

Laboratorio 4, Metadati intelligenti

Bozza versione finale 2022-05-15

Linee guida per l'utilizzo di wikibase.cloud in un progetto di gestione dei (meta)dati rilevanti in una istituzione MAB

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 \(minimale\)](#)

0. Perché Metadati intelligenti e il Laboratorio 4

L'uso del termine *intelligente* per formare la figura retorica della **personificazione** viene qui applicato ai metadati in analogia a quanto 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 tema trasversale a tutti gli obiettivi.

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.

I componenti del *Laboratorio 4* in ordine alfabetico:

- Angelica Bertellini
- Anna Molino
- Cinzia Martone
- Denise Biagiotti
- Federica Olivotto
- Laura Borsi
- 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 *WBstack* "*Wikibase as a service*".¹

Il risultato della sperimentazione collettiva è accessibile qui:

<<https://smartmeta.wiki.opencura.com/>>.²

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 Qui notizie sul progetto WBStack <<https://addshore.com/2019/11/an-introduction-to-wbstack/>>.

² Diventerà dopo la migrazione su [wikibase.cloud](https://smartmeta.wikibase.cloud): <<https://smartmeta.wikibase.cloud>>. Per notizie sulla migrazione <<https://addshore.com/2022/02/wbstack-close-and-migration/>>.

1. Perché wikibase e wikibase.cloud

Mettendo insieme due definizioni - quella di P. Caplan (Metadata fundamentals for all librarians, 2003) e quella di K. Coyle (Metadata: data with a purpose, 2004) - 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 2012³ 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.⁴ 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.

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*⁵ 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.⁶ 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

³ Resoconto dettagliato di tutte le tappe <<https://addshore.com/2022/02/wikibase-a-history/>>

⁴ Per la *notability* in Wikipedia: <<https://en.wikipedia.org/wiki/Wikipedia:Notability>>; in Wikidata: <<https://www.wikidata.org/wiki/Wikidata:Notability/en>>

⁵ <<https://www.wikimedia.de/the-wikilibrary-manifesto/>>

⁶ <<https://www.go-fair.org/fair-principles/>>

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.⁷

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 Deutschland (WMDE) dell'avvio di *wikibase.cloud*.

Nel documento *Linked Open Data/Strategy 2021*⁸ viene presentata la nuova strategia di sviluppo per quanto riguarda i *linked open data*.

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.⁹

»

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);¹⁰
- 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*.

