

Florence–Washington Round Trip: Ways and Intersections between Semantic Indexing Tools in Different Languages

Anna Lucarelli & Elisabetta Viti

To cite this article: Anna Lucarelli & Elisabetta Viti (2015) Florence–Washington Round Trip: Ways and Intersections between Semantic Indexing Tools in Different Languages, *Cataloging & Classification Quarterly*, 53:3-4, 414-429, DOI: [10.1080/01639374.2014.1004008](https://doi.org/10.1080/01639374.2014.1004008)

To link to this article: <http://dx.doi.org/10.1080/01639374.2014.1004008>



Published online: 19 Mar 2015.



Submit your article to this journal [↗](#)



Article views: 95



View related articles [↗](#)



View Crossmark data [↗](#)

Florence–Washington Round Trip: Ways and Intersections between Semantic Indexing Tools in Different Languages

ANNA LUCARELLI and ELISABETTA VITI
National Central Library of Florence, Florence, Italy

This article presents an Italian experience of developing streamlined semantic interoperability between the Italian Thesaurus of Nuovo soggettario and the Library of Congress Subject Headings (LCSH). This ongoing project must take into consideration the differences between the two indexing tools, while the criteria on which the resulting actions are based are being clarified continually. Reciprocal interoperability, thanks to the Simple Knowledge Organization System format, enables us to create links with English language subject headings. The National Central Library of Florence is studying methods of automatically catching LCSH equivalents and the question of how to take advantage of both Semantic Web outputs and the multilingual dataset of Wikidata.

KEYWORDS *linked data, Library of Congress Subject Headings (LCSH), Nuovo soggettario, indexing vocabularies, interoperability, subject access*

INTRODUCTION

This article outlines the Italian perspective on interoperability between subject indexing tools in different languages, with a particular focus on the work undertaken at the National Central Library of Florence (BNCF) toward

© National Central Library of Florence (Italy)

Received June 2014; revised December 2014; accepted December 2014.

Address correspondence to Elisabetta Viti, Research and Tools for Subject Indexing and Classification, National Central Library of Florence, Piazza Cavallotti 1, Florence 50122, Italy. E-mail: elisabetta.viti@beniculturali.it

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/wccq.

linking the vocabulary of the *Nuovo soggettario* (NS) and the *Library of Congress Subject Headings* (LCSH).

The NS is a system consisting of codified rules (concerning terminology and syntax), and of a general Thesaurus; it is edited by the BNCF.¹ It is based on the analytico-synthetic model and uses syntactic role analysis for the building of subject strings that are created using both controlled vocabulary terms and the rules of a conventional syntax. It is a flexible and modular system to be applied both in general and specialized information environments, characterized by different types of resources (graphic material, audiovisual, etc.) or by different domains. It is made up of some interactive components: the general Thesaurus (compliant with International Standards Organization [ISO] standards), the rules for vocabulary control and the construction of subject strings, and the Manual for users. The Thesaurus is the core of the whole system and has been accessible on the web since 2007. It currently consists of approximately 52,000 terms and is continuously evolving. It is steadily growing thanks to the efforts of the BNCF and, since 2009, also by the external collaboration of several institutions.² In contrast to the LCSH, which explicitly list simple and complex subject headings and precoordinated strings in only one verbal indexing scheme, the NS system separates semantic and syntactic components.

Both lexical alignment and the reconciliation of subject vocabularies can facilitate the semantic enrichment and integration of heterogeneous collections. The conversion of controlled vocabularies to mark up formats, suitable for the Semantic Web and Linked Data world, will improve their function as interlinking hubs for direct searching by different professional communities.

From an international perspective, many projects have concentrated on linking different subject vocabularies in the same language or in different languages, among various classification schemes, as well as between vocabularies and classification schemes. Some of these projects deal with the interdisciplinary, special, and multilingual domains. Although the question has been studied for many years, work on the topic has advanced significantly during the last few years.³ We can also cite various scientific meetings in Europe, for example the meeting in Cologne in 2010: *Concepts in Context. Cologne Conference on Interoperability and Semantics in Knowledge Organization*,⁴ and the international meeting in Florence in 2012: *Global Interoperability and Linked Data in Libraries*.⁵ Furthermore the topic of integration between different cataloging tools is closely connected with linked data services managed by National libraries. Not all interoperability projects have always achieved success conceptually, structurally, and technically. However, some of them were a reference for the latest standards.⁶

Since 1985, with the introduction of the ISO 5964 standard, the scientific community has worked to put forward specific standards and guidelines for multilingual thesauri and for interoperability between thesauri in

different languages. ISO 5964 is regarded as an extension of the scope of the monolingual guidelines covered by ISO 2788-1986, and is restricted to the problems of multilingualism. It is designated as a reference point for subsequent standards such as ANSI/NISO Z39.19-2005 and BS 8723-2005/2008. The latest standard, ISO 25964, looks at the question of multilingualism and mapping from the point of view of the Semantic Web and linked data.⁷

It is well known that these topics require further clarification. It is important not to confuse problems regarding multilingualism with questions of interoperability. Furthermore, the establishment and development of multilingual thesauri is not the same as the linking of different vocabularies or thesauri that might include one or more languages.

