLONASS o Galileo. Ésta es la llamada Red de Estaciones de Referencia GNSS (ERGNSS) cuyos datos están disponibles de forma gratuita. Algunas de estas estaciones pertenecen, a su vez, a redes internacionales como la red de estaciones permanentes de EUREF (EPN) o la red del Servicio Internacional GNSS (IGS).

4. METODOLOGÍA PARA LA SUBIDA DE LOS VÉRTICES GEODÉSICOS DE ESPAÑA A WIKIDATA

A continuación, se procederá a mostrar la metodología seguida para la introducción de los vértices geodésicos de España en la base de conocimiento de Wikidata, con la intención que pueda ser replicable para cualquier tipo de información geográfica. Por ello, se ha planteado dividir esta metodología en seis fases consecutivas, que se describen a continuación:

4.1. Selección del origen de la información

Con carácter previo a la introducción de los datos en la plataforma es conveniente identificar cuatro aspectos esenciales sobre los mismos, como son su naturaleza pública, su estructura, la naturaleza de los datos geográficos (vectorial o ráster) y la disponibilidad de los datos en la plataforma.

Así pues, solamente son compatibles con Wikidata los datos que dispongan de una licencia Creative Commons o sean de dominio público, en coherencia con la naturaleza de esta plataforma libre. Asimismo, la información geográfica a almacenar en Wikidata debe ser de tipo vectorial, mientras que la información ráster debe introducirse en el almacén de archivos multimedia Wikimedia Commons. Por este motivo se trabajará con información vectorial en esta metodología propuesta.

Aunque se pueden introducir en Wikidata elementos geográficos de diversos tipos se va a plantear una metodología que introduce elementos geográficos puntuales. La información geográfica más simple es la que se representa como un punto, pues solamente precisa dos coordenadas. Sin embargo, la plataforma no trabaja con coordenadas planas, sino esféricas, por lo que es conveniente que se disponga de la información geográfica en este formato. Además, Wikidata opera únicamente con coordenadas decimales en lugar de grados, minutos y segundos, por lo que, si procede, es necesario realizar la transformación. Ello se puede realizar mediante programas como QGIS o ARCGIS, así como con aplicaciones libres de internet como Canadensys (<https://data.canadensys.net/tools/coordinates>).

Adicionalmente, la aplicación de esta metodología requiere introducir datos que no existen en la base de conocimiento, por lo que, con carácter previo a todas las fases se ha planteado el análisis, revisando previamente los datos disponibles en ese momento en Wikidata.



Igualmente es conveniente analizar el formato en el que se encuentran los datos, ya que disponer de archivos con la información en formato no textual o sin datos estructurados es incompatible con el almacenamiento de información en Wikidata.

Una vez que se ha tomado en consideración todo lo anterior, se inicia una consulta en Wikidata para comprobar la existencia de la información geográfica a subir, a fin de no generar duplicados y garantizar que no se había realizado previamente el trabajo. Dicha consulta se puede realizar desde Wikidata Query Service (<https://query.wikidata.org/>), esto es, el servicio de consultas de la base de conocimiento que emplea el lenguaje de consultas denominado SPARQL, un lenguaje estandarizado similar a SQL, aunque adaptado a la consulta de grafos RDF (Resource Description Framework).

De este modo, en primer lugar, se comprobó que la base de datos de vértices geodésicos es de carácter público, contenía información espacial y, más concretamente, de tipo puntual con coordenadas esféricas. Posteriormente, se procedió a realizar la consulta sobre la disponibilidad de la información en Wikidata, mediante la herramienta desde Wikidata Query Service.

Tras realizar ambas consultas, se observó que los vértices geodésicos encontrados a escala global no fueron muy numerosos. No alcanzaban los 500 elementos, y principalmente se concentraban en unos países concretos de Europa, Oceanía y Sudamérica. En el caso de España, únicamente se encontraron 12 vértices geodésicos (Rivilla, Hazas, San Bartolomé, Canaleja, Toranzo, Cárdenas, Boca del Río, Cabo de Gata, Retamar, Viso, Limaria y Mágina), que estaban posicionados tal y como se muestra en la figura 3.

Una vez comprobado que la página del Instituto Geográfico Nacional (IGN) cuenta con varios repositorios sobre vértices geodésicos y que esta información tiene una licencia compatible con CC-BY 4.0., se inició el procedimiento. Así pues, tras haber realizado las comprobaciones previas, se identificó la información buscada en dos repositorios del Instituto Geográfico Nacional (IGN). Una primera fuente de datos (fuente 1), está disponible en la siguiente dirección web: <https://centrodedescargas.cnig.es/CentroDescargas/catalogo.do?Serie=CAANE#> (sección Redes geodésicas). La página de descarga ofrece dos formatos para obtener los archivos: .kmz o .xls. Se descargó el segundo (archivo .xls), con el nombre de "Coordenadas Red Geodésica" que contaba con dos hojas en el mismo archivo. La primera incluía 10849 vértices de la península ibérica y Baleares, mientras la segunda contenía 294 de las Islas Canarias.



Figura 3. Consulta a Wikidata Query Service durante la fase de selección del origen de la información, con el objetivo de mostrar la posición de los vértices geodésicos en España. Fuente: Elaboración propia a partir de Wikidata Query Service.



La información de este archivo es de tipo geográfico y temático e incluye información sobre la Red Regente y la Red de Orden Inferior. Asimismo, dado que se aportan coordenadas en dos sistemas distintos (ED50 y ETRS89), se adoptará este último al ser más reciente y, por lo tanto, más preciso. Asimismo, al proporcionar la información en grados, minutos y segundos fue necesaria su transformación en decimales, ya que Wikidata solamente puede obtener las coordenadas en dicho formato.

Tabla 1. Columnas del archivo Coordenadas de la Red de Orden Inferior.

Variables		Definición	
Número		Número de hoja	
Nombre		Nombre de la localidad del vértice	
REGENTE		Coincide con la base de datos REGENTE (sí/no)	
Inoperativo		Está inoperativo el vértice	
ETRS89	Geográficas	Latitud	Latitud en el sistema ETRS (Grados, minutos y segundos)
		Longitud	Longitud en el sistema ETRS (Grados, minutos y segundos)
		helip(BP)	Altura elipsoidal calculada en la base de pilar (BP)
		helip(CF)	Altura elipsoidal en la parte de arriba del pilar (CF)
	Desv. Típica (m)	Lat.	Desviación típica de la latitud
		Lon.	Desviación típica de la longitud
		h	desviación típica de la altura
	UTM	X	Coordenadas UTM X
		Y	Coordenadas UTM Y
		Huso	Hoja número
ED50	Geográficas	Latitud	Latitud en el sistema ED50 (Grados, minutos y segundos)
		Longitud	Longitud en el sistema ED50 (Grados, minutos y segundos)
	UTM	X	Coordenadas X planas UTM según el sistema ED50
		Y	Coordenadas Y planas UTM según el sistema ED50
		Huso	Hoja número
H ortom (BP)		Altura ortométrica	
Pilar		altura del pilar	

Fuente: Elaboración propia a partir de IGN (2023).

Esta información inicial pudo ser complementada con información de la red de vértices geodésicos en otra fuente (fuente 2), perteneciente a la misma institución (IGN) ubicado en la siguiente dirección web: <https://www.ign.es/web/ign/portal/gds-vertices/-/vertices-geodesicos/setTabNumber>. Esta fuente es un buscador de los vértices geodésicos que incluye información de dos tipos de redes de vértices: la Red REGENTE y de la Red ROI (Red de coordenadas de Orden Inferior). En esta base de datos se aloja un archivo en formato PDF por cada uno de los vértices de España, en el que se ofrece una información descriptiva más detallada, como se puede observar en la figura 4.



