

## Elena Cencetti, Camillo Carlo Pellizzari di San Girolamo, Elisabetta Viti\*

Termini, dati e collegamenti:

'conversazioni' tra il Thesaurus del *Nuovo soggettario* e Wikidata

### 1. Cenni storici: il Thesaurus del *Nuovo soggettario* e i progetti Wikimedia

#### 1.1 Il Thesaurus del *Nuovo soggettario* e Wikipedia in italiano

Il progetto di interoperabilità tra il Thesaurus del *Nuovo soggettario* (di seguito Thesaurus NS) e Wikidata (di seguito WD) si inserisce nell'ambito di una più ampia collaborazione tra la Biblioteca nazionale centrale di Firenze (di seguito BNCF)<sup>1</sup> e l'associazione Wikimedia Italia, istituzionalizzata nell'ottobre 2013 con una lettera di intenti in cui si sono definiti gli obiettivi di cooperazione.

Il *Nuovo soggettario*, curato dalla BNCF, è lo strumento ufficiale impiegato per indicizzare il contenuto delle opere di ogni campo disciplinare e di qualsiasi natura (testi, immagini, suoni, ecc.) su supporti vari (cartacei, digitali, ecc.) nell'ambito di biblioteche, archivi, mediateche, ecc. In particolare, il suo Thesaurus<sup>2</sup> è la componente terminologica del sistema che costituisce un 'ponte semantico' con i differenti patrimoni di molteplici *database* dell'eredità culturale.

È proprio in questa panoramica che nel 2005 si inserisce il primo colloquio semantico tra il Thesaurus NS e Wikipedia, cioè quando nel campo "Fonte" del termine 'RFID <Radio frequency identification>' del Thesaurus NS è stato inserito un riferimento statico a Wikipedia in italiano (la sigla Wikipedia(IT))<sup>3</sup>.

Nel 2007 il Thesaurus NS è stato il primo thesaurus multidisciplinare, a livello internazionale, ad aver attivato dei *deeplink* verso Wikipedia tramite il formato SKOS/RDF<sup>4</sup>. In particolare, sono stati creati collegamenti alle voci di Wikipedia in italiano e di Wikipedia in inglese (citate nel campo "Fonte" dei termini con le sigle Wikipedia(IT) e Wikipedia(ENG)), cominciando a considerare Wikipedia non più solo come fonte per l'attestazione della forma e del significato dei termini, ma anche come link esterno. Negli anni successivi la BNCF ha continuato ad aggiungere manualmente ai termini del Thesaurus NS *deeplink* alle voci di Wikipedia, che nell'ottobre 2012 erano 9905.

Il 2013 è stato un anno particolarmente significativo per la collaborazione tra BNCF e Wikimedia Italia, poiché ha portato alla sottoscrizione di una lettera d'intenti, a un incremento rilevante dei *deeplink* dal Thesaurus NS a Wikipedia in italiano (di seguito itWP) e all'implementazione di *crosswalk* da itWP al Thesaurus NS.

Nei primi mesi del 2013 la BNCF ha cominciato a usare l'*endpoint* SPARQL della DBpedia Italiana<sup>6</sup> (questo progetto, realizzato tra l'aprile 2012 e l'aprile 2019, aveva lo scopo di estrarre dati semistrutturati dalle voci di itWP e renderli disponibili come dati strutturati, ispirandosi a quanto fatto a partire dal 2007 da DBpedia<sup>7</sup>) per abbinare i termini del Thesaurus NS e le voci di itWP in modo automatico, con una successiva revisione manuale: in questo modo i *deeplink* verso itWP sono cresciuti molto velocemente, arrivando a quota 12.350 nell'ottobre 2013 e poi a quota 13.530 nell'aprile 2014<sup>8</sup>.

## Filologia

Da Wikipedia, l'enciclopedia libera.



Fig. 1. Un esempio di link da itWP al Thesaurus NS, nella voce 'Filologia'.

Il 12 maggio 2013 itWP si è dotata di un *template* Thesaurus BNCF<sup>9</sup> che consentiva di inserire, nella sezione Collegamenti esterni delle voci, un link al corrispettivo termine del Thesaurus NS.

Da questo momento i link tra il Thesaurus NS e itWP sono diventati reciproci, con un notevole incremento delle visite del Thesaurus<sup>10</sup>. Questo *template* è stato poi cancellato il 26 marzo 2020, quando i link a termini del Thesaurus NS sono stati spostati all'interno del *template* Controllo di autorità<sup>11</sup> posto al di sotto della sezione Collegamenti esterni (fig. 1).