Mapping should have real and practical aims:

1. Mapping is useful to improve and amplify search results. When librarians establish crosswalks (mapping tables) between terms/concepts in two or more structured vocabularies (equivalence relationships are established between terms/notations/phrases representing the same concepts in different subject systems and/or languages) they support the ability to search effectively across heterogeneous resource databases, even though the works are indexed with different indexing tools;
2. Mapping links metadata schemes together in a way that is useful to the Semantic Web.

SOME PROBLEMS

Standards are only guidelines and as such do not solve all problems. Some semantic and technical aspects emerge only when we try to create indexing tools. For example, the establishment and development of multilingual thesauri involves some semantic and structural problems that are connected with different professional, technical, social, linguistic, and cultural contexts. The interoperability between two different indexing tools involves several aspects, such as the conceptual structure of the tools and languages that are connected; for example, the links between the thesauri and subject list, subject heading authorities or pre/post-coordinated languages. Other aspects focus on compatibility between specific and general domains,⁸ the level of domain specificity of the thesauri converted, the meaning of terms (polysemy and monosemy), the conventions of grammar and syntax of a language in regional or national use, the percentage of linking between source thesaurus and target thesaurus, the type of links established (exact or partial equivalence; broader or narrower equivalence), and so on.

Fundamental progress has been made with ISO 25964 because it analyzes and solves these problems, above all through an approach that favors

“concepts” over “terms.” In particular, it clarifies the difference between the concepts of a search for information and the terms in which the query is expressed. Stella Dextre and Marcia Lei Zeng point out that “to perform on the Semantic Web, computer software needs an explicit data model that distinguishes between terms and concepts. . . . All the while a concept is inside our heads, it can be independent of words or language. But as soon as we try to communicate it to another person or to a search system, we have to represent it in some way—usually by words or codes or pictures. The only way a thesaurus can list concepts in alphabetical order is by representing them as a term. . . .”⁹

The Simple Knowledge Organization System (SKOS) data model¹⁰ allows us to explicitly distinguish between concepts and the labels used to represent them. In this data model the concept is a formal entity identified by Uniform Resource Identifier (URI) and label `rdf:type`; it is independent from natural language, documents, and users.

THE ITALIAN EXPERIENCE

In Italy, the most important practical experience of interoperability between different indexing tools is the one set up for NS with LCSH and maintained by the National Library of Florence.

The architecture of the NS system consists of various components, of which the first is the Thesaurus.¹¹ Its open source software was founded from AGROVOC,¹² the multilingual thesaurus of the Food and Agriculture Organization (FAO). To date, the NS thesaurus contains over 52,000 terms, almost all of them in Italian, and has been used by various types of Italian libraries and cultural institutions. Since 2010, all metadata (just available by Zthes protocol) are available in SKOS/Resource Description Framework (RDF) format, which establishes tags for management interoperability.¹³ The NS metadata can be downloaded free of charge and used in accordance with the Creative Commons Attribution guidelines.¹⁴

A recent innovation is that some terms are linked to the English equivalent established by the LCSH,¹⁵ the most important indexing system in Anglo-American countries and also in other national versions or translations in other countries. The links establish equivalence relationships between some NS preferred terms and some LCSH preferred terms. To date, there are no links between non-preferred terms or variants (synonyms or quasi-synonyms and earlier established forms). The two systems are quite different, and it is impossible that there will ever be a complete set of links for all terms as the terms and their structures vary greatly. These differences do not prevent the implementation of partial linking. Additionally LCSH is linked with other thesauri, for example the Art & Architecture Thesaurus (AAT),¹⁶ which enables links to NS through these additional thesauri.

The project fits in with the wider scope of the BNCF's work to integrate the NS with other Knowledge Organization Systems (KOS) available on the web, such as *Répertoire d'autorité-matière encyclopédique et alphabétique unifié* (RAMEAU), AGROVOC, and Eurovoc thesauri, Treccani.it,¹⁷ and Wikipedia.it encyclopedias (the latter having over 12,000 reciprocal links with NS) (Table 1). Recently, these links have increased thanks to RDF/SKOS formats.

The NS Thesaurus includes descriptors but not subject heading strings. In this, it differs from other controlled vocabularies and from other enumerative schemes such as LCSH or RAMEAU, which are built on authority headings for subject, complex subjects, proper name, title, and so on.

In the NS system, subject headings are established by a single descriptor, or through combining many terms into pre-coordinated strings, using syntactic rules.¹⁸ These are clearly visible in the subject headings archives and in the online public access catalogs (OPACs) of the libraries that have adopted the NS system. In addition, the Thesaurus does not include personal or corporate names or title headings; but it includes the names of historical events.¹⁹

LINKING BETWEEN THE NS THESAURUS AND LCSH:
PECULIARITIES AND PROBLEMS

Our aim is not to establish a multilingual thesaurus, but to develop streamlined semantic and technical interoperability between the NS Thesaurus and the LCSH Linked Data Service.