Reseña Vértice Geodésico 1-jul-2022

Número.....: **51797**
 Nombre.....: **Abadía**
 Municipios: Utrillas
 Provincias: Teruel
 Fecha de Construcción.....: 06 de agosto de 1986
 Pilar sin centrado forzado...: 1,20 m de alto, 0,30 m de diámetro.
 Último cuerpo.....: 1,00 m de alto, 1,00 m de ancho.
 Total cuerpos.....: 1 de 1,00 m de alto.

— Coordenadas Geográficas: —

Sistema de Ref.:	ED 50	ETRS89
Longitud.....:	- 0° 52' 57,0980"	- 0° 53' 01,53149" ±0.094 m
Latitud.....:	40° 47' 59,5467"	40° 47' 55,39407" ±0.061 m
Alt. Elipsoidal...:		1473,316 m ±0.079 (BP)
Compensación..:	01 de septiembre de 1991	01 de noviembre de 2009 Elipse de error al 95% de confianza.

— Coordenadas UTM. Huso 30 : —

Sistema de Ref.:	ED 50	ETRS89
X.....:	678634,39 m	678525,472 m
Y.....:	4518777,42 m	4518568,298 m
Factor escala....:	0,999992746	0,999992297
Convergencia...:	1° 23' 02"	1° 22' 59"
Altitud sobre el nivel medio del mar:		1421,605 m. (BP)

— Situación: —
 En la Sierra de San Just en la zona denominada Abadía al borde del Cortado de la cara N., próximo a la línea límite entre Utrillas y Escucha.

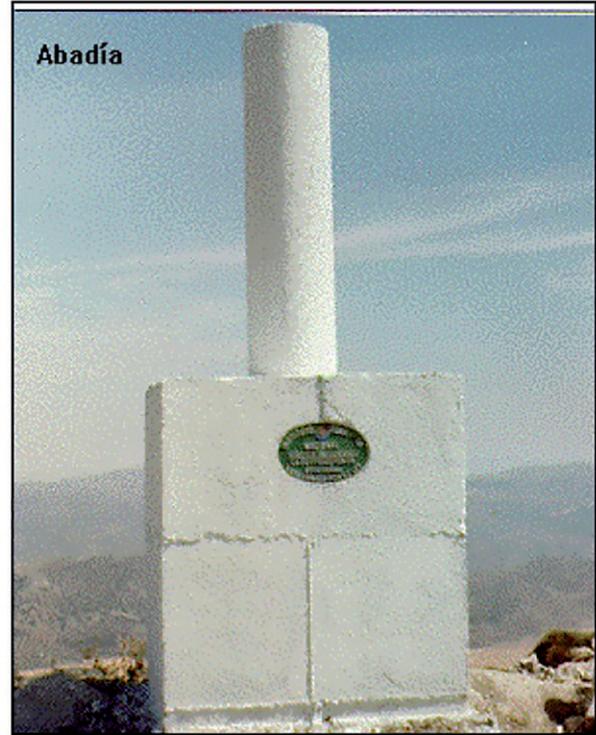


Figura 4. Ejemplo de la información de detalle disponible en un archivo correspondiente al vértice geodésico de Abadía en la provincia de Teruel. Fuente: IGN (2022).

Sin embargo, esta información solamente se puede proporcionar en Wikidata a través del enlace. Por lo tanto, solamente se han podido tabular algunos datos de la fuente 2 que se obtendrían a través de la página del buscador. Así, para obtener la información del buscador es necesario proporcionar entre qué pares de código de vértices debía extraerse la consulta (es decir, todos los vértices de la base de datos). Tras revisar la fuente 1 se identificó el código del vértice de mayor magnitud y se procedió a consultar en la página todos los vértices existentes entre 1 y 107.829. Dicha consulta proporcionó una lista de 11.127 vértices, con accesos directos a la misma cantidad de documentos en PDF con información detallada, como la que se muestra en la figura 4. Esta información se ha podido incluir en la tabla 2.

La tabla obtenida de la fuente 2 se insertó en la hoja de cálculo, junto a los datos de la fuente 1, y se ordenaron ambas tablas para ser comparadas y ver si coincidían respecto al campo que tienen en común, el campo "Número". Ambos listados deberían ser coincidentes, pero había 16 vértices (Angarillones, Arcela, Bandera, Cornialto, Coroa, Gestosa, La Rhune, La Rhune (antena), Larouco, Loma Negra, Morrón, Pedrouzos, Raso de Vitoria, San Xoán, Sistalto, Tripazul) que aparecían en la fuente 1 y no en el buscador (fuente 2).

Tras la unificación de ambas fuentes de datos mediante el único campo en común que tienen (campo "número"), se unificó toda la información en una sola tabla que se muestra en la tabla 3. Así, se obtuvo un fichero con 15 columnas y 11143 filas, correspondientes a la misma cantidad de vértices geodésicos.



Tabla 2. Columnas del buscador de vértices geodésicos de la Red REGENTE y de la Red ROI (Red de Orden Inferior).

Variable	Descripción
Número	Código del vértice
Nombre	Nombre del vértice
Municipio	Municipio del vértice
Provincia	Provincia del Vértice
Nº Hoja MTN	Hoja del Mapa Topográfico Nacional
Coordenadas geográficas	Coordenadas geográficas del vértice en el sistema ETRS89 y ED50
Coordenadas UTM	Coordenadas planas del vértice en el huso correspondiente según los sistemas ETRS89 y ED50

Fuente: Elaboración propia a partir de IGN (2023).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1						ETRS89									
2						Geográficas			Desv. Típica (m)			UTM			
3	Número	Nombre	REGENTE	Inoperativo	Latitud	Longitud	helig(BP)	helig(CF)	Lat.	Lon.	h	X	Y	Huso	
4986	51767	Cantera de las Umbrías			40° 47' 55.17122"	0° 57' 33.10333" W	1419,72		0,03	0,03	0,04	672026,535	4518410,51	30	
4987	51791	Cantera			40° 41' 12.30023"	0° 51' 31.38869" W	1499,88		0,03	0,03	0,04	680941,298	4506189,07	30	
4988	51797	Abadía			40° 47' 55.39407"	0° 53' 1.53149" W	1473,32		0,03	0,04	0,04	678525,472	4518568,3	30	
4989	51816	San Just			40° 46' 10.21933"	0° 48' 5.20168" W	1574,96		0,01	0,03	0,03	685551,273	4515495,84	30	
4990	51818	Muela de Montalbán			40° 48' 53.2132"	0° 47' 46.40885" W	1345,46		0,04	0,03	0,04	685865,507	4520533,34	30	
4991	51821	Mojones			40° 41' 53.43614"	0° 46' 6.19729" W	1479,46		0,03	0,03	0,04	688542,736	4507647,62	30	
4992	51826	Trinidad	Sí		40° 45' 59.7										
4993	51843	Peñalisa			40° 43' 7.32										
4994	51850	Collado Alto			40° 40' 17.9										

Red de Infraestructuras Geodésicas
Subdirección General de Astronomía, Geodesia y Geofísica

Reseña Vértice Geodésico 1-jul-2022

<p>Número: 51797 Nombre: Abadía Municipios: Umbrías Provincias: Teruel Fecha de Construcción: 06 de agosto de 1988 Pilar sin centrado forzado: 1,20 m de alto, 0,30 m de diámetro. Último cuerpo: 1,00 m de alto, 1,50 m de ancho. Total cuerpos: 1 de 1,00 m de alto.</p>	
---	---

Figura 5. Representación del campo en común de ambas bases de datos. Fuente: Elaboración propia a partir de IGN (2023).

Tabla 3. Características de la tabla definitiva a subir a Wikidata.

