

Archivi, biblioteche e musei toscani in Wikidata

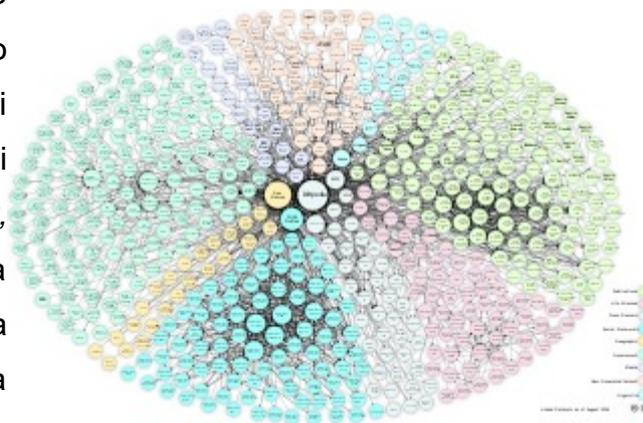
Susanna Giaccai

Nell'autunno 2019 Francesca Navarria (Direzione cultura e ricerca di Regione Toscana)¹ ha proposto a un gruppo di wikimediani toscani di portare in Wikidata il contenuto della banca dati *Sistema cultura*, l'anagrafica delle istituzioni e dei luoghi della cultura presenti in Toscana². Si tratta di un *repository* di raccolta di dati anagrafici e di servizio nato nel 2008, inizialmente per gestire le informazioni sui musei ed esteso successivamente anche ad archivi e biblioteche; il data base è stato aggiornato annualmente direttamente dagli operatori di archivi, biblioteche e musei toscani attraverso uno specifico modulo web con accesso vincolato da password. L'archivio, gestito con software proprietario, è collocato nei server di Regione Toscana ed è utilizzato internamente anche per fini amministrativi. La scelta di portare questi dati in un ambiente aperto è coerente con la politica open access di Regione Toscana; in questo caso specifico la scelta ha anche l'obiettivo di favorire una manutenzione capillare e tempestiva dei dati, affidandola direttamente agli stessi istituti del territorio attraverso l'uso di un archivio di dati aperti facile da aggiornare. La scelta ha anche l'obiettivo di rendere i dati di archivi, biblioteche e musei toscani molto più visibili nel Web. Wikidata è infatti il maggior contenitore mondiale di dati ed è il cuore del mondo dei linked open data e del Web semantico.

¹ Responsabile del settore Promozione, comunicazione e valorizzazione dei patrimoni e servizi delle reti documentarie toscane (biblioteche e archivi) e della lettura, Regione Toscana.

² I soci di Wikimedia Italia, che hanno lavorato all'iniziativa, sono stati oltre alla sottoscritta anche Alessandro Marchetti, Camillo Pellizzari e Manuela Musco (Wikimediani di lunga data).

Wikidata, la più recente delle piattaforme Wikimedia, è nata nel 2012 con l'obiettivo primario di unificare i collegamenti tra le voci enciclopediche nelle diverse versioni linguistiche di Wikipedia, chiamati *interlink*, fino a quel momento creati a mano; si trattava di un'operazione complessa e soggetta naturalmente a molti errori. Con Wikidata invece si ha una centralizzazione dei dati che velocizza l'operazione e soprattutto ne garantisce la coerenza.



Negli anni successivi oltre ai dati strutturati provenienti da Wikipedia e dagli altri progetti Wikimedia, sono stati caricati in Wikidata un numero considerevole di risorse provenienti da archivi istituzionali di qualità. Tra questi i dati dell'authority file VIAF, identificativi, riferimenti e riferimenti biografici dalla Biblioteca nazionale di Francia, il thesaurus del British Museum e di moltissimi altri enti. Ad oggi Wikidata contiene oltre 90 milioni di dati, denominati Elementi, con licenza Creative Commons CC 0, sono quindi di pubblico dominio; chiunque può incrementare questo archivio e tutti i dati in esso contenuti possono essere riutilizzati senza limitazione alcuna.

L'operazione proposta da Regione Toscana era molto interessante perché aveva lo scopo di importare nel cuore del Web tre archivi omogenei di istituti culturali presenti sul suo territorio; d'altra parte essendo il primo caso in Italia di caricamento di dati di una Regione in Wikidata, l'operazione si presentava complessa non avendo esempi di riferimento. La complessità risiedeva nel dover trattare tre insiemi di dati relativi a tipologie diverse di istituti (archivi, biblioteche e musei) caricati nel database regionale in tempi differenti e a

cura di uffici distinti. È stato per questo svolto un approfondito lavoro di analisi dei circa 3.000 dati forniti e di confronto con le proprietà disponibili in Wikidata per descrivere archivi, biblioteche e musei. Un lavoro minuzioso è stato necessario per verificare la qualità dei dati e per selezionare, di comune accordo con gli uffici competenti di Regione Toscana, quali dati trasferire. È stato deciso di non caricare informazioni che negli anni sono state soggette a frequenti modifiche (ad esempio gli URL del sito web della biblioteche, le e-mail di riferimento ecc.) e di rimandare l'inserimento a una fase successiva del progetto, che prevedeva un corso di formazione per i referenti delle Reti documentarie e museali organizzato con AIB Toscana. Il compito di mantenere aggiornati i dati veniva quindi delegato ai centri rete o agli archivi, alle biblioteche e ai musei interessati a gestire direttamente le proprie risorse in Wikidata.