What is this type of realized interoperability? First of all, it enables us to recognize the English equivalent of an Italian term and to display the full

TABLE 1 Links between NS Thesaurus and Other KOS

Links with Other KOS (April 2014)	Links	Subject Indexing Tools (Thesauri, Classification Systems, Etc.)		
			Encyclopedias	Other
Wikipedia	12122		X	
Treccani.it	9356		X	
LCSH	5144			
AGROVOC Thesaurus	1724	X		
RAMEAU	1639	X		
ThESS Thesaurus	740	X		
EUROVOC Thesaurus	165	X		
DOGI (Ittig)	162	X		
AAT	74	X		
Accademia della Crusca	8			X

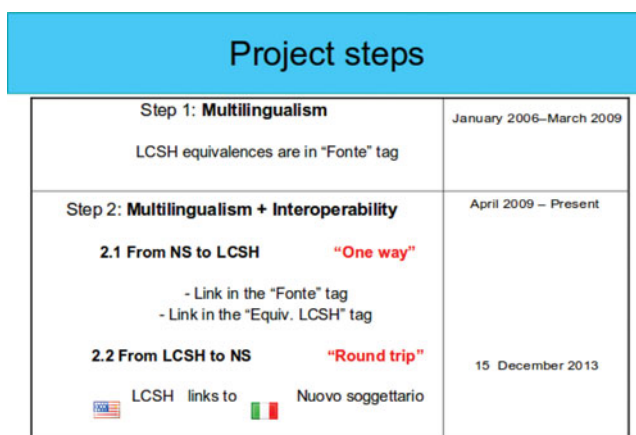


FIGURE 1 *Nuovo soggetto* and LCSH Connections.

authority record for the term, and vice versa. Thus, it provides a “round trip” between subject indexing tools that differ in language and in structure (i.e., English and Italian). The advantages of this interoperability are many and it is anticipated that some are yet to be discovered.

The beneficiaries of the project obviously are not only the users of the NS and LCSH but also the users of web services that use KOS to organize their contents.

The level of development of these equivalences in recent years has grown considerably. As of October 2014, over 7,000 English equivalents have been assigned to Italian terms, thanks to the intellectual work of translation and multilingual task of NS software. The terms are shown in the NS system in the same style (order of words and double dashes) as the American tool.

The percentage of LCSH equivalents is around 25% of all NS preferred terms.

Different domains have different percentages of linking. In some specialized sectors, such as Economics, Library and Information Science, History, and Literature, the LCSH equivalences occur more than in others. For example, the percentage of linked terms in Economics is higher than other disciplines, thanks to the partnership of Bocconi University. Their librarians suggest new terminology for NS in Economics and Financial disciplines and establish intellectual cross-concordances between NS and LCSH.

STEPS OF THE PROJECT

This project has evolved through several steps since 2006 (see Figure 1). From 2006–March 2009 the NS included LCSH equivalent terms in the “Fonte”

field as a statement to refer users to the alternate terminology without links.

From 2009 to 2011, hyperlinks to the online catalog of the Library of Congress (LC) were added. By clicking on the arrow before “LCSH” in the “Fonte” field, a link was activated to the LC online catalog (<http://catalog.loc.gov/>) to launch a query within the LC system (see Figure 2). LCSH equivalents also have been assigned to some NS terms using the multilingual task of the thesaurus software (original software: Food and Agriculture Organization of the United Nations [FAO], AGROVOC).

In this step, equivalences were established between single terms, rather than with LCSH complex subjects, even in cases of semantic equivalence, and only for new Italian terms added to the Thesaurus.

From 2011 to 2013, LCSH equivalences were shown in the “Fonte” field and in the new dedicated field: “Equiv. LCSH.”

In 2014, the equivalent term is indicated only in this new field, “Equiv. LCSH,” and completely deleted from the “Fonte” field. Thanks to a deep link, activated by the term, one can see the exact equivalent of the target tool (Figure 3).

Not all NS terms have LCSH equivalent terms. This may be for many reasons (e.g., an equivalent term cannot be identified and assigned; an equivalent term does not exist; more than one term is needed to express the concept in the alternative language). In none of these cases does the NS give any indication, contrary to RAMEAU, which gives the clear explanation: “Pas d’équivalents.”²⁰

Ricerca

- per termine
- per termine apicale
- per categoria
- classificazione Dewey
- ricerche combinate
- per variante storica

Soggettario(1956)
Liste di aggiornamento
(1956-1985)

- ricerca
- termini soppressi
- crediti
- aiuto

Sistema
Nuovo soggettario

- home page

Parchi naturali

Categoria/Faccetta: Cose:Spazio
Nota d'ambito: Aree terrestri, fluviali, lacuali o marine che contengono uno o più ecosistemi intatti o anche parzialmente alterati da interventi antropici, una o più formazioni fisiche, geologiche, geomorfologiche, biologiche, di rilievo internazionale o nazionale per valori naturalistici, scientifici, estetici, culturali, educativi e ricreativi tali da richiedere l'intervento pubblico ai fini della loro conservazione per le generazioni presenti e future

TT Spazio
BT Aree protette
NT Parchi a fini multipli
Parchi fluviali
Parchi nazionali
Parchi regionali
RT Ecosistemi
Ecoturismo
Riserve naturali