Variable	Descripción
Código	Código de vértice
Nombre	Nombre del vértice
Descripción en español	Descripción de qué se inserta
Descripción en inglés	Descripción en inglés de lo que se inserta
Instancia de	Tipo de información que se introduce
País	País donde está situado
Alias	Otro nombre alternativo
Municipio	Municipio del vértice



Variable	Descripción
Provincia	Provincia del Vértice
REGENTE	Coincide con la base de datos REGENTE (sí/no)
Inoperativo	Si está inoperativo el vértice
Coordenada	Coordenadas ETRS en formato decimal
Altitud	Altura en m respecto al nivel del mar
Altura	Altura de la infraestructura
URL de origen	Url del archivo con información detallada sobre el vértice

Fuente: Elaboración propia a partir de IGN (2023)

4.2. Limpieza de los datos

Esta fase consiste en la corrección o eliminación de registros para lo cual hay que conocer la estructura de la plataforma de destino, donde van a ser insertados los datos, ya que ésta impone la forma en la que se almacenará la información. Así, una vez que se dispone de toda la información será necesario descartar los valores que no serán ingresados y adaptar los que vayan a utilizarse, para que se puedan insertar correctamente.

Igualmente, en esta etapa resulta conveniente conocer detalladamente los formatos que puede comprender Wikidata para introducir la información de un modo correcto. De este modo, cada elemento introducido en la base de conocimiento dispone de un identificador único que comienza por una "Q" seguida de un código numérico. Por lo tanto, en este caso, cada vértice geodésico debe tener un identificador "Q" único.

Los elementos están compuestos por declaraciones, también denominadas afirmaciones, que describen qué es el elemento. En este caso, las declaraciones aportarán información geográfica y están compuestas por una pareja formada por una propiedad, que comienzan por la letra "P" seguida de un código numérico y su valor correspondiente. Por ejemplo, el vértice geodésico del municipio de A Carba tendrá varias declaraciones, siendo una de ellas la propiedad de la elevación sobre el nivel del mar (P2044) y otra su valor 907,99 metros.

A este respecto debe tomarse en consideración que cada base de conocimiento dispone de sus propiedades para almacenar información como, por ejemplo, el país, el municipio, la altura o la provincia. Por lo tanto, cuando se introduzca una nueva declaración en un elemento, se deben revisar los valores que puede tomar la propiedad en cuestión, para que la plataforma entienda qué se va a introducir. Por ejemplo, la propiedad elevación sobre el nivel del mar (P2044) espera un valor numérico como el indicado anteriormente (907,99 metros), y se sitúa en la entidad territorial administrativa (P131), es decir, en un municipio, ya introducido previamente en la base de conocimiento con su correspondiente código territorial. De esta manera, al introducir un municipio existente se podrá relacionar con toda la información que ya existe en Wikidata para el mismo municipio.

Así pues, una vez que han sido agrupados todos los datos en un mismo fichero, se constató que en algunos casos el nombre no coincidía, principalmente porque la denominación estaba en gallego en la fuente 2 y en español en la fuente 1. En aquellos casos donde el nombre coincidía se borró el dato, y los restantes se utilizaron como "Alias", para ser introducidos como tal en Wikidata.

Al copiar los datos de la fuente 2 en la hoja de cálculo, el nombre del vértice geodésico contenía el enlace al archivo PDF explicativo, pero este no se mostraba como una URL, por lo que se programó la siguiente función en Visual Basic for Applications (VBA) para extraerlas:

- Function Enlace(Rango As Range)
- Dim Hipervinculo As String
- Hipervinculo = Rango.Hyperlinks(1).Address
- Enlace = Hipervinculo
- End Function



Por ejemplo, tras poner =Enlace(H2) en una celda de la hoja de cálculo, esta devuelve la dirección web del vértice que se encuentra en la celda H2, el de A Carba: https://datos-geodesia.ign.es/Red_Geodesica/Hoja0023/002355.pdf. Una vez conseguidos todos los enlaces en una columna, se renombró esta como "Origen".

Una vez comprobada la coincidencia de todos los valores en las dos fuentes, se borraron aquellas columnas que no se iban a insertar en Wikidata, como la hoja MTN de la fuente 2 o todos los datos de coordenadas de la fuente 1.

Una vez que se ha unificado la información en un solo fichero se ha iniciado el proceso de limpieza de datos y, por lo tanto, es necesario un cotejo previo para incluir los identificadores que existen en la plataforma. Por ejemplo, y siguiendo el ejemplo expuesto en la metodología, el vértice A Carba tendrá la siguiente declaración: situado en la entidad territorial administrativa (P131) con el valor Villalba (Q1605437).

Así pues, la adaptación del contenido de la tabla 3 a lo que requiere Wikidata se ha realizado con la plataforma OpenRefine, cuyo propósito es limpiar los datos y proporcionar los identificadores correspondientes al usuario que sube información a Wikidata de un modo intuitivo y rápido. De este modo, a la tabla 3 inicial le corresponden en Wikidata los valores que se muestran en la tabla 4.

Tabla 4. Datos con los que se trabajó en OpenRefine.

Columna tabla 2	Propiedad en WD	Ejemplo (Q115498684)	Elemento en WD
Num	P528	2355	
Nombre	Etiqueta	A Carba	
Descripción en español	Descripción	vértice geodésico en Vilalba, España	
Descripción en inglés	Descripción	triangulation station in Vilalba, Spain	
Alias	También conocido como	Carba	
Instancia de	P31	Vértice geodésico	Q131862
País	P17	España	Q29
Municipio	P131	Vilalba	Q1605437
Provincia		Ourense	
REGENTE	P361	Regente	Q115497793
Inoperativo	P5817		
Coordenada	P625	43.4203692, -7.6586297	
Altitud	P2044	907,99	
Altura	P2048	1,2	
URL de origen	P973	https://datos-geodesia.ign.es/Red_Geodesica/Hoja0023/002355.pdf	

Fuente: Elaboración propia (2023).

La columna provincia se introdujo en OpenRefine, pero no con la finalidad de subir el dato a Wikidata, sino como ayuda para el cotejamiento, ya que existen municipios con el mismo nombre en provincias diferentes, por lo que resultó imprescindible añadir el campo provincia.

Como se observa en la tabla 4, algunas de las columnas fueron insertadas directamente en Wikidata, como el número de identificación, las descripciones, el alias, la coordenada, la altitud, la altura o el origen de los datos. Sin embargo, el resto debía ser cotejado con Wikidata para introducir los elementos concretos de esta base de conocimiento. El cotejamiento de la columna "inoperativo" y la columna "REGENTE" fue muy sencillo. En el caso de inoperativo contenían la expresión "fuera de servicio" y en el caso de "REGENTE" contenían el nombre de la "red" para indicar que forma parte de dicha red.



