

Wikidata: la soluzione wikimediana ai linked open data

di Luca Martinelli

Introduzione

Wikidata è il più giovane dei progetti gestiti dalla Wikimedia Foundation, la fondazione non-profit di diritto statunitense che gestisce, fra gli altri, l'enciclopedia libera Wikipedia. Lo scopo principale di Wikidata è raccogliere e strutturare i dati fondamentali delle voci e delle pagine degli oltre 800 progetti Wikimedia (come il numero di abitanti di un comune, le date di nascita e di morte di un personaggio storico, l'anno di pubblicazione di un libro, la tassonomia di una specie vegetale o animale e così via), in modo tale che possano essere letti, tradotti, modificati e riutilizzati da chiunque (macchine comprese) in ciascuna delle 285 lingue ufficiali dei progetti Wikimedia.

Nonostante la sua giovane età (il progetto è nato il 29 ottobre 2012) e il fatto di essere relativamente sconosciuto al grande pubblico, Wikidata è già il terzo progetto Wikimedia più attivo e il suo potenziale è stato già riconosciuto da importanti istituzioni che operano nel settore dei dati aperti. Il 4 novembre 2014, è stata insignita dall'Open Data Institute (fondato da Tim Berners Lee, il 'papà' del Web, e da Nigel Shadbolt, uno dei maggiori esperti nel campo della scienza cognitiva e dell'intelligenza artificiale) dell'ODI Award 2014 come miglior sito per la pubblicazione di dati¹. Inoltre, lungo il 2015, grossi aggregatori di dati come VIAF² o MusicBrainz³ hanno scelto di interfacciarsi direttamente, laddove Google ha addirittura scelto di chiudere il proprio progetto sui dati aperti, Freebase, e di donarle l'intero database⁴.

Ma cos'è davvero Wikidata e come funziona?

LUCA MARTINELLI, Wikimedia Italia – Associazione per la diffusione della conoscenza libera, via Bergamo 18, 20900 Monza, e-mail: luca.martinelli@wikimedia.it.

Ultima consultazione siti web: 22 gennaio 2016.

1 *First ODI Open data awards presented by Sirs Tim Berners-Lee and Nigel Shadbolt*, «Open data institute news», 4 novembre 2014, <<http://theodi.org/news/first-odi-open-data-awards-presented-by-sir-tim-berners-lee-and-sir-nigel-shadbolt>>.

2 Thom Hickey, *Moving to Wikidata*, «Outgoing», 26 marzo 2015, <<http://outgoing.typepad.com/outgoing/2015/03/moving-to-wikidata.html>>.

3 *Downstream Wikipedia link usage and migration to Wikidata*, «MusicBrainz blog», 8 maggio 2015, <<http://blog.musicbrainz.org/2015/05/08/downstream-wikipedia-link-usage-and-migration-to-wikidata/>>.

4 Jason Douglas, *Announcement: from Freebase to Wikidata*, 16 dicembre 2014, <https://groups.google.com/d/msg/freebase-discuss/s_BPoL92edc/Y585r7_2E1Y>.

Un passo indietro: cosa sono gli *open data*?

Gli *open data* o dati aperti sono dati forniti da un'istituzione, pubblica o privata, resi liberamente accessibili a chiunque, con la sola restrizione dell'obbligo di citazione della fonte in caso di loro riutilizzo. La decisione di fornire libero accesso alle proprie banche dati si richiama alla più ampia disciplina dell'*open government*, in base alla quale la pubblica amministrazione dovrebbe osservare una serie di misure volte a garantire la trasparenza e la partecipazione diretta al processo di creazione dei dati, anche attraverso il ricorso alle nuove tecnologie dell'informazione e della comunicazione.

Nonostante la pratica e l'ideologia che caratterizzano i dati aperti siano da anni ben consolidate, con la locuzione *open data* si identifica un'accezione ancora piuttosto recente e fondamentalmente legata a Internet come canale principale di diffusione dei dati stessi, che condivide l'etica di base con molti altri movimenti e comunità di sviluppo (anch'essi legati prevalentemente a Internet) di tipo *open*, come l'*open source*⁵, l'*open access*⁶ e l'*open content*⁷. Per esempio, Wikipedia, Wikidata e tutti gli altri progetti Wikimedia – di cui diffusamente si parlerà in seguito – sono uno degli esempi di maggiore successo del movimento per l'*open content*.

