

**Maria Grazia Pepe\****Nuovo soggettario* – Modalità di implementazione dell'interoperabilità

La Biblioteca nazionale centrale di Firenze (BNCF) sin dall'inizio ha concepito il Thesaurus del *Nuovo soggettario* (NS)<sup>1</sup> come uno strumento interoperabile, promuovendo la collaborazione con altre istituzioni nazionali e internazionali anche attraverso la sua pubblicazione in formato di dati aperti secondo la licenza Creative Commons (CC BY 4.0)<sup>2</sup>. Nell'ottica di favorire uno scambio di informazioni flessibile e ampio il Thesaurus è disponibile fra i *dataset* della nuvola dei *linked open data* (LOD cloud)<sup>3</sup> (fig.1) e sulla piattaforma dati.cultura.gov.it del Ministero della Cultura (MiC)<sup>4</sup>; il modello dei dati è lo SKOS/RDF<sup>5</sup>.

Le modalità di interazione con altri strumenti implementate nel corso del tempo sono differenti e vanno dalla semplice citazione al collegamento puntuale da termine a termine. Saranno qui descritte in sintesi le procedure utilizzate per realizzare l'interoperabilità fra le diverse banche dati, evidenziando le problematiche emerse e i possibili sviluppi, legati alla sempre maggiore standardizzazione nell'uso degli strumenti del web semantico da parte delle istituzioni culturali.