11143 filas											Extensiones	Wikidata					
Mostrar como: filas registros											Mostrar: 5 10 25 50 100 500 1000 filas		« primera	< anterior	1	siguiente >	última »
▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼					
Todo	num	Nombre	es descripción	en descripción	Nombre alternativo	Municipio	Provincia	Regente	Inoperativo	coord	Altitud						
☆	1.	30207	A Aguioncha <small>Escoger nueva coincidencia</small>	vértice geodésico en Calvos de Randín, España	triangulation station in Calvos de Randín, Spain	Aguioncha	Calvos de Randín <small>Escoger nueva coincidencia</small>	Ourense		41.9621464, -7.831315	1238.923						
☆	2.	12542	A Area <small>Escoger nueva coincidencia</small>	vértice geodésico en Pedrafita do Cebreiro, España	triangulation station in Pedrafita do Cebreiro, Spain	Area	Pedrafita <small>Escoger nueva coincidencia</small>	Lugo		42.7029672, -7.0498329	1394.269						
☆	3.	18945	A Arma <small>Escoger nueva coincidencia</small>	vértice geodésico en Castro Caldelas, España	triangulation station in Castro Caldelas, Spain	Arma	Castro Caldelas <small>Escoger nueva coincidencia</small>	Ourense		42.4170531, -7.3716485	1016.483						
☆	4.	2355	A Carba <small>Escoger nueva coincidencia</small>	vértice geodésico en Vilalba, España	triangulation station in Vilalba, Spain	Carba	Vilalba <small>Escoger nueva coincidencia</small>	Lugo	Red Geodésica Nacional por Técnicas Espaciales <small>Escoger nueva coincidencia</small>	43.4203692, -7.6586297	907.99						
☆	5.	9316	A Carbaliza <small>Escoger nueva coincidencia</small>	vértice geodésico en Dumbría, España	triangulation station in Dumbría, Spain	Carbaliza	Dumbría <small>Escoger nueva coincidencia</small>	A Coruña		42.9398298, -9.1222537	474.581						
☆	6.	851	A Carballoso <small>Escoger nueva coincidencia</small>	vértice geodésico en Ourot, España	triangulation station in Ourot, Spain		Orol <small>Escoger nueva coincidencia</small>	Lugo		43.5237529, -7.6786648	729.212						
☆	7.	9821	A Chaira <small>Escoger nueva coincidencia</small>	vértice geodésico en O Páramo, España	triangulation station in O Páramo, Spain		Páramo <small>Escoger nueva coincidencia</small>	Lugo		42.8555208, -7.4450524	513.156						
☆	8.	22494	A Chan de Moeiras <small>Escoger nueva coincidencia</small>	vértice geodésico en Crecente, España	triangulation station in Crecente, Spain	Chan do Moiras	Crecente <small>Escoger nueva coincidencia</small>	Pontevedra		42.2395614, -8.2087744	797.308						

Figura 6. OpenRefine tras cotejar la información con Wikidata. Fuente: Elaboración propia (2023).

Cotejar el nombre también fue sencillo, ya que únicamente había 12 vértices geodésicos insertados, por lo que el resto tenían que crearse. Esos 12 fueron cotejados con los correspondientes elementos en Wikidata. La columna más costosa fue la del municipio. Había 4990 municipios diferentes, y más de la mitad no fueron cotejados automáticamente porque existían dudas sobre su coincidencia con los datos existentes en Wikidata. Uno a uno y con ayuda de la columna provincia se cotejaron todos los municipios con los correspondientes de Wikidata.

Para concluir este proceso, además de la información que se coteja, hay que controlar el resto que se va a introducir en Wikidata. Por ejemplo, el archivo de datos contenía numerosos vértices con una altura de 0 metros. Wikidata muestra un error tras grabar la edición, pero permite hacerla, a pesar de que no tiene sentido. Por ello, estos valores deben localizarse y borrarse para que no se inserten erróneamente en la base de conocimiento.

4.3. Inserción

El almacenamiento de información masiva hace que sea imposible guardar uno por uno miles de datos en una base de conocimiento, por el contrario, es necesario emplear herramientas o lenguajes de programación que permitan hacerlo a gran escala. Así pues, una vez limpiados los datos con la hoja de cálculo y OpenRefine, y tras haberlos ajustado a la información que contiene Wikidata, es necesario crear un esquema, esto es, una plantilla de ediciones de Wikidata que se aplica a cada fila que se va a insertar. En la parte superior derecha del programa existe un enlace denominado "Editar esquema de Wikidata". En él aparecen todas las columnas insertadas, y con una línea inferior verde, aquellas columnas que ya fueron cotejadas anteriormente, esto es, que contienen información que se solapa con la que vamos a introducir, por ejemplo, el nombre de los municipios, que habían sido introducido previamente la base de conocimiento. El enlace "añadir elemento" permite crear un elemento nuevo y, una vez pulsado se pide insertar una columna cotejada previamente.

En las declaraciones hay que indicar en la parte izquierda la propiedad, y en la derecha la columna con la variable que se va a introducir en cada caso. En la parte inferior existe un botón para añadir referencia, donde se inserta el origen de los datos, e incluye información como la siguiente: la dirección URL, el título, el idioma y la fecha de consulta, tal y como muestra la figura 7.



Figura 7. Esquema realizado en OpenRefine para añadir la coordenada y su referencia. Fuente: Elaboración propia (2023).

Una vez finalizado el esquema con todas las columnas que se quieren introducir, se debe acceder a la pestaña “Problemas” que se encuentra en el menú de la parte superior, ya que esta opción indica las incongruencias de los datos trabajados. Por ejemplo, es habitual que indique que el formato de algunas coordenadas no es correcto, que no se han añadido referencias en algunas declaraciones, que falta por revisar alguna celda de las columnas que han sido cotejadas donde se ha producido la inserción de dos valores en la misma declaración u otros múltiples errores posibles.

Una vez respondido a los problemas que se indica en esta opción, se puede previsualizar en la pestaña que se encuentra al lado de problemas qué ediciones se van a realizar en Wikidata. Cuando todo esté correcto, se puede pulsar en “Cargar ediciones en Wikidata” en el menú superior derecho. Esta opción pedirá el nombre de usuario de Wikidata para realizar las ediciones en nuestro nombre, aunque justo en la parte inferior proporciona otra posibilidad para la inserción, la exportación a QuickStatements.

Esta última es una herramienta que sirve para insertar valores masivos en Wikidata, lo que permite separar las ediciones en lotes de varios miles, permite detener la subida de cada lote o reactivarlo más tarde. Con esta herramienta también es posible previsualizar los datos de subida e ir comprobando posibles errores. QuickStatements es una opción interesante cuando el volumen de los datos a subir es muy alto, ya que OpenRefine podría sufrir algún tipo de problema y dejar la inserción sin finalizar, no permitiendo saber qué datos han sido insertados y cuáles no. Por el contrario, QuickStatements sí permite retomar la subida de información donde se había detenido.

A continuación, siguiendo la metodología descrita para esta fase se ha procedido a crear un esquema, esto es, una plantilla de ediciones de Wikidata que se aplica a cada fila que se va a insertar. Tras pulsar el enlace “añadir elemento” se comprobó que la columna vértice ya había sido cotejada, (había 12 creados y el resto fueron nuevos). A continuación, se introdujeron los términos y declaraciones. En los términos se añadió la etiqueta, descripción y alias, mientras que en las declaraciones el resto de columnas que se muestran en la tabla 4.

Por lo tanto, en la parte izquierda de las declaraciones se indicó la propiedad, que se corresponde con la segunda columna de la tabla 4, y en la parte de la derecha, la columna o variable correspondiente. Finalmente se introdujo la referencia de los datos, esto es la dirección URL de origen, el título, el idioma y la fecha de consulta, tal y como muestra la figura 8.



11143 filas Esquema Problemas Previsualización Extensiones Wikidata

Actualmente estás trabajando contra Wikidata. Primero debe conciliar sus datos con los servicios de conciliación vinculados a Wikidata (como \$3). El esquema a continuación especifica cómo se transformarán sus datos tabulares en ediciones de Wikidata. Puede arrastrar y soltar los nombres de las columnas a continuación en la mayoría de los cuadros de entrada: para cada fila, las ediciones se generarán con los valores de estas columnas.

Guardar esquema Descartar cambios

num Nombre es descripción en descripción Nombre alternativo Municipio Provincia Regente Inoperativo coord Coordenada Altitud Altura Origen

Nombre

Términos
no se añadió ninguna etiqueta, descripción ni alias

Declaraciones
coordinate location coord

1 referencias

reference URL	Origen
title	es Reseña Vértic
language of w	Spanish
retrieved	2022-12-01

Figura 8. Esquema realizado en OpenRefine para añadir la coordenada y su referencia. Fuente: Elaboración propia (2023).

Una vez finalizado el esquema con todas las columnas que se quieren introducir, se accedió a la pestaña “Problemas”, donde se identificaron algunas incongruencias de los datos a introducir y se corrigieron. Posteriormente, se subió la información mediante la exportación a QuickStatements.

