

Wikipedia



Nuovo soggettario

*Lo stato dell'arte e gli sviluppi futuri*

Assemblea di Wikimedia Italia (Firenze, aprile 2014)



Maria Grazia Pepe  
(Biblioteca nazionale Centrale di Firenze)

## Argomenti trattati

- Il Nuovo soggettario BNCF
- Wikipedia e Nuovo soggettario
- Multilinguismo e Wikidata

Assemblea Wikimedia Italia (Firenze, 5 aprile 2014)

## Il Nuovo soggettario BNCF

Il Nuovo soggettario nasce dal Soggettario 56 ed è lo strumento impiegabile nell'indicizzazione per soggetto realizzato a cura della Biblioteca nazionale centrale di Firenze.

- Aderente ai principi IFLA ed agli standard internazionali (da ISO 5964 a ISO 25964-2).
- Rivolto a biblioteche ma anche musei, mediateche, archivi, centri di documentazione.
- Strumento in continua evoluzione e accrescimento, costruito in cooperazione.
- La Bibliografia nazionale italiana lo utilizza dal 2007.

*Assemblea Wikimedia italia (Firenze, 5 aprile 2014)*

## Il Nuovo soggettario BNCF

### Bibliotecari

- **Norme** Biblioteca nazionale centrale di Firenze, *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto*. Prototipo del Thesaurus, Milano, Editrice Bibliografica, © 2006 + aggiornamenti
- **Corredo sintattico-applicativo** (note sintattiche, manuale applicativo).
- Thesaurus multidisciplinare.

### Utenti

- Opac BNCF (ricerca per soggetto).
- Cataloghi delle organizzazioni culturali che adottano il Nuovo soggettario.

*Assemblea Wikimedia italia (Firenze, 5 aprile 2014)*

## Il Nuovo soggettario BNCF

### Ambiti di cooperazione - proposte di nuova terminologia

#### Biblioteche

- ◆ [BNI](#) (Bibliografia nazionale italiana)
- ◆ [CoBis](#) (Coord. Bibl. speciali e specialistiche Torino)
- ◆ [Biblioteca Carlo Cattaneo LIUC \(Castellanza\)](#)
- ◆ [Biblioteca Centrale Giuridica](#)
- ◆ [Biblioteca dell'Accademia della Crusca](#)
- ◆ [Biblioteche della CEI](#) (polo PBE in SBN)
- ◆ [Biblioteca Polimoda](#) (Centro di documentazione Matteo Lanzoni)
- ◆ [Biblioteca Università Bocconi](#)
- ◆ [Biblioteche Università di Milano](#)
- ◆ [Biblioteche Università di Pisa](#)
- ◆ [SDIAF](#) (Sistema documentario integrato dell'Area fiorentina)

#### Istituti di ricerca

- ◆ CNR - [Istituto di teorie e tecniche dell'informazione giuridica](#)
- ◆ Università di Bologna - [Dipartimento Scienze economiche](#)

#### Archivi

- ◆ [Soprintendenza Archivistica per la Toscana](#)

#### Privati

- ◆ [Idest s.r.l.](#)

*Assemblea Wikimedia italia (Firenze, 5 aprile 2014)*

## Il Nuovo soggettario BNCF

### Ambiti di cooperazione - interoperabilità

Per il controllo terminologico viene utilizzato un apparato di [fonti](#). Il thesaurus è interoperabile con alcune tra queste:

- ◆ [AAT](#) (Art & architecture thesaurus online , The J. Paul Getty Trust)
- ◆ [AGROVOC](#) (thesaurus agricolo multilingue, FAO)
- ◆ [LIUC](#) (Thesaurus di economia e scienze sociali LIUC, Castellanza, Università Carlo Cattaneo)
- ◆ [EUROVOC](#)
- ◆ [Crusca](#) (Accademia della Crusca, "Parole nuove")
- ◆ [DoGi](#) (ITTIG, DoGi-Dottrina Giuridica. Classificazione)
- ◆ [Treccani.it](#)
- ◆ [Wikipedia](#)

In modo analogo è interoperabile con strumenti simili in altre lingue:

- ◆ [LCSH](#) (Library of Congress Subject Headings)
- ◆ [RAMEAU](#) (BNF, Répertoire d'autorité-matière encyclopédique et alphabétique unifié)

Dal 2010 il [thesaurus è disponibile in formato SKOS/RDF](#) secondo la licenza CC BY 2.5 IT

*Assemblea Wikimedia italia (Firenze, 5 aprile 2014)*

# NS e Wikipedia

## Fasi della cooperazione:

- 2011 Wikipedia come fonte.
- 2012 Link verso Wikipedia
- 2013
  - Sperimentazione automatismo per la creazione dei collegamenti.
  - Inizio della collaborazione con Wikimedia Italia.
  - Link reciproci verso Wikipedia.
- 2014 Multilinguismo e Wikidata

*Assemblea Wikimedia Italia (Firenze, 5 aprile 2014)*

# NS e Wikipedia

## Similitudini

In modo analogo le voci del NS e di Wikipedia:

- assumono forme accettate e forme non accettate che rinviano alle prime;
- possono avere una qualificazione;
- presentano delle relazioni di tipo associativo.

## Differenze

- Il Nuovo soggetto non contiene nomi propri e nomi geografici;
- in Wikipedia esistono le pagine di disambiguazione;
- Nel caso di sostantivi numerabili il Nuovo soggetto usa il plurale come forma accettata, Wikipedia il singolare.