Cos'è e come funziona Wikidata

Wikidata è una base di conoscenza secondaria, libera, collaborativa e multilingua. Più nello specifico:

- 1) secondaria. Wikidata contiene, oltre ai dati strutturati a cui si è accennato in formato 'proprietà: valore', anche le fonti da cui essi vengono tratti, in modo da riflettere la diversità delle stesse nella descrizione di un determinato argomento e ribadire la centralità che ha la nozione di verificabilità in tutti i progetti Wikimedia;
- 2) libera. Tutti i dati sono pubblicati con licenza Creative Commons zero (CC0), ossia rilasciati in pubblico dominio, in modo da garantirne il più ampio riuso possibile. In particolare, ciò significa che chiunque può liberamente copiare, modificare, distribuire e utilizzare in ogni forma i dati di Wikidata, anche per uso commerciale, con l'unico obbligo di citare la fonte da cui sono stati prelevati, esattamente come accade in tutti gli altri progetti Wikimedia;
- 3) multilingua. La navigazione, la fruizione, la modifica e il riuso dei dati è garantito nelle 285 lingue ufficiali riconosciute dalla Wikimedia Foundation, oltre che in formati *machine-readable*. L'inglese ha un ruolo preminente nelle comunicazioni fra membri, ma nell'ambito dati è fortemente incoraggiato l'uso di qualsiasi lingua;
- 4) collaborativa. L'inserimento e la gestione dei dati è supervisionato dalla comunità di Wikidata, che decide inoltre sulle linee guida generali del progetto e su quelle specifiche sui contenuti. In particolare, la comunità gestisce l'importazione di materiale proveniente da fonti affidabili attraverso dei programmi automatici (chiamati *bot*).

5 Per *open source* si intende il software e, per estensione, il movimento che spinge i detentori dei diritti su un software a renderne pubblico il codice sorgente (ossia gli algoritmi alla base di un programma), favorendone il libero studio e permettendo a programmatori indipendenti di apportarvi modifiche ed estensioni.

6 Per *open access* si intende una modalità di pubblicazione del materiale prodotto dalla ricerca (articoli scientifici pubblicati in riviste accademiche o atti di conferenze, ma anche capitoli di libri, monografie, o dati sperimentali) che ne consente accesso libero e senza restrizione, in contrapposizione al modello prevalente di accesso chiuso (per esempio, tramite registrazione a pagamento).

7 Per *open content* si intende un qualsiasi tipo di lavoro o contenuto creativo, pubblicato con una licenza d'uso che permette esplicitamente la copia e la modifica da parte di qualcun altro. Per molti versi, la locuzione è simile a quella di *open source*, con la differenza che la locuzione *open content* può estendersi a qualunque contenuto, anche non di natura informatica.

Come tutti i progetti Wikimedia, Wikidata si basa sul software MediaWiki con l'aggiunta dell'estensione Wikibase⁸, nata nel 2011 da un gruppo di lavoro guidato da Denny Vrandečić e finanziato originariamente da Wikimedia Deutschland (la corrispondente tedesca della Wikimedia Foundation)⁹.

L'idea di fondo dell'estensione Wikibase è quella della tavoletta di cera, ossia di un database privo di un'ontologia predeterminata dall'alto. Come detto, Wikidata raccoglie i dati fondamentali di voci provenienti da 805 progetti Wikimedia, afferenti a differenti famiglie di progetti molto differenti fra loro. Considerando, dunque, che sarebbe stato praticamente impossibile stabilire in anticipo un'ontologia che potesse descrivere una tale varietà di voci e pagine, si è deciso fin dall'inizio di demandare alla comunità stessa di Wikidata tutte le decisioni riguardo quale ontologia adottare, adattare o addirittura creare.

Gli obiettivi di Wikidata

Gli obiettivi di Wikidata sono fondamentalmente quattro:

1) centralizzazione dei link. In passato i link fra voci e pagine dei progetti Wikimedia (chiamati *sitelink* o *interlink*) venivano scritti a mano al loro interno, per poi essere sincronizzati tramite dei programmi automatizzati. Per esempio, per connettere la voce 'Italia' sulla versione italiana di Wikipedia con la voce 'Italy' su quella inglese, alla fine della voce italiana andava inserito il codice `[[en:Italy]]`, che avrebbe poi provveduto a fornire il collegamento al corrispettivo inglese.