4.4. Revisión

Los datos ingresados pueden contener datos erróneos, aunque la limpieza se realizara correctamente. Por ello, se deben hacer consultas a la base de conocimiento para revisar si existen valores extraños, duplicaciones o ausencia de información que se esperase encontrar.

En la revisión de la información introducida, tanto con OpenRefine como con QuickStatements se deben realizar miles de ediciones. En caso de que se inserten todos los elementos a la vez con OpenRefine serían tantas revisiones como finales tengamos en el archivo, cada una de ellas de muchos bytes, ya que contienen las etiquetas, descripciones, declaraciones y referencias. Por el contrario, si la inserción se realiza con QuickStatements el tamaño de cada edición sería mucho menor, ya que se realiza una edición para cada etiqueta, descripción, declaración y referencia, sin embargo, el número de ediciones se multiplicaría. Se escoja un sistema u otro, el número de errores podría ser considerable.

En el caso de QuickStatements los errores se pueden visualizar en cada lote creado. Cuando termina con un lote se puede repetir para solucionar los problemas mostrados. Sin embargo, es posible que este procedimiento genere una duplicación de los datos involuntaria. Por lo tanto, una vez terminada la inserción en Wikidata, deben realizarse consultas a la base de conocimiento para revisar que todos los elementos contienen la información correcta.

La consulta para ver si existían vértices con dos valores dentro de la propiedad situada en la entidad territorial administrativa (P131) fue la siguiente:



```
SELECT DISTINCT ?item ?itemLabel ?count ?sample1 ?sample2 WHERE {  
  {  
    SELECT ?item (COUNT(?value) AS ?count) (MIN(?value) AS ?sample1) (MAX(?value) AS ?sample2) WHERE {  
      ?item wdt:P31 wd:Q131862;  
      wdt:P17 wd:Q29;  
      p:P131 ?stat.  
      ?stat ps:P131 ?value;  
      rdf:type wikibase:BestRank.  
    }  
    GROUP BY ?item  
    HAVING (?count > 1 )  
  }  
  SERVICE wikibase:label { bd:serviceParam wikibase:language "de,en". }  
}  
ORDER BY DESC (?count)
```

(En la siguiente dirección se puede visualizar la consulta en Wikidata Query Service: <https://w.wiki/69yR>). Después se comprobó que los valores numéricos no contenían errores en los datos, como valores extremos. También, se realizó la comprobación de que todos los vértices se localizaban dentro del territorio nacional. Igualmente, se buscaron los 16 vértices que aparecían en la fuente 1, pero no en la fuente 2, ya que éstos no podían tener la misma referencia, al no disponer del enlace URL que se obtuvo de la página web.

4.5. Visualización y análisis

El almacenamiento de los datos en sí mismo no aporta nada. Éstos deben ser útiles, es decir, deben permitir la visualización en tablas, la obtención de estadísticas, la realización de cálculos, la realización de gráficos e infografías o su representación en mapas. Es, por tanto, información que será analizada posteriormente. El instrumento esencial para consultar la información en Wikidata es el Wikidata Query Service cuyas consultas deben formularse en el lenguaje SPARQL, como se había mencionado en fases anteriores. Esta aplicación permite visualizar la información en formato tabla, gráfico o mapa, siempre que la información a representar incluya coordenadas geográficas en este último caso.

No obstante, la información espacial a consultar también puede ser introducida en un GIS, multiplicando así las opciones de análisis espacial que permiten ir más allá de la mera visualización. Concretamente, en este trabajo se plantea el uso del plugin de QGIS "SPARQLing Unicorn", el cual permite realizar consultas en Wikidata mediante el lenguaje de consultas SPARQL.

Esta importación de datos de Wikidata se realiza seleccionando "Wikidata" como "endpoint" e incluyendo el texto de la consulta de la herramienta Wikidata Query Service, tras su ejecución se podrá disponer de datos en la aplicación QGIS. Sin embargo, para poder realizar las consultas correctamente es imprescindible que los nombres de los campos se adapten a lo que entiende SPARQLing, como las coordenadas (geo) y el elemento (item). La denominación del resto de campos es libre. Por otra parte, esta conexión directa con un repositorio público como Wikidata es posible toda vez que QGIS es un software de código abierto, por lo que son abundantes los accesos a servidores de acceso público. Asimismo, aunque no existe una herramienta de conexión directa entre Wikidata y ARCGIS siempre es posible exportar los datos en el formato csv desde Wikidata Query Service y añadir la información espacial a ArcGIS según dicho formato.



```
1 SELECT ?item ?itemLabel ?geo WHERE {
2   ?item wdt:P31 wd:Q39715 .
3   ?item wdt:P625 ?geo .
4   SERVICE wikibase:label {
5     bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en".
6   }
7 }
```

Figura 9. Ejemplo de consulta en el Wikidata Query Service.
Fuente: Elaboración propia a partir de Wikidata (2022).

Projecto Editar Ver Capa Configuración Complementos Vectorial Ráster Base de datos Web Malla MMQGIS Procesos Ayuda

Navegador

Capas

Editor de vértices

SPARQLing Unicorn QGIS Plugin

Linked Data Processing Query Helpers Settings Help

Query Interlink English (Español)

Select endpoint: Wikidata -> ?item

Quick Add RDF Resource

Filter Concept Results:

Query Templates: Bem+Label layer name:

Valid Query

```
1 SELECT ?item ?itemLabel ?geo WHERE {
2   ?item wdt:P31 wd:Q39715 .
3   ?item wdt:P625 ?geo .
4   SERVICE wikibase:label {
5     bd:serviceParam wikibase:language
6     "[AUTO_LANGUAGE],en".
7   }
8 }
```

Saved Queries: Christs: Load Query Query Name: Save Query

add layer

Language for query results: English (en)

Figura 10. Ejemplo de consulta en el SPARQLing Unicorn. Elaboración propia a partir de Wikidata (2022)

4.6. Replicación

Una vez que se ha completado la metodología es deseable revisar si el mismo proceso puede ser repetido con datos similares. Este sería el último paso y tiene por objeto validar la metodología, comprobando que el procedimiento es igualmente válido y aplicable a otros datos que tengan unas características similares.



Dentro de la página del IGN también se constató que existía información de la Red Geodésica Nacional de Estaciones de Referencia GNSS (ERGNSS), compuesta por 119 estaciones permanentes. Se disponía de la localidad, el código IDN, la ficha y en otra página enlazada en esta las coordenadas, que habían sido calculadas con software científico de precisión. Siguiendo la misma metodología que la descrita hasta el momento, se realizó la inserción de todas las estaciones de referencia GNSS, comprobando que el modelo es válido para datos geográficos similares, como puede verse en la consulta siguiente: <https://w.wiki/65Zm>. Misma información que puede encontrarse en la página oficial: <http://ntrip.rep-gnss.es>.

Dado que la cantidad de estaciones permanentes no es muy alta, permitió utilizar esta información para crear un mapa interactivo en la propia Wikipedia, en el artículo dedicado al término vértice geodésico, tal y como puede verse aquí: [https://es.wikipedia.org/wiki/V%C3%A9rtice_geod%C3%A9sico#Red_de_Estaciones_de_Referencia_GNSS_\(ERGNSS\)](https://es.wikipedia.org/wiki/V%C3%A9rtice_geod%C3%A9sico#Red_de_Estaciones_de_Referencia_GNSS_(ERGNSS)). También fue posible recoger los datos de la base de conocimiento para crear un listado en Wikipedia. En este caso se utilizó una plantilla denominada Wikidata list, que recibe una consulta de Wikidata y muestra la información en forma de tabla, tal y como se realizó en un anexo: https://es.wikipedia.org/wiki/Anexo:Estaciones_permanentes_GPS_en_Espa%C3%B1a

5. RESULTADOS

Tras haber detallado el procedimiento metodológico y aplicado dicha metodología, y tras haber empleado la base de vértices geodésicos como fuente de datos, se obtuvieron los siguientes resultados.