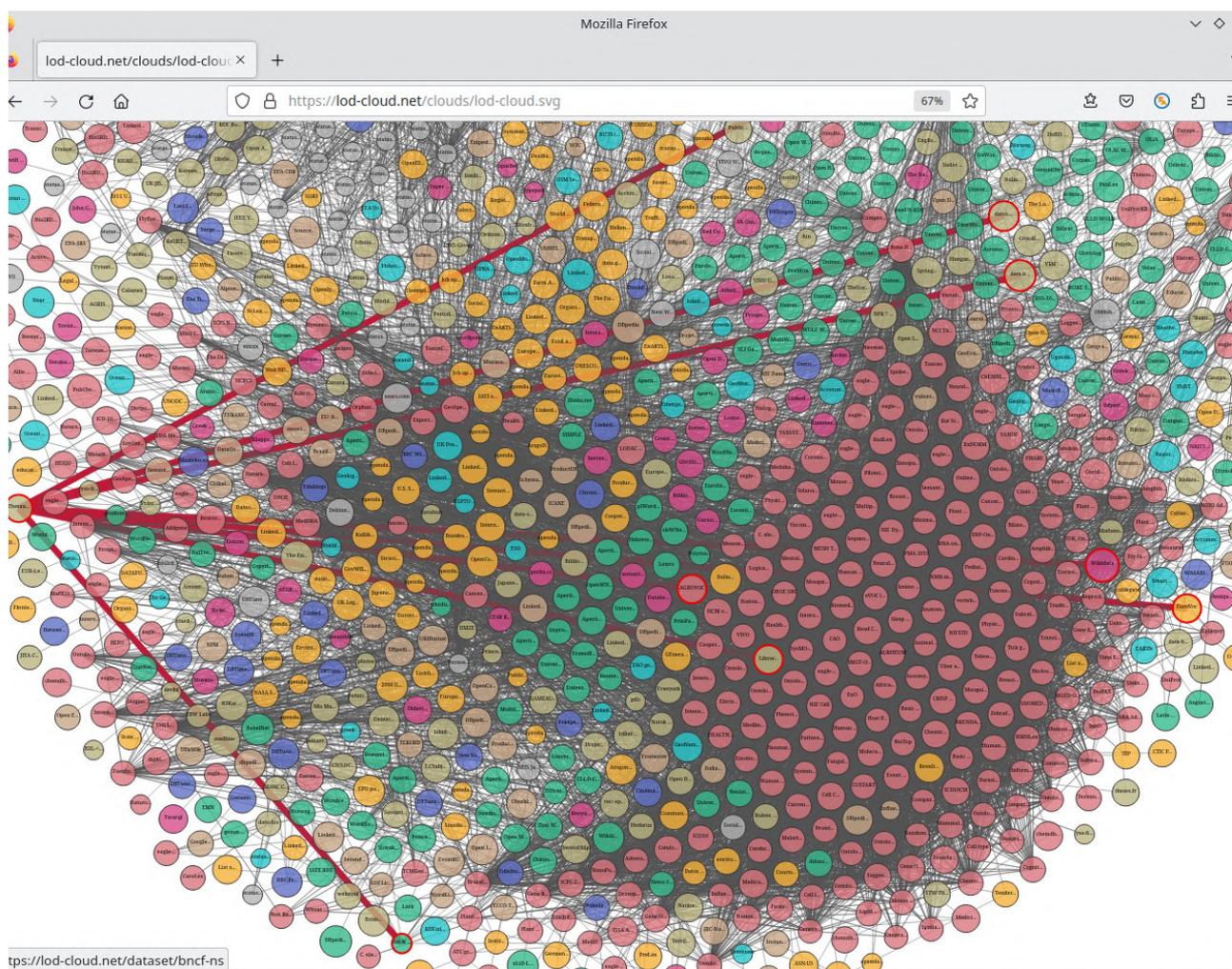


Fig. 1a. Il Thesaurus del *Nuovo soggettario* nella nuvola dei *linked open data*.

