

My First Ten Years: *Nuovo soggettario* Growing, Development and Integration With Other Knowledge Organization Systems †

Elisabetta Viti

Biblioteca Nazionale Centrale di Firenze, Piazza Cavalleggeri 1, 50122 Firenze (Italy),
<elisabetta.viti@beniculturali.it>



Elisabetta Viti has a PhD in library and information science from Udine University, she is a member of the Regional Executive Committee of the Italian Library Association (AIB), and librarian at the National Central Library of Florence (BNCF). She works at the BNCF's Department of Research and Tools for Subject Indexing and Classification and works in collaboration with the Classification and Indexing Section of the Italian National Bibliography.

Viti, Elisabetta. 2017. "My First Ten Years: *Nuovo soggettario* Growing, Development and Integration with other Knowledge Organization Systems." *Knowledge Organization* 44(8): 624-637. 33 references.

Abstract: The *Nuovo Soggettario* is a subject indexing system edited by the Biblioteca Nazionale Centrale di Firenze. It was presented to librarians from across Italy on 8 February 2007 in Florence as a new edition of the *Soggettario* (1956), and it has become the official Italian subject indexing tool. This system is made up of two individual and interactive components: the general thesaurus, accessible on the web since 2007 and the rules of a conventional syntax for the construction of subject strings. The *Nuovo soggettario* thesaurus has grown significantly in terms of terminology and connections with other knowledge organization tools (e.g., encyclopedias, dictionaries, resources of archives and museums, and other information data sets), offering the users the possibility to browse through documents, books, objects, photographs, etc. The conversion of the *Nuovo soggettario* thesaurus into formats suitable for the semantic web and linked data world improves its function as an interlinking hub for direct searching and for organizing content by different professional communities. Thanks to structured data and the SKOS format, the *Nuovo soggettario* thesaurus is published on the Data Hub platform, thus giving broad visibility to the BNCF and its precious patrimony.

Received: 16 September 2017; Accepted: 19 September 2017

Keywords: thesaurus, *Nuovo soggettario*, terms, Italian, linked open data

† Paper presented at ISKO-Italy: 8' Incontro ISKO Italia, Università di Bologna, 22 maggio 2017, Bologna, Italia. The author would like to thank the Library of Congress staff, in particular Mark Sweeney (Director of Preservation directorate), Beecher Wiggins (Director for Acquisitions & Bibliographic Access), and Janis L. Young (Policy and Standard Division); also many thanks to Ms. Kristin Kolshus of the AGROVOC team and Piero Cavaleri, the Librarian in Chief of the *Biblioteca Mario Rostoni*.

1.0 Introduction

This article presents an overview of the growth and the development of the *Nuovo soggettario* in the last ten years not only as a semantic knowledge organization system (KOS) but also in terms of semantic interoperability with other tools. Nowadays, national libraries must satisfy the expectations of a user who is skilled in navigating the internet, utilizing search engines, and approaching the OPACs with a new attitude and needs. For this reason, libraries must make available not only works and catalogs but also metadata in machine-readable formats in order to promote visibility, recovery, data export, and reusability on the network (Lucarelli 2015).

For the past few years, the National Central Library of Florence (*Biblioteca Nazionale Centrale di Firenze*, BNCF)

has been experimenting open data with the *Nuovo soggettario* thesaurus (Bergamin and Lucarelli 2013, 4). Its metadata, freely downloadable and reusable, have been released in SKOS format (World Wide Web Consortium 2009), which is a W3C recommendation particularly applicable to the semantic web and the linked data developments, and is encoded using XML (*Extensible Markup Language*) and RDF (Resource Description Framework). This format combines concepts expressed by terms to specific classes by assigning unique identifiers, thus facilitating mapping and interoperability between KOSs. The thesaurus's terms can link to other subject indexing tools, encyclopedias, dictionaries, resources of archives and museums, and other information data set, offering the users the possibility to browse through documents, books, objects, photographs, etc.

2.0 The *Nuovo soggettario*

The *Nuovo soggettario* is a subject indexing system, edited by the National Central Library of Florence (BNCF). After three years of prototype work, it was presented to librarians on 8 February 2007 in Florence, as a new edition of the *Soggettario* (Biblioteca Nazionale Centrale di Firenze, ed. 1956), and it has become the official Italian subject indexing tool (Biblioteca Nazionale Centrale di Firenze c2006). It is a flexible and modular system that can be applied in different domains, in both general and specialized information environments, and can be used in libraries, archives, multimedia libraries, and all other bibliographic and documentary information centers for indexing different bibliographic resources (e.g., books and periodicals), regardless of their support, whether paper or digital, and materials such as audiovisuals, graphic materials, etc.