Fonte: BNI 1956-1985; PRNI; → AGROVOC: National parks; → LCSH: National parks and reserves; → TRT; I. 6/12/1991
Classificazione Dewey (Ed. 21): 333.783
Agenzia catalografica/Proponente: BNI

FIGURE 2 The LCSH Acronym in the “Fonte” Field. <http://thes.bncf.firenze.sbn.it/termine.php?id=1205&menuR=2&menuS=2>

+ Ricerca per termine del Nuovo soggetto

+ Soggettario(1956)
Liste di aggiornamento (1956-1985)

Sistema Nuovo soggetto

- home page
- manuale applicativo

Cartolarizzazione

Categoria/Faccetta: Azioni:Attività

UF: Cartolarizzazione di crediti, Securitization, Securitizzazione

TT: Attività

BT: Operazioni finanziarie

NT: Collateralized debt obligation

RT: Cessione di crediti

Titoli di rendita

HSF: Cartolarizzazione di crediti

Definizione: Cessione dei propri crediti pecuniari a una società appositamente costituita perché si occupi della loro trasformazione in titoli negoziabili sul mercato finanziario (DeMagg)

Nota storica: Precedentemente: Cartolarizzazione di crediti (BNI 1999-2005)

Fonte: BNI 1999-2005: Cartolarizzazione di crediti; *Crusca*; DeMagg; PTsuppl; NDBBF: Securitization; *RAMEAU*: Titrisation; *Wikipedia*(IT)

Classificazione Dewey (Ed. 21): 332.645

Equiv. LCSH: Asset-backed financing

FIGURE 3 A New Field, “Equiv. LCSH,” to Deep Link.

ADDITIONAL DEVELOPMENTS

In the course of the project new functions have been added. In fact, since 2012 LCSH equivalent terms are displayed near the Italian terms in the list of the results generated by clicking on “Ricerca per termine” and are directly searchable (Figure 4). Also, all terms can be shown in alphabetic order by clicking on “equivalenti LCSH” in the left-hand column.

LCSH represents the principal hub, using the Multilingual Access to Subjects (MACS) project to establish equivalences among the three indexing languages: LCSH, RAMEAU, and Schlagwortnormdatei (SWD).²¹ Therefore linking to LCSH allows the NS to benefit from the semantic links available through the MACS project.

acido s **esegui**

A|B|C|D|E|F|G|H|I|J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z

termini trovati: 8

Acido salicilico (LCSH: Salicylic acid)	Cose:Materia
Acido solfamico	Cose:Materia
Acido solfamnico	Cose:Materia
Acido solfidrico	Cose:Materia
Acido solforico (LCSH: Sulfuric acid)	Cose:Materia
Acido solforoso	Cose:Materia
Acido sorbico (LCSH: Sorbic acid)	Cose:Materia
Acido succinico (LCSH: Succinic acid)	Cose:Materia

FIGURE 4 LCSH Equivalents are Displayed Near the Italian Term in the List of the Results.

Financial crises

<http://id.loc.gov/authorities/subjects/sh89002668.html>

Closely Matching Concepts from Other Schemes

-  [Börsenkrach](#) 
-  [Crises financières](#) 
-  [Crisi finanziarie](#) 
-  [Kurssturz](#) 

FIGURE 5 An Example of Equivalence Links among LCSH, SWD, RAMEAU, and *Nuovo soggettario*.

In mid-December 2013, thanks to our contacts with the LC's Linked Data Service, we began creating round-trip data linking. The Library of Congress Linked Data Service hooks to the NS terms that have LCSH equivalents, using the SKOS format. In the records of the LCSH database, among the “Closely Matching Concepts from Other Schemes,” Italian terms are clearly visible beside French (RAMEAU) and German (SWD) terms (Figure 5). In this way, reciprocal interoperability has been opened between LCSH descriptors and their Italian equivalents.

The links between NS and LCSH terms/concepts can be grouped into two main categories, as laid out by the ISO standard 25964-2, § 7–11:

1. exact or almost exact semantic equivalent;
2. partial equivalent.

In addition, we can identify additional and more specific types of equivalence.

- The LCSH concept is clearly broader than the NS concept. In this case, the NS uses *skos:closeMatch*, for non-restrictive equivalences (e.g., Cerimonie [NS] EQ Rites and ceremonies [LCSH]).
- The LCSH concept is clearly narrower than the NS concept. Also in this case, the NS uses *skos:closeMatch* (e.g., Cellule neoplastiche neoplastiche [NS] EQ Cancer cells [LCSH]).
- The LCSH term expresses a concept belonging to a different category or hierarchical structure from the NS term (e.g., Banks [NS] EQ Banks and Banking [LCSH]).
- The NS concept is expressed in LCSH by two LCSH terms. In such cases, the correspondence is activated in the NS by one of the two terms only: the term that is more semantically similar. The decision is based on careful lexical and linguistic research (e.g., Chiese ortodosse [NS] EQ Orthodox Eastern Church or Oriental Orthodox churches). We selected Oriental Orthodox churches.