L'attività di caricamento in Wikidata è stata complessa perché il risultato cui si puntava era la creazione di un insieme di dati di qualità. Frequenti sono infatti in questo data base importazioni massive, effettuate con scarsa attenzione e analisi della banca dati di provenienza. Nel 2018 sono stati importati i dati degli edifici storici tratti dalle banche dati del Ministero dei beni e delle attività culturali e del turismo (MIBACT),³ ma la scarsa attenzione prestata alla complessità della struttura di questi dati, provenienti dal cumulo di archivi diversi, ha portato alla creazione in Wikidata di molti elementi scarsamente identificabili: per es. centinaia di San Francesco o di San Giuseppe sono arrivati in Wikidata, privi di un termine qualificatore che identificasse con precisione l'oggetto, specificando se si trattasse di una chiesa, di una scuola o di un convento. Solo un successivo lavoro certosino su ogni singolo dato, con la aggiunta di un qualificatore (chiesa o abazia o convento ecc.), ha potuto renderli attendibili e riusabili. Un altro esempio di bassa qualità nel lavoro di importazione dei dati è quello svolto nei primi mesi del 2020 da un gruppo di wikimediani svedesi, che ha trasferito in Wikidata migliaia di dati estratti da *Archives portal Europe*,⁴ un aggregatore europeo con dati provenienti da centinaia di fonti diverse. Il risultato è stato ovviamente la creazione di diverse migliaia di elementi con nomi improbabili, indirizzi errati, geolocalizzazione approssimativa o errata

3 Ad oggi, Ministero della cultura (MiC).

4 <tinyurl.com/5bh2rt6n>.

ecc. Nel maggio 2013, un'altra importazione di massa ha portato in Wikidata migliaia di scuole italiane prese dagli archivi del MIUR⁵; anche in questo caso l'errore macroscopico è stata l'assegnazione al dato importato di un nome troppo generico e ambiguo per Wikipedia, seppur nel contesto del data base d'origine avesse avuto una propria significatività. Anche in questo caso, Wikidata si è così popolato di elementi con nome San Francesco, Giovanni Pascoli, Dante Alighieri relativi una volta a una scuola un'altra volta a una chiesa o a una piazza senza che ciò sia evidente dal nome. Negli anni '80 quando le biblioteche iniziavano a lavorare sullo scambio dei record bibliografici, tra i bibliotecari era nota l'espressione *garbage in garbage out* (se importi dati sporchi la ricerca darà risultati di bassa qualità)⁶.

Per evitare di ripetere questi errori, il gruppo di wikimediani toscani ha svolto un lavoro attento di analisi dei dati di provenienza durato diverse settimane, verificando spesso direttamente con l'istituto locale la loro attendibilità. Una volta completata l'importazione, i dati sono ora riutilizzabili in diverse modalità di visualizzazione.

Lista di alcune biblioteche sensi

 wd:Q86537223	Biblioteca comunale Ranuccio Bianchi Bandinelli	biblioteca toscana sita nel comune di Castelnuovo Berardenga e parte di REDOS - Rete documentaria senese	Castelnuovo Berardenga
 wd:Q86537186	Biblioteca comunale San Giovanni d'Asso - Montalcino	biblioteca toscana sita nel comune di Montalcino e parte di REDOS - Rete documentaria senese	Montalcino
 wd:Q86538386	Biblioteca comunale di San Gimignano	biblioteca del Comune di San Gimignano	San Gimignano
 wd:Q86537226	Biblioteca comunale di Monteriggioni	biblioteca toscana sita nel comune di Monteriggioni e parte di REDOS - Rete documentaria senese	Monteriggioni
 wd:Q86537219	Biblioteca comunale di Trequanda	biblioteca toscana sita nel comune di Trequanda e parte di REDOS - Rete documentaria senese	Trequanda
 wd:Q86537191	Biblioteca comunale Angelo Ferrazzani	biblioteca toscana sita nel comune di Piancastagnaio e parte di REDOS - Rete documentaria senese	Piancastagnaio

come mappa

⁵ Cfr. la mia richiesta di chiarimenti e la risposta dell'ideatore del bot che ha effettuato il caricamento delle scuole <tinyurl.com/kje2543k>.