Come già accennato, l'inizio ufficiale della collaborazione tra la BNCF e Wikimedia Italia è stato sottoscritto nel 2013 con una lettera di intenti, ampliata e integrata da due successivi rinnovi nel 2014 e nel 2017<sup>12</sup>, con l'obiettivo di "proseguire la collaborazione per lo sviluppo del colloquio fra Wikipedia e il Thesaurus del *Nuovo soggettario* e per individuare ulteriori strategie nell'ambito del web semantico" (in particolare l'interoperabilità tra Thesaurus NS e Wikidata)<sup>13</sup>.

## 1.2 Il Thesaurus del Nuovo soggettario e Wikidata

WD è un progetto Wikimedia nato il 29 ottobre 2012<sup>14</sup>. È una base di conoscenza libera, che raccoglie dati strutturati riguardo a tutte le entità (argomenti, concetti, oggetti, persone, luoghi, eventi, ecc.) aventi una voce in almeno una versione di Wikipedia e, più in generale, riguardo a qualsiasi entità che soddisfi almeno uno dei criteri di rilevanza<sup>15</sup>.

In WD ogni entità, denominata elemento (*item*), è identificata univocamente dalla lettera Q seguita da un numero crescente assegnato in automatico dalla piattaforma (es. Q1); per ciascuna lingua gli elementi possono avere un'etichetta (*label*, ossia il nome preferito per l'entità), una descrizione (*description*, ossia una breve descrizione/definizione per l'entità) e uno o più *alias* (ossia ulteriori nomi con i quali l'entità è indicata nelle fonti), una o più dichiarazioni (*statements*, ossia coppie proprietà-valore, potenzialmente dotate di uno o più qualificatori e di uno o più riferimenti), uno o più identificativi esterni (*identifiers*, ossia collegamenti a siti web esterni ai progetti Wikimedia) e uno o più *sitelink* (ossia collegamenti a pagine dei progetti Wikimedia)<sup>16</sup>.

A partire dal 2013, ogni voce di itWP è collegata a un elemento di WD; le nuove voci di itWP vengono collegate periodicamente a WD, manualmente o tramite bot. Il 12 maggio 2013 è stata creata in WD la proprietà P508 (identificativo Thesaurus BNCF) per collegare gli elementi di WD ai termini del Thesaurus NS. Nei mesi successivi si è cominciato a importare i valori del *template* Thesaurus BNCF presenti nelle voci di itWP come valori della proprietà P508 nei corrispondenti elementi WD.

## filologia (Q40634)

insieme di discipline che studia i testi letterari al fine della ricostruzione della loro forma originaria

 [modifica](#)

▼ In altre lingue

Lingua	Etichetta	Descrizione	Anche conosciuto come
italiano	filologia	insieme di discipline che studia i testi letterari al fine della ricostruzione della loro forma originaria	
inglese	philology	study of language in written historical sources	

identificativo Thesaurus BNCF  6960  [modifica](#)

tipo di relazione corrispondenza esatta

soggetto indicato come Filologia

▼ 1 riferimento

fonte dell'affermazione Nuovo soggettoario

[+ aggiungi riferimento](#)

[+ aggiungi valore](#)

Fig. 2. Un esempio di link da WD al Thesaurus NS, nell'elemento 'Filologia' (Q40634).

Ora tutti i link verso il Thesaurus NS sono gestiti solo su WD; di conseguenza, le modifiche su questo *database* sono immediatamente visibili anche su itWP (dove, come detto nel precedente paragrafo, dal 2020 il link ai termini del Thesaurus NS è incluso nel *template* Controllo di autorità) (fig. 2).

A partire dal 2020 la BNCF, che fino ad allora ha preso in considerazione esclusivamente itWP nel compilare i termini del Thesaurus NS, ha corredato i termini preferiti di un campo dedicato a WD nel quale inserire gli identificatori dei corrispondenti elementi, successivamente (dal 2023) resi visibili e linkabili nell'interfaccia utente del Thesaurus NS dal campo *Equiv. Wikidata*<sup>17</sup> (fig. 3).