5.1. Visualización y análisis

Una vez completada la inserción y revisión se comprobó que se actualizaron o insertaron 1143 vértices geodésicos, que pueden verse tras ejecutar en Wikidata Query Service la siguiente consulta en SPARQL:

```
SELECT ?vértice_geodésico ?vértice_geodésicoLabel WHERE {  
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es". }  
  ?vértice_geodésico wdt:P31 wd:Q131862.  
  ?vértice_geodésico wdt:P17 wd:Q29.  
}
```

224

Los resultados pueden verse en el siguiente enlace: <https://w.wiki/64t6>. Asimismo, con los datos introducidos se pueden realizar consultas para conocer información de utilidad, como los vértices geodésicos que se encuentran a mayor elevación sobre el nivel del mar (<https://w.wiki/64yH>), los que su estructura es más alta (<https://w.wiki/64yK>), los que forman parte de REGENTE (<https://w.wiki/64yL>) o los que se encuentran fuera de servicio (<https://w.wiki/64yP>).

Sin embargo, seguramente, los datos más interesantes están relacionados con sus coordenadas, que pueden verse en un mapa, que muestra un punto rojo por cada uno de los vértices geodésicos (<https://w.wiki/6Dg8>). Estos mismos datos pueden consultarse también con la siguiente sentencia en SPARQL:

```
SELECT ?vértice_geodésico ?vértice_geodésicoLabel ?coordenadas WHERE {  
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }  
  ?vértice_geodésico wdt:P31 wd:Q131862;  
  wdt:P17 wd:Q29;  
  wdt:P625 ?coordenadas.  
}
```



Figura 11. Mapa obtenido con Wikidata Query Service que muestra un punto por cada vértice geodésico introducido. Fuente: Elaboración propia (2023).

Además, esta información puede ser utilizada por otras aplicaciones externas o de la propia Fundación Wikimedia, como WikiShootMe, herramienta que analiza Wikidata para indicar aquellos elementos que tienen una fotografía o que necesitan ser fotografiados.

De los 12 vértices geodésicos que existían antes de la inserción masiva, únicamente había fotografía para cinco. Para intentar aumentar esta cantidad se pidió a los socios de Wikimedia España, asociación que promueve la mejora de los proyectos de la Fundación Wikimedia en España, su ayuda para recuperar fotografías existentes. Para ello, en primer lugar se ordenó la categoría en Wikimedia Commons que contenía los vértices geodésicos, denominada "Category:Trig points in Spain". Se pidió que fuesen colocadas las fotografías de esta categoría en Wikidata, en la propiedad de imagen (P18), además de subir aquellas fotografías que pudieran tener realizadas con anterioridad. Después de unas horas había más de 100 fotografías asociadas a su elemento en Wikidata.

Además, se pidió a Wikimedia España su colaboración para difundir en sus redes sociales el trabajo realizado en este artículo con el objetivo de conseguir más fotografías de los elementos que no disponían. Todas las imágenes de los vértices geodésicos insertados pueden verse en esta consulta: <https://w.wiki/64yW>.

Finalmente, también se procedió a consultar la información introducida en Wikidata mediante el plugin de QGIS SPARQLing Unicorn, adaptando la consulta anterior a los términos que maneja el plugin, concretamente sustituyendo "vértice_geodésico" por "item" y "coordenadas" por "geo". A continuación, se le ha añadido una sentencia opcional, para incorporar el valor elevación ("elev").

La consulta a ejecutar fue la siguiente:



```
SELECT ?item ?itemLabel ?geo ?elev WHERE {  
  
  ?item wdt:P31 wd:Q131862;  
  
  wdt:P17 wd:Q29;  
  
  wdt:P625 ?geo.  
  
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }  
  
  OPTIONAL { ?item wdt:P2044 ?elev. }  
  
}
```

De este modo, se obtuvo una cobertura puntual con información sobre la altura de los vértices geodésicos respecto al nivel del mar. Esta información fue clasificada en función de la altura de dichos vértices y, por lo tanto, se obtuvieron cuatro categorías de vértices, tal y como se muestra en la siguiente figura 12:

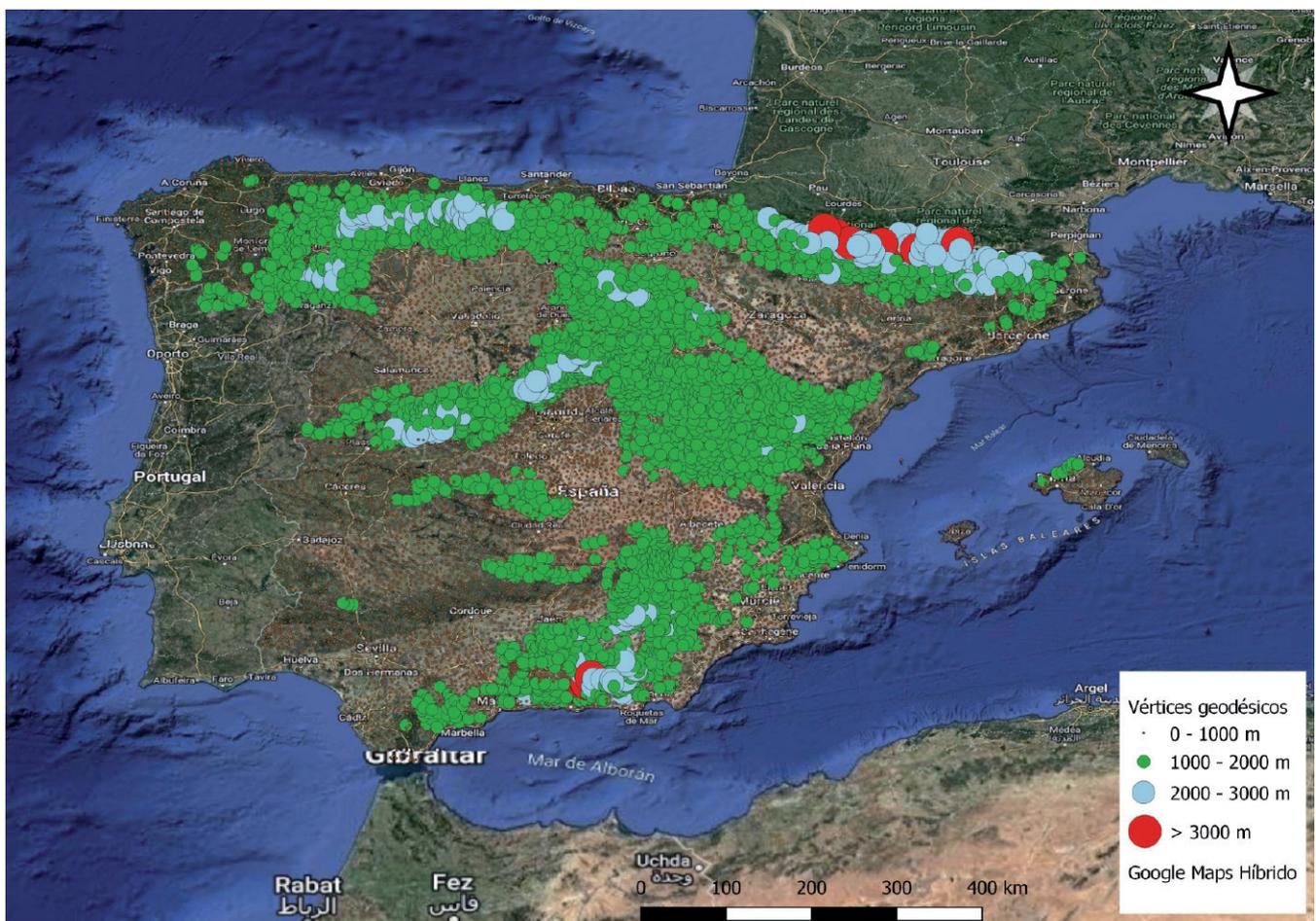


Figura 12. Mapa obtenido con los datos subidos a Wikidata y procesados con QGIS. Fuente: Elaboración propia (2023).

