Linee guida per l'utilizzo di wikibase.cloud in un progetto di gestione dei (meta)dati nelle biblioteche

Tipo di documento	Risultato finale del Laboratorio 4, Metadati intelligenti
Contesto	10 Laboratori virtuali per biblioteche e sviluppo sostenibile organizzati da • AIB • Rete delle Reti • EBLIDA
Riferimenti del Contesto	 Associazione Italiana Biblioteche https://tinyurl.com/10-lab-AIB. Rete delle Reti EBLIDA Presentazione dell'iniziativa dei 10 laboratori di Anna Bernabè alla Conferenza di Atene 2022 di EBLIDA
Autori	 Stefano Bargioni Giovanni Bergamin Valentina Lepore
Versione finale	6 luglio 2022
Condizioni di disponibilità	CC BY

Laboratorio 4, Metadati intelligenti

Indice

- 0. Perché Metadati intelligenti e il Laboratorio 4
- 1. Perché wikibase e wikibase.cloud
- 2. L'utilizzo di Wikibase oggi
- 3. Cosa si può fare con wikibase e con wikibase.cloud
- 4. Il (meta)modello di dati Wikibase
- 5. Come si può adattare wikibase ad un progetto
- 6. Riepilogando
 - 6.1. Il problema
 - 6.2. Le alternative sono poco sostenibili
 - 6.3. La soluzione Wikibase.cloud
 - 6.4. Benefici attesi
 - 6.5. Casi d'uso e tipologie di utenti
 - 6.6. Struttura dei costi
 - 6.7. Fonti di sostenibilità economica
- 7. Bibliografia (minimale)
 - 7.1 Documentazione ufficiale:
 - 7.2 Studi e presentazioni utili

0. Perché Metadati intelligenti e il Laboratorio 4

Il termine *intelligente* per formare la figura retorica della personificazione viene usato anche nel titolo del Laboratorio 4 in analogia a quanto oggi viene comunemente fatto per *intelligenza artificiale, smart cities, smartphone* ecc.

Per questo Laboratorio i metadati *intelligenti* sono quei metadati che contribuiscono al raggiungimento degli obiettivi dell'Agenda 2030: in particolare per l'obiettivo 4, ma l'organizzazione della conoscenza è un prerequisito per tutti gli obiettivi dell'Agenda.

Il *Laboratorio 4* è stato *coordinato* da Stefano Bargioni e *facilitato* da Giovanni Bergamin e Valentina Lepore.

Scopo del laboratorio era quello di approfondire l'uso di Wikibase come strumento per gestire i (meta)dati rilevanti per i servizi di una biblioteca o di una istituzione MAB.¹

I componenti del *Laboratorio 4* in ordine alfabetico:

- Angelica Bertellini
- Anna Molino
- Cinzia Martone
- Denise Biagiotti
- Federica Olivotto
- Marina Gatti
- Stefania Puccini

I laboratori hanno privilegiato la metodologia della *sperimentazione sul campo* grazie alla disponibilità di una istanza *Wikibase* di lavoro *Metadati intelligenti* installata su *wikibase.cloud* (inizialmente WBstack).²

Il risultato della sperimentazione collettiva è accessibile qui:

https://smartmeta.wikibase.cloud/>.

¹ < https://it.wikipedia.org/wiki/GLAM#Terminologia_italiana>

² Qui notizie sul progetto WBStack

https://addshore.com/2019/11/an-introduction-to-wbstack/.

Tutti i membri del Laboratorio hanno creato il proprio account su questa istanza e hanno contribuito alla sperimentazione.

Sono stati effettuati 5 incontri di lavoro. Per ogni incontro è stato redatto un resoconto del lavoro svolto.

1. Perché wikibase e wikibase.cloud

Mettendo insieme due definizioni - quella di Priscilla Caplan³ e quella di Karen Coyle⁴ - possiamo definire i *metadati* come *informazione strutturata secondo un determinato modello (o schema) con l'obiettivo di rendere possibile o facilitare una determinata attività umana*. Ad esempio i metadati di un catalogo ci aiutano a trovare il libro che ci serve così come i metadati di *Google maps* ci possono guidare alla destinazione che abbiamo scelto.