Modifica di Unione europea

Attenzione: non sei un utente registrato oppure non hai effettuato l'accesso. Ogni contributo, o Tieni comunque conto che il tuo indirizzo IP (visibile qui) verrà registrato nella cronologia di que

The screenshot shows the edit interface for the Wikidata item 'Unione europea'. It displays a list of interlink code blocks for various languages. The following code blocks are highlighted with an orange border:

```

[[ab:Esponeașun Așunama]]
[[af:Europese Unie]]
[[als:Europäische Union]]
[[am:አድርጋ ህብረት]]
[[an:Unión Europea]]
[[ang:Europiscæ Gesamnung]]
[[ar:الاتحاد الأوروبي]]
[[arc:ܘܨܬܘܬܐ ܘܪܘܨܝܬܐ]]
[[arz:الاتحاد الأوروبي]]
[[ast:Xunión Europea]]
[[az:Avropa İttifaqı]]
[[ba:Espona Bepəşməşe]]

```

Other visible code blocks include:

```

[[Categoria:Unione europea]]
[[Categoria:Premi Nobel per la pace]]
{{Link Vq|de}}
{{Link Vq|en}}
{{Link Vq|es}}
{{Link Vq|fr}}
{{Link Vq|no}}
{{Link Ad|sco}}
{{Link Ad|sw}}
{{Link Ad|ta}}
{{Link Vq|es}}

```

At the bottom of the page, there are buttons for 'Salva la pagina', 'Visualizza anteprima', 'Mostra modifiche', and 'Annulla'.

Figura 1 – Un esempio di *interlink* inseriti con il vecchio metodo (evidenziati in arancione)

8 Wikibase, a sua volta, è l'evoluzione dell'estensione Semantic mediaWiki, che permette l'annotazione semantica dei dati su pagine *wiki* e la loro esportazione in RDF e CSV. Fra gli allora sviluppatori di Semantic mediaWiki, c'erano Markus Krötzsch e Denny Vrandečić, che poi saranno fra i principali sviluppatori di Wikibase.

9 Lungo il primo anno, il team di sviluppo di Wikidata ebbe anche un sostegno finanziario da Google, dall'Allen Institute for Artificial Intelligence e dalla Gordon and Betty Moore Foundation.

Questo sistema ha funzionato per circa dodici anni (dal 2001 al 2013) non senza problemi, soprattutto per quanto riguarda i collegamenti con gli altri progetti. A partire dal 6 marzo 2013, Wikidata è diventata il *repository* centrale di tutti questi link, permettendo ai singoli progetti di richiamarli automaticamente nelle voci;

2) raccolta di dati strutturati. Wikidata, come detto, raccoglie in forma strutturata i dati fondamentali su ciascun argomento trattato dalle voci e pagine dei progetti Wikimedia, oppure su altri argomenti a essi correlati e/o importanti per la loro trattazione. Tali dati sono organizzati in dichiarazioni e sono registrati in un formato *machine-readable*, che permette anche ai computer di leggere e capire le informazioni contenute;

3) creazione di *query* automatiche. In futuro i dati di Wikidata saranno resi disponibili in *query* predefinite, in modo da poterli sfruttare, per esempio, per creare automaticamente nuove voci o per tradurre automaticamente liste di voci. Tuttavia, al momento è già possibile sfruttare un *endpoint* SPARQL¹⁰, così come numerosi altri strumenti che permettono di effettuare *query*, anche complesse, sul database;

4) fornire supporto a parti terze. Come già accennato nell'introduzione, Wikidata viene utilizzata da siti esterni (per esempio Dbpedia, VIAF, MusicBrainz) per lo scambio di dati: per esempio, nel caso di VIAF (che già utilizzava Wikipedia in inglese come fonte dati), lo spostamento verso Wikidata ha permesso di ampliare notevolmente le possibilità di catturare nomi (anche in forme non latine) e di migliorare sensibilmente le proprie sincronizzazioni.

I namespace di Wikidata

Wikidata si divide in vari *namespace*, ossia in diversi gruppi di pagine che hanno ciascuno una propria funzione. Per motivi di spazio e di rilevanza, ci si concentrerà soltanto su due di questi *namespace*: quello principale e quello riguardante le proprietà.

Il namespace principale

Il *namespace* principale (chiamato anche *namespace* zero o ns0) è l'area del sito in cui sono contenuti gli elementi (*item*), ossia le pagine che contengono i dati strutturati, provenienti dalle pagine di Wikipedia e degli altri progetti, su un singolo oggetto, soggetto o concetto. Ciascun elemento, inoltre, raggruppa anche tutti i *sitelink* verso tutti i progetti che contengono una voce o una pagina riguardante quel dato argomento. Tutti gli elementi sono caratterizzati dalla lettera Q seguita da un numero progressivo¹¹ e sono composti di cinque parti:

- 1) una etichetta (per esempio, Unione Europea);
- 2) una descrizione (per esempio, entità politica di carattere sovranazionale e inter-governativo);
- 3) uno o più alias, ossia nomi alternativi (per esempio, UE);
- 4) un certo numero di *sitelink*;
- 5) un certo numero di dichiarazioni.

¹⁰ Per chi fosse interessato, l'*endpoint* è disponibile all'indirizzo <<https://query.wikidata.org/>>.