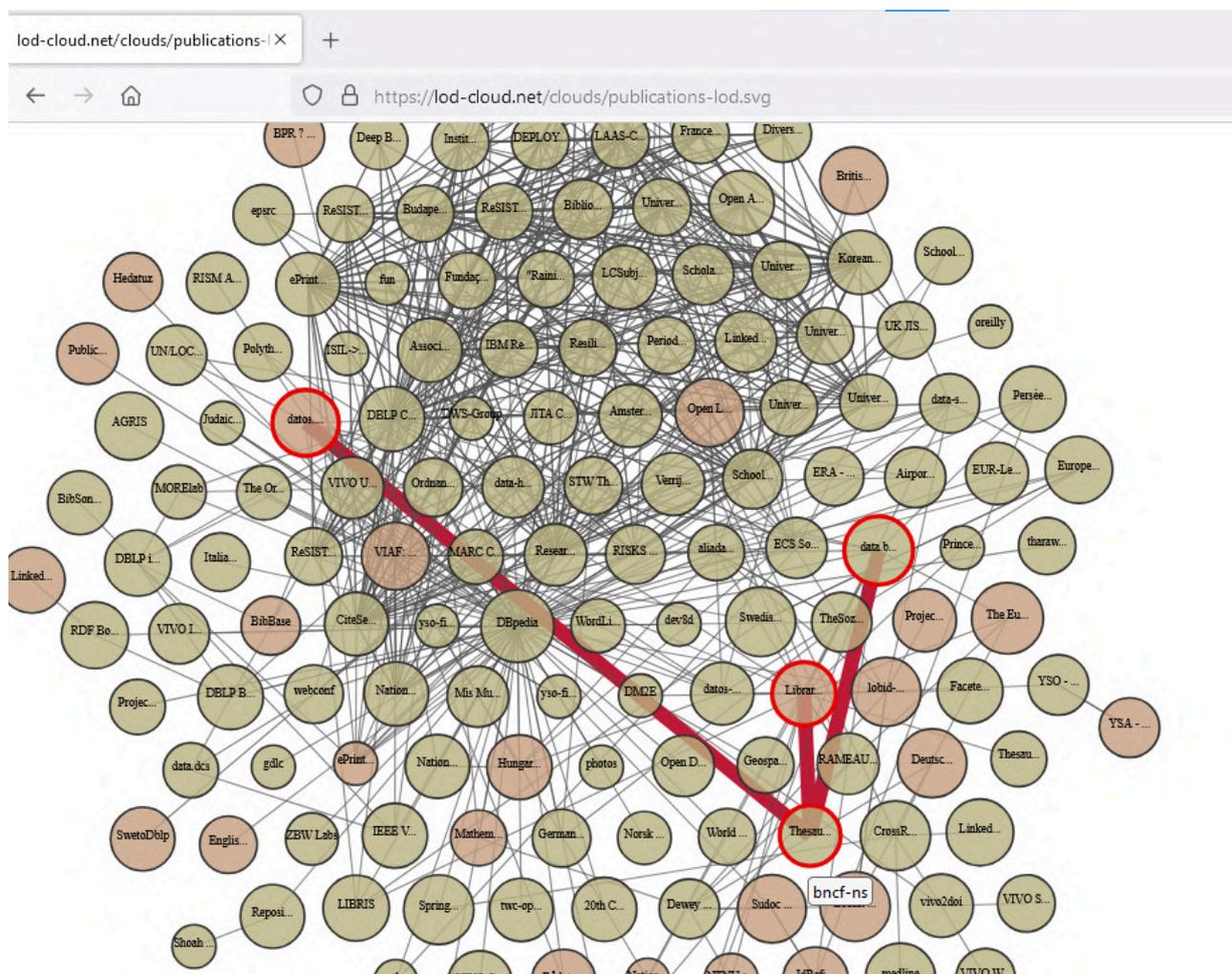


Fig. 1b. Il Thesaurus del Nuovo soggetto nella nuvola dei linked open data.

### Modalità di implementazione

L'interoperabilità del NS con altri strumenti e sistemi di organizzazione della conoscenza (KOS)<sup>6</sup> è realizzata sostanzialmente in due modalità, a seconda della disponibilità o meno della risorsa in versione *linked open data* (LOD)<sup>7</sup>.

Per gli strumenti che mettono a disposizione la versione LOD dei dati, già in fase di strutturazione della terminologia, è possibile:

- effettuare una ricerca sul termine da collegare;
- proporre ai bibliotecari i risultati di questa ricerca;
- consentire il collegamento puntuale attraverso l'uso di identificatori univoci.

I collegamenti così attivati sono presenti in parallelo nell'interfaccia web del Thesaurus e nel formato SKOS/RDF<sup>8</sup> secondo il tipo *skos:closeMatch*.

Nel caso di assenza di dati in versione LOD i collegamenti vengono attivati solo nell'interfaccia web del NS, simulando una ricerca in modalità HTTP GET<sup>9</sup> sullo strumento di riferimento. Il termine da ricercare corrisponde a quello accettato presente nel Thesaurus oppure è indicato esplicitamente dopo il controllo dei bibliotecari in fase di strutturazione.

Una problematica che si presenta spesso in questi casi è relativa all'uniformità nell'uso del plurale, che nel NS è riservato ai lemmi che rappresentano concetti numerabili.

In entrambi i casi le corrispondenze con altri strumenti sono presenti nel record del termine nell'interfaccia web del Thesaurus<sup>10</sup> a partire da:

- gli equivalenti linguistici;
- le fonti, di ambito generale o specialistico, utilizzate per il controllo terminologico;
- la sezione di menù dedicata a risorse di archivi e musei.

### **Equivalenti linguistici in altri strumenti di indicizzazione semantica**

La terminologia del NS è corredata di equivalenti semantici in altre lingue, presenti nei vocabolari controllati curati dalle biblioteche nazionali di altri paesi; i corrispettivi linguistici consentono all'utente sia di navigare trasversalmente all'interno degli strumenti stessi, sia di consultare direttamente altri OPAC (*online public access catalogue*) a essi integrati.

Già nel 2011 il Thesaurus riportava numerose corrispondenze con gli equivalenti inglesi di *Library of Congress Subject Headings* (LCSH)<sup>11</sup>. Oltre a questi attualmente sono gestiti anche gli equivalenti previsti dal sistema tedesco *Gemeinsame Normdatei* (GND), dal francese *Répertoire d'autorité-matière encyclopédique et alphabétique unifié* (RAMEAU) e dallo spagnolo *Encabezamientos de Materia de la Biblioteca Nacional de España* (EMBNE).

In tutti questi casi per l'attivazione dei link il punto di partenza è sempre la versione dei dati esposti in formato *linked open data*, più nello specifico:

- il formato RDF secondo la GND Ontology<sup>12</sup> per gli equivalenti in tedesco;
- il formato SKOS/RDF dei dati di LCSH<sup>13</sup>, RAMEAU<sup>14</sup> ed EMBNE<sup>15</sup> per gli equivalenti in inglese, francese e spagnolo.

La pubblicazione dei dati in formato LOD, secondo ontologie ben definite, ha consentito l'incremento dell'interoperabilità tra i diversi strumenti terminologici, l'attivazione di contatti tra i gruppi di lavoro delle diverse biblioteche e la sperimentazione di procedure per la condivisione dei risultati<sup>16</sup>.