⁶ Cfr. Wikipedia, voce: *Garbage in, garbage out* <tinyurl.com/kzmxr4hx>.



biblioteche senesi



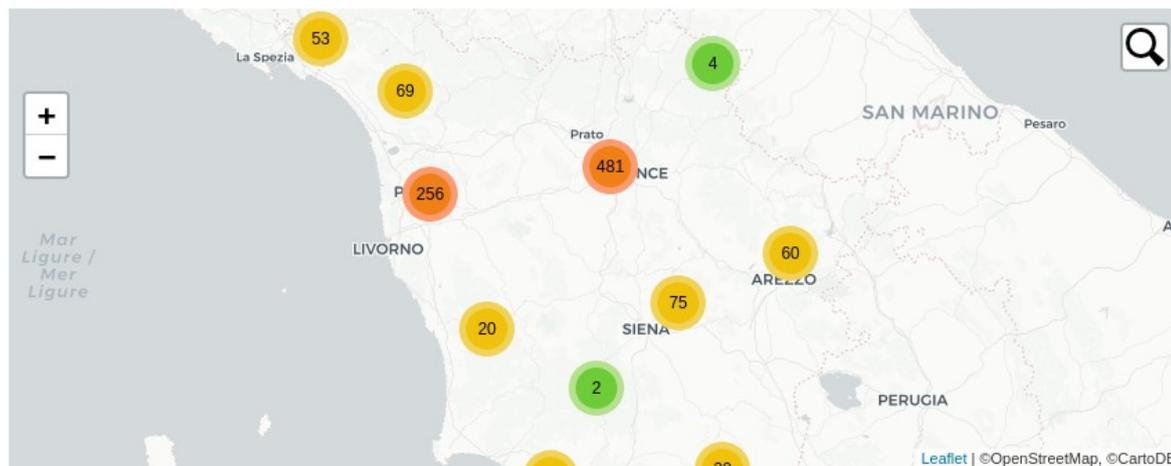
biblioteche toscane

Aumentando il livello descrittivo delle biblioteche, con ulteriori informazioni quali anno di fondazione, dimensione del patrimonio, utenti annuali, prestiti annuali ecc., i tipi di visualizzazioni posso moltiplicarsi. La particolare facilità di aggiornamento immediato dei dati e la loro granularità rende lo strumento Wikidata estremamente interessante.

In Toscana questi dati sono già stati riutilizzati come fonte di un servizio dal catalogo collettivo BiblioToscana per localizzare le biblioteche, facilitare l'accesso ai loro cataloghi e al sito web⁷. Affidare alle singole Reti e alle biblioteche la possibilità di mantenere aggiornate le proprie informazioni in tempo reale, le responsabilizza rispetto alla propria presenza nel Web e favorisce sicuramente un'elevata qualità informativa del servizio.

7 Cfr. BiblioToscana, <<https://biblio.toscana.it/>>

LE BIBLIOTECHE DELLA TOSCANA



Apri l'elenco di tutte le biblioteche della Toscana

BiblioToscana. Mappa delle biblioteche

Qualsiasi modifica avvenga nell'indirizzo web della biblioteca o nella sua e-mail, in tempi brevissimi gli operatori possono intervenire con facilità su Wikidata e aggiornare il dato.

Wikidata offre alle biblioteche anche altre possibilità, come caricare nel data base la lista dei beni culturali del proprio territorio, facilitandone la visibilità anche a scopo turistico, grazie alle proprie competenze e alle fonti presenti nella Sezione locale. Quanto più i beni culturali sono visibili come open data nel Web, tanto più si favorisce la creazione autonoma da parte degli utenti di percorsi turistico-tematici e la valorizzazione di risorse meno conosciute⁸.

Nel 2018, per favorire la collaborazione delle biblioteche a Wikidata, con la sezione AIB Toscana ho organizzato 2 corsi su Wikidata e, nella primavera 2020, il corso rivolto ai centri rete documentari e museali per garantire la manutenzione dei dati degli istituti estratti dal *Sistema cultura* e caricati in Wikidata. La comunità dei volontari ha avviato diversi progetti in Wikidata per promuovere i beni culturali. Nel mondo anglosassone segnalò il Progetto Wikidata:WikiProject cultural heritage e in Italia un progetto simile promosso dalla associazione Wikimedia Italia. Recentemente si è anche costituito il Gruppo Wikidata per Musei, Archivi e Biblioteche, promosso da Carlo Bianchini

⁸ Cfr. Susanna Giaccari, [Wiki loves monuments: il concorso fotografico del mondo wiki](#), *Bibelot*, 25, (2019), n. 3.

dell'Università di Pavia assieme ad altri bibliotecari e wikimediani, che sta lavorando sulle problematiche relative all'inserimento di dati bibliografici in Wikidata⁹.

La biblioteca, con le sue competenze nella organizzazione della conoscenza, può garantire la creazioni in Wikidata di informazione di qualità e acquisire così autorevolezza nell'ecosistema della conoscenza aperta.

Susanna Giaccai
volontaria wikimediana
susanna.giaccai@gmail.com

⁹ <<https://tinyurl.com/c3s8ndum>>