- The NS concept has an equivalent LCSH term with two different IDs. There are two circumstances in which this can occur:
 - when in LCSH there is an identifier for the term as a topic and another for the term as a topic-general subdivision (e.g., Propaganda; Histology)
 - when in LCSH there is an identifier for the term as an LC Subject Heading and another for the same term as it is present in other LC vocabularies, such as LC Children’s Subject Headings (e.g., Basque language). In such cases, the NS activates the match with the ID used in the LC Subject Headings. In the SKOS interface the links to both IDs are active and visible.
- The compound terms in NS are equivalent to LCSH terms consisting of:
 - two elements separated by a comma (e.g., Ebrei francesi [NS] EQ Jews, French [LCSH])
 - two elements connected by a conjunction (*And headings*) (e.g., Turismo enogastronomico [NS] EQ Tourism and gastronomy [LCSH])
- The NS term is equivalent to an LCSH subject string (complex subject in the form of a compound heading), with its own ID. The NS term links to the compound heading and not to the single component parts. In this case, as for all the other equivalents, the LC subject headings are visualized in the NS in their original graphic form (e.g., Inquinamento atmosferico [NS] EQ Air – Pollution [LCSH]).

It is even possible that the equivalents between LCSH, NS, RAMEAU, and SWD could coincide. For example, the term “A priori,” a philosophical concept, has been expressed in the same way in all of these resources. In other cases, the only multilingual link available to LCSH is that of the NS: “pasta factories” is a particularly interesting example of such an occurrence, as previously the preferred LCSH term was “Macaroni factories.”

A SELECTION OF CASES AND PROBLEMS

Since 2012, the methodology followed to assign English equivalents has been documented by a constantly updated report. This technical report is very interesting, because it details the semantic cases encountered and the criteria adopted in deciding the methodology to be used in creating links.

To date, many aspects of the BNCf project are progressing and there are also some critical points still requiring decisions. For example, decisions must be made about:

- establishing equivalences between variants and preferred terms;
- finding a method to lay out links between concepts that change their preferred forms over time; for example, “Tensor calculus” has been changed

with “Calculus of tensors,” while “Ribonucleic acid” has been changed with “RNA.” The problem could be solved with manual labor based on recurrent and expensive revision, or by making the most of automatic functions allowed by SKOS;

- developing a new style of display.

CHARACTERISTICS AND ADVANTAGES

In summary, let us review the characteristics and advantages of the BNCf project:

- this kind of linking is challenging, because the indexing tools themselves are different not only in language but also in structure;
- interoperability has been possible because the NS and LCSH are both based on shared standards;
- thanks to the work of the BNCf, the Library of Congress has established reciprocal equivalence;
- thanks to mapping, a number of semantic and morphological problems have been discovered. We could consider the possibility to evaluate corrections of form and style in certain NS terms;
- users of other languages and cultures may discover and explore the Italian OPAC through the use of LCSH terms;
- BNCf and Servizio Bibliotecario Nazionale users can search the Library of Congress OPAC and any OPACs that use the LCSH system;
- these types of links between subject indexing tools encourage other links; for example, between OPACs of different libraries;
- other Italian institutions that use LCSH (such as the Vatican Library) could also take part in the BNCf project.

ONGOING AND FUTURE WORKS: QUESTIONS AND PERSPECTIVES

The development of these interoperability projects opens up new horizons and prospects. Is it useful to establish multilingual thesauri or would it be better to take advantage of pre-existing links between controlled vocabularies in different languages, above all in multidisciplinary contexts?

The BNCf is studying ways to automatically catch LCSH equivalents. We would like to develop automated methods to increase links, even if we know that they would require some human quality control. Our approach requires both machines and human intervention for the evaluation of the quality of these links.

It would also be useful to identify a way in which to synchronize the updates made to both databases. At present LCSH is updated on a weekly

basis; the NS search interface (where users are able to view links to LCSH), however, is updated each semester. It would therefore be advantageous to implement ways in which changes or variants and “Earlier Established Forms” may be reflected in real time.

Certainly we could take advantage of pre-existing links between the NS and others tools (e.g., AGROVOC), which have just been linked to LCSH, to import links automatically.

A link between the NS and the Italian WebDewey (under construction)²² could also have an important role. Indeed, WebDewey already allows us to visualize LC Subject Headings and could, as a result, provide a “bridge” for the acquisition of further equivalents within the NS.

Certainly, we will work to take advantage of Semantic Web capabilities. The BNCf has drawn up an agreement with Wikimedia Italia²³ in order to optimize NS development through to data sharing.

We would like to expand interoperability between subject access data regarding institutions’ library collections to an interoperability based on other types of structures of free, collaborative, and multilingual data such as Wikidata,²⁴ a knowledge base of all languages.

COULD THE NS AUTOMATICALLY HOOK LCSH EQUIVALENTS USING WIKIDATA?

Our preliminary aim was to take advantage of pre-existing links between the NS, Italian Wikipedia, and the English Wikipedia, in order to match automatically with LCSH equivalents and to connect these to non-linked NS terms.

On the other side of the world (in Pennsylvania), the librarian John Mark Ockerbloom²⁵ has tested the crosswalks between LCSH and DBpedia. Using the same procedure, we have matched those crosswalks with our terms, acquiring, as a result, a further 600 new links for the NS.