**Filologia** GERARCHIA

Macrocategoria: Categoria Azioni:Discipline

**Nota d'ambito** Insieme di discipline intese alla ricostruzione di documenti linguistici e letterari e alla loro corretta interpretazione e comprensione

**Fonti** [Soggettario](#); [Treccani.it](#); [DeM](#); [PT](#); [VT](#); [WebDewey\(IT\)](#); [Wikipedia\(IT\)](#)

DDC (WebDewey) 400

**Equiv. in altri strumenti di indicizzazione**

-  LCSH: [Philology](#)
-  RAMEAU: [Philologie](#)
-  GND: [Philologie](#)
-  EMBNE: [Filologie](#)

**Equiv. Wikidata** [Q40634](#)

Fig. 3. Didascalia: un esempio di link dal Thesaurus NS a itWP e a WD, nel termine 'Filologia' (6960).

A fine agosto 2024 le crosswalk tra Thesaurus NS, itWP e WD erano le seguenti:

	da Thesaurus NS	da itWP	da WD
a Thesaurus NS		15129	17007
a itWP	24743		
a WD	12893		

## 2. Da Wikipedia a Wikidata: due paradigmi diversi

È importante sottolineare come i ruoli di WD e delle varie versioni linguistiche di Wikipedia siano complementari e come sia pertanto utile per il Thesaurus NS collegarsi a entrambe.

Wikipedia (nelle sue varie versioni linguistiche) è un'enciclopedia e le sue voci raccolgono informazioni riguardo alle entità in forma testuale, pensata per un lettore umano; generalmente contengono informazioni più ampie e una bibliografia più estesa rispetto ai corrispettivi elementi di WD; tuttavia, essendo pensate per la comodità del lettore umano, possono talvolta trattare più argomenti correlati all'interno della medesima voce, anche se logicamente dovrebbero essere trattati in voci distinte (così infatti come si fa, per necessità classificatoria, in WD e nel Thesaurus NS).

WD invece è una base di conoscenza e i suoi elementi raccolgono informazioni riguardo alle entità in forma strutturata, pensata tanto per un lettore umano quanto per i *bot*; gli elementi generalmente contengono informazioni più schematiche e riferimenti molto essenziali rispetto alle corrispettive voci delle varie versioni linguistiche di Wikipedia, ma in aggiunta forniscono anche gli identificatori di molteplici siti (altri thesaurus, enciclopedie, *authority file*, ecc.) riguardanti quelle medesime entità; inoltre, in linea di principio in WD ogni entità ha un elemento a sé stante, anche laddove invece una o più versioni linguistiche di Wikipedia uniscono più entità in una stessa voce per comodità del lettore umano. Essendo WD multilingue (a differenza delle varie Wikipedie, una per ogni lingua), esiste anche l'ulteriore difficoltà di costruire elementi in sé coerenti, che trattino la stessa entità nelle loro etichette, descrizioni e *alias* nelle varie lingue, o comunque entità quasi completamente sovrapponibili, e non invece entità omonime ma i cui significati variano sensibilmente da una lingua all'altra.

In conclusione, essendo il Thesaurus NS uno strumento di organizzazione della conoscenza, è naturale che si colleghi sia ad altri KOS (*Knowledge organization system*) in varie lingue (compresa WD) sia a opere di riferimento in lingua italiana (compresa itWP).

## 3. Il colloquio semantico e tecnico tra il Thesaurus del Nuovo Soggettario e Wikidata

### 3.1 Obiettivi

Il primo obiettivo di questa collaborazione è stata la piena interoperabilità reciproca tra i due strumenti, ormai raggiunta dal punto di vista tecnico: gli elementi di WD linkavano ai termini del Thesaurus già dal 2013 tramite la proprietà P508 (identificativo Thesaurus BNCf); il Thesaurus NS ha introdotto nel 2023 il campo *Equiv. Wikidata* che permette di navigare dai suoi termini ai corrispondenti elementi di WD. Nello stesso anno 2023 è stata creata in WD una pagina di coordinamento<sup>18</sup> relativa alla collaborazione tra i bibliotecari della BNCf e la comunità di WD sulle *crosswalk* tra Thesaurus NS e WD. Questo è, a livello internazionale, il secondo progetto documentato di interconnessione tra un thesaurus di rilevanza nazionale e WD, dopo il progetto iniziato nel 2020 relativo al thesaurus finlandese YSO<sup>19</sup>.

Ora l'obiettivo di lungo periodo è migliorare i collegamenti reciproci tra i termini del Thesaurus NS e gli elementi di WD, da un punto di vista sia qualitativo (rivedendo e correggendo i legami già esistenti qualora si rivelino essere imprecisi) sia quantitativo (rivedendo i legami unidirezionali già esistenti per renderli eventualmente reciproci e attivando *ex novo* crosswalk).

