

Observando la curación colaborativa de conocimiento estructurado en *Wikidata*

Looking at collaborative curation of structured knowledge on *Wikidata*

Juan-Antonio Pastor-Sánchez; Tomás Saorín

Pastor-Sánchez, Juan-Antonio; Saorín, Tomás (2024). "Observando la curación colaborativa de conocimiento estructurado en *Wikidata*". *Anuario ThinkEPI*, v. 18, e18e35.

<https://doi.org/10.3145/thinkepi.2024.e18a35>

Publicado en *IweTel* el 4 de diciembre de 2024

Juan-Antonio Pastor-Sánchez

<https://orcid.org/0000-0002-1677-1059>
<https://www.directorioexit.info/ficha1964>

Universidad de Murcia

Departamento de Información y Documentación
pastor@dum.es

Tomás Saorín

<https://orcid.org/0000-0001-9448-0866>
<https://www.directorioexit.info/ficha1039>

Universidad de Murcia

Departamento de Información y Documentación
tsp@dum.es



Resumen: Se revisan aspectos concretos relevantes para entender el funcionamiento de *Wikidata* desde la perspectiva de la organización colaborativa del conocimiento. La curación colectiva presenta retos sobre modelado y calidad de los datos. *Wikidata* tiene la capacidad de incorporar taxonomías y vocabularios controlados, pero se identifican prácticas problemáticas en el uso de clases/subclases e instancias. También se describen las dificultades para establecer esquemas de metadatos, y en la realización de descripciones consistentes de elementos, y el uso como base de datos bibliográfica, así como un panorama de las

paradojas que crea el licenciamiento en dominio público y su uso como conjunto de datos para enriquecimiento de las aplicaciones de nueva generación de inteligencia artificial.

Palabras clave: *Wikidata*; Organización del conocimiento; Taxonomías; Esquemas de metadatos; Descripción de recursos; Reutilización de datos.

Abstract: Specific aspects relevant to understanding the functioning of *Wikidata* from the perspective of the collaborative organization of knowledge are reviewed. Collective curation poses challenges in data modeling and quality. *Wikidata* has the capacity to incorporate taxonomies and controlled vocabularies, but problematic practices are identified in the use of classes/subclasses and instances. The difficulties in establishing metadata schemas, creating consistent descriptions of elements, and using it as a bibliographic database are also described. Additionally, an overview is provided of the paradoxes created by licensing in the public domain and its use as a dataset for enriching next-generation artificial intelligence applications.

Keywords: *Wikidata*; Knowledge organization; Taxonomies; Metadata schema; Resource description; Data reuse.

En esta nota se realiza una revisión de algunos aspectos concretos que son relevantes para entender el funcionamiento de *Wikidata* desde la perspectiva concreta de la organización colaborativa del conocimiento. Aunque recogemos numerosas investigaciones e informes recientes, no se pretende hacer una revisión sistemática o informe de estado del arte en bibliotecas, GLAM o investigación experimental por campos (humanidades digitales, procesamiento del lenguaje natural, grafos de conocimiento, etc.) y remitimos a otros trabajos específicos con este enfoque (**Farda-Sarbas; Müller-Birn, 2019; Mora-Cantalops; Sánchez-Alonso; García-Barriocanal, 2019; Tharani, 2021; Zhao, 2022; Turki et al., 2023; Candela et al., 2024**).

Desde su concepción y lanzamiento hace ya más de una década, este proyecto de base de datos multilingüe nativa que respalda y cohesiona los datos factuales de las más de 300 enciclopedias libres, ha sido un interesante campo de experimentación y aprendizajes que también afectan a la forma en la que se entienden los instrumentos clásicos para la organización del conocimiento (vocabularios, clasificaciones), los esquemas de metadatos, la descripción del contenido y, con respecto a los catálogos y bases de datos bibliográficas, su forma y alcance. *Wikidata* prueba cosas, cambia y se adapta. Su propio proceso de ideación y puesta en marcha ha sido evolutivo y basado más en interacciones entre personas y proyectos que en una planificación estratégica (**Vrandečić; Pintscher; Krötzsch, 2023**).

Abordaremos los siguientes temas que ilustran la complejidad de la curación colectiva de datos, entendiendo esta curación tanto como modelado de datos (**Piscopo; Simperl, 2018; Krötzsch, 2018**) como calidad de los datos (**Shenoy et al., 2022**). En primer lugar, hablaremos de la aplicación de la diferenciación entre instancias y clases. En segundo lugar, abordaremos algunas aproximaciones al uso elástico de esquemas de metadatos o instrumentos parecidos. En tercer lugar, hablaremos de casos problemáticos de descripción o tipificación de elementos, teniendo en cuenta las dificultades de su explotación en consultas. En cuarto lugar, aspectos sobre reutilización y participación en el maremoto de los *datasets* para la inteligencia artificial generativa. Por último, incluiremos unas reflexiones sobre el uso de *Wikidata* como plataforma de control bibliográfico en un sentido genérico (universo bibliográfico) y específico (referencias y citas en proyectos *Wikimedia*).