Moreover, we now aim to establish crosswalks between the two subject indexing tools by means of Wikidata. The picture featured in Figure 6 shows the potential connection of the NS and LCSH with Wikidata, DBpedia, and Wikipedia and their roles. In this figure, the link from the NS to Wikipedia and LCSH is shown in a single direction but, of course, it could be possible in two directions. Figure 7 presents the idea of using DBpedia and Wikidata to build a bridge between the NS and LCSH, automatically or semi-automatically. The uninterrupted line arrows show direct links; the broken lines mark indirect links; the dotted lines show the various linguistic versions of Wikipedia; the blue arrows stand for the correspondence only for a subset of terms; the red arrows represent what we need: automatic links between the NS Thesaurus and LCSH, and vice versa.

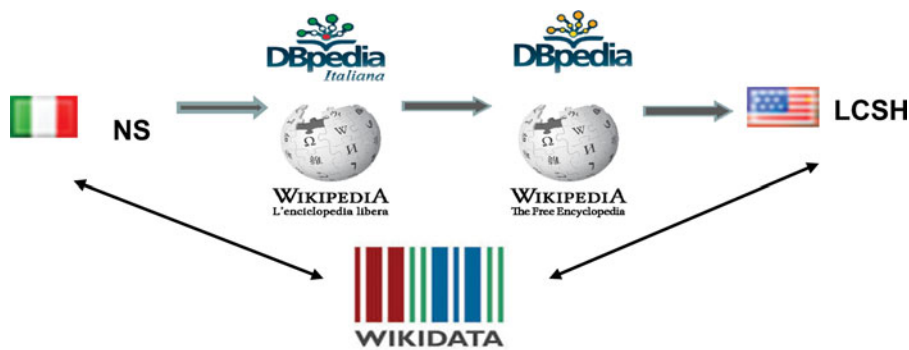


FIGURE 6 Could NS Automatically Hook LCSH Equivalents Using Wikidata?

Wikipedia.it and its English-language counterpart are two versions of Wikipedia in different languages and are part of a global metadata system called Wikidata. The latter is under development and its data are not marked up in RDF format.

However, both It.DBpedia.org and DBpedia.org contain Wikipedia’s data in RDF format and represent an interlinking hub from the thesauri to Wikipedia and vice versa. So, one can connect the NS with LCSH either using SKOS format and the DBpedia hub, or using Wikidata.

Wikidata lets us identify a concept expressed in multiple languages. Using a tool developed by Magnus Manske,²⁶ showing all the terms in Wikidata that are linked to the NS, we can display the equivalents in other languages

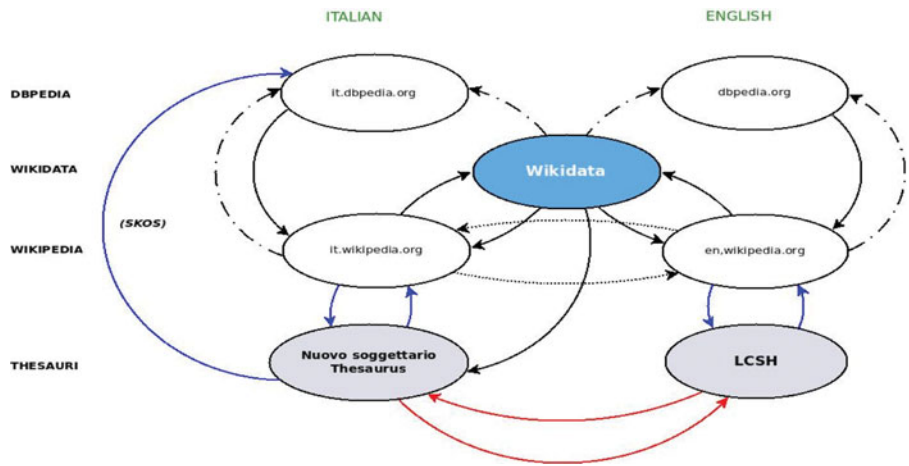


FIGURE 7 Using DBpedia and Wikidata to Build a Bridge between NS and LCSH. © Cristian Consonni. Reproduced by permission of Cristian Consonni (Fondazione Bruno Kessler and Vice President of Wikimedia Italia), cristian.consonni@wikimedia.it. Permission to reuse must be obtained from the rightsholder.



FIGURE 8 From “Universe” of Wikidata (Connected with *Nuovo soggettario*) to the Term in the Chinese Version of Wikipedia.

and their connected versions of Wikipedia. In the example in Figure 8, we can see the Chinese term in the Chinese version of Wikipedia for the Italian NS term, “Universe.”

The National Central Library of Florence intends to take its future steps in all of these directions and develop the potential offered by the Wikidata data store, not only for the implementation of a multilingual thesaurus but also numerous functionalities and applications in the metadata universe, an aim that is shared by specialists across the world of information science and libraries.

NOTES

1. *Nuovo soggettario* system, <http://thes.bncf.firenze.sbn.it/> (accessed December 29, 2014). For a description of the *Nuovo soggettario* system, see: Alberto Cheti, Anna Lucarelli, and Federica Paradisi, “Subject Indexing in Italy: Recent Advances and Future Perspectives,” *International Cataloguing and Bibliographic Control* 39, no. 3 (2010): 47–52.