A marzo 2024, infatti, su 33.920 termini preferiti del Thesaurus NS solo 12.822 linkavano a WD (il 37,8% del totale)<sup>20</sup>, mentre erano 16.823 i collegamenti da WD al Thesaurus NS il 6 marzo 2024<sup>21</sup> (e c'erano quindi almeno 4.001 link unidirezionali da WD al Thesaurus NS, da verificare ed eventualmente rendere bidirezionali).

### 3.2 Differenze strutturali

Il Thesaurus NS è uno strumento con struttura ad albero classificatoria basata su Macrocategorie/Categorie/Faccette dove i termini hanno un'unica posizione all'interno della scala gerarchica basata sulle caratteristiche definitorie essenziali del concetto stesso<sup>22</sup>. Il vocabolario è prevalentemente monogerarchico, essendo la poligerarchia ammessa solo in pochi casi e con particolari cautele, quindi i termini hanno un solo BT (Broader Term, termine semanticamente più ampio), anche se possono avere più NT (Narrower Term, termine semanticamente più specifico) e RT (Related Term, termine semanticamente associato), come si può osservare ad esempio nel termine 'Geografia fisica'<sup>23</sup>.

WD permette di aggiungere ai vari elementi diverse migliaia di proprietà (12.076 a metà agosto 2024), la maggior parte delle quali (8.975) sono attualmente identificativi esterni; un ridotto numero di proprietà riguarda le relazioni ontologiche tra gli elementi, che si possono così schematizzare: il legame NT, essendo l'esatto inverso del legame BT, non è esplicitato; i legami BT vengono creati tramite la P279 (sottoclasse di); i legami RT vengono creati tramite una molteplicità di proprietà specifiche, ad esempio P3095 (praticato da) per collegare una disciplina (es. 'filologia') alla professione relativa (es. 'filologo') oppure P2578 (studia) per collegare una disciplina (es. 'biologia') all'argomento studiato (es. 'vita').

Le due proprietà classificatorie fondamentali in WD<sup>24</sup> sono la già citata P279 (sottoclasse di) e P31 (istanza di): la prima lega l'elemento alla tipologia più ampia a cui appartiene (es. 'filologo classico' è sottoclasse di 'filologo', ossia è un tipo di filologo), la seconda definisce l'elemento come singolo esponente di un gruppo (es. 'filologo classico' è istanza di 'professione', ossia è una professione). Gli elementi che rappresentano entità singole (es. singoli esseri umani, singole montagne, ecc.) hanno soltanto P31 e non P279 (es. 'Napoleone' è istanza di 'umano', il 'Monte Bianco' è istanza di 'montagna') e, in generale, sono elementi che non sono presenti nel Thesaurus NS, appunto perché si tratta di nomi propri<sup>25</sup>; al contrario, tutti gli elementi che rappresentano concetti (es. discipline, professioni, ecc.) devono avere P279 e, risalendo nella catena delle P279, devono essere in ultima analisi sottoclassi di 'entità' (Q35120), che è l'elemento apicale dell'intera ontologia di WD. Il miglior strumento per visualizzare graficamente le relazioni ontologiche tra gli elementi di WD è il Wikidata Graph Builder<sup>26</sup>.

È necessario accennare, sebbene non sia questa la sede adeguata per trattare estesamente questo tema, che l'ontologia di WD ha rilevanti problemi di coerenza interna di cui da svariati anni si sta discutendo gravità e possibili soluzioni, sia all'interno di Wikidata<sup>27</sup> sia nella letteratura scientifica<sup>28</sup>. Ad alcuni di questi problemi si farà cenno nei seguenti paragrafi.

### 3.3 Attività in corso

La prima attività svolta durante la collaborazione tra Thesaurus NS e WD ha riguardato l'aggiunta semiautomatica del campo Equiv. Wikidata ai termini del Thesaurus NS, laddove fosse già presente itWP nel campo Fonte, e controllo intellettuale da parte di BNCF per verificare l'esattezza semantica del legame.

Non è stato tuttavia possibile ricavare l'elemento WD laddove la voce di itWP citata non aveva una corrispondenza esatta con il termine del Thesaurus NS. Questo arricchimento si è ormai concluso ed è stato il principale mezzo che ha permesso di raggiungere gli attuali 12.893 link dal Thesaurus NS a WD.

