

Wikidata per la ricerca (nelle università)

Camillo Carlo Pellizzari di San Girolamo (alias [Epìdosis](#))

Scuola Normale Superiore, Wikimedia Italia

Perché Wikidata?

Ossia, per quali motivi un'università può avere interesse a creare un progetto su Wikidata?

Cosa Wikidata ha da offrire (lista non esaustiva)

1. apertura

- visibilità dei propri dati in uno dei punti nodali del web semantico
- integrazione dei propri dati con oltre 109 milioni di entità già presenti in Wikidata; questo, tra l'altro, facilita notevolmente la traduzione dei propri dati, essendo Wikidata un progetto multilingue, e li rende quindi più accessibili al pubblico non italofono e non anglofono
- interoperabilità dei propri dati con tutti i database e servizi che usano Wikidata
- vasto riuso dei propri dati nel web semantico

Tutto ciò sarebbe impossibile mantenendo questi dati soltanto in un database proprio di piccole dimensioni.

Cosa Wikidata ha da offrire (lista non esaustiva)

2. collaborazione

- confronto dei propri dati con quelli di Wikidata, così da individuare errori e incongruenze sia nei propri dati sia in quelli di Wikidata
- confronto dei propri dati, tramite Wikidata, anche con quelli di altri database già riconciliati con Wikidata, così da individuare ulteriori errori e incongruenze sia nei propri dati sia altrove
- collaborazione diretta colla comunità di volontari (cfr. citizen science) e di studiosi che contribuiscono a Wikidata

Cosa Wikidata ha da offrire (lista non esaustiva)

3. analisi

- approfondite possibilità di analizzare i propri dati, con vari metodi (query SPARQL, API, dump ecc.), per ricavare nuova conoscenza
- alcuni servizi per visualizzare i propri dati (su mappa, in grafici ecc.); trattandosi in generale di servizi open source, è possibile adoperarsi per migliorarli se lo si ritiene utile

4. vantaggi tecnici

- open source; in costante sviluppo ma fondamentalmente stabile
- mantenuta da Wikimedia Deutschland e Wikimedia Foundation (quindi senza nessun costo di gestione per l'università)

Due premesse introduttive

su Wikidata a confronto con Wikipedia

Premessa 1: Wikidata è largamente sconosciuta

- sia all'opinione pubblica in generale
 - sia (in misura solo lievemente minore) nelle università
- perché:
- esiste (2012-) da meno tempo di Wikipedia (2001-)
 - soprattutto, perché i suoi principali lettori non sono gli umani ...
 - ... sebbene gli umani ricevano indirettamente (senza rendersene conto) dati provenienti da Wikidata in mille modi diversi (tramite Google, assistenti automatici, IA, ecc.)

Conseguenza 1: bisogna innanzitutto farla conoscere

- soprattutto nelle università (ma anche in istituzioni GLAM, pubblica amministrazione ecc.)

quindi:

- la prima e fondamentale richiesta di Wikimedia Italia alle università italiane dovrebbe essere quella di ospitare dei brevi corsi introduttivi a Wikidata, distinti in almeno due percorsi distinti: per il personale di ricerca (professori, ricercatori, dottorandi, anche studenti) e per il personale non di ricerca (bibliotecari, archivisti, amministrativi ecc.)

Conseguenza 1: bisogna innanzitutto farla conoscere

- sarebbe inoltre interessante includere in questi corsi esempi di progetti già svolti (interamente o parzialmente) di diverse tipologie, come fonte di ispirazione per nuovi progetti
- i corsi dovrebbero essere almeno in parte in presenza, perché svolgere sessioni laboratoriali a distanza è largamente inefficace

inoltre:

- i progetti svolti devono avere documentazione il più possibile chiara e completa in Wikidata, e possibilmente dovrebbero essere descritti anche in pubblicazioni scientifiche

Premessa 2A: Wikidata appassiona meno di Wikipedia

(nella mia percezione)

- scrivere voci enciclopediche è un'attività che, come hobby, suscita la passione potenzialmente di numerose persone
- al contrario, compilare schede di un database è un'attività che molte meno persone trovano spontaneamente interessante

quindi:

- una larga parte di chi scrive su Wikipedia lo fa per passione come hobby, ma ciò è significativamente più raro su Wikidata

Conseguenza 2A: l'attività su Wikidata richiede un chiaro scopo, ossia un progetto

chi contribuisce a Wikidata molto spesso lo fa principalmente:

- come parte della propria attività lavorativa
- oppure in occasione di un singolo progetto pagato
- oppure per uno scopo accademico (preparare un seminario, un intervento a un convegno, una tesi, una pubblicazione)

quindi:

- affinché uno di queste fattispecie possa verificarsi, serve che l'attività si inserisca in un preciso progetto definito a priori

Premessa 2B: il valore aggiunto di Wikidata non sta nei singoli dati

- una voce enciclopedica ha un suo senso compiuto, principalmente perché può essere letta da più umani
- al contrario, una scheda di un database è raramente letta (direttamente) da umani e in generale ha una ricchezza di informazioni inferiore rispetto alla corrispondente trattazione dell'argomento in forma testuale (che di solito ne è la fonte)

quindi:

- singoli elementi di Wikidata difficilmente si possono considerare un risultato compiuto per un progetto

Conseguenza 2B: un progetto su Wikidata deve riguardare un dataset coerente

- il valore aggiunto di Wikidata in quanto collezione di dati strutturati è permettere di analizzare questi dati (principalmente tramite query SPARQL), ricavando **nuova conoscenza**
- analizzare un dataset incompleto o incoerente non dà risultati apprezzabili; invece, analizzare un dataset che sia coerente e completo permette di ottenere risultati significativi (quindi degni di ricevere un finanziamento e/o di essere pubblicati)
- quindi, ogni progetto deve partire dallo stabilire dei criteri che permettano di individuare l'insieme di elementi su cui lavorare

Due conclusioni preliminari

1. la priorità è far conoscere Wikidata nelle università
2. lavorare su Wikidata diventa interessante nel momento in cui si ha un progetto riguardante un dataset chiaramente definito

Quali progetti su Wikidata per le università

Classificazione di possibili progetti su Wikidata con università

principalmente per il personale non di ricerca (bibliotecari, archivisti, amministrativi ecc.)

1. progetti con biblioteche, archivi e musei universitari
2. progetti sui dati relativi all'università

principalmente per il personale di ricerca (professori, ricercatori, dottorandi, anche studenti)

3. progetti di ricerca riguardanti (περὶ τοῦ) Wikidata
4. progetti di ricerca tramite (διὰ τοῦ) Wikidata

Postilla: Wikibase Cloud

oltre che a Wikidata, il personale di ricerca può essere interessato anche alla possibilità di creare delle istanze di Wikibase Cloud (<https://www.wikibase.cloud/>) per i propri progetti di ricerca; è quindi utile, nella presentazione generale di Wikidata, includere almeno un cenno a Wikibase Cloud

❖ esempio di progetto di ricerca in Wikibase Cloud:

<https://hypotheseis.wikibase.cloud/> (filologia greca e bizantina)

1) Progetti con biblioteche, archivi e musei universitari

- affini, per molti aspetti, ai progetti GLAM con istituzioni simili non legate a università
[quindi, per motivi di tempo, non verranno trattati qui]
- possono avere come tematica privilegiata dati relativi alla storia dell'università stessa (es. prosopografia di professori passati)

2) Progetti sui dati relativi all'università

- possono avere come interlocutori gli uffici dedicati alla valutazione della ricerca e all'open science; o gli uffici dei singoli dipartimenti

possono avere come oggetto elementi riguardanti:

- repository della ricerca e database gestiti dall'università, o ai quali l'università collabora/contribuisce
- progetti di ricerca cui l'università partecipa (se rilevanti, es. perché finanziati da stati od organizzazioni internazionali)
- suddivisioni amministrative dell'università (dipartimenti, centri di ricerca, laboratori)
- singoli membri del personale di ricerca dell'università