¹¹ Per esempio, l'elemento dedicato all'Unione Europea (mostrato in Figura 2) è disponibile all'indirizzo <<https://www.wikidata.org/wiki/Q458>>.

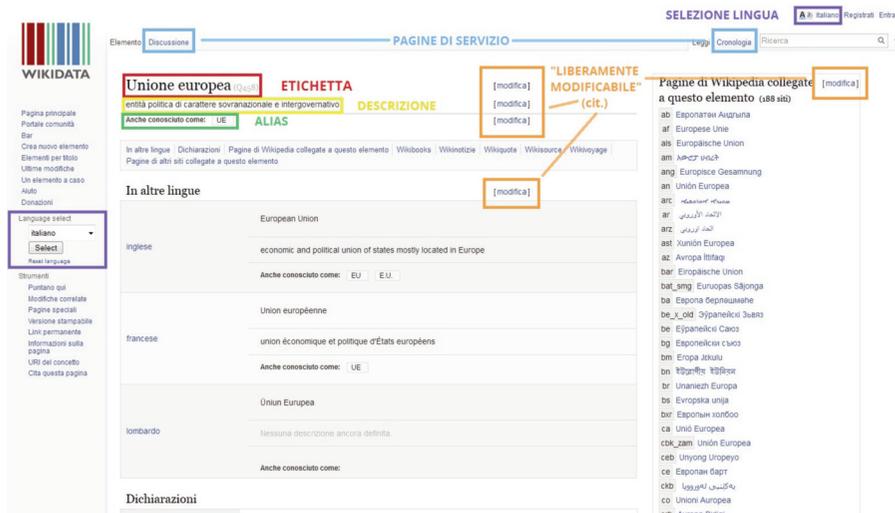


Figura 2 – Un esempio commentato di elemento

Una dichiarazione è una parte dei dati riguardanti un elemento contenuti in una pagina. Essa consiste in un'affermazione (la coppia proprietà-valore come 'Stato: Italia', insieme a uno o più qualificatori eventuali)¹², corredata da riferimenti (ossia le fonti da cui i dati provengono) e da una valutazione (usata per distinguere diverse affermazioni contenute nella stessa proprietà)¹³. Wikidata, come tutti i progetti Wikimedia, non garantisce la correttezza *tout court* di tali dichiarazioni, ma si limita soltanto a raccoglierle e riportarle con un riferimento a una fonte (come un libro, una pubblicazione, un database di autorità e così via), in ossequio al principio cardine di verificabilità.

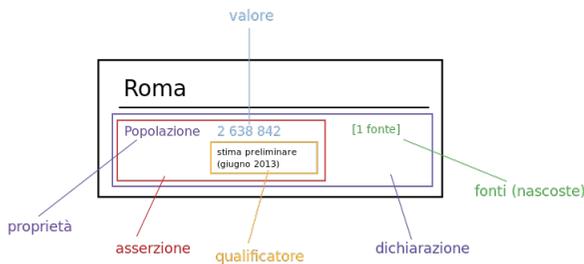


Figura 3 – Un esempio commentato di dichiarazione

12 Un qualificatore è una affermazione che è da ritenersi subordinata all'affermazione principale. Per esempio, per indicare che Sergio Mattarella è il Presidente italiano dal 3 febbraio 2015, Wikidata riporterà le triple 'Capo dello Stato: Sergio Mattarella' (affermazione principale) e 'Data di inizio: 3 febbraio 2015' (qualificatore).

13 Le valutazioni disponibili sono tre: preferibile, che indica un dato più aggiornato rispetto ad altri; normale, il grado di *default* per ogni dato; deprecato, che indica un dato che è da considerare errato oppure obsoleto. In particolare, quest'ultima caratteristica permette contemporaneamente di registrare comunque il dato come relativo a un certo argomento e di segnalarlo pubblicamente come errato o obsoleto.

Il namespace *proprietà*

Una proprietà è un descrittore dei valori, di una relazione, di caratteristiche composite o di un valore eventualmente mancante. Ogni dichiarazione in una pagina di un elemento ha la connessione a una proprietà e le assegna uno o più valori, relazioni, valori composti o, eventualmente, valori mancanti. Tutte le proprietà sono caratterizzate dalla lettera P seguita da un numero progressivo¹⁴ e sono composte in maniera molto simile agli elementi, possedendo anche loro un'etichetta, una descrizione, uno o più alias e una o più dichiarazioni descrittive della proprietà.