ABC de la organización del conocimiento en Wikidata

Wikidata es un grafo de conocimiento cuyo funcionamiento es relativamente sencillo. Los diferentes objetos (ítems) que se describen en el grafo se identifican mediante un código único que comienza por la letra 'Q'. Por ejemplo, el ítem Q29 se corresponde con España o Q60 con Nueva York. Desde 2018 *Wikidata* también se utiliza como base de datos lexicográfica que permite definir lexemas en diferentes idiomas junto con sus formas y los posibles significados e ítems vinculados. Pero en esta nota únicamente se abordarán los aspectos del grafo de conocimiento con las descripciones de los ítems. Los ítems están descritos mediante propiedades que comienzan por la letra 'P'. Por ejemplo, la propiedad P27 se refiere a la nacionalidad del objeto o la propiedad P18 que permite vincular una imagen representativa del mismo.

Entre estas propiedades hay tres que cumplen una función especial: P31, P279 y P361.

- La propiedad P31 (es instancia de) define una relación de clase/instancia, por ejemplo: Q91 “Abraham Lincoln” es una instancia Q5 “humano”.
- La propiedad P279 (subclase de) define las relaciones de genero/especie. Por ejemplo, Q349 “deporte” es una subclase de Q747883 “actividad física”, que a su vez es una subclase de Q61788060 “actividad humana”. Consecuentemente, la propiedad P279 permite definir taxonomías de clases.

- La propiedad P361 (parte de) permite definir relaciones de todo/parte, de tal forma que un Q595871 “sistema estelar” forma parte de Q318 “galaxia”. Se excluye de este análisis la propiedad P1279 (faceta de) cuyo uso resulta prácticamente insignificante (y algo complejo) en relación con las tres propiedades anteriormente mencionadas.

El grafo de conocimiento de *Wikidata* adopta la forma de tripletas que pueden representarse siguiendo el modelo RDF. Tanto los ítems como las propiedades tienen su propio espacio de nombres que suelen abreviarse con los siguientes prefijos:

wd: <http://www.wikidata.org/entity/>

wdt: <http://www.wikidata.org/prop/direct/>

De esta forma las anteriores afirmaciones podrían representarse como se muestra en la figura 1.

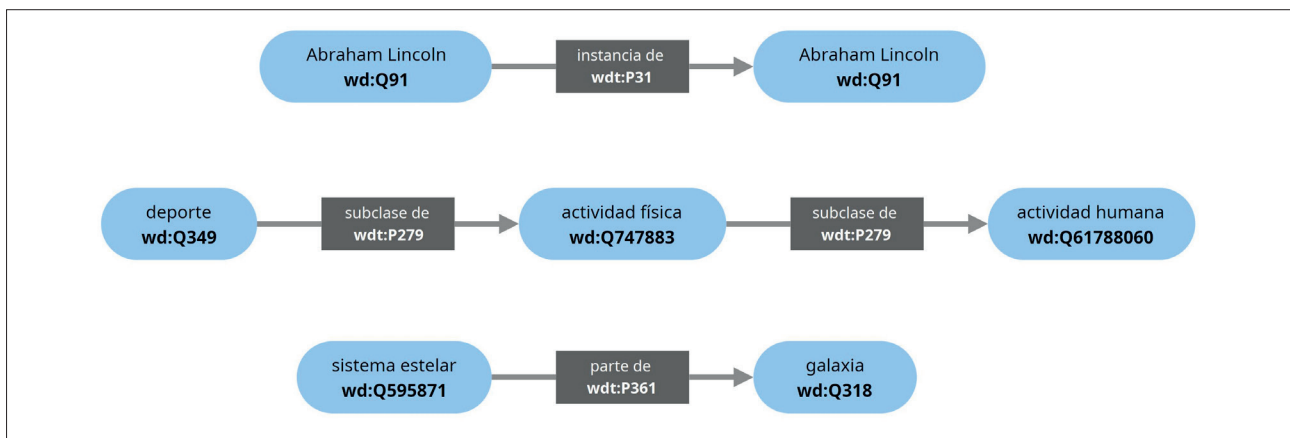


Figura 1. Representación de ítems y propiedades de *Wikidata*.

Desde el punto de vista de la organización del conocimiento, la aplicación en el grafo de *Wikidata* de las propiedades anteriormente mencionadas, debería regirse por criterios de clasificación y estructuración. La finalidad de esto sería recuperar los datos sobre los diferentes ítems del modo más eficiente y con la menor tasa de error posible. Por lo tanto, deberían seguirse algunos principios básicos:

- Las clases deben estar organizadas en taxonomías coherentes mediante relaciones recursivas género/especie (P279), carentes de inconsistencias que puedan derivarse de las polijerarquías, ciclos de relaciones jerárquicas, etc.
- Los individuos se distinguen claramente como instancias (P31) de alguna (o algunas) de las clases de las taxonomías. Metafóricamente hablando: el tronco y las ramas del árbol serían las taxonomías de clases, mientras que las flores y frutos serían las instancias de dichas clases.
- Las relaciones partitivas (P361), cubran aspectos físicos o conceptuales, deberían estar limitados a casos específicos: disciplinas o campos del discurso, localizaciones geográficas o temporales, sistemas y órganos biológicos, estructuras sociales y organizativas, etc. En general las partes son componentes, componentes o localización de algo más amplio (norma ISO 25964-1:2011) [Aenor, 2014].

Sin embargo, la realidad es que el proceso de edición de *Wikidata* tiene una naturaleza abierta y colaborativa y el grafo mezcla clases, partes e instancias ya que todas ellas son ítems. Esto significa que las decisiones de organización del conocimiento se centran única y exclusivamente en la aplicación que se haga de las propiedades P31, P279 y P361.