### **Fonti**

Per il controllo terminologico il NS si avvale di un apparato di numerose fonti<sup>17</sup>: repertori di ambito generale o specialistico, vocabolari controllati e altri thesauri, fruibili in diversi formati. I repertori consultati nel record del termine vengono indicati con una sigla apposita che vale come attestazione di fonte ma anche come possibile link alla risorsa, a seconda della sua disponibilità o meno sul Web.

### **Fonti di ambito generale**

Tra le fonti del NS c'è EUROVOC<sup>18</sup>, il thesaurus multilingue e multidisciplinare dell'UE, utilizzato per indicizzare il contenuto dei documenti ufficiali. I suoi dati sono disponibili anche in formato SKOS/RDF: è possibile quindi attivare collegamenti puntuali da termine a termine<sup>19</sup>.

Per la gestione dei dati in formato LOD il NS ha fatto riferimento sin dagli inizi a Wikipedia, grazie anche a un accordo di collaborazione tra Wikimedia Italia e BNCF<sup>20</sup>. L'interoperabilità tra i due strumenti è stata realizzata in modalità differenti nel corso del tempo<sup>21</sup>.

Nella fase iniziale la creazione di collegamenti veniva attivata utilizzando la versione LOD dei dati di Wikipedia, gestiti su DBPedia in versione italiana<sup>22</sup>. Dal NS venivano attivati in parallelo collegamenti a Wikipedia (interfaccia web) e DBPedia (versione LOD). Oltre che il confronto tra lemmi è stato utile anche l'uso di ontologie presenti su DBPedia per la definizione di terminologia di genere diverso, nomi propri di persone, luoghi, ecc.), da escludere nel collegamento al Thesaurus (ad esempio, in caso di sinonimia con nomi comuni, 'Maglie' presente nel NS e 'Maglie' comune della provincia di Lecce presente in Wikipedia). La pubblicazione nel 2012 di Wikidata<sup>23</sup>, che raccoglie in modo centralizzato i dati delle diverse versioni linguistiche di Wikipedia, ha portato progressivamente a modificare le procedure per l'attivazione dei collegamenti e a renderli reciproci. La versione LOD di Wikidata ha sostituito DBPedia nell'automatismo per la ricerca di corrispondenze tra i due strumenti. La nuova procedura segue la falsariga della precedente, utilizzando inoltre la proprietà P508<sup>24</sup>, appositamente attivata su Wikidata, e lo strumento QuickStatement<sup>25</sup> per i collegamenti reciproci verso il NS.

Nella versione LOD del Thesaurus i collegamenti a Wikidata inizialmente erano affiancati da quelli a DBPedia (eliminati da aprile 2020); nei LOD di Wikidata sono presenti riferimenti alla versione SKOS/RDF del NS. Dal 2023 i collegamenti a Wikidata sono visibili anche nell'interfaccia web del Thesaurus.

La collaborazione della comunità di Wikidata con i bibliotecari della BNCF è stata fondamentale per la verifica di qualità sui collegamenti attivati dagli automatismi; inoltre, il multilinguismo di Wikidata e i riferimenti agli strumenti di indicizzazione semantica nelle diverse lingue hanno consentito di fare controlli anche sui collegamenti agli equivalenti linguistici del NS.

I vocabolari e le enciclopedie realizzati a cura dell'Istituto dell'Enciclopedia Treccani sono sempre stati fondamentali per i bibliotecari della BNCF durante le fasi di controllo per l'attestazione della terminologia del Thesaurus. Già dal 2012 sono attivi collegamenti al sito web Treccani.it<sup>26</sup> e sono presenti link da termine a lemma, realizzati attraverso una ricerca effettuata in modalità HTTP GET sull'oggetto di riferimento.

### **Fonti di ambito specialistico**

Tra i thesauri di tipo specialistico è importante segnalare AGROVOC<sup>27</sup> della *Food and Agriculture Organization of the United Nations* (FAO), riferimento importante anche per quanto ha riguardato lo sviluppo del software gestionale del NS. AGROVOC è disponibile in versione SKOS/RDF: è quindi stato possibile attivare l'interoperabilità reciproca fra i due strumenti di indicizzazione nelle modalità già descritte; la stessa procedura è stata utilizzata per ThESS, il sistema di

indicizzazione della Biblioteca Mario Rostoni (LIUC)<sup>28</sup> e la classificazione DoGi<sup>29</sup> dell'Istituto di Informatica Giuridica e Sistemi Giudiziari del CNR.