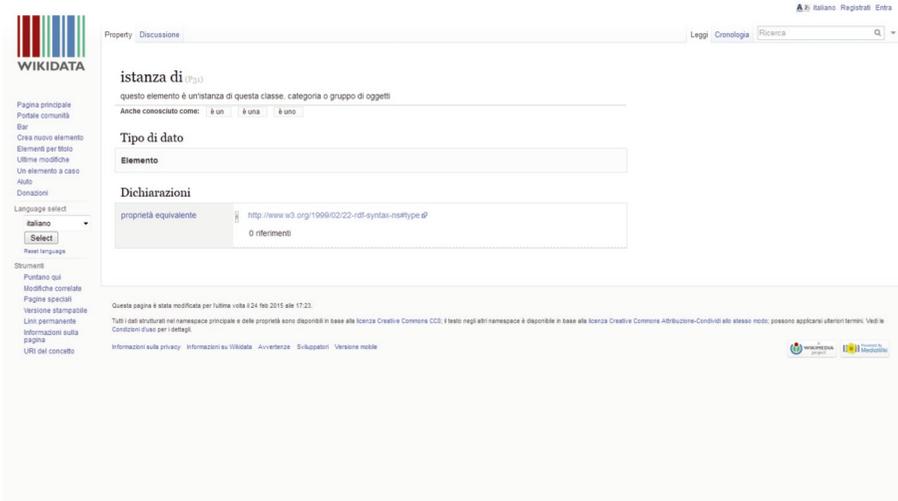


Figura 4 – Un esempio di proprietà

Esistono attualmente dieci tipologie di proprietà, a seconda di quelli che sono i valori da inserire nell'affermazione, che possono essere: collegamenti a elementi; stringhe alfanumeriche; immagini da Wikimedia Commons; coordinate temporali; coordinate geografiche; indirizzi URI; quantità; quantità con unità di misura; stringhe mono-lingua; stringhe multi-lingua. Ad oggi sono state create più di 2.200 proprietà che coprono un'impressionante varietà di dati, dalle date di nascita alle presenze di un giocatore in un torneo, dal nome ufficiale di un ente nella sua lingua originaria agli identificatori di un determinato soggetto in un tesoro o in un database di autorità¹⁵.

Lo stato di Wikidata in relazione ai progetti Wikimedia

Wikidata, come accennato all'inizio, è nata ufficialmente il 29 ottobre 2012. Alla data del 22 gennaio 2016, sono stati creati più di 15,8 milioni di elementi, di cui poco più del 52% presenta tre o più dichiarazioni. In totale, sono state create più di 70 milioni di dichiarazioni e si registrano oltre 294 milioni di modifiche complessive, rendendo così Wikidata il terzo progetto Wikimedia più attivo, dopo Wikipedia in inglese e Wikimedia commons.

¹⁴ Per esempio, la proprietà 'istanza di' (mostrata in Figura 4) è disponibile all'indirizzo <<https://www.wikidata.org/wiki/P31>>.

¹⁵ Per navigare fra le proprietà, si veda <https://www.wikidata.org/wiki/Wikidata:List_of_properties>.

I primi dati a essere riutilizzati da Wikidata sono stati i collegamenti fra progetti: nei primi tre mesi del 2013¹⁶, le 291 versioni esistenti di Wikipedia hanno progressivamente abbandonato il vecchio sistema di gestione dei *sitelink*, portando alla rimozione di oltre 300 milioni di righe di codice (ormai inutile).

A oggi le famiglie di progetti attualmente supportate sono, in ordine di abilitazione: Wikipedia (enciclopedia), Wikivoyage (raccolta di guide turistiche), Wikimedia commons (il più grande *repository* di file multimediali liberi al mondo, con circa 30 milioni di file), Wikisource (biblioteca di testi editi in pubblico dominio o in licenza libera), Wikiquote (raccolta di aforismi e citazioni), Wikinews (sito di *citizen journalism*), Wikibooks (biblioteca di manuali creati da volontari), Wikispecies (tassonomia libera di tutte le forme di vita esistenti) e Meta-Wiki (progetto di coordinamento per la comunità wikimediana).

Dal 27 marzo al 25 aprile 2013, inoltre, è stato abilitato in più fasi anche il riutilizzo dei dati strutturati su Wikipedia e, successivamente, anche sugli altri progetti Wikimedia. A partire dal 16 settembre 2015, è stato abilitato anche il cosiddetto accesso arbitrario ai dati: in altri termini, una voce o pagina può accedere ai dati del proprio elemento collegato, così come a quelli di un altro elemento, laddove necessario¹⁷.

In totale, i progetti a oggi supportati sono 617, compresi Wikidata stessa e il sito ufficiale di MediaWiki. Mancano ancora all'appello le 16 versioni linguistiche di Wikiversity (raccolta di corsi universitari generati dagli utenti) e le 172 versioni di Wikizionario (dizionario libero); per le prime l'integrazione avverrà nel corso del 2016, mentre per le seconde è allo studio un progetto apposito di integrazione, che difficilmente vedrà la luce prima del 2017¹⁸.