Le biblioteche hanno una lunga tradizione in questo campo anche se - come molti hanno osservato - tutte le potenzialità del lavoro delle biblioteche nel campo dei metadati non sono ancora adeguatamente valorizzate nell'ecosistema del web. Stiamo parlando ovviamente di un web conforme alla visione del *web semantico* e alla proposta tecnologica dei *linked data*.

L'iniziativa Wikidata nasce nel 2012⁵ come base di dati di supporto alle varie versioni linguistiche di Wikipedia ma anche con la prospettiva di servire come base di conoscenza generale "for the world at large". In altre parole non è necessario che una "informazione strutturata" rispetti il canone della *notability* che caratterizza Wikipedia: un (meta)dato può trovare posto in Wikidata se si riferisce a una entità concettuale o materiale chiaramente identificabile e descritta utilizzando fonti affidabili e disponibili al pubblico.⁶ La struttura di Wikidata è stata progettata per essere del tutto conforme alle esigenze del *web semantico* e della tecnologia dei *linked data* e si presenta come uno strumento di collaborazione alla portata di tutti.

³ Priscilla Caplan, *Metadata fundamentals for all librarians*, 2003.

⁴ Karen Coyle, *Metadata: data with a purpose*, 2004.

http://www.kcoyle.net/meta_purpose.html>.

⁵ Resoconto dettagliato di tutte le tappe https://addshore.com/2022/02/wikibase-a-history/

⁶ Per la *notability* in Wikipedia: https://en.wikipedia.org/wiki/Wikipedia:Notability; in Wikidata: https://www.wikidata.org/wiki/Wikidata:Notability; in

Nel 2018 *Wikibase* ovvero il motore software di *Wikidata* è stato distribuito come software open source. Tutti possono contribuire a Wikidata ma possono anche installare una istanza di *wikibase* e popolarla con (meta)dati relativi a un dominio specifico.

L'interesse del mondo delle biblioteche sia per Wikidata che per Wikibase è cresciuto nel tempo. Esistono numerose iniziative nel campo delle biblioteche che utilizzano Wikidata o Wikibase. L'IFLA alla fine del 2019 ha creato un gruppo di lavoro per esplorare "l'integrazione di Wikidata e Wikibase con i sistemi bibliotecari e l'allineamento dell'ontologia Wikidata con i formati dei metadati delle biblioteche come BIBFRAME, RDA e MARC").

A fine 2020 Deutsche Bibliothek e Wikimedia Deutschland hanno pubblicato il *Wikilibrary manifesto*⁷ che propone la collaborazione tra biblioteche e il mondo wiki per la costruzione di una rete aperta per l'arte, la cultura e la scienza basata sui *linked data*. La proposta è aperta a tutte le istituzioni interessate a un ecosistema di dati basati sui principi FAIR.⁸ La prima parte del Manifesto è dedicata ai principi, mentre la seconda è dedicata alle azioni (measures) che rendono concretamente applicabili quei principi. In particolare il progetto Wikibase viene individuato come una promettente infrastruttura tecnica per la gestione, condivisione e fruizione dei dati (non solo bibliografici). Il 15 giugno 2021 il Manifesto è stato sottoscritto anche da IFLA.⁹

Un svolta significativa nella modalità di gestione di Wikibase arriva nell'ottobre 2021 in occasione dell'evento *WikidataCon 2021* con l'annuncio da parte di Wikimedia Deutchland (WMDE) dell'avvio di *wikibase.cloud*.

Nel documento *Linked Open Data/Strategy 2021*¹⁰ viene presentata la nuova strategia di sviluppo per quanto riguarda i *linked open data.*

⁷ < https://www.wikimedia.de/the-wikilibrary-manifesto/>

^{8 &}lt;https://www.go-fair.org/fair-principles/>

^{9 &}lt; https://www.ifla.org/node/93952>

¹⁰ < https://meta.wikimedia.org/wiki/LinkedOpenData/Strategy2021>

Questa la *visione* che ispira la strategia WMDE

«Lavoriamo per un futuro in cui le persone condividono il potere di raccogliere e organizzare i dati che permettono la comprensione del mondo da parte dell'umanità.

Comunità di tutte le tipologie in tutto il mondo partecipano a Wikidata e a una rete di istanze Wikibase specializzate, co-creando un grafico della conoscenza globale aperto e gratuito nella ricchezza del web caratterizzato dai Linked Open Data.