2. For more details on the thesaurus, see: Luciana Franci et al., “The Nuovo Soggettario Thesaurus: Structural Features and Web Application Projects,” in *Subject Access: Preparing for the Future*, ed. Patrice Landry et al. (Berlin: De Gruyter Saur, 2011), 155–168.

3. Several studies and reports demonstrate that there has been interest in this topic for many years; for example, Stella G. Dextre Clarke and Marcia Lei Zeng, “From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards Interoperability and Data Modeling,” *Information Standards Quarterly* 24, no. 1 (2012): 20–26, http://www.niso.org/apps/group_public/download.php/8289/SP_clarke_zeng_isqv24no1.pdf; Stella G. Dextre Clarke, “Thesauri, Interoperability and the Role of ISO 25964” (paper presented at the meeting *Search Solutions 2012: Innovations in Web & Enterprise Search*, Thursday, November 29, Covent Garden, London), <http://irsg.bcs.org/SearchSolutions/2012/sse2012.php>;

Stella G. Dextre Clarke, "ISO 2788 + ISO 5964 + Much Energy = ISO 25964," *Bulletin of the American Society for Information Science and Technology* 35, no. 1 (2008): 31–33, http://www.asis.org/Bulletin/Oct-08/OctNov08_DextreClark.pdf; Stella G. Dextre Clarke, "ISO 25964: A Standard in Support of KOS Interoperability" (paper presented at the meeting, *Facets of Knowledge Organization: A Tribute to Professor Brian Vickery 1918–2009*, ISKO UK Biennial Conference, July 4–5, 2011, London), <http://www.iskouk.org/conf2011/papers/dextreclarke.pdf>; Leonard Will, "The ISO 25964 Data Model for the Structure of an Information Retrieval Thesaurus," in *Categories, Contexts and Relations in Knowledge Organization Proceedings of the Twelfth International ISKO Conference, 6–9 August 2012, Mysore, India*, eds. A. Neelamegham and K. S. Raghavan (Würzburg: Ergon, 2012), 284–290, http://www.willpowerinfo.co.uk/Will_ISKO2012_paper.doc (accessed October 15, 2014).

4. "Concepts in Context. Cologne Conference on Interoperability and Semantics in Knowledge Organization," http://linux2.fbi.fh-koeln.de/cisko2010/index_en.html (accessed October 15, 2014).

5. "Global Interoperability and Linked Data in Libraries," <http://www.linkedheritage.eu/linkeddataseminar/> (accessed October 15, 2014).

6. Stella G. Dextre Clarke, "In Pursuit of Interoperability," in *Concepts in Context. Cologne Conference on Interoperability and Semantics in Knowledge Organization, 19–20 July* (Cologne, Germany: Ergon Verlag, 2011), http://ixtrieve.fh-koeln.de/cisko2010/presentationen/2010-07-20_dextre_clarke.ppt [slides], http://linux2.fbi.fh-koeln.de/cisko2010/mp3/2-01_dextre_clarke.mp3 [mp3] (accessed October 15, 2014).

7. International Organization for Standardization, *Information and Documentation, ISO 25964-1: 2011: Thesauri and Interoperability with Other Vocabularies. Part 1: Thesauri for Information Retrieval* ([Geneva]: International Organization for Standardization, 2011); International Organization for Standardization, *Information and Documentation, ISO 25964-2: 2013 Thesauri and Interoperability with other Vocabularies. Part 2: Interoperability with Other Vocabularies* (Geneva: International Organization for Standardization, 2013). For more information on Part 2, see Stella Dextre Clarke, "Applying ISO25964 to Thesaurus Mapping and Other Forms of Linkage," in the 13th European Networked Knowledge Organization Systems (NKOS) Workshop held during the DL2014 Conference in London, September 11–12, 2014, <https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2014/content/NKOS2014-presentation-dextreclarke-ISO25964Mappings.pdf> (accessed October 15, 2014).

8. For more about the interoperability between an Italian universal thesaurus and a specific thesaurus, see: Elisabetta Viti, "Interoperabilità fra thesauri generali e thesauri specialistici in ambito economico-finanziario. Il caso del *Nuovo soggettario*" (PhD diss. Università degli studi di Udine, 2012), <http://dspace.uniud.cilea.it/bitstream/10990/72/1/Tesi%20di%20dottorato%20in%20Scienze%20bibliografiche%20%28XXIV%20ciclo%29-Universit%C3%A0%20degli%20studi%20di%20Udine-Elisabetta%20Viti.pdf> (accessed October 15, 2014); Elisabetta Viti, "Interoperabilità fra thesauri generali e thesauri specialistici in ambito economico-finanziario. Il caso del *Nuovo soggettario* Sintesi e risultati di un progetto di ricerca," in *1. Seminario nazionale di biblioteconomia: didattica e ricerca nell'università italiana e confronti internazionali, Roma, 30–31 maggio 2013*, (Milano: Ledizioni, 2013), http://w3.uniroma1.it/seminario-biblioteconomia/wp-content/uploads/2013/05/Viti_Elisabetta.pdf (accessed October 15, 2014).