The *Nuovo soggettario* system is made up of two individual and interactive components: the general thesaurus (compliant with ISO standard 25964) (ISO 25964 2011 and 2013) and the rules of a conventional syntax for the construction of subject strings. It is based on the analytic-synthetic model that allows for the creation of subject strings using thesaurus terms, which are interconnected in compliance with the rules of the syntax (Buizza 2011). The thesaurus (<http://thes.bncf.firenze.sbn.it/ricerca.php?menuR=1&menuS=2>) can be applied in both pre-coordinated (the terms are combined in subject strings) and post-coordinated indexing environments (the terms are used as keywords). The BNCF is also trying to find ways to associate the thesaurus with terms extracted from

texts based on their semantic relevance with the purpose of generating keywords related to the content of digital documents. The Thesaurus is run by the open source software produced for *AGROVOC Multilingual Agricultural Thesaurus* (<http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>) edited by the Food and Agriculture Organization of the United Nations (FAO).

The primary aim of the BNCF has been to increase not only the number of terms of the thesaurus but also its technical development with regard to the semantic web. In addition, the BNCF intends to promote bibliographic metadata by publishing its linked data online, free of charge and used in accordance with the Creative Commons Attribution 2.5 license, owing to formats and protocols, such as Zthes v. 1.0 (since 2009), SKOS/RDF (since 2010), and Marc 21 markup language (since 2015) (http://thes.bncf.firenze.sbn.it/thes-dati_eng.htm).

2.0 Terminological heritage

The general thesaurus, the core of the entire system, has been accessible on the web since 2007. Its concepts, represented by terms, are organized into facets according to five basic categories (agents, actions, things, time, and space), which contain a homogeneous class of concepts and whose members share common characteristics that distinguish them from the members of other classes. A rich apparatus of notes clarifies the meaning of the terms, also establishing connections with formerly preferred terms (historical variants) with the correspondent

The screenshot shows the web interface of the *Nuovo soggettario* Thesaurus. At the top, there is a header with the BNCF logo and the text 'Biblioteca Nazionale Centrale Firenze' and 'Nuovo soggettario THESAURUS'. Below the header, there is a navigation bar with links: 'Criteri', 'Aiuto alla ricerca', 'Sigle e simboli', 'Fonti', 'Novità', and 'Download'. The main content area is divided into several sections:

- Search and Navigation:** Includes a search box, a 'Gerachia' button, and a 'Notizie bibliografiche' link. There are also links for 'Ricerca per termine del Nuovo soggettario' and 'Soggettario(1956) Liste di aggiornamento (1956-1985)'. A 'Sistema Nuovo soggettario' section contains links for 'home page' and 'manuale applicativo'.
- Record Details:**
 - Abachi:** The main term being viewed.
 - Categoria/Faccetta:** Cose:Oggetti
 - Nota d'ambito:** Antichi strumenti di calcolo, utilizzati come ausilio per effettuare operazioni matematiche, oggi si conservano come mezzi didattici per la prima infanzia nella forma del pallottoliere, e anche, in taluni paesi, come strumenti ausiliari per i contabili, i commessi di negozio, ecc.
 - UF (User Facet):** Abbachi
 - TT (Term Type):** Oggetti
 - BT (Broader Term):** [Oggetti nell'ambito della matematica]
 - RT (Related Term):** Abacisti, Operazioni aritmetiche, Sussidi didattici
 - HSF (Historical Variant):** Abaco
- Additional Information:**
 - Nota storica:** Precedentemente al singolare: Abaco (Soggettario)
 - Fonte:** Soggettario; Treccani.it; Zin; DTS; ESG; Nicosia; WebDewey(IT); Wikipedia(IT)
 - DDC (WebDewey):** 513.0284
 - Equiv. in altri strumenti di indicizzazione:** LCSH: Abacus; RAMEAU: Abaques (mathématiques)

Figure 1. An example of a *Nuovo soggettario* thesaurus record.

numbers of the *Dewey Decimal Classification (DDC23 2011)* and with sources employed for the control of morphologies and meaning, etc.

The tool inherited most of its terminology from the *Soggettario*, which has been reanalyzed in agreement with ISO standards and enriched with further terminology acquired for semantic relationships and, of course, respecting the literary warrant.

The thesaurus includes common single or compound terms but not proper names and titles; however, it accepts the name of wars, battles, historical events, artistic movements, sporting events, etc. The pre-coordinated subject strings (formed either by single terms or by combining many terms) are searchable in the subject heading archives and in the OPACs of the libraries that have adopted the *Nuovo soggettario* system, with particular regard to the libraries of the National Library Service (Servizio Bibliotecario Nazionale or SBN), the network of Italian libraries promoted by the Ministero dei Beni e delle Attività Culturali e del Turismo and coordinated by the Istituto Centrale per il Catalogo Unico, ICCU (<http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/index.html?jsessionid=F7E582F270F1D9C42BF1F38F6B67E951>). It is continuously growing thanks to the efforts