Più di recente sono stati attivati collegamenti all'*Art & Architecture Thesaurus* (AAT)<sup>30</sup> del Getty Research Institute, disponibile in versione LOD così come il repertorio *Medical Subject Headings* (MESH)<sup>31</sup>, pubblicato in versione italiana dall'Istituto Superiore di Sanità<sup>32</sup>.

Il Thesaurus della Biblioteca Polimoda<sup>33</sup> ha invece consentito l'attivazione di collegamenti puntuali attraverso una ricerca in modalità HTTP GET, come anche l'*Elenco delle parole nuove*<sup>34</sup> dell'Accademia della Crusca. Oltre a quelli elencati, tra le fonti del NS sono presenti circa 180 repertori disponibili su web, attualmente segnalati con una semplice citazione (link alla *home page* del repertorio)<sup>35</sup>.

### **Classificazione Dewey**

I numeri della Classificazione decimale Dewey, attribuiti ai termini del Thesaurus NS, sono riportati in un campo apposito. La WebDewey in italiano<sup>36</sup> ha realizzato una mappatura con la terminologia del Thesaurus, favorita dal formato MARC, attraverso l'individuazione di concetti equivalenti. Sviluppi futuri potrebbero riguardare l'attivazione di collegamenti utilizzando le funzionalità messe a disposizione da una possibile versione italiana di WebDeweySearch<sup>37</sup>.

### **Risorse di Archivi e Musei**

Grazie alla partecipazione della BNCf alle attività del gruppo MAB Toscana è stata avviata una sperimentazione sull'interoperabilità tra il NS e le risorse di Archivi e Musei<sup>38</sup>.

#### **Archivi**

La collaborazione con gli archivi è stata avviata nel 2016 nell'ambito del progetto *Archivi di personalità*<sup>39</sup> della Soprintendenza archivistica per la Toscana. La sperimentazione è proseguita poi grazie alla collaborazione con l'Archivio Storico Ricordi<sup>40</sup> e con gli Archivi storici dell'UE<sup>41</sup>. Si tratta sempre di terminologia relativa a ruoli di persone con collegamenti puntuali realizzati con una ricerca in modalità HTTP GET. I link sono attivati dall'interfaccia web del NS a partire da una sezione del record del termine appositamente predisposta.

Si guarda con attenzione al lavoro intrapreso dall'Istituto Centrale per gli Archivi (ICAR) per la definizione di un'ontologia specifica per il Sistema archivistico nazionale<sup>42</sup>: l'obiettivo è rendere più efficienti le procedure per l'attivazione di collegamenti e renderli disponibili anche nei dati in versione LOD.

#### **Musei**

La collaborazione con le Gallerie degli Uffizi ha dato il via nel 2018 alla sperimentazione per collegare, tramite il NS, dati su risorse in ambito museale. Il lavoro è partito dalla descrizione delle opere d'arte presenti nelle schede di catalogo<sup>43</sup>. Sono stati attivati link reciproci, realizzati con una ricerca in modalità HTTP GET. Il flusso di lavoro per lo scambio di informazioni nel tempo è stato reso sistematico attraverso l'uso di file in formato testuale.

La presenza delle Gallerie degli Uffizi nel sistema SIGECWeb<sup>44</sup> dell'Istituto Centrale per il Catalogo e la Documentazione (ICCD) ha consentito di avere una visione più ampia degli standard di catalogazione in ambito museale.

Sono stati analizzati alcuni tra gli strumenti terminologici pubblicati dall'ICCD, in particolare i thesauri per la compilazione del paragrafo OG delle scheda OA<sup>45</sup> e RA<sup>46</sup>. Il lavoro si è poi arricchito della collaborazione con il Museo della Fondazione Scienza e Tecnica di Firenze<sup>47</sup> che ha dato il suo prezioso contributo per l'analisi della terminologia utilizzata per la definizione del patrimonio scientifico e tecnologico (scheda PST)<sup>48</sup>.