*Assemblea Wikimedia Italia (Firenze, 5 aprile 2014)*

# NS e Wikipedia

## Automatismo per la creazione dei collegamenti

È stato utilizzato il formato RDF/XML disponibile sulla [Dbpedia italiana](#) ed accessibile attraverso un [endpoint SPARQL](#).

A partire dalle forme accettate del Nuovo soggetto, la creazione dei link avviene per corrispondenza esatta tra lemmi con attenzione particolare ai seguenti casi:

- presenza di una qualificazione;
- pagine di disambiguazione;
- **ontologie** presenti in DBpedia (**nomi propri**, **nomi geografici**, **titoli**, ecc..)

In ogni caso viene creato un link verso una forma accettata di Wikipedia.

Da giugno 2013 un automatismo realizzato dal gruppo di lavoro di Wikipedia attiva i collegamenti reciproci verso il Nuovo soggetto. La sincronizzazione è realizzata attraverso il campo P508 di Wikidata.

*Assemblea Wikimedia Italia (Firenze, 5 aprile 2014)*

# NS e Wikipedia

## Criticità dell'automatismo

### Forma accettata

<b>NS:</b> Pasquinate	<b>Wikipedia:</b> Pasquino
<b>NS:</b> Agricoltori	<b>Wikipedia:</b> Agricoltore
<b>NS:</b> Contadini	<b>Wikipedia:</b> Agricoltore

### Qualificazione

<b>NS:</b> Arbitraggio <Operazioni finanziarie>	<b>Wikipedia:</b> Arbitraggio
<b>NS:</b> Canto <Versi degli animali>	<b>Wikipedia:</b> Canto

### Pagine di disambiguazione

<b>NS/Wikipedia:</b> Banchieri
<b>NS/Wikipedia:</b> Fabbri

### Nomi propri

<b>NS:</b> Autogrill ( <i>nome comune</i> )	<b>Wikipedia:</b> Autogrill ( <i>nome proprio</i> )
<b>NS:</b> Monotype ( <i>nome comune</i> )	<b>Wikipedia:</b> Monotype Corporation ( <i>nome proprio</i> )

### Titoli

<b>NS:</b> Vittime di guerra	<b>Wikipedia:</b> Vittime di guerra ( <i>titolo di film</i> )
<b>NS:</b> Esercizi spirituali	<b>Wikipedia:</b> Esercizi spirituali ( <i>titolo di un libro</i> )
<b>NS:</b> Teatro di burattini	<b>Wikipedia:</b> Teatro dei burattini

Necessaria verifica manuale. Possibile inviare commenti al gruppo di lavoro del Nuovo soggetto attraverso una maschera accessibile a partire dalla scheda di un termine .

*Assemblea Wikimedia Italia (Firenze, 5 aprile 2014)*

# NS e Wikipedia

## Statistiche

- Numero di link sul totale delle forme accettate (41%)
- Numero di link creati con automatismo (85%)
- Numero di link verso pagine di disambiguazione (9%)

Da una verifica di qualità su un campione di circa 1000 termini è risultata una percentuale di errore non superiore al 10%.

*Assemblea Wikimedia italia (Firenze, 5 aprile 2014)*

# Multilinguismo e Wikidata

## Wikipedia e LCSH

Il multilinguismo di Wikipedia, unito all'interoperabilità con il NS può essere di supporto per la creazione di nuovi legami con strumenti di indicizzazione in altre lingue.

Con la collaborazione del gruppo di lavoro di Wikimedia Italia, dal gennaio 2014 è in corso una sperimentazione su Wikipedia (versione inglese) ed LCSH.

Sono stati seguiti due differenti percorsi:

- dai link tra il NS e Wikipedia (IT), attraverso Wikipedia(EN) a LCSH;
- dalla [mappatura tra LCSH e Wikipedia \(EN\)](#) eseguita da Mark Ockerbloom (Università della Pennsylvania), attraverso Wikipedia (IT) al NS.

La sperimentazione ha prodotto circa 1800 possibili link tra NS ed LSCH.

I potenziali errori nei collegamenti creati in automatico tra NS e Wikipedia (IT) impongono una verifica di qualità su questo risultato.

Circa 100 tra questi link tuttavia saranno già disponibili nella versione di marzo 2014 del Nuovo soggettario.

*Assemblea Wikimedia italia (Firenze, 5 aprile 2014)*



# Multilinguismo e Wikidata

## DBPedia o Wikidata

Gli strumenti di lavoro della sperimentazione sono state le versioni RDF/XML di Wikipedia, rese disponibili da [DBPedia\(IT\)](#) e [DBPedia\(EN\)](#) e la [versione SKOS/RDF di LCSH](#).

[Wikidata](#) fornisce un accesso centralizzato ai dati di Wikipedia, presenti in forma strutturata, individuati univocamente da un identificatore e descritti da etichette espresse in tutte le lingue.

Pur non utilizzando ancora l'RDF, il multilinguismo di Wikidata, unito al formato aperto dei dati ed all'uso di identificatori univoci, potrebbe farla diventare una piattaforma comune di lavoro in vista della creazione di una più ampia rete di collegamenti tra liste di autorità in diverse lingue.

*Assemblea Wikimedia Italia (Firenze, 5 aprile 2014)*

Grazie per l'attenzione!!!



Gruppo di lavoro del Nuovo sogettoario

[nuovosoggetario@bncf.firenze.sbn.it](mailto:nuovosoggetario@bncf.firenze.sbn.it)