La seconda attività svolta nell'ambito di questa collaborazione, e tuttora in corso, riguarda i link problematici tra Thesaurus NS e WD. Come si è detto, in linea di principio, la corrispondenza tra i termini del Thesaurus NS e gli elementi di WD dovrebbe essere biunivoca: ogni termine preferito del Thesaurus NS dovrebbe collegarsi a un solo elemento di WD e tale elemento dovrebbe essere collegato soltanto a quel termine del Thesaurus NS. Non dovrebbero quindi esistere né A) casi in cui più elementi WD si collegano al medesimo termine del Thesaurus NS, né B) casi in cui un solo elemento WD si collega a più termini del Thesaurus NS. Tuttavia, alcune centinaia di questi casi problematici si sono accumulate nel corso del tempo, per cui dal 2023 i bibliotecari della BNCF hanno cominciato, in collaborazione con la comunità di WD, ad analizzare queste criticità e a risolverle manualmente (a metà agosto 2024 si registrano ancora 196 casi del tipo A<sup>29</sup> e 48 casi del tipo B<sup>30</sup>; il 5 giugno 2023, all'inizio della collaborazione, erano 272 casi del tipo A e 129 casi del tipo B<sup>31</sup>).

La terza attività svolta nell'ambito di questa collaborazione, destinata a essere permanente, riguarda l'aggiunta del link a un elemento WD da parte dei bibliotecari della BNCF per il maggior numero possibile di termini del Thesaurus NS di nuova introduzione. In particolare va sottolineata l'intenzione di adottare un *modus operandi* diverso dal passato: non più cercare la voce di itWP corrispondente al termine del Thesaurus NS e, qualora esista, ricavare da essa anche l'elemento di WD, ma cercare direttamente in WD se esiste già un elemento per il concetto rappresentato dal termine del Thesaurus NS e, in caso non esista, provare a crearlo *ex novo* sulla base di altri thesauri e di ricerche dedicate.

### 3.4 Esempi di legami in corso di revisione

Le equivalenze tra due strumenti di organizzazione della conoscenza possono essere nel significato e/o nella forma dei termini (es. il Thesaurus NS e WD per l'italiano, RAMEAU e WD per il francese, ecc.). In particolare, abbiamo considerato equivalenze esatte nella forma i casi in cui per lo stesso concetto il significante ha medesima radice ma una variazione nella desinenza, es. laddove

il Thesaurus NS sceglie sistematicamente la forma al plurale (es. ‘Filologi’), WD invece ha sempre l’etichetta in italiano al singolare (es. ‘filologo’; si nota anche come WD abbia come indicazione l’uso dell’iniziale maiuscola soltanto dove grammaticalmente necessaria, mentre tutti i termini del Thesaurus NS hanno l’iniziale maiuscola).

Le equivalenze nel significato si possono invece stabilire confrontando le note d’ambito e le note di definizione di un thesaurus con le descrizioni di WD e soprattutto con le sue dichiarazioni e con i suoi identificativi; questo perché, come le etichette e gli *alias*, anche le descrizioni sono molto più soggette a variazioni rispetto al resto dell’elemento (dichiarazioni, identificativi esterni e *sitelink*); è in ogni caso fondamentale che il loro contenuto sia ragionevolmente coerente con quello delle dichiarazioni e degli identificativi, e in caso contrario bisogna renderlo tale prima di procedere all’associazione con un thesaurus.

Nello specifico, ai fini del collegamento tra i termini preferiti del Thesaurus NS e gli elementi di WD si è concordato che l’equivalenza debba essere esatta, o quasi esatta, nel significato del termine, non necessariamente nella forma. Il caso standard è la corrispondenza biunivoca tra Thesaurus NS e WD: ‘Biologia’ (1541) = ‘biologia’ (Q420); ‘Cani’ (285) = ‘cane’ (Q144); ‘Abeti greci’ (31099) = ‘Abies cephalonica’ (Q854892).

Sono invece problematiche, come già detto, A) le corrispondenze due elementi WD con un termine del Thesaurus NS e B) le corrispondenze un elemento WD con due termini del Thesaurus NS, alla cui risoluzione si sta lavorando dal giugno 2023. Si presenta sinteticamente di seguito una disamina delle casistiche più frequenti.