La pubblicazione della nuova versione del Catalogo generale dei Beni Culturali<sup>49</sup> e dei thesauri ICCD<sup>50</sup>, anche in versione LOD secondo l'ontologia ARCO<sup>51</sup>, ha consentito di rendere più efficienti le procedure per l'attivazione dei collegamenti, rendendoli disponibili anche nella versione LOD del NS.

Problematiche emerse riguardano il livello di specificità della terminologia utilizzata dai musei per la definizione degli oggetti e il riferimento a più strumenti terminologici<sup>52</sup>. Ad esempio, concetti che nel Thesaurus sono espressi da termini con aggettivazione geografica (stampe francesi, inglesi ecc.), nei thesauri ICCD sono resi con un'espressione che si riferisce all'ambito culturale<sup>53</sup>; da un punto di vista tecnico queste corrispondenze non possono essere gestite direttamente dai tipi di collegamento previsti dallo SKOS/RDF.

Inoltre, importanti collaborazioni sono state attivate con il Museo Galileo<sup>54</sup> e il Museo Egizio di Torino<sup>55</sup>. Il risultato in entrambi i casi è l'attivazione di link con la terminologia utilizzata per la definizione degli oggetti nei rispettivi cataloghi. Nel caso del Museo Galileo la cooperazione si è estesa alle collezioni digitali e all'indicizzazione delle risorse iconografiche<sup>56</sup>. Si tratta anche qui di collegamenti realizzati con ricerche in modalità HTTP GET.

## Conclusioni

Il lavoro svolto per rendere interoperabile il NS ha evidenziato quanto i dati strutturati in formato *linked open data* e il riferimento a ontologie standard abbiano permesso di integrare risorse eterogenee, prodotte nell'ambito di progetti differenti e provenienti da fonti diverse. Queste sperimentazioni hanno inoltre dimostrato come la collaborazione tra istituzioni di varia natura e professionalità differenziate sia stata fondamentale per accrescere in modo proficuo e razionale l'accesso alle informazioni.

## NOTE

---

\* Biblioteca nazionale centrale di Firenze.

<sup>1</sup> <https://www.bncf.firenze.sbn.it/biblioteca/nuovo-soggettario/>

<sup>2</sup> <https://creativecommons.org/licenses/by/4.0/deed.it>

<sup>3</sup> <https://lod-cloud.net/dataset/bncf-ns>

<sup>4</sup> [http://dati.cultura.gov.it/descrizione\\_dataset/](http://dati.cultura.gov.it/descrizione_dataset/)