Il riutilizzo dei dati sui progetti Wikimedia

I principali riutilizzatori dei dati di Wikidata sono, ovviamente, i progetti Wikimedia: nel corso di questi due anni, si è assistito a una lenta e inesorabile trasformazione dei *template*¹⁹, anche attraverso l'implementazione di un nuovo linguaggio di programmazione chiamato Lua, che garantisce una maggiore facilità di riutilizzo dei dati. Solo sulla versione italiana di Wikipedia, al momento, sono 117 i *template* che utilizzano dati provenienti da Wikidata²⁰, un *trend* in costante crescita.

I vantaggi di questa centralizzazione dei dati sono molteplici: innanzitutto, si riduce notevolmente il peso per la comunità di volontari dell'aggiornamento di

16 La prima versione di Wikipedia a sperimentare l'accesso centralizzato ai link fu la versione ungherese il 14 gennaio 2013, seguita dalle versioni in ebraico e in italiano il 30 gennaio, dalla versione in inglese il 15 febbraio e da tutte le altre 287 versioni di Wikipedia il 6 marzo.

17 A causa della complessità del progetto, Wikimedia commons è l'unico progetto a oggi che ha abilitato il solo accesso ai *sitelinks*, ma non ai dati strutturati. Per chi è interessato a visionare la discussione sul come permettere il riutilizzo dei dati strutturati, è disponibile (in inglese) all'indirizzo <https://www.wikidata.org/wiki/Wikidata:Wikimedia_Commons/Development>.

18 Le varie proposte di integrazione sono disponibili (in inglese) all'indirizzo <<https://www.wikidata.org/wiki/Wikidata:Wiktionary/Development>>.

19 Un *template* è una pagina speciale creata per essere inclusa in altre pagine, in base al meccanismo dell'inclusione, per diversi scopi, come per esempio generare delle tabelle sinottiche standard (chiamate *infobox*) o delle tabelle di navigazione per argomento (chiamate *navbox*).

20 Per una lista aggiornata, si veda <https://it.wikipedia.org/wiki/Categoria:Template_che_usano_dati_di_Wikidata>.

determinate informazioni. La standardizzazione riduce, inoltre, il rischio di discrepanze nei dati fra versioni, permettendo al contempo ai volontari di potersi concentrare solo sui casi che effettivamente presentano problemi risolvibili con un intervento umano. Un altro vantaggio è dato, per esempio, dalla possibilità di poter creare nuovi abbozzi di voce, chiamato informalmente *stub*, contenenti le informazioni fondamentali riguardo un dato argomento.

Questi vantaggi diventano ancora più evidenti nel caso dei cosiddetti progetti minori, ossia di quelle comunità che non dispongono di un sufficiente numero di volontari atto a garantire un certo tasso di crescita. La riduzione della manutenzione spicciola e la possibilità di concentrarsi sulla generazione di valore aggiunto, sotto forma di traduzioni o adattamenti di altre voci o di creazione di nuove voci, è infatti uno degli obiettivi che fin dall'inizio è stato messo in conto durante lo sviluppo di Wikidata.

Uno sviluppo assolutamente inaspettato, che si è venuto man mano palesando fin dai primi mesi del 2013, è stato quello relativo ai codici di autorità: Wikidata conta oggi almeno 40 proprietà che fanno riferimento a codici di autorità di enti internazionali (VIAF, ULAN, ISNI) o dei servizi bibliotecari nazionali, come la Library of Congress, la BNF, il GND e, ovviamente, SBN. In particolare, quest'ultima è richiamata da 16.266 elementi di Wikidata (su 64.502 record di autorità giudicati di livello AUF). Con riferimento a quest'ultimo punto, ICCU e Wikimedia Italia stanno conducendo una collaborazione per migliorare la sincronizzazione fra SBN e altri cataloghi e i progetti Wikimedia²¹.

A questi si aggiungono poi numerosi identificatori di database nazionali e internazionali, pubblici e privati, riguardanti ambiti che spaziano dai monumenti allo sport, dalla politica ai tesauri. In particolare, vale la pena segnalare come l'Italia sia presente tramite il *Thesaurus* della Biblioteca nazionale centrale di Firenze, anch'esso oggetto di una collaborazione in corso con l'Associazione Wikimedia Italia²².

Altri esempi di riutilizzo e visualizzazione dei dati al di fuori dei progetti Wikimedia

In aggiunta ai già citati casi di riutilizzo dei dati da parte di progetti terzi, esistono vari altri esempi di *tool* – spesso generati dagli stessi utenti di Wikidata – che danno l'idea di come questo progetto, e ancora di più i suoi dati, possano essere sfruttati per generare nuova conoscenza.