Quando un elemento WD X linka a due termini A e B del Thesaurus NS, si rilevano le seguenti casistiche:

1) raramente, è corretto che X punti sia ad A sia a B; questa eccezione riguarda in particolare le professioni, per le quali generalmente WD ha un unico elemento, mentre il Thesaurus NS differenzia due elementi in base al genere; quindi è corretta la corrispondenza ‘archeologo’ *alias* ‘archeologa’ (Q3621491) = ‘Archeologi’ (1079) + ‘Archeologhe’ (1081);

2) più spesso, X punta correttamente ad A ma non a B; in questo caso si può semplicemente rimuovere il link da X a B per risolvere il problema; tuttavia, la soluzione migliore prevede anche di cercare quale sia l’elemento WD Y da collegare al termine B del Thesaurus NS. Si procede quindi a cercare tale elemento Y: se si è ragionevolmente sicuri che non esista, lo si crea e lo si collega a B; se esiste e non presenta problemi di coerenza interna, lo si collega direttamente a B; se esiste ma presenta problemi di coerenza interna, lo si collega a B soltanto dopo averlo adeguatamente sistemato nelle sue varie parti (etichette, descrizioni e *alias*; dichiarazioni, in particolare P279; identificativi esterni).

### 3.5 Benefici reciproci

Si prevede che una miglior interconnessione reciproca tra i termini del Thesaurus NS e gli elementi di WD possa avere rilevanti benefici per entrambi gli strumenti.

Il Thesaurus NS e WD possono fare uso l'uno dell'altro per creare nuovi termini/elementi mancanti, per arricchire il proprio apparato sinonimico (ossia i termini non preferiti nel Thesaurus NS, gli *alias* in WD) e per migliorare le definizioni dei concetti (ossia per rivedere le note d'ambito e le note di definizione nel Thesaurus NS, le descrizioni in WD).

Il Thesaurus NS può usare WD per arricchire i propri termini traendone link ad altri strumenti di indicizzazione semantica di biblioteche nazionali straniere (LCSH = P244, RAMEAU = P268, GND = P227, EMBNE = P950), a thesauri specialistici (es. AAT = P1014, AGROVOC = P8061), a fonti di reference (per arricchire il file *Fonti*<sup>32</sup> usato per la validazione della forma e del significato dei termini), a *database* di archivi e musei e a voci di *itWP*.

WD può usare il Thesaurus NS per introdurre etichette in italiano mancanti, per rivedere i propri link ad altri thesauri e soprattutto per controllare e migliorare i legami gerarchici tra i propri elementi.

### 3.6 Progetti futuri

Nei prossimi mesi la collaborazione tra i bibliotecari della BNCf e la comunità di WD dovrebbe svilupparsi principalmente lungo le seguenti direttrici: concludere la risoluzione dei legami problematici tra WD e Thesaurus NS, in modo da mantenere solo i (pochi) legami 2:1 effettivamente corretti; procedere alla graduale verifica manuale dei legami unidirezionali da WD al Thesaurus NS e dal Thesaurus NS a WD, al fine di renderli gradualmente reciproci; accrescere i link bidirezionali tra WD e Thesaurus NS, ad esempio con progetti tematici che possano considerare in parallelo come uno stesso corpus concettuale è trattato in WD e nel Thesaurus NS, cercando il più possibile di allineare i due strumenti. Si sta infine indagando anche la possibilità di usare WD per arricchire la terminologia del Thesaurus NS, in particolare valutando gli elementi di WD che contengono già numerosi link a thesauri specialistici e a quelli di biblioteche nazionali considerati rilevanti dal NS<sup>33</sup>.

## NOTE

---

\* Elisabetta Viti è Funzionario bibliotecario presso la BNCf ed Elena Cencetti lavora per Ales S.p.A. presso la stessa biblioteca; Camillo Carlo Pellizzari di San Girolamo è dottorando presso la Scuola Normale Superiore (Pisa).

<sup>1</sup> Il progetto è curato dai settori Ricerche e strumenti di indicizzazione semantica e Servizi informatici.

<sup>2</sup> <https://thes.bncf.firenze.sbn.it/>. Questa pagina, e tutte le pagine citate nell'articolo, sono state consultate il 30 agosto 2024; dove si specifica "l'URL originale è" si intende che la pagina in tale data non era più accessibile e viene quindi fornita una versione di tale pagina archiviata nella Wayback Machine.

<sup>3</sup> Lucarelli 2014, p. 250.

<sup>4</sup> *Ivi*, p. 253.

---

<sup>5</sup> Ivi, pp. 250-252.