Questa vuole essere la base su cui persone, aziende e istituzioni di tutte le dimensioni possono produrre nuove informazioni, creare nuove applicazioni e servizi e cambiare il mondo in meglio».¹¹

La visione si traduce in due linee strategiche:

- potenziare Wikidata e il suo ruolo di mettere a disposizione "general-purpose data about the world" (con la consapevolezza emersa chiaramente all'evento WikidataCon 2021- dell'impossibilità di gestire con un unico servizio tutti i (meta)dati del mondo;¹²
- favorire la nascita di una rete di istanze wikibase specialistiche mettendo a disposizione:
 - Wikibase suite ovvero la possibilità per tutti di installare e gestire e personalizzare una propria istanza specialistica wikibase;
 - Wikibase.cloud ovvero la possibilità per tutti di fruire di una istanza
 Wikibase in modalità Saas e gestita direttamente in cloud da WMDE.

_

¹¹ https://meta.wikimedia.org/wiki/LinkedOpenData/Strategy2021/Joint_Vision">https://meta.wikimedia.org/wiki/LinkedOpenData/Strategy2021/Joint_Vision

¹² <https://tinyurl.com/3zt8z2bm>

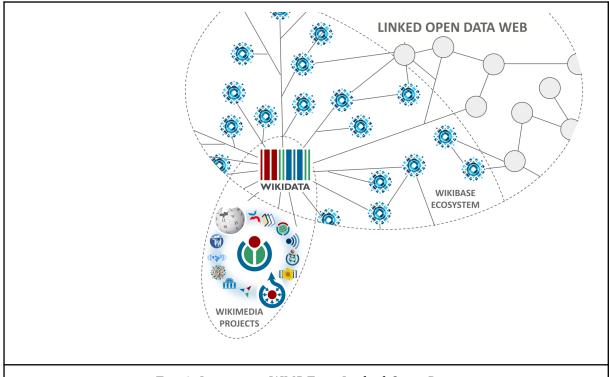


Fig. 1: La visione WMDE sui Linked Open Data

Image credits: Dan Shick (WMDE), CC BY-SA 4.0, via Wikimedia Commons

Wikidata con le istanze Wikibase (*suite* e *cloud*) formeranno l'ecosistema wikibase che si candida ad essere un parte fondamentale del web dei LOD (vedi figura 1).

Così come pianificato nel febbraio 2022 il servizio Wikibase.cloud è attivo.¹³ In questo momento (luglio 2022) è possibile prenotarsi per ottenere un invito.

2. L'utilizzo di Wikibase oggi

Senza la pretesa di completezza qualche dato sull'iniziativa Wikibase:¹⁴

- Wikibase per la ricerca scientifica: circa 63 progetti in corso;
- Wikibase in generale: 148 utenti e 510 istanze di Wikibase su WBstack (ora wikibase.cloud);

¹³ < https://meta.wikimedia.org/wiki/Wikibase/Wikibase.cloud>.

¹⁴ I dati si trovano in https://tinyurl.com/3hcpdnpe.

- WMDE sta dedicando al progetto Wikibase 14 persone a tempo pieno;
- La comunità interessata allo sviluppo di Wikibase può contare su oltre 200 membri sul gruppo Telegram Wikibase e 182 sviluppatori contribuiscono direttamente su Github;
- esiste un gruppo Wikibase Stakeholder Group formato a oggi da 18 organizzazioni e da 24 membri a titolo personale - che collabora direttamente allo sviluppo, alla pianificazione e alla crescita dell'iniziativa Wikibase

Un elenco ragionato di realizzazioni in corso

1. Enslaved

- «Search over numerous datasets and reconstruct the lives of people involved in the historical slave trade. Browse interconnected data, generate visualizations, and explore short biographies of enslaved and freed peoples»
- o Interfaccia wiki

2. Artbase

- «Welcome to the Rhizome ArtBase. This is an archive of born-digital artworks from 1999 to the present day. A small sample of artworks are shown below; you can also browse the archive by date or by artist name.»
- o note:
 - i. si basa su una personalizzazione di MediaWiki per l'interfaccia di fruizione
 - ii. vedi per esempio la query SPARQL Artworks in the ArtBase with more than one artist, including the artist names (visualized as a graph with images)