Fra i *tool* di pura visualizzazione dei dati, i primi a essere sviluppati sono stati Reasonator²³, una sorta di Wikidata graficamente più graziosa, e Wikidata temporal display²⁴, un *tool* che permette di visualizzare contemporaneamente su una cartografia e su una *timeline* una determinata serie di eventi (come, per esempio, le differenti fasi di una guerra). Un'evoluzione di questa categoria di strumenti è data, per esempio, da Histropedia²⁵, dove chiunque può scegliere fra oltre 300.000 percorsi tematici, oppure crearne uno proprio «giocando ai Lego con la Storia», come recita lo slogan del sito.

21 Per maggiori informazioni, si veda <<https://it.wikipedia.org/wiki/Progetto:GLAM/ICCU>>.

22 Per maggiori informazioni, si veda <https://it.wikipedia.org/wiki/Progetto:GLAM/Biblioteca_Nazionale_Centrale_di_Firenze>.

23 Disponibile all'indirizzo <<https://tools.wmflabs.org/reasonator/?>>.

24 Disponibile all'indirizzo <https://tools.wmflabs.org/wikidata-todo/tempo_spatial_display.html?>.

25 Disponibile all'indirizzo <<http://histropedia.com/>>.

Il primo vero *tool* che però ha dimostrato le potenzialità di Wikidata è stato Sum of all paintings, poi trasformatosi in un vero e proprio progetto²⁶ di Wikidata²⁷, dove viene coordinato il lavoro sugli elementi relativi ai dipinti e ai quadri realizzati da un artista famoso e/o detenuti da un museo o un'istituzione famosa. Al momento, il progetto ha superato le 100.000 opere descritte, anche se il lavoro di approfondimento continua a non mancare.

Un *tool* molto simile, stavolta per appassionati bibliofili, è Inventaire²⁸, dove gli utenti possono da un lato mappare la propria biblioteca, inserendo i metadati relativi all'edizione in loro possesso di un libro, e dall'altro scambiare queste informazioni con persone vicine a loro. Ovviamente, anche Inventaire sfrutta i dati provenienti da Wikidata in maniera biunivoca.

Il dialogo ha avuto successo anche in ambito scientifico: verso la fine di ottobre 2014, alcuni giovani ricercatori tedeschi hanno completato l'importazione di tutti i dati dello Human Genome Project all'interno di Wikidata, allo scopo di mettere in correlazione i due database e, tramite Wikidata, mostrare questi dati direttamente su Wikipedia²⁹. Anche sulla scorta di questi risultati, Wikidata intende ulteriormente questo discorso attraverso un ambizioso progetto chiamato Wikidata for research, che si pone lo scopo di creare un ambiente virtuale di ricerca dove la comunità scientifica con le comunità che si occupano di *open data* e *open knowledge*³⁰ possano collaborare.

In ultimo, la comunità di Wikidata ha creato dei *tool* per semplificare quello che i wikimediani chiamano lavoro sporco, ossia quel lavoro di riordino e manutenzione dei contenuti che necessita dell'impegno e dell'attenzione di un umano. Larga parte di questi strumenti la si deve a Magnus Manske, programmatore tedesco che è possibile considerare una delle colonne portanti della comunità wikimediana internazionale.

In particolare, fra i suoi principali contributi c'è il Wikidata game³¹, uno strumento che permette di affrontare dei compiti importanti (come il determinare se un certo elemento riguarda o meno un essere umano) con la semplicità di un *click*. Il Wikidata game è poi evoluto nel Distributed game³², che raggruppa altri *tool* (fatti sia da Manske, sia da altri utenti wikimediani) che condividono fundamentalmente lo spirito del gioco iniziale.

26 Un progetto è un luogo di coordinamento tra utenti che discutono come gestire le informazioni di uno specifico ramo della conoscenza: come devono essere formattate, cosa devono comprendere, in che ordine devono essere esposte, eccetera.

27 Disponibile all'indirizzo <https://www.wikidata.org/wiki/Wikidata:WikiProject_sum_of_all_paintings>.

28 Disponibile all'indirizzo <<https://inventaire.io>>.

29 Jens Ohlig, *Establishing Wikidata as the central hub for linked open life science data*, «Wikimedia Deutschland Blog», 22 ottobre 2014, <<http://blog.wikimedia.de/2014/10/22/establishing-wikidata-as-the-central-hub-for-linked-open-life-science-data/>>.