Taxonomías sospechosas e “hiperinstanciación”

Como se ha visto en uno de los ejemplos anteriores, la propiedad P279 es el mecanismo que permite definir taxonomías de clases. La “pertenencia” de un ítem a una de estas clases en forma de instancia se realiza mediante la propiedad P31. La pregunta es ¿a qué nivel de una taxonomía se vincula un ítem como instancia de una clase? Lo lógico sería hacerlo con la clase más específica posible que mejor se adapte a la instancia. Sin embargo, tomemos el ejemplo del ítem correspondiente a Nueva York (Q60). Podemos ver que este ítem se ha definido como instancia de: ciudad global, asentamiento portuario, ciudad de Estados Unidos, ciudad, gran ciudad, ciudad mayor, ciudad del estado de Nueva York, metrópoli y megaciudad.

<https://www.wikidata.org/entity/Q60>

Para facilitar la lectura, el siguiente gráfico toma únicamente las etiquetas en español de dichas clases (obviando el identificador del ítem) para representar parte de la taxonomía de clases implicadas en la descripción de Nueva York. El sentido de la flecha va desde la clase más específica a la más amplia.

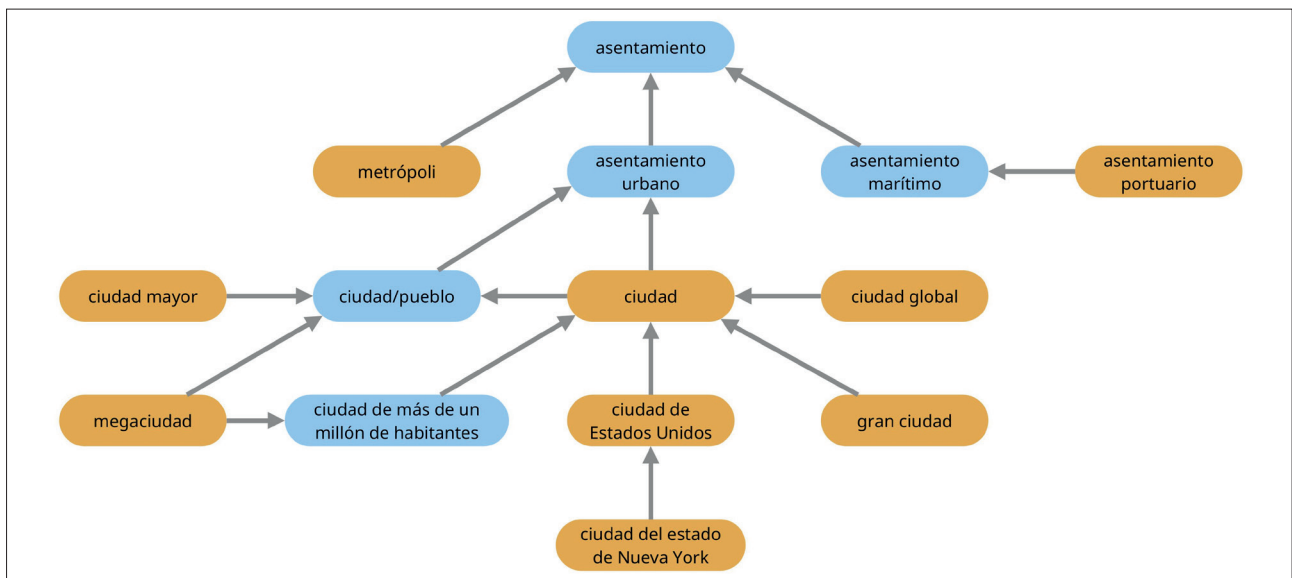


Figura 2. Jerarquía de clases vinculadas al ítem Q60 de Wikidata correspondiente a Nueva York.

En color naranja se muestran las clases utilizadas para definir Nueva York como instancia (P31) de una clase. En color azul otras clases implicadas en la taxonomía. Llama la atención que se utilicen el 65% de clases de esta taxonomía para instanciar un ítem. Sería un buen ejercicio rehacer el diagrama anterior eliminando la relación entre asentamiento urbano y ciudad y entre ciudad/pueblo y megaciudad. También es muy discutible las relaciones entre ciudad/pueblo y ciudad mayor y entre asentamiento y metrópoli: ciudad mayor podría ser una subclase de ciudad; metrópoli podría ser una subclase de asentamiento urbano.

Por otro lado, hubiese sido suficiente, definir Nueva York como instancia de ciudad del estado de Nueva York sin necesidad de definirla como instancia de ciudad de Estados Unidos y de ciudad, puesto que las tres clases están en la misma línea jerárquica. En estos casos, utilizar la clase más específica sería lo más adecuado.

Aunque pueda aparentar que se trata de dos problemas distintos, en realidad están muy relacionados entre sí. La asignación de un número excesivo de clases a una instancia podría deberse a la propia evolución de la taxonomía para responder a necesidades de descripción de entidades. Por otro lado, la ampliación de la taxonomía incorporando nuevas clases puede requerir la compro-

bación de relaciones P279 preexistentes cuya revisión se soslaye por la propia dinámica de edición apresurada del grafo de *Wikidata*.