9. Dextre Clarke and Zeng, "From ISO 2788 to ISO 25964," 20–21.

10. "SKOS Simple Knowledge Organization System—Home Page," 2014, <http://www.w3.org/2004/02/skos/> (accessed October 15, 2014).

11. "*Nuovo Soggettario* Thesaurus," http://thes.bncf.firenze.sbn.it/ricerca_eng.php (accessed October 15, 2014).

12. "AGROVOC," <http://aims.fao.org/website/AGROVOC-Thesaurus/sub> (accessed October 15, 2014).

13. About the conversion testing, see: Giovanni Bergamin and Anna Lucarelli, "The *Nuovo soggettario* as a Service for the Linked Data World" (paper presented at the meeting *Global Interoperability and Linked Data in Libraries*, University of Florence, June 18–19, 2012); Giovanni Bergamin and Anna Lucarelli, "La conversione dei metadati del Thesaurus del *Nuovo soggettario* nel formato SKOS/RDF" (paper presented at the Università La Sapienza, Rome May 3, 2013). About the conversion of RAMEAU and LCSH to the RDF Semantic Web format, see: Antoine Isaac and Thierry Bouchet, "RAMEAU et SKOS: une expérimentation," *Arabesques*, no. 54 (2009), http://rameau.bnf.fr/informations/pdf/arabesques54_art_isaac_bouchet.pdf (accessed October 15, 2014); Ed Summers, Antoine Isaac, Clay Redding, and Dan Krech, "LCSH, SKOS and Linked Data" (papers and project reports presented at the *International Conference on Dublin Core and Metadata Applications*,

Berlin, September 22–26, 2008), <http://dcpapers.dublincore.org/pubs/article/view/916/912> (accessed October 15, 2014). For another implementation of the SKOS model, see: Sirje Nilbe and Tiit Tarkpea, "Using the Estonian Subject Thesaurus in the Digital Environment," *Cataloging & Classification Quarterly* 52, no. 1 (2014): 32–41.

14. "Thesaurus Availability by Formats and Standard Protocols," http://thes.bncf.firenze.sbn.it/thes-dati_eng.htm (accessed October 15, 2014).

15. "Library of Congress Subject Headings," <http://id.loc.gov/authorities/subjects.html> (accessed October 15, 2014).

16. Seth van Hooland et al., "Evaluating the Success of Vocabulary Reconciliation for Cultural Heritage Collections," *Journal of the American Society for Information Science and Technology* 64, no. 3 (2013): 464–479.

17. Treccani is the most prestigious universal Italian encyclopedia, and has been linked to the NS since 2009. As of October 2014, there are nearly 10,000 links.

18. Biblioteca nazionale centrale di Firenze, *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto. Prototipo del Thesaurus* (Milano: Editrice Bibliografica, © 2006).

19. "Criteri (§ Termini non appartenenti ai nomi comuni inseriti nel Thesaurus)," <http://thes.bncf.firenze.sbn.it/criteri.php?menuR=2&menuS=2> (accessed October 15, 2014).

20. For example, *Futurisme (cinéma)*, <http://catalogue.bnf.fr/servlet/autorite?ID=15925840&id-Noeud=1.1&host=catalogue> (accessed October 15, 2014).

21. "MACS: Multilingual Access to Subjects," http://www.nb.admin.ch/nb_professionnel/projekt-arbeit/00729/00733/index.html?lang=en (accessed October 15, 2014). In 2008 the Biblioteca nazionale centrale di Firenze took an interest in this project and the Bibliothèque Nationale Suisse (BN) tested the compatibility of Italian subject strings with headings of other indexing languages in the MACS link management interface (LMI). BNCf placed 116 headings of Music, 146 headings of Religion, and 203 of Natural science and Mathematics, which linked with target languages of RAMEAU and SWD. The results of this mapping were satisfactory with approximately 85% equivalents. In the majority of cases the equivalences were exact but also there were some partial equivalences and some equivalences between one term from *Nuovo Soggettario* and one subject string from the other indexing languages (RAMEAU, LCSH, and SWD).

22. Piero Cavaleri, "WebDewey Italiana," *Biblioteche oggi* XXXII, no. 1 (2014), <http://www.bibliotecheoggi.it/2014/20140102301.pdf> (accessed October 15, 2014).

23. "WIKIMEDIA Italia," <http://wiki.wikimedia.it/wiki/Info> (accessed October 15, 2014).

24. The Introduction to Wikidata states, "Multilingual. Editing, consuming, browsing, and reusing the data is fully multilingual. Data entered in any language is immediately available in all other languages; editing in any language is possible and encouraged." 2014, <http://www.wikidata.org/wiki/Wikidata:Introduction> (accessed October 15, 2014).

25. John Mark Ockerbloom, "Forward to Libraries: Experiences Connecting Digital Libraries, Local Libraries, and Wikipedia" (presented at the Digital Library Federation Forum, Austin, Texas, November 2013), http://works.bepress.com/john_mark_ockerbloom/16/ (accessed October 15, 2014).

26. "Magnus' Tools on wmflabs," <http://tools.wmflabs.org/magnustools/> (accessed October 15, 2014).