30 Il progetto è disponibile all'indirizzo <https://www.wikidata.org/wiki/Wikidata:WikiProject_Wikidata_for_research>.

31 Disponibile all'indirizzo <<https://tools.wmflabs.org/wikidata-game/>>.

32 Disponibile all'indirizzo <<https://tools.wmflabs.org/wikidata-game/distributed/>>.

Infine, sempre a opera di Manske, c'è Mix'n'match³³, un *tool* che contiene varie liste riguardo vari argomenti provenienti da istituzioni e fonti affidabili (come il Servizio bibliotecario nazionale, l'*Oxford dictionary of National Biography*, l'*Australian dictionary of biography*, il catalogo della National Portrait Gallery e così via). Tramite questo *tool*, si possono abbinare gli elementi di queste liste agli elementi di Wikidata, identificando quali di questi esistono sui progetti Wikimedia e quali no.

Conclusioni

Wikidata è ancora lontana dall'essere perfetta, al pari di qualsiasi altro progetto Wikimedia. Tuttavia, lo sforzo infuso in questi tre anni dai suoi volontari ha permesso di rivelare che l'intuizione del gruppo di lavoro iniziale era corretta, che era possibile creare una tavoletta di cera senza ontologia e potenzialmente infinita che potesse raccogliere e strutturare i dati fondamentali delle voci e delle pagine degli oltre 800 progetti Wikimedia.

Questo risultato, ovviamente, non avrebbe potuto essere raggiunto senza l'aiuto (diretto e indiretto) delle istituzioni culturali che hanno man mano pubblicato il proprio lavoro su Internet, permettendo ad appassionati e dilettanti³⁴ di renderlo qualcosa di più di un semplice catalogo di dati, utilizzandolo per integrare e correggere il lavoro svolto dai wikimediani.

Wikidata è, dunque, un buon punto di incontro fra l'anarchica comunità wikimediana e le comunità di studiosi, bibliotecari e scienziati, un luogo virtuale dove le due parti possono riconoscere le tante similitudini e lavorare insieme per diffondere conoscenza. Tutto ciò è ancora più vero se si tiene a mente che la visibilità e la diffusione di dati affidabili e verificati da una fonte ufficiale è oggi più che mai importante, data la facilità con cui è possibile reperire qualsiasi tipo di informazione sulla Rete e considerando la presenza di tanti, troppi segnali di disturbo causati da un utilizzo non sempre consapevole dello strumento.

Articolo proposto il 19 dicembre 2015 e accettato il 22 gennaio 2016.

ABSTRACT

AIB studi, vol. 56 n. 1 (gennaio/aprile 2016), p. 75-85. DOI 10.2426/aibstudi-11434

LUCA MARTINELLI, Wikimedia Italia – Associazione per la diffusione della conoscenza libera, via Bergamo 18, 20900 Monza, e-mail: luca.martinelli@wikimedia.it.

Wikidata: la soluzione wikimediana ai linked open data

Wikidata è il progetto più giovane fra quelli gestiti dalla Wikimedia Foundation, la fondazione non-profit che gestisce Wikipedia e i suoi progetti fratelli. Il suo scopo principale è raccogliere e strutturare i dati fondamentali delle voci e delle pagine degli oltre 800 progetti Wikimedia, in modo tale che possano essere letti, tradotti, modificati e riutilizzati da chiunque (macchine comprese) in ciascuna delle 285 lingue ufficiali

³³ Disponibile all'indirizzo <<https://tools.wmflabs.org/mix-n-match/>>.

³⁴ Il termine dilettanti è qui da intendere nel suo etimo originale, ossia di persona che si diletta nello svolgere un'attività.

dei progetti Wikimedia. In questo articolo si affrontano i vari aspetti di Wikidata: si analizzano le caratteristiche tecniche e teoriche del progetto; si elencano gli obiettivi fondamentali; ci si concentra sulle nozioni e il significato, fondamentali per Wikidata, di 'elemento' e 'proprietà'; si analizza lo stato di Wikidata a tre anni dal lancio, sia in relazione ai progetti Wikimedia, sia in relazione a parti terze; infine, si propone una panoramica di tutti i principali tool di riutilizzo dei dati di Wikidata finora ideati.

Wikidata: Wikimedia's answer to linked open data

Wikidata is the latest project run by Wikimedia Foundation, the non-profit organization sustaining Wikipedia and several related projects. The aim of Wikidata is to collect and structure the key data of every article and page of the 800+ Wikimedia projects, so that they can be read, translated, edited and re-used by anyone (including machines) in every official language of Wikimedia projects. The paper examines Wikidata's technical and theoretical features; its main objectives; the essential notion and meaning of 'item' and 'property'; the state of the art three years after its launch – with regard to Wikimedia projects as well as to third parties; and finally it provides an overview of the main tools created so far for the reuse of Wikidata's data.