Olvidando la transitividad (y la reciprocidad) de las relaciones partitivas

Un problema diferente se produce con la propiedad P361 (parte de). La finalidad de esta propiedad es reflejar relaciones todo-parte. En este caso se detecta un problema de definir relaciones partitivas de un elemento con respecto a elementos más amplios en varios niveles en vez de hacerlo únicamente con el elemento inmediatamente superior. Para ilustrar este problema es posible comprobar como Q18 “América del Sur” forma parte de Q828 “América”, al tiempo que Q736 “Ecuador” tiene sendas relaciones P361 con los dos ítems anteriores.

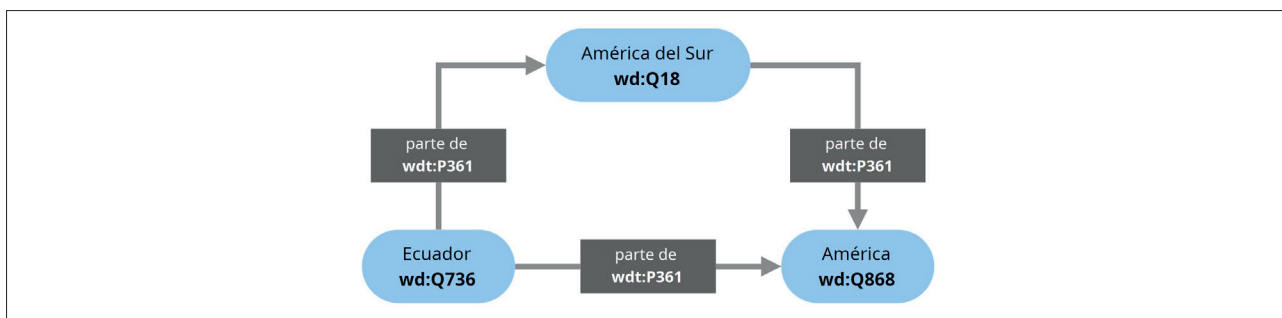


Figura 3. Redundancia de uso de propiedades P361 (parte de) en el ítem Q736 (Ecuador).

En este caso, la relación de Ecuador como parte de América es superflua, ya que dicha pertenencia se realiza indirectamente a través de la relación con América del Sur y la propiedad P361 está definida como una relación transitiva.

Otro punto reseñable es el uso de la propiedad P527 (compuesto por) que *Wikidata* define como inversa de P361. Cabría esperar que al tratarse de una relación inversa el número de relaciones definidas con P361 fuese exactamente igual que las definidas con P527. Sin embargo, podemos ver cómo la propiedad P361 se utiliza en 5.175.841 relaciones, mientras que P527 hace lo propio solamente en 2.433.802 relaciones. Por lo tanto, una pregunta queda en el aire respecto a los editores: ¿existe un abuso en la aplicación de la propiedad o P361 o una infrautilización de su relación inversa P527? Aparentemente la relación P361 no se utiliza del todo correctamente, pero habría que realizar un análisis más profundo.

Instancia o clase... esa es la cuestión

Uno de los principales problemas que pueden encontrarse en *Wikidata* es que muchos editores, combinan o confunden la noción de instancia con la de subclase. Es frecuente encontrar casos en los que un ítem se define simultáneamente como instancia de un ítem A y como subclase de un ítem B y posteriormente los ítems A y B forman parte de una taxonomía de clases. Sirva como ejemplo el caso del ítem Q4202 (estrella de neutrones) (figura 4).

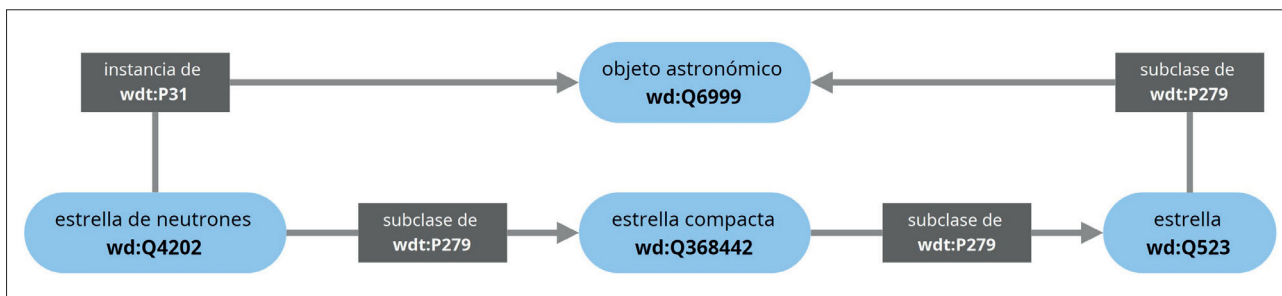


Figura 4. El ítem Q4202 (estrella de neutrones) definido como instancia y como subclase dentro de la misma taxonomía de clases.

El ejemplo de la figura 4 ilustra multitud de casos de taxonomías confusas que podemos encontrar en muchos otros ámbitos: organización de disciplinas científicas, obras creativas, actividades humanas, etc.

Estas estructuras pueden plantear inconsistencias durante el proceso de recuperación, ya que cuando se intentan recuperar únicamente taxonomías de clases también se recuperan instancias que introducen ruido en los resultados y que a veces hacen difícil de utilizar los datos obtenidos.