⁷ <<https://www.ifla.org/node/93952>>

⁸ <<https://meta.wikimedia.org/wiki/LinkedOpenData/Strategy2021>>

⁹ <https://meta.wikimedia.org/wiki/LinkedOpenData/Strategy2021/Joint_Vision>

¹⁰ <<https://tinyurl.com/3zt8z2bm>>

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

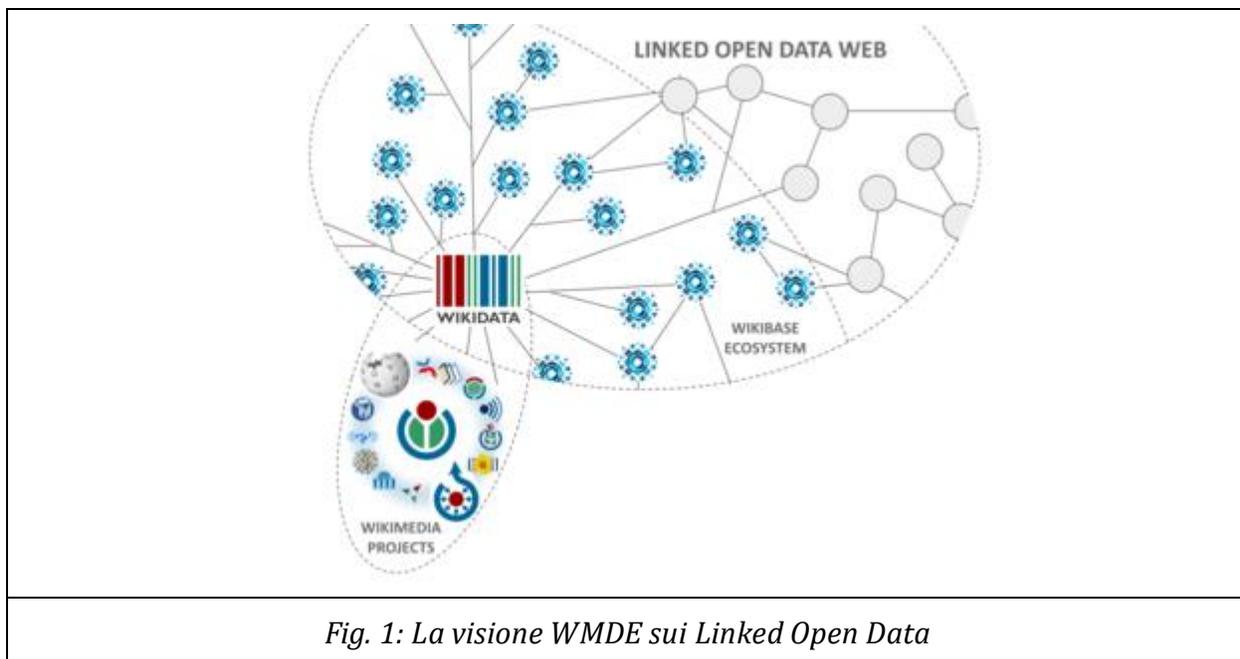


Fig. 1: La visione WMDE sui Linked Open Data

2. L'utilizzo di Wikibase oggi

Senza la pretesa di completezza qualche dato sul fenomeno Wikibase:¹¹

- Wikibase per la ricerca scientifica: circa 63 progetti in corso;
- Wikibase in generale: 148 utenti e 510 istanze di Wikibase su WBstack (l'infrastruttura WMDE di test su cloud che sta migrando - aprile, maggio 2022 - su [wikibase.cloud](#));
- 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](#)

¹¹ I dati si trovano in <https://tinyurl.com/3hcpdnpe> e sono stati aggiornati da queste *Linee guida* a maggio 2022

- «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»
- [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.»
- 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

- [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](#);
- [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 metalexigraphic 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).¹² Dal punto di vista tecnico - come prevede RDF - ogni tripla può essere “reificata” e diventare soggetto di un’altra tripla;¹³
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](#)).¹⁴

4. Il (meta)modello di dati Wikibase

Il Wikibase data model è il modello concettuale che presiede l’archiviazione e il management centralizzati delle informazioni sulle entità che a partire dal repository di Wikidata popolano tutti i progetti Wikimedia. È da sottolineare, in particolare, che il

¹² 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*

¹³ Per Wikibase come (meta)modello vedi il paragrafo 3

¹⁴ Fonte con integrazioni: <https://tinyurl.com/3hcpdnpe> (Fare voce in Bibliografia)

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 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 metamodel. 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 fra loro interrelabili, 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. Non sembra esistere sul mercato uno strumento non legato a un particolare che consente l'immediata pubblicazione di linked open data.

Normalmente gli strumenti per pubblicare LOD sono legati a un particolare dominio (hanno una ontologia predeterminata) oppure sono dei lavori di traduzione in LOD si quando viene immesso nei database gestionali.

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 (e con questo principio che Wikidata ha superato a oggi il milione di entità e conta su diecimila proprietà).

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

Che cosa manca in Wikibase

- una interfaccia controllata per l'immissione dati.
 - Esistono strumenti come [ShEx](#) o [cradle](#) che offrono una interfaccia di inserimento personalizzata senza bisogno di programmi esterni (seguendo le istruzioni con un po' di pazienza).
 - se ci sono esigenze particolari l'interfaccia di inserimento può essere fatta con un'applicazione ad hoc che usa le API per l'accesso a Wikibase attraverso programmi in lettura e scrittura
- una interfaccia amichevole di fruizione (tipo Enslaved o Kohesio)
 - si può fare con API per l'accesso a Wikibase attraverso programmi in lettura e scrittura.

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 [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 viene di solito abbandonato al momento della chiusura del progetto: mancano finanziamenti per mantenere il software. La scelta di progettare il software all'interno dell'ecosistema wikibase consente di proteggere l'investimento fatto.

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* e perdita del *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;
- facilità di intervento in una piattaforma di grande diffusione;
- modello *sostenibile* di *linked data* e rispetto della *ontodiversità*.¹⁵
- aumentare la *partecipazione culturale* oltre che il *riuso* attraverso uno dei più vasti ecosistemi di gestione dei (meta)dati.

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 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 massiva di dati);
6. utente che aggiorna tramite *form* (es. cradle).

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);
- è possibile investire per interfacce di fruizione e/o di aggiornamento diversificate per particolari categorie di utenti (caso d'uso 1);
- l'investimento sui dati è pensato per il lungo periodo (nella previsione che il mondo Wikibase non esista più): tutti i metadati sono scaricabili in qualsiasi momento in formati e serializzazioni standard.

6.7. Fonti di sostenibilità economica

- se anche l'investimento nelle interfacce viene messo a fattor comune (es open source con github) i costi sono del tutto affrontabili dalle istituzioni che adottano *wikibase.cloud*

¹⁵ neologismo creato sul modello di *biodiversità*. Ogni ontologia nel web semantico è *un punto di vista* e wikibase può ospitare la condivisione e il riuso dei dati tra punti di vista/ontologie differenti.

7. Bibliografia (minimale)

Il metamodello dei dati wikibase:

<https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer>

Sito ufficiale Wikibase: <https://wikiba.se/>

Introduzione a Wikibase video di Wikimedia Deutschland

<https://www.youtube.com/watch?v=dCAjhjeJpgY>

Wikibase, starting from scratch / by Luca Mauri

<https://www.youtube.com/watch?v=7kUUarBN2vM>

De Dauw, Jeroen. 2022. *Fast bulk import into Wikibase*. <https://wikibase.consulting/fast-bulk-import-into-wikibase/> “We have used this approach with our clients. Including with the German National Library, where our script can import their ~10 million item GND dataset in about a day.” (molto tecnico)