<sup>6</sup> L'URL originale è <http://it.dbpedia.org/> (la prima versione archiviata dalla Wayback Machine è <https://web.archive.org/web/20120414174718/http://it.dbpedia.org/>).

<sup>7</sup> <https://www.dbpedia.org/>. Il progetto DBpedia ha come obiettivo l'estrazione di dati strutturati e multilingue dalle varie versioni linguistiche di Wikipedia rendendole liberamente disponibili sul Web, grazie alle tecnologie del Web semantico e dei *Linked Data*. Queste informazioni strutturate assomigliano a un grafo della conoscenza aperto, cioè un tipo speciale di *database* che memorizza i dati in una forma leggibile dalla macchina e fornisce un mezzo per raccogliere, organizzare, condividere, cercare e utilizzare le informazioni, rendendole disponibili per tutti sul web.

In pratica DBpedia, prima di Wikidata, è stato il centro di interconnessione più importante nella nuvola dei *Linked Open Data* (LOD): oltre a mappare le infobox di 27 edizioni linguistiche diverse di Wikipedia a un'unica ontologia condivisa composta da 320 classi e 1.650 proprietà, ha stabilito collegamenti RDF (circa 27 milioni) a 30 fonti di dati esterne e ha costituito il centro di puntamento verso cui centinaia di *dataset* sul Web hanno creato collegamenti RDF. Cfr. Lehmann et alii 2015.

<sup>8</sup> Lucarelli 2014, p. 252.

<sup>9</sup> L'URL originale è [https://it.wikipedia.org/wiki/Template:Thesaurus\\_BNCF](https://it.wikipedia.org/wiki/Template:Thesaurus_BNCF) (la prima versione archiviata dalla Wayback Machine è [https://web.archive.org/web/20140331074111/http://it.wikipedia.org/wiki/Template:Thesaurus\\_BNCF](https://web.archive.org/web/20140331074111/http://it.wikipedia.org/wiki/Template:Thesaurus_BNCF)).

<sup>10</sup> Lucarelli 2014, pp. 253-254; cfr. annuncio dato nel settembre 2013 su Wikimedia Outreach: [https://outreach.wikimedia.org/wiki/GLAM/Newsletter/September\\_2013/Contents/Italy\\_report](https://outreach.wikimedia.org/wiki/GLAM/Newsletter/September_2013/Contents/Italy_report).

<sup>11</sup> [https://it.wikipedia.org/wiki/Template:Controllo\\_di\\_autorit%C3%A0](https://it.wikipedia.org/wiki/Template:Controllo_di_autorit%C3%A0).

<sup>12</sup> [https://www.bncf.firenze.sbn.it/wp-content/uploads/2019/11/Convenzione\\_BNCF\\_Wikimedia.pdf](https://www.bncf.firenze.sbn.it/wp-content/uploads/2019/11/Convenzione_BNCF_Wikimedia.pdf).

<sup>13</sup> L'evoluzione della collaborazione tra Thesaurus NS e itWP fino a inizio 2014 è ottimamente riassunta in una presentazione di Maria Grazia Pepe intitolata *Wikipedia e Nuovo soggetto: Lo stato dell'arte e gli sviluppi futuri*, tenutasi a Firenze presso l'assemblea di Wikimedia Italia il 5 aprile 2014 (le slide sono disponibili su Wikimedia Commons: [https://commons.wikimedia.org/wiki/File:NuovoSoggetto\\_e\\_Wikipedia\\_05-04-2014.pdf](https://commons.wikimedia.org/wiki/File:NuovoSoggetto_e_Wikipedia_05-04-2014.pdf)).

<sup>14</sup> Cfr. Vrandečić et alii 2023.

<sup>15</sup> <https://www.wikidata.org/wiki/Wikidata:Notability>.

<sup>16</sup> Per maggiori informazioni sulle varie parti di un elemento Wikidata, cfr. <https://www.wikidata.org/wiki/Wikidata:Glossary>.

<sup>17</sup> Per quanto riguarda i criteri con cui è stata attivata l'interoperabilità tra Thesaurus NS e altri strumenti di organizzazione della conoscenza cfr. BNCF 2021, pp. 168-170.

<sup>18</sup> [https://www.wikidata.org/wiki/Wikidata:Gruppo\\_Wikidata\\_per\\_Musei\\_Archivi\\_e\\_Biblioteche/Nuovo\\_soggetto](https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei_Archivi_e_Biblioteche/Nuovo_soggetto).