La solución es relativamente sencilla puesto que la dinámica editorial en *Wikidata* refleja (como se ha indicado anteriormente) gran parte de la problemática detectada en la edición de artículos y estructura de categorías de *Wikipedia*. Lo ideal sería que el sistema detectase cuando se intenta definir un ítem como instancia de una clase cuando dicho ítem ya ha sido declarado previamente como clase dentro de la misma estructura taxonómica. También habría que comprobar si se desea definir dicho ítem como una subclase de otro si previamente ha sido declarado como instancia de una clase de la misma taxonomía.

Abordar la corrección de este problema en los datos actuales de *Wikidata* implicaría soluciones por áreas temáticas: adoptar un criterio y elaborar un *bot* que permita la corrección automática en los casos en los que sea posible.

Los problemas de curación de datos son amplios, y requieren análisis más sistemáticos que los aquí esbozados. Existen muchos casos de conjuntos de datos del mismo tipo (películas, libros) en los que se aprecia la ausencia de propiedades básicas. Existen casos en los que los editores de datos (denominados a veces *wikidatians* (Piscopo; Phethean; Simperl, 2017) oscilan entre varias propiedades para representar un mismo aspecto del elemento descrito. Existen taxonomías muy enrevesadas, que presentan tanto lagunas como hipertrofia en alguna de sus ramas.

Sin esquemas no hay futuro

Lo más llamativo del modelo abierto de *Wikidata* es la ausencia total de cualquier esquema. Si bien es cierto que conforme evoluciona la práctica del enriquecimiento colaborativo de datos aparecen diversas aproximaciones, sigue siendo cierto que, al crear un ítem, casi lo único reclamado por el sistema es asignarle una etiqueta y, si viene bien, una declaración de tipificación o instanciación (P31).

En *Wikidata* existe la posibilidad de definir restricciones y validaciones para ayudar a la consistencia y detectar errores u omisiones. De este modo, el propio modelo de datos ha evolucionado a lo largo de los años para permitir la creación de restricciones que marquen a los editores posibles inconsistencias, como por ejemplo que una relación con la propiedad P941 (inspirado por) tiene que apuntar a una obra creativa o a una persona, y no a ítems de otros tipos.

Más centrados en el esquema, entendido como un conjunto de propiedades aplicables para un cierto tipo de contenidos, son las extensiones de “predicción de propiedades” (Luggen *et al.*, 2021), que analizan las propiedades usadas con más frecuencia en cierto tipo de ítems, como podrían ser libros o películas, y que recomiendan al editor posibles propiedades que usar (por ejemplo, *Recoin*, que calcula y marca en cada ítem su *Relative Completeness Indicator*). Se trata más de un esquema construido sobre la marcha por el propio uso de la comunidad y que está muy lejos de la concepción habitual de *element sets* usados en la descripción de recursos.

<https://www.wikidata.org/wiki/Wikidata:Recoin>

Existen proyectos avanzados como los *Schemas* que permiten formalizar estructuras de datos y procedimientos de validación y completitud aplicando *Shape Expressions* (ShEx).

<https://www.wikidata.org/wiki/Wikidata:Schemas>

https://www.wikidata.org/wiki/Wikidata:WikiProject_Schemas

Por ejemplo, para un ser humano, existe ya una entidad de tipo que delimita las propiedades aplicables, como podrían ser lugar de nacimiento, hijos, parientes, nombre, ocupación, etc.

<https://www.wikidata.org/wiki/EntitySchema:E10>

Se trata de aplicaciones muy parciales y con cierto aire de experimentalidad a las que les queda bastante camino por recorrer. Pese a todo, existe un interés por comprender cómo la comunidad, a través también de *Wikiproyectos* que reúnen a los interesados en una clase o tipo de información, trabajan en subesquemas y en reglas de representación adecuada del contenido (Baroncini *et al.*, 2022).

Wikidata como base de datos bibliográfica: catálogo universal o motor de citas

Desde bien pronto, diferentes comunidades exploraron la capacidad de usar *Wikidata* para sistematizar la información bibliográfica vinculada a las *Wikipedias* e, incluso, como herramienta para construir una base de datos bibliográfica global (Bianchini; Sardo, 2022). En *Wikidata* se han cargado masivamente conjuntos de artículos científicos, incluyendo sus redes de citas, para reproducir la funcionalidad de las bases de datos comerciales *Scopus* o *Web of Science*. También se han desarrollado scripts para la inserción sistematizada de referencias en artículos de *Wikipedia*, tratando de superar la circunstancia de que las citas y referencias de la enciclopedia son meramente texto marcado. Es decir, carecen de una verdadera capacidad de trazar redes y ser consultadas. Alrededor del proyecto *WikiCite* se han reunido todos estos ensayos, con mayor o menor fortuna.

<https://meta.wikimedia.org/wiki/WikiCite>

En general, podemos decir que las pruebas de concepto aportan *insights* valiosos, pero que presentan una gran dificultad para su escalado y conformar sistemas exhaustivos y viables. Su consulta a través de *Wikidata Query Service* mediante consultas *SPARQL* es poco eficiente y consume gran cantidad de recursos del sistema. Esto posiblemente se debe a la estructura monolítica del grafo de *Wikidata* que utiliza un único grafo para representar todo el conocimiento que alberga. Esto puede resultar muy problemático para la ejecución de consultas y recuperación de datos en determinados dominios.