Esta información de partida permite identificar y jerarquizar los vértices con cuencas visuales de distinto tamaño. Así, la red de vértices de más de 3000 m tiene la cuenca visual más amplia, mientras que las que figuran entre 0 y 1000 m cuentan con la cuenca visual de menor tamaño. En el siguiente enlace interactivo es posible visualizar la clasificación de los vértices con mayor detalle https://slovenianman.github.io/vertices_geodesicos/.



6. DISCUSIÓN Y CONCLUSIONES

Una de las principales contribuciones de esta investigación consiste en el planteamiento de una metodología accesible para investigadores no especializados en programación, que permite introducir información geográfica a la base de conocimiento libre Wikidata y que, tanto estos como el conjunto de la comunidad científica pueda disponer de dicha información. Asimismo, esta información se ha empleado para realizar análisis geográficos a través de un plugin de interfaz con QGIS. Tal y como se había avanzado anteriormente, la revisión bibliográfica constató que, aunque cada vez hay más trabajos sobre la creación de repositorios en Wikidata, no son muchas las aportaciones que usan la plataforma como repositorio de información geográfica, por lo que puede considerarse que esta investigación es una aportación para la literatura científica de su campo.

Adicionalmente, otro aspecto destacable de este estudio consiste en el desarrollo de una metodología capaz de introducir en la plataforma información geográfica en grandes cantidades, en coherencia, con la creciente necesidad de datos por parte de los algoritmos de modelización para realizar sus predicciones. Por lo tanto, ello también representa otra contribución metodológica de esta investigación.

Igualmente, la expansión del repositorio Wikidata en el ámbito de la geografía es creciente, ha sido exponencial considerando los 10 años de existencia del mismo y su crecimiento será aún mayor gracias a metodologías como esta. Ello resulta coherente con las conclusiones de Andra et al. (2020) y Turki et al. (2019).

Por otra parte, aunque los vértices geodésicos han quedado en desuso con el advenimiento de la tecnología GPS, el modo en que ha sido seleccionada la ubicación de los vértices es interesante para un apreciable número de trabajos relacionados con la geografía y, más concretamente con los estudios de paisaje. Concretamente, tal y como se ha referido en el apartado 2, existen tres redes de vértices geodésicos jerarquizadas según varios criterios, siendo el esencial la obligación de que los vértices sean visibles entre sí. Ello suele ir relacionado con la altura, los vértices de mayor altura son visibles entre sí, mientras que los de baja altura serán más o menos visibles dependiendo de la orografía en las capas inferiores de la geografía. Por lo tanto, la cobertura de vértices geodésicos permite obtener un escaneo básico de la cuenca visual compartida por al menos tres puntos en tres rangos: entre 30 y 70 km, entre 10 y 25 km y entre 5 y 10 km. Gracias a ello es posible identificar cuencas visuales de distinta extensión en todo el territorio nacional, lo que ciertamente debe ser incluido en cualquier estudio de paisaje de ámbito municipal o comarcal. Todo ello puede comprobarse en la dirección mostrada en el apartado anterior https://slovenianman.github.io/vertices_geodesicos/.

Asimismo, dado que Wikidata ha incorporado en una propiedad las imágenes de Wikimedia Commons de los vértices geodésicos, como se ha mostrado en el enlace: <https://w.wiki/64yW>, es posible utilizar dicha información para análisis adicionales en la disciplina de estudios de paisaje. Entre ellos cabría destacar un inventario de imágenes de paisajes naturales, rurales y urbanos de toda la geografía española. A este respecto, cabe decir que existen investigaciones (Law et al., 2017; Li et al., 2017) con modelos de IA de redes convolucionales, que identifican objetos geográficos a partir de repositorios de imágenes de paisaje urbano. Estos investigadores emplean como fuente imágenes disponibles en Google StreetView, lo que excluye la disponibilidad de información sistemática de espacios naturales y rurales. Por lo tanto, la existencia de repositorios de imágenes como los existentes en Wikimedia Commons, son susceptibles de seguir expandiéndose, ofreciendo esta posibilidad para entornos naturales y rurales. Además, su acceso es totalmente libre, a diferencia de las imágenes que se ubican en Google Maps, por ejemplo.

Asimismo, la posibilidad de trabajar desde QGIS o ARCGIS con repositorios masivos y públicos como Wikidata favorece el proceso de análisis al incrementarse la disponibilidad y accesibilidad a los datos. Por otra parte, el hecho de que los datos geográficos de partida para realizar los análisis sean públicos y libres facilitará a cualquier investigador o revisor replicar la metodología de un autor y comprobar la veracidad de los resultados obtenidos. Igualmente dispondrá de la posibilidad de realizar análisis derivados o alternativos, que podrían enriquecer o alterar las conclusiones obtenidas. La introducción masiva de datos a escala planetaria favorece el enriquecimiento de la capacidad predictiva de los modelos y las conclusiones alcanzadas, en coherencia con los resultados de Andra et al. (2020).



Por otra parte, la metodología de introducción y análisis de datos propuesta no solo tiene implicaciones en la investigación de geografía, también en la difusión general de información geográfica en otras disciplinas de investigación o en el ámbito de la difusión por internet, ya sea institucional o personal. Wikidata es un repositorio que tiene la capacidad de proveer a los proyectos de la Fundación Wikimedia de información geográfica y de cualquier otro tipo, pero también a otras páginas de difusión públicas y privadas.

Entre las limitaciones más destacables cabe destacar el hecho de que el valor de los repositorios de Wikidata depende de la comunidad de editores que realizan contribuciones desinteresadamente. En este sentido, el beneficio percibido de la creación de estos repositorios es inmediato cuando se trata de una investigación específica que comparte su información de partida, pero el valor general del repositorio se incrementa en la medida en que las aportaciones sean más numerosas, lo que suele requerir un número de editores mayor. De este modo, los vértices geodésicos existentes en Wikidata son casi en su totalidad los aportados en este artículo. El conjunto de vértices geodésicos a escala planetaria en la base de conocimiento son 11.616, incluyendo algunas aportaciones significativas en Alemania y Australia (ver <https://w.wiki/69o9>). Sin embargo, en la medida en que la disponibilidad de información se extienda a otros países contiguos, continentes enteros y finalmente a escala global, el valor del repositorio se incrementará considerablemente.

Financiación

Este artículo es resultado de la colaboración de Ángel Obregón Sierra, Antonio Gavira Narváez y Rafael Vega Pozuelo, miembros del grupo de investigación "Nuevas Tecnologías de la Información y Comunicación Aplicadas al Territorio, Difusión del Conocimiento, Movilidad y Desarrollo Sostenible", con el investigador Javier López Otero miembro del Proyecto I+D+i PID2019-107993GB-I00 financiado por MCIN/AEI/10.13039/501100011033.

Declaración responsable y conflicto de intereses

Los cuatro autores declaran que no existe ningún conflicto de interés con relación a la publicación de este artículo. Todos los autores han participado en la revisión bibliográfica, la propuesta metodológica, la redacción del documento y los procesos de revisión requeridos previos a su aceptación definitiva.