<sup>5</sup> <https://www.w3.org/2004/02/skos/>

- 
- <sup>6</sup> [https://en.wikipedia.org/wiki/Knowledge\\_organization\\_system](https://en.wikipedia.org/wiki/Knowledge_organization_system)
- <sup>7</sup> <https://www.w3.org/DesignIssues/LinkedData.html>
- <sup>8</sup> <https://thes.bncf.firenze.sbn.it/thes-dati.htm>
- <sup>9</sup> [https://www.w3schools.com/tags/ref\\_httpmethods.asp](https://www.w3schools.com/tags/ref_httpmethods.asp)
- <sup>10</sup> <https://thes.bncf.firenze.sbn.it/help.php#rec>
- <sup>11</sup> <https://thes.bncf.firenze.sbn.it/stat.php>
- <sup>12</sup> <https://d-nb.info/standards/elementset/gnd>
- <sup>13</sup> <https://id.loc.gov>
- <sup>14</sup> <https://data.bnf.fr>
- <sup>15</sup> <https://datos.bne.es>
- <sup>16</sup> <https://wiki.dnb.de/display/GND/Mapping+GND+-+NSogg>
- <sup>17</sup> <https://thes.bncf.firenze.sbn.it/fonti.php>
- <sup>18</sup> <http://publications.europa.eu/it/web/eu-vocabularies>
- <sup>19</sup> <https://thes.bncf.firenze.sbn.it/collegamenti.php>
- <sup>20</sup> <https://www.bncf.firenze.sbn.it/amministrazione/convenzione-tra-bncf-e-wikimedia-italia/>
- <sup>21</sup> <https://www.jlis.it/index.php/jlis/article/view/428>
- <sup>22</sup> <https://it.wikipedia.org/wiki/DBpedia>
- <sup>23</sup> [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)
- <sup>24</sup> <https://www.wikidata.org/wiki/Property:P508>
- <sup>25</sup> <https://www.wikidata.org/wiki/Help:QuickStatements>
- <sup>26</sup> <https://www.treccani.it>
- <sup>27</sup> <https://www.fao.org/agrovoc/>
- <sup>28</sup> <https://biblio.liuc.it/pagineita.asp?codice=210>
- <sup>29</sup> <http://dati.igsg.cnr.it/dogi/doc-classificazione>
- <sup>30</sup> <https://www.getty.edu/research/tools/vocabularies/aat/>
- <sup>31</sup> <https://id.nlm.nih.gov/mesh/>
- <sup>32</sup> <https://mesh.iss.it>
- <sup>33</sup> [https://polimoda.sebina.it/opac/termlist/KF\\_ThesDescr2](https://polimoda.sebina.it/opac/termlist/KF_ThesDescr2)
- <sup>34</sup> <https://accademiadellacrusca.it/it/lingua-italiana/parole-nuove/>
- <sup>35</sup> Fra questi è importante segnalare BIANET, motore di ricerca semantico a supporto della *Bibliotheca Iuris Antiqui*: <https://bia.igsg.cnr.it/>.
- <sup>36</sup> <http://deweyit.pansoft.de/webdewey/login/login.html>
- <sup>37</sup> Si veda P. Cavaleri, *L'Indice del Nuovo soggetto nella WebDewey Italiana*, pubblicato in questo stesso numero di "Imagines".
- <sup>38</sup> <https://thes.bncf.firenze.sbn.it/collegamentiMAB.php>
- <sup>39</sup> <https://siusa-archivi.cultura.gov.it/cgi-bin/pagina.pl?RicProgetto=personalita>

---

<sup>40</sup> <https://www.archivioricordi.com>

<sup>41</sup> <https://www.eui.eu/en/academic-units/historical-archives-of-the-european-union>

<sup>42</sup> <https://icar.cultura.gov.it/attivita-e-progetti/progetti-icar-1/linked-open-data>

<sup>43</sup> <https://catalogo.uffizi.it/it/1/home>

<sup>44</sup> <http://www.iccd.beniculturali.it/it/sigec-web>

<sup>45</sup> <http://www.iccd.beniculturali.it/it/ricercanormative/139/thesaurus-per-la-definizione-dei-beni-storici-artistici>

<sup>46</sup> <http://www.iccd.beniculturali.it/it/ricercanormative/108/thesaurus-per-la-definizione-dei-reperti-archeologici>

<sup>47</sup> <https://www.fstfirenze.it>

<sup>48</sup> <http://www.iccd.beniculturali.it/getFile.php?id=1702>

<sup>49</sup> <https://catalogo.beniculturali.it/progetto>

<sup>50</sup> [https://dati.cultura.gov.it/dataset\\_locali](https://dati.cultura.gov.it/dataset_locali)

<sup>51</sup> <https://dati.cultura.gov.it/progetto-arco-architettura-della-conoscenza>

<sup>52</sup> Si veda C. Manco, E. Viti, *Biblioteca nazionale centrale di Firenze e Gallerie degli Uffizi: convergenze di obiettivi e saperi per l'interoperabilità dei dati e la condivisione dei patrimoni*, pubblicato in questo stesso numero di "Images".

<sup>53</sup> <http://www.iccd.beniculturali.it/it/ricercanormative/192/vocabolario-ambito-culturale>

<sup>54</sup> <https://www.museogalileo.it>

<sup>55</sup> <https://collezioni.museoegizio.it>

<sup>56</sup> <https://bibdigtematiche.museogalileo.it/Prospettiva/index.php/esplora/ambiente/Iconografia%20/43271>



Direzione generale  
Biblioteche e diritto d'autore



Direzione generale  
Biblioteche e diritto d'autore

Biblioteca nazionale centrale di Firenze

# Nuovo soggettario

Guida al sistema italiano di indicizzazione per soggetto

*Seconda edizione interamente rivista e aggiornata*



ISSN n. 2533-2015

Images  
è pubblicata a Firenze  
dalle Gallerie degli Uffizi