Un ejemplo claro de ello es el reciente anuncio de los responsables técnicos de las tecnologías que dan soporte a esta *Wikibase* y a su *SPARL EndPoint*, del *WDQS graph split*, es decir, la subdivisión del grafo actual en varios, separando la información bibliográfica de la enciclopédica (Pintscher, 2024). El volumen de información meramente bibliográfica –artículos, libros, documentos– tiene un crecimiento varias magnitudes por encima del de la información enciclopédica y sus datos.

https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/WDQS_graph_split

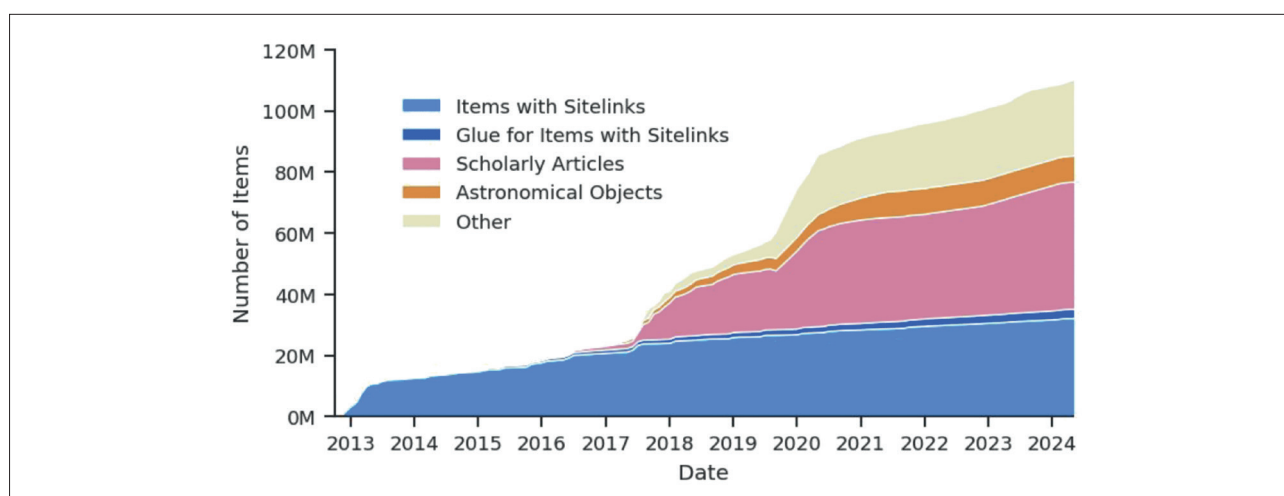


Figura 5. los ítems bibliográficos arrojan el *triplestore BlazeGraph* de *Wikidata*, junto a otros elementos no enciclopédicos (sin *sitelinks*) como objetos astronómicos. Fuente: Pintscher, 2024.

El modelo en grafo que propone *Wikidata* es sugerente para producir bases de datos bibliográficas con mayor capacidad de descubrimiento y relación entre elementos. Sin embargo, el volumen de datos de la producción científica y cultural no es abarcable ni por la capacidad de trabajo de la comunidad voluntaria *Wikimedia*, ni por una tecnología de base de conocimiento concebida para respaldar la información enciclopédica. Por estas razones se advierte un creciente movimiento hacia la creación de proyectos autónomos basados en la tecnología *Wikibase*, donde realizar descripciones profundas de determinados ámbitos. Algunos proyectos de investigación en humanidades e historia, como *FactGrid* y *MimoTextBase*, comienzan a registrar sus conjuntos de datos usando *Wikibase*, contando, además, con la capacidad de vincularse con *Wikidata* para la realización de consultas *SPARQL* federadas.

https://database.factgrid.de/wiki/Main_Page

https://data.mimotext.uni-trier.de/wiki/Main_Page

Podría ser una tendencia a tener muy en cuenta la de la proliferación de bases de conocimiento especializadas, diseñadas y mantenidas por los propios grupos de interés o instituciones estables, que se pueden interconectar a través de sus *query services* y mediante la estrategia de datos enlazados y vinculación a través de identificadores estables.

Se ha de tener en cuenta, además, que han sido los identificadores de entidades (autoridades) y el uso de *Wikidata* como *hub* de interconexión entre diferentes bases de datos y sistemas de información el que ha recibido más atención y presenta un desarrollo más estable. Se ha trabajado en todos los dominios, tanto en especies botánicas, autoras orientales, obras creativas, patrimonio cultural, etc.

Licencia CC0, los riegos de la barra libre de datos

Tampoco queremos dejar de señalar un aspecto problemático, presente desde el primer momento y puesto de manifiesto recientemente por **McDowell** y **Vetter** (2024), en su trabajo sobre la realienación del bien común y la ética de los datos libres. Identifican lo que llaman el *Wikidata's Turn*, que supone que, al ser un proyecto de datos, su licencia es CC0, lo que supone que se puede reutilizar sin ninguna limitación. Esta barra libre tiene consecuencias que pueden romper la “cadena de sentido” de la producción colaborativa de buena fe. El trabajo de editores voluntarios, pensado para el bien común y la continuidad del proyecto, puede convertirse en materia prima para que los gigantes tecnológicos construyan un nuevo modelo de negocio avanzado, derivado de la potencia de la IA generativa, que a su vez genere grandes diferencias entre quienes se la puedan permitir y quienes no. Los valiosos datos multidominio de *Wikidata*, vienen con licencia de cesión al dominio público. Esto pone de relieve que el debate informado sobre cómo la elección de una licencia para el contenido, datos y metadatos tiene un potencial crítico para asegurar la sostenibilidad y equidad del conocimiento libre. En el caso de *Wikidata*, nos enfrentamos al desafío adicional de la comodificación del conocimiento. La transformación del trabajo colaborativo voluntario en una materia prima para corporaciones tecnológicas pone de relieve cuestiones éticas importantes. ¿Estamos asistiendo a una “realienación” del conocimiento comunitario, donde los creadores pierden el control sobre su trabajo? ¿Cómo podemos garantizar que el conocimiento libre no se convierta simplemente en otra fuente de explotación por parte de las grandes corporaciones tecnológicas? (**Saorín; Claes**, 2024)