3. Kohesio

- «Kohesio: discover EU projects in your region»
- L'installazione Wikibase che supporta Kohesio

4. Wikibase nel mondo GLAM

- o Il progetto FNE (BnF e Abes) Wikibase per l'archivio di autorità nazionale
- Il progetto NOEMI (BnF) uso di Wikibase come strumento interno di catalogazione (oltre 300 catalogatori). L'acronimo NOEMI sta per «Nouer les Œuvres, Expressions, Manifestations et Items».
- GND: 1) registrazione di "How to open Authority Control system The GND & Wikibase (WikidataCon 2021)" 2) slide 30.6.2022;
- Lexbib: «Welcome to LexBib wikibase at Elexis. LexBib is a digital bibliography project for the domain of Lexicography and Dictionary Research. General aims are LOD-ification of publication metadata, and content description of metalexicographic article full texts, as a service to the lexicographic community. LexBib wikibase brings together LexBib Zotero bibliographical data and a vocabulary of lexicographic terms, LexVoc. We are working on extending the scope of LexBib towards Dictionaries.»
- Smithsonian Libraries and Archives & Wikidata: Using Linked Open Data to Connect Smithsonian Information

3. Cosa si può fare con wikibase e con wikibase.cloud

Con una installazione Wikibase (basata su Wikibase suite o su Wikibase.cloud) un progetto sui (meta)dati ha a disposizione:

- 1. una interfaccia grafica (GUI) per aggiornare un database in modalità collaborativa e con strumenti di controllo della versione dei dati;
- 2. un database organizzato come triplestore ovvero costruito appositamente per il salvataggio e il recupero di triple soggetto-predicato-oggetto. Una particolarità molto interessante è che il modello dei dati Wikibase prevede che ogni dichiarazione (statement) espresso sotto forma di *tripla* possa essere arricchita

con qualificatori (qualifiers) e/o da riferimenti (references). ¹⁵ Dal punto di vista tecnico - come prevede RDF - ogni tripla può essere "reificata" e diventare soggetto di un'altra tripla; ¹⁶

- 3. un endpoint SPARQL con una interfaccia grafica (GUI) per l'accesso alle triple inserite:
- 4. la possibilità di ricercare i dati inseriti attraverso l'integrazione con Elasticsearch;
- 5. API per l'accesso a Wikibase attraverso programmi in lettura e scrittura;
- 6. disponibilità di script come PyWikibot, WikidataIntegrator;
- 7. disponibilità di strumenti di facile uso per l'importazione o l'esportazione (soprattutto massiva) dei dati:
 - a. quickstatements;
 - b. cradle;
 - c. OpenRefine reconciliation service (non ancora disponibile su wikibase.cloud).¹⁷

4. Il (meta)modello di dati Wikibase

Il *Wikibase data model* è il modello concettuale che rende possibile la centralizzazione dell'archiviazione e della gestione delle informazioni sulle entità che a partire dal repository di Wikidata popolano tutti i progetti Wikimedia. È da sottolineare, in particolare, che il *Wikibase data model* delinea una modalità di strutturazione dei dati applicabile a qualsiasi tipo di dato purché esso sia singolo e disaggregato, nonché esprimibile come entità atomica suscettibile di interrelazione con altre entità atomiche a loro volta trattabili secondo il *Wikibase data model*. In breve, il *Wikibase data model* può essere adottato per descrivere qualsiasi oggetto di qualsiasi ambito di realtà, quindi è utilizzabile anche per la descrizione di qualsiasi entità rientri nel perimetro del mondo

¹⁵ Si tratta di un requisito fondamentale per Wikidata. Per poter essere utilizzata come fonte secondaria Wikidata deve documentare le sue dichiarazioni (triple) con il contesto (qualifier) e la fonte (references) di una determinata dichiarazione o tripla. Ad esempio (vedi https://www.wikidata.org/wiki/Q38) la tripla *L'Italia ha 50.199.700 abitanti* ha come contesto *nel 1960* e come fonte *World Bank Open Data* mentre la tripla *L'Italia ha 60.317.000 abitanti* ha come contesto *nel 2020* e come fonte *ISTAT*

¹⁶ Per Wikibase come (meta)modello vedi il paragrafo 4. Per la reificazione si può vedere anche http://ceur-ws.org/Vol-1457/SSWS2015_paper3.pdf.