REFERENCIAS

- Andra, W., Stupp, G., Sebastian, B. M., Good, B. M., Malachi, G., Griffith, O. L., ..., & Su, A. I. (2020). Wikidata as a knowledge graph for the life sciences. *eLife*. <https://doi.org/10.7554/eLife.52614>
- Castañón, J.C., Puyo, J.Y., & Quirós F. (2008). La herencia cartográfica y el avance en el conocimiento geográfico de España. En F. Quirós & J.C. Castañón (Eds.), *Madrid 1808 - Guerra y territorio - Mapas y planos 1808-1814* (pp. 109-127). Museo de Historia. https://geografia.uniovi.es/c/document_library/get_file?uuid=8405df4b-ee9c-4c25-919e-c8844a00bbb4&groupId=1700038
- Farda-Sarbas, M., & Müller-Birn, C. (2019). *Wikidata from a Research Perspective - A Systematic Mapping Study of Wikidata*. *arXiv*. <https://doi.org/10.48550/arxiv.1908.11153>
- Gaceta de Madrid*, núm. 257, de 14/09/1870, páginas 1 a 2. https://www.boe.es/diario_gazeta/comun/pdf.php?p=1870/09/14/pdfs/GMD-1870-257.pdf
- Garijo, D., & Poveda-Villalón, M. (2020). *Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web*. <https://doi.org/10.48550/arXiv.2003.13084>
- González, G., & Evans, C. L. (2019). Biomedical Image Processing with Containers and Deep Learning: An Automated Analysis Pipeline: Data architecture, artificial intelligence, automated processing, containerization, and clusters



- orchestration ease the transition from data acquisition to insights in medium-to-large datasets. *BioEssays*, 41(6), 1900004. <https://doi.org/10.1002/bies.201900004>
- Gurtovoy, D., & Gottschalk, S. (2022). *Linking Streets in OpenStreetMap to Persons in Wikidata* [Sesión de conferencia]. Web Conference 2022 (WWW '22). Nueva York. <https://doi.org/10.1145/3487553.3524267>
- Gutiérrez Puebla, J., García Palomares, J. C., & Salas Olmedo, M. H. (2016). Big (geo) data en ciencias sociales: retos y oportunidades. *Revista de estudios andaluces*, (33), 1-23. <https://dx.doi.org/10.12795/rea.2016.i33.0>
- Heindorf, S., Potthast, M., Stein, B., & Engels, G. (2016). *Vandalism Detection in Wikidata* [Sesión de conferencia]. 25th ACM International on Conference on Information and Knowledge Management (CIKM '16). Nueva York. <https://doi.org/10.1145/2983323.2983740>
- Instituto Geográfico Nacional [IGN] (1870). *Mapa de España. Triangulación geodésica. 1870*. Cartografía de España en la Biblioteca Nacional. Siglos XVI al XIX. Tomo I, p. 151, ficha 337. <https://www.ign.es/web/catalogo-cartoteca/resources/html/016641.html>
- Law, S., Shen, Y., & Seresinhe, C. (2017). An application of convolutional neural network in street image classification: The case study of London. In *Proceedings of the 1st Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery* (pp. 5-9). <http://dx.doi.org/10.1145/3149808.3149810>
- Lemus-Rojas, M., & Pintscher, L. (2018). Wikidata and Libraries: Facilitating Open Knowledge. En M. Proffitt (Ed.), *Leveraging Wikipedia: Connecting Communities of Knowledge* (pp. 143-158). ALA Editions. <https://hdl.handle.net/1805/16690>
- Ley 11/1975, de 12 de marzo, sobre Señales Geodésicas y Geofísicas. Jefatura del Estado. BOE núm. 63, de 14 de marzo de 1975. <https://www.boe.es/buscar/pdf/1975/BOE-A-1975-5295-consolidado.pdf>
- Li, X., Ratti, C., & Seiferling, I. (2017). Mapping urban landscapes along streets using google street view. In *International cartographic conference* (pp. 341-356). Springer, Cham. https://doi.org/10.1007/978-3-319-57336-6_24
- Martín López J. (1999). *Francisco Coello: su vida y su obra*. Ministerio de Fomento, Centro Nacional de Información Geográfica.
- Martín López J. (2002). *Historia de la Cartografía y de la Topografía*. Ministerio de Fomento, Centro Nacional de Información Geográfica.
- Mittermeier, J. C., Correia, R., Grenyer, R., Toivonen, T., & Roll, U. (2021). Using Wikipedia to measure public interest in biodiversity and conservation. *Conservation Biology*, 35(2), 412-423. <https://doi.org/10.1111/cobi.13702>
- Mora-Cantalops, M., Sánchez-Alonso, S., & García-Barriocanal, E. (2019). A systematic literature review on Wikidata. *Data Technologies and Applications*, 53(3), 250-268. <https://doi.org/10.1108/DTA-12-2018-0110>
- Obregón, Á. (2022). Inserción de metadatos de las bibliotecas españolas en Wikidata: un modelo de datos abiertos enlazados. *Revista Española De Documentación Científica*, 45(3), a330. <https://doi.org/10.3989/redc.2022.3.1870>
- Piscopo, A., & Simperl, E. (2019). *What we talk about when we talk about wikidata quality: a literature survey* [Sesión de conferencia]. 15th International Symposium on Open Collaboration (OpenSym '19). Nueva York. <https://doi.org/10.1145/3306446.3340822>
- Quirós Linares, F. (2010). *La cartografía de la metrópoli en el Atlas de España y sus posesiones de Ultramar (1847-1870), de Francisco Coello. Características, fuentes y colaboradores*. *Ería*, 81, 63-92. <https://reunido.uniovi.es/index.php/RCG/article/view/1673/1571>
- Regidor, J., Prieto, J.F., Sanz, J.M., Quirós, R., & Barbadillo, A. (2000). *El proyecto REGENTE*. VII Congreso Nacional de Topografía y Cartografía. https://oa.upm.es/13829/1/Regidor2000_TOPCAR_libro.pdf
- Registered user. (14 de enero de 2023). En *Wikimedia*. https://meta.wikimedia.org/w/index.php?title=Registered_user&oldid=24370729
- Shenoy, K., Ilievski, F., Garijo, D., Schwabe, D., & Szekely, P. (2022). A study of the quality of Wikidata. *Journal of Web Semantics*, 72. <https://doi.org/10.1016/j.websem.2021.100679>
- Stark, C., Breikreutz, B. J., Regulý, T., Boucher, L., Breikreutz, A., & Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic acids research*, 34(suppl_1), D535-D539. <https://doi.org/10.1093/nar/gkj109>
- Shan, P., & Sun, W. (2021). Research on 3D urban landscape design and evaluation based on geographic information system. *Environmental Earth Sciences*, 80(17), 1-15. <https://doi.org/10.1007/s12665-021-09886-y>
- Tharani, K. (2021). Much more than a mere technology: A systematic review of Wikidata in libraries. *The Journal of Academic Librarianship*, 47(2), 102326. <https://doi.org/10.1016/j.acalib.2021.102326>



- Turki, H., Shafee, T., Taieb, M. A. H., Aouicha, M. B., Vrandečić, D., Das, D., & Hamdi, H. (2019). Wikidata: A large-scale collaborative ontological medical database. *Journal of Biomedical Informatics*, 99, 103292. <https://doi.org/10.1016/j.jbi.2019.103292>
- Urteaga, L., & Nadal, F. (2001). *Las series del mapa topográfico de España a escala 1:50.000*. Ministerio de Fomento, Dirección General del Instituto Geográfico Nacional, 397
- Van Veen, T. (2019). Wikidata: Wikidata: From „an” Identifier to „the” Identifier. *Information Technology and Libraries*, 38, 2. <https://doi.org/10.6017/ital.v38i2.10886>
- Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10), 78-85. <http://dx.doi.org/10.1145/2629489>
- Wikidata Mainpage (2022). *Wikidata Stats*. Wikidata. <https://wikidata-todo.toolforge.org/stats.php>
- Wikidata (2023). *Wikidata Notability* <https://www.wikidata.org/w/index.php?title=Wikidata:Notability&oldid=1806697286>
- Wu, J., Orlandi, F., O’Sullivan, D., & Dev, S. (2022). LinkClimate: An interoperable knowledge graph platform for climate data. *Computers & Geosciences*, 169, 105215. <https://doi.org/10.1016/j.cageo.2022.105215>