Wikipedia y *Wikidata* están siendo exprimidas de forma sistemática para los modelos de lenguaje sobre los que se construyen nuevas propuestas de valor críticas, y existe una preocupación en las comunidades promotoras del contenido y datos abiertos, como *Creative Commons*, *Open Future*, *Wikimedia* o la *OKFN* sobre las paradojas que está produciendo lo abierto cuando se aplica para tecnologías no equitativas y sin tener en cuenta la ética de la colaboración que origina los datos explotados por empresas tecnológicas (*Open Knowledge*, 2024). Pero también las propias comunidades

están redefiniendo el tipo de *dataset* apropiados que pueden aportar para sus propias implementaciones de mejoras basadas en IA generativa (Johnson; Kaffee; Redi, 2024). Los *dumps* tradicionales ya no son óptimos para estas tareas, y se está produciendo un giro hacia una ingeniería de la reutilización para la IA generativa que tendrá muchos impactos.

Conclusiones

Wikidata se enfrenta a los retos del paso a la vida adulta, tras una adolescencia llena de expectativas y aprendizajes. No hemos comentado otras derivas significativas como su uso para la descripción de imágenes (*Structured Commons*) o las sinergias con los nuevos proyectos paralelos de *Wikifunctions* y *Abstract Wikipedia*, surgidos también del magín de Vrandečić y que apunta hacia ensoñaciones de lenguaje y escritura automática.

Lo que sí es relevante es su voluntad de afrontar la diversidad de conocimientos que afecta tanto al modelado como a las taxonomías como a las fuentes verificables usadas para realizar afirmaciones más o menos factuales (Vrandečić, 2020). Es un reto, con conflictos y debilidades, pero que *eppur si muove*. Además, la posibilidad de participar en *Wikidata* a diferentes niveles, tanto en la creación de descripciones, esquemas, negociación de propiedades, explotación de datos o la organización de taxonomías, la convierte en un entorno óptimo de aprendizaje en situación real de la idea de la web semántica o web de datos (Sigalov; Nachmias, 2023).

Wikidata no es solo un recurso de datos, es también una comunidad abierta y colaborativa. Esta apertura se manifiesta tanto en un modelo de datos muy elástico que permite modelar con agilidad casi cualquier dominio, y tiene como contrapartida las incoherencias y dificultades de sistematización. Supone un reto la resolución colaborativa de conflictos en la construcción de instrumentos de organización del conocimiento y es un caso inédito de este tipo de empresas. Los comités para la normalización de vocabularios o para la actualización de sistemas de clasificación, a todos los niveles y ámbitos, son maquinarias pesadas. Sin embargo, en *Wikidata* los medios clásicos de organización del conocimiento adoptan unas formas con mayor capacidad de reacción y permiten probar la dimensión real de los mecanismos de construcción colaborativa no solo de los datos factuales, sino también de las taxonomías. La situación actual, bastante caótica e inestable, merece investigaciones de mayor calado para un diagnóstico adecuado de este modelo de trabajo.

Referencias

- Baroncini, Sofia; Martorana, Margherita; Scrocca, Mario; Śmiech, Zuzanna; Polleres, Axel (2022). "Analysing the evolution of community-driven (sub-)schemas within *Wikidata*". *Wikidata'22: Wikidata workshop at ISWC*. https://wikidataworkshop.github.io/2022/papers/Wikidata_Workshop_2022_paper_9582.pdf
- Bianchini, Carlo; Sardo, Lucia (2022). "*Wikidata*: a new perspective towards universal bibliographic control". *Italian journal of library, archives and information science*, v. 13, n. 1, pp. 291-311. <https://doi.org/10.4403/jlis.it-12725>
- Candela, Gustavo; Cuper, Mirjam; Holownia, Olga; Gabriëls, Nele; Dobрева, Milena; Mahey, Mahendra (2024). "A systematic review of *Wikidata* in GLAM institutions: a labs approach". In: Antonacopoulos, Apostolos; Hinze, Annike; Piwowarski, Benjamin; Coustay, Mickaël; Di-Nunzio, Giorgio-Maria; Gelati, Francesco; Vanderschantz, Nicholas. *Linking theory and practice of digital libraries. TPD 2024. Lecture Notes in Computer Science*, v. 15178. Springer, Cham. https://doi.org/10.1007/978-3-031-72440-4_4
- Farda-Sarbas, Mariam; Müller-Birn, Claudia (2019). *Wikidata from a research perspective - A systematic mapping study of Wikidata*. arXiv. <https://doi.org/10.48550/arxiv.1908.11153>
- Johnson, Isaac; Kaffee, Lucie-Aimée; Redi, Miriam (2024). "Wikimedia data for AI: a review of Wikimedia datasets for NLP tasks and AI-assisted editing". *Wikipedia Workshop at EMNLP '24*. <https://doi.org/10.48550/arXiv.2410.08918>