2) Progetti sui dati relativi all'università

- hanno come principale scopo aumentare la visibilità dell'attività di ricerca dell'università presso un pubblico più ampio
- riguardano di solito la situazione presente dell'università, ma potrebbero anche riguardare la sua storia (recente o meno recente)
- possono essere svolti da:
 - personale amministrativo dell'università (dopo apposita formazione)
 - utenti di Wikidata (non dipendenti dell'università) pagati per il progetto
 - studenti dell'università (dopo apposita formazione) appositamente pagati per lo scopo, oppure che si dedicano all'attività in occasione di stage/tirocini curriculari

2) Progetti sui dati relativi all'università – alcuni esempi

- Università di Stanford:

[https://www.wikidata.org/wiki/Wikidata:WikiProject_Stanford_Libraries/](https://www.wikidata.org/wiki/Wikidata:WikiProject_Stanford_Libraries/Data_models)

[Data models](https://www.wikidata.org/wiki/Wikidata:WikiProject_Stanford_Libraries/Data_models) (il più completo tentativo di mappare su Wikidata tutto ciò che riguarda un'università: suddivisioni amministrative, personale di ricerca, progetti di ricerca ecc.)

esistono progetti simili, di estensione minore (spesso limitati a singoli dipartimenti), per numerose altre università statunitensi

2) Progetti sui dati relativi all'università – alcuni esempi

- Italia (IRIS):

https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/IRIS (tentativo di mappare su Wikidata il personale di ricerca delle università italiane, sulla base dei repository della ricerca IRIS)

- Svizzera:

https://www.wikidata.org/wiki/Wikidata:WikiCite/Researchers_in_Switzerland (tentativo di mappare su Wikidata il personale di ricerca delle università svizzere)

3) progetti di ricerca riguardanti Wikidata

hanno come focus:

- il funzionamento di Wikidata (diacronicamente e/o sincronicamente)
- i dati già presenti in Wikidata (presi di per sé e/o messi a confronto con uno o più altri database)
- a seconda dell'estensione, possono essere oggetto di seminari (es. per un esame universitario), tesi, interventi a conferenze, articoli scientifici, monografie, progetti di ricerca annuali o pluriennali
- possono avere un taglio più o meno informatico

3) progetti di ricerca riguardanti Wikidata

- si differenziano dai progetti *tramite* Wikidata perché, a differenza di questi ultimi, hanno un approccio principalmente passivo a Wikidata, ossia studiano ciò che è presente in Wikidata senza modificarla

esempi potenziali (di varia complessità):

- studio su come in Wikidata vengono strutturati i dati relativi alle biblioteche in quanto istituzioni e agli edifici che le ospitano
- studio statistico sui cittadini svedesi descritti in Wikidata (per genere, per epoca, per professione, per numero di voci di Wikipedia ecc.)
- studio su un set definito di scrittori italiani: confronto tra la completezza dei dati di Wikidata, di SBN, del GND, della BNF e della LOC

3) progetti di ricerca riguardanti Wikidata – alcuni esempi

- <https://zenodo.org/records/8107888>: articolo sulla modellizzazione dei dibattiti tra studiosi di scienze umanistiche in Wikidata
- <https://doi.org/10.36253/JLIS.IT-573>: mio articolo sulla storia della riconciliazione tra l'authority file di SBN e Wikidata
- https://aic.ai.wu.ac.at/~polleres/supervised_theses/Marco_Marsoner_BSc_2022.pdf: tesi sulla scientificità dei riferimenti in Wikidata
- WiGeDi (<https://wigedi.com/>): progetto pluriennale di studio sulla diversità di genere in Wikidata (diacronicamente e sincronicamente); ha dato luogo a numerose pubblicazioni e presentazioni in convegni

3bis) progetti di ricerca riguardanti il riutilizzo dei dati di Wikidata

- cfr. buona panoramica offerta da <https://doi.org/10.1093/llc/fqac083>
(*A systematic review of Wikidata in Digital Humanities projects*, 2023)
- tra le conclusioni, importante notare:
«Most projects tend to consume data from Wikidata, whereas there is more potential to utilize it as a platform and a technology stack to publish data on Wikidata or to create an ecosystem of data exchange»
- per possibili progetti che lavorino su Wikidata, cfr. ora punto 4

4) progetti di ricerca tramite Wikidata

si svolgono nel seguente modo:

1. scelta di una o più fonti di riferimento
2. definizione di un dataset su cui lavorare, secondo precisi criteri
3. lavoro (manuale ed eventualmente anche semiautomatico) su Wikidata per modificare gli elementi rientranti nel dataset sulla base della/e fonte/i scelta/e
4. uso di query per analizzare il dataset in Wikidata [questa parte non è sempre presente, ma dovrebbe esserlo, perché è un valore aggiunto cruciale]
5. redazione di una pubblicazione che contiene informazioni sul dataset scelto (punti 1-2), sul lavoro svolto in Wikidata (punto 3) e sui risultati emersi dalle query utilizzate (punto 4); è rilevante notare come il lavoro su Wikidata abbia spinto a una riflessione critica sulla/e fonte/i utilizzata/e

4) progetti di ricerca tramite Wikidata

- a seconda dell'estensione, possono essere oggetto di seminari (es. per un esame universitario), tesi, interventi a conferenze, articoli scientifici, monografie, progetti di ricerca annuali o pluriennali
- possono usare fonti cartacee, o comunque testuali (il cui utilizzo richiede ovviamente una fatica maggiore, ma ha anche un valore aggiunto maggiore, poiché si tratta di dati che non sono ancora stati strutturati e quindi non fanno parte del web semantico), oppure database già esistenti se in licenza CC0 (nota: il fatto che le università appongano **licenza CC0** ai loro database è condizione necessaria affinché i dati di tali database possano essere importati in Wikidata)

4) progetti di ricerca tramite Wikidata

- nello svolgimento di questi progetti, che implicano un'importante mole di lavoro in Wikidata, è utile che il personale di ricerca (soprattutto dottorandi e studenti) sia affiancato da utenti esperti di Wikidata, che possono dare indicazioni utili sulla fattibilità dei progetti e su come svolgerli nel modo al tempo stesso qualitativamente migliore e praticamente più funzionale
- l'utente di Wikidata dovrebbe essere preferibilmente (anche se non necessariamente) esperto dell'area disciplinare in cui si svolge la ricerca, o almeno di un'area affine; l'affinità di area disciplinare può aiutare l'utente a comprendere meglio le necessità della ricerca in termini di modellizzazione dei dati e terminologia; d'altro canto, la collaborazione tra persone aventi formazione diversa può recare spunti interessanti (considerato anche che spesso nelle discipline umanistiche, rispetto a quelle scientifiche, le ricerche sono condotte fin troppo spesso da singoli individui e non sempre c'è un confronto approfondito sui temi trattati, specie con persone non specializzate nel medesimo settore disciplinare)

4) progetti di ricerca tramite Wikidata

esempi potenziali (cfr.

https://www.wikidata.org/w/index.php?oldid=2096108009#Possibili_ricerche):

- lavoro su un dataset prosopografico
- lavoro su un dataset bibliografico
 - spoglio di una o più riviste (bassa complessità)
 - spoglio di miscellanee o atti di convegno (media complessità; sarebbe pionieristico, e di fondamentale importanza per le discipline umanistiche)
 - inserimento di una bibliografia tematica (alta complessità)
- lavoro su un dataset di concetti (usando come fonte/i uno o più thesaurus)
- + lavoro su un dataset di oggetti materiali
- ...

4) progetti di ricerca tramite Wikidata – alcuni esempi

- <https://riviste.aib.it/index.php/bibelot/article/view/13809>: articolo sull'importazione dell'Anagrafe delle Biblioteche Italiane in Wikidata (cfr. https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/Anagrafe_ICCU)
- <https://www.researchgate.net/publication/358246309>: articolo sull'importazione delle persone presenti nell'authority file del Schoenberg Database of Manuscripts in Wikidata
- <https://dx.doi.org/10.2426/AIBSTUDI-13893>: articolo sull'importazione di un corpus di articoli di riviste di biblioteconomia italiane in Wikidata (cfr. https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/Riviste_di_biblioteconomia)
- <https://zenodo.org/records/5126913>: tesi di laurea (in informatica) sull'importazione di un corpus bibliografico in Wikidata

Grazie per l'attenzione!

camillo.pellizzaridisangirolamo@sns.it