¹⁷ Fonte con integrazioni: https://tinyurl.com/3hcpdnpe.

biblioteconomico, o, in senso più ampio, del trattamento delle risorse delle istituzioni MAB. Ciò è reso possibile perché esso è aperto e non limitato ontologicamente, ovvero è interoperabile.

Va del resto detto, più esattamente, che il *Wikibase data model* più che un modello concettuale è un metamodello. Vale a dire che, rispetto ai modelli di dati comuni, il *Wikibase data model* consta di un livello di astrazione ulteriore in ragione del quale, sul piano della rappresentazione della realtà in concetti correlabili, esso comporta soltanto la definizione della macrostruttura sintattica (cioè delle modalità di relazione fra le entità di interesse) ma non stabilisce né un'ontologia né un dominio di riferimento che al livello semantico esigano un vincolo sul piano dei contenuti da gestire.

In altre parole, il *Wikibase data model* predispone soltanto le istruzioni su come organizzare le informazioni sui contenuti ritenuti rilevanti allo scopo di strutturare enunciati significativi e di senso compiuto. Ma i contenuti ritenuti rilevanti sono di volta in volta individuati e ricontrattati dai contributori di Wikidata in modo scalabile e dal basso secondo un approccio bottom-up, nonché descritti e interrelati fra loro tramite proprietà anch'esse non predefinite. Ciò permette la massima interoperabilità semantica, ovvero la possibilità di formulare e interpretare qualsiasi enunciato su qualsiasi oggetto, perché il *Wikibase data model* non impone dall'alto, in modalità top-down, nessun vocabolario dominante.

5. Come si può adattare wikibase ad un progetto

Wikibase può essere conveniente se l'obiettivo finale è quello di pubblicare Linked Open Data (LOD). Non sembra esistere uno strumento liberamente disponibile che non sia legato a un particolare fornitore commerciale e che consenta l'immediata pubblicazione di linked open data.

Normalmente gli strumenti per pubblicare LOD sono legati a un particolare dominio (hanno una ontologia predeterminata) oppure prevedono una conversione periodica in LOD i dei metadati prodotti con strumenti tradizionali.

Esistono strumenti open per costruire ontologie come ad esempio Protégé. Ma non esistono strumenti che permettono di costruire progressivamente l'ontologia mentre si immettono i dati. Questo comporta anche rischi da valutare, ma in principio la *costruzione collaborativa di una ontologia* è un obiettivo sensato: a oggi gli "item" creati in Wikidata sono intorno ai 98 milioni (con 1,6 miliardi di aggiornamenti o "edit"). ¹⁸

Wikibase ha una barriera di ingresso molto bassa se lo mettiamo al confronto con altri strumenti nel mondo dei linked open data.

Che cosa potrebbe mancare in Wikibase:

- una interfaccia controllata ed ergonomica per l'immissione dati:
 - esistono strumenti come ShEx o cradle che offrono una interfaccia di inserimento personalizzata senza bisogno di programmi esterni;
 - se ci sono esigenze particolari l'interfaccia di inserimento può essere costruita con un'applicazione ad hoc che usa le API (in lettura e scrittura) per l'accesso a Wikibase (come ad esempio Enslaved e Kohesio).

Per l'installazione di Wikibase sono necessarie competenze informatiche di base, ma per l'uso di wikibase.cloud (gestione, configurazione, progettazione di nuove funzionalità) sono utili percorsi di formazione (magari sullo stile dei corsi proposti da Library Carpentry). Molte sono le possibilità offerte in rete.

In generale è noto che il software open source sviluppato nei progetti di ricerca e nei progetti europei è a rischio quando un progetto termina: mancano finanziamenti per sviluppare e mantenere il software. La scelta di progettare e di sviluppare il software all'interno dell'ecosistema wikibase consente di proteggere l'investimento fatto (se vengono seguite le regole dello sviluppo del software open source la comunità lo prende in carico e lo mantiene).

-

¹⁸< https://www.wikidata.org/wiki/Wikidata:Statistics> (alla data del 1.6.2022).

6. Riepilogando

6.1. Il problema

- le biblioteche hanno di solito strumenti specifici per gestire i metadati relativi a un determinato dominio;
- mancano strumenti per gestire le particolarità di una istituzione (tipologie di raccolte) o per dare vita a particolari iniziative (es. repertori tematici, raccolta di dati che documentano eventi ecc);
- mancanza di strumenti per interoperare semanticamente con altri domini.