<sup>19</sup> Cfr. <https://www.wikidata.org/wiki/Q89345680>.

<sup>20</sup> <https://thes.bncf.firenze.sbn.it/stat.php>.

<sup>21</sup> <https://www.wikidata.org/w/index.php?oldid=2095615843>.

<sup>22</sup> A ogni termine è assegnato un identificativo univoco numerico leggibile dalla macchina; esempio: 'Costellazioni' (26906).

<sup>23</sup> <https://thes.bncf.firenze.sbn.it/termine.php?id=22142>.

<sup>24</sup> [https://www.wikidata.org/wiki/Help:Basic\\_membership\\_properties](https://www.wikidata.org/wiki/Help:Basic_membership_properties).

<sup>25</sup> Nel Thesaurus del *Nuovo soggettoario* sono presenti solo nomi comuni, insieme ad alcune limitate tipologie di nomi propri (cfr. BNCf 2021, p. 82).

<sup>26</sup> <https://angryloki.github.io/wikidata-graph-builder/>.

<sup>27</sup> Cfr. materiali raccolti in [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Ontology](https://www.wikidata.org/wiki/Wikidata:WikiProject_Ontology) e soprattutto [https://www.wikidata.org/wiki/Wikidata:Ontology\\_issues\\_prioritization](https://www.wikidata.org/wiki/Wikidata:Ontology_issues_prioritization).

<sup>28</sup> Cfr. Brasileiro *et alii* 2016, Piscopo – Simperl 2018, Dadalto *et alii* 2024.

<sup>29</sup> <https://w.wiki/AvEa>.

<sup>30</sup> <https://w.wiki/AvEc>.

<sup>31</sup> <https://www.wikidata.org/w/index.php?oldid=1881201349>.

<sup>32</sup> <https://thes.bncf.firenze.sbn.it/fonti.php>.

<sup>33</sup> Cfr. <https://w.wiki/AvOq> (query ricca di falsi positivi, ma contenente anche alcuni potenziali termini mancanti nel Thesaurus NS, oppure già presenti ma ancora privi di link bidirezionale con WD).

## BIBLIOGRAFIA

BNCf 2021: Biblioteca nazionale centrale di Firenze, *Nuovo soggettoario. Guida al sistema italiano di indicizzazione per soggetto*, 2. ed. interamente rivista e aggiornata, Roma-Firenze 2021. [https://www.bncf.firenze.sbn.it/wp-content/uploads/2020/01/Nuovo-soggettoario\\_Guida.pdf](https://www.bncf.firenze.sbn.it/wp-content/uploads/2020/01/Nuovo-soggettoario_Guida.pdf) (pagina consultata il 22 agosto 2024).

Brasileiro *et alii* 2016: F. Brasileiro, J. P. A. Almeida, V. A. Carvalho, G. Guizzardi, *Applying a Multi-Level Modeling Theory to Assess Taxonomic Hierarchies in Wikidata*, in AA.VV., *WWW '16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web*, 2016, pp. 975-980. DOI:10.1145/2872518.2891117

Dadalto *et alii* 2024: A. A. Dadalto, J. P. A. Almeida, C. M. Fonseca, G. Guizzardi, *Evidence of large-scale conceptual disarray in multi-level taxonomies in Wikidata*, in “Semantic Web”, 2024, pp. 1-18. DOI:10.3233/SW-243562

Lehmann *et alii* 2015: J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, C. Bizer. *DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia*, in “Semantic Web”, VI, 2015, pp. 167-195. DOI:10.3233/SW-140134

Lucarelli 2014: A. Lucarelli, “Wikipedia loves libraries”: in Italia è un amore corrisposto..., in “AIB studi”, LIV, 2014, pp. 241-259. DOI:10.2426/aibstudi-10108

Piscopo – Simperl 2018: A. Piscopo, E. Simperl, *Who Models the World? Collaborative Ontology Creation and User Roles in Wikidata*, in AA.VV., *Proceedings of the ACM on Human-Computer Interaction 2*, CSCW, 2018, art. 141. DOI:10.1145/3274410

Vrandečić *et alii* 2023: D. Vrandečić, L. Pintscher, M. Krötzsch, *Wikidata: The Making Of*, in AA.VV., *WWW '23: The ACM Web Conference 2023*, Companion Proceedings of the ACM Web Conference 2023, Austin 2024, pp. 615-624. DOI:10.1145/3543873.3585579.