- Krötzsch, Markus** (2018). "Ontological modelling in *Wikidata*". In: *Knowledge-based systems, TU Dresden. Workshop on Ontology Design and Patterns*.
https://iccl.inf.tu-dresden.de/w/images/e/ed/Ontology_modelling_Wikidata_Markus_Kroetzsch_WOP2018.pdf
- Luggen, Michael; Audiffren, Julien; Difallah, Djellel; Cudré-Mauroux, Philippe** (2021). "Wiki2Prop: A multimodal approach for predicting *Wikidata* properties from *Wikipedia*". In: *Proceedings of the Web Conference 2021 (WWW '21)*. Association for Computing Machinery, New York, pp. 2357–2366.
<https://doi.org/10.1145/3442381.3450082>
- McDowell, Zachary J.; Vetter, Matthew A.** (2024). "The realienation of the Commons: *Wikidata* and the ethics of 'free' data". *International journal of communication*, v. 18, pp. 590–608.
<https://ijoc.org/index.php/ijoc/article/view/20807>
- Mora-Cantalops, Marçal; Sánchez-Alonso, Salvador; García-Barricocal, Elena** (2019). "A systematic literature review on *Wikidata*". *Data technologies and applications*, v. 53, n. 3, pp. 250–268.
<https://doi.org/10.1108/DTA-12-2018-0110>
- Open Knowledge* (2024). *Open movement's common(s) causes: Report from a Wikimania 2024 side event*. Open Future; Creative Commons; Open Knowledge Foundation; Wikimedia Europe.
<https://blog.okfn.org/2024/11/18/report-open-movements-commons-causes/>
- Pintscher, Lydia** (2024). "Wikidata query service - the way forward for getting the most out of Wikimedia's knowledge graph". In: *Wikimania 2024*.
https://upload.wikimedia.org/wikipedia/commons/3/31/Wikimania_2024_Wikidata_Query_Service_-_the_way_forward_for_getting_the_most_out_of_Wikimedia%27s_knowledge_graph.pdf
- Piscopo, Alessandro; Phethean, Christopher; Simperl, Elena** (2017). "*Wikidatians* are born: paths to full participation in a collaborative structured knowledge base". In: *Proceedings of the 50th Hawaii International Conference on System Sciences 2017*.
<http://hdl.handle.net/10125/41688>
- Piscopo, Alessandro; Simperl, Elena** (2018). "Who models the world?: Collaborative ontology creation and user roles in *Wikidata*". *Proceedings of the ACM on human-computer interaction*, v. 2.
<https://doi.org/10.1145/3274410>
- Saorín, Tomás; Claes, Florencia** (2024). "Lectura artificial para una enciclopedia escrita por máquinas: reflexiones sobre una Wikipedia hecha a mano ante el vértigo generativo". *DIFF Wikimedia*, 19 septiembre.
<https://diff.wikimedia.org/es/2024/09/19/lectura-artificial-para-una-enciclopedia-escrita-por-maquinas-reflexiones-sobre-una-wikipedia-hecha-a-mano-ante-el-vertigo-generativo/>
- Shenoy, Kartik; Ilievski, Filip; Garijo, Daniel; Schwabe, Daniel; Szekely, Pedro** (2022). "A study of the quality of *Wikidata*". *Journal of web semantics*, v. 72, 100679.
<https://doi.org/10.1016/j.websem.2021.100679>
- Sigalov, Shani-Evenstein; Nachmias, Rafi** (2023). "Investigating the potential of the semantic web for education: Exploring *Wikidata* as a learning platform". *Education and information technologies*, v. 28, pp. 12565–12614.
<https://doi.org/10.1007/s10639-023-11664-1>
- Tharani, Karim** (2021). "Much more than a mere technology: A systematic review of *Wikidata* in libraries". *The journal of academic librarianship*, v. 47, n. 2.
<https://doi.org/10.1016/j.acalib.2021.102326>
- Turki, Houcemeddine; Hadj-Taieb, Mohamed-Ali; Ben-Aouicha, Mohamed; Raspberry, Lane; Mietchen, Daniel** (2023). "Ten years of *Wikidata*: A bibliometric study". In: *Wikidata '23: Wikidata Workshop at ISWC*.
<https://ceur-ws.org/Vol-3640/paper13.pdf>
- Aenor* (2014). *UNE 25964:2014. Tesoros e interoperabilidad con otros vocabularios. Parte 1: Tesoros para la recuperación de la información*. Madrid: Aenor, 2014. Equivalente a ISO 25964-1:2011.
- Vrandečić, Denny; Pintscher, Lydia; Krötzsch, Markus** (2023). "Wikidata: The making of". In: *WWW'23 Companion: Companion Proceedings of the ACM Web Conference 2023*, April 30–May 04. ACM. ISBN: 978 145039419
<https://doi.org/10.1145/3543873.3585579>
- Vrandečić, Denny** (2020). "Collaborating on the sum of all knowledge across languages". *Wikipedia@20*.
<https://wikipedia20.mitpress.mit.edu/pub/svu3uy1z/release/2>
- Zhao, Fudie** (2022). "A systematic review of *Wikidata* in digital humanities projects". *Digital Scholarship in the Humanities*, v. 38, n. 2, pp. 852–874.
<https://doi.org/10.1093/llc/fqac083>