6.2. Le alternative sono poco sostenibili

- forzare gli applicativi esistenti orientati al dominio dei metadati bibliografici a trattare altri tipi di metadati;
- soluzioni *ad hoc* con rischio di perdere il *controllo* d'insieme.

6.3. La soluzione Wikibase.cloud

- per le istituzioni è senza costi diretti per l'acquisizione e il mantenimento delle tecnologie di base: viene fornita come servizio (Wikibase as a Service);
- permette di far parte di un ecosistema di gestione dei metadati diffuso (federazione e uso delle tecnologie del web semantico);
- modello wiki: barriera di ingresso molto bassa, commenti, trasparenza delle modifiche, possibilità di ripristino.

6.4. Benefici attesi

- buon rapporto risorse (umane, finanziarie ecc) investite/risultati;
- in caso di problemi è più facile trovare soluzioni considerato il numero di installazioni di wikibase;
- modello *sostenibile* di *linked data* e rispetto della *ontodiversità*. 19
- aumento della *partecipazione culturale* oltre che il *riuso* attraverso uno dei più vasti ecosistemi di gestione dei (meta)dati.

¹⁹ Neologismo qui creato sul modello di *biodiversità*. Ogni ontologia nel web semantico è *un punto di vista*: wikibase può ospitare la condivisione e il riuso dei dati tra punti di vista/ontologie differenti.

6.5. Casi d'uso e tipologie di utenti

- 1. utente che cerca e ottiene come risultato solo pagine web (non è necessario che conosca il funzionamento di wikibase);
- 2. utente che usa funzionalità SPARQL;
- 3. macchine (agenti software) che interrogano, indicizzano dati (es. motori di ricerca) o riusano i dati per altri servizi, applicazioni e utenti;
- 4. utente che aggiorna tramite interfaccia wikibase;
- 5. utente che aggiorna tramite interfaccia quickstatements (importazione, modifiche e cancellazioni massive di dati);
- 6. utente che aggiorna tramite *form liberamente personalizzabili* con controlli sui dati immessi (es. craddle).

6.6. Struttura dei costi

- wikibase.cloud è senza costi se i dati sono in libera fruizione (ogni istituzione può in ogni caso decidere chi può aggiornare i dati);
- i costi opzionali riguardano la creazione di interfacce per l'utente finale o per l'immissione dei dati;
- nel caso l'istituzione decida di non usare più wikibase, tutti i (meta)dati immessi in una istanza wikibase sono scaricabili in qualsiasi momento in formati e serializzazioni standard

6.7. Fonti di sostenibilità economica

Se vengono seguite le buone pratiche della produzione del software open source tutti gli investimenti nel miglioramento e nella personalizzazione di wikibase possono essere condivisi e migliorare la sostenibilità economica complessiva di tutti gli utilizzatori di wikibase

7. Bibliografia (minimale)

7.1 Documentazione ufficiale:

- il sito ufficiale Wikibase: https://wikiba.se/;
- video introduttivo di Wikimedia Deutchland
 ;">https://www.youtube.com/watch?v=dCAjhje]pgY>;
- il metamodello dei dati wikibase:
 https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer;
- Wikibase suite https://www.mediawiki.org/wiki/Wikibase/Install;
- Wikibase cloud
 - o <https://wikibase.cloud>
 - <https://meta.wikimedia.org/wiki/Wikibase/Wikibase.cloud>

7.2 Studi e presentazioni utili

Lepore, Valentina. 2019. Il Wikibase data model per la pubblicazione dei dati bibliografici sul web semantico. Una sperimentazione presso la Biblioteca Nazionale Centrale di Firenze. Il level Master thesis, Università degli studi di Firenze. http://eprints.rclis.org/39494/

Mauri, Luca. 2021. *Wikibase, starting from scratch* https://www.youtube.com/watch?v=7kUUarBN2vM>. Come installare da soli la suite Wikibase.

Rossenova, Lozana. 2022. *Examining Wikidata and Wikibase in the context of research data management applications*. https://tinyurl.com/3hcpdnpe.

Shorland, Adam. 2022. Wikibase a history.

https://addshore.com/2022/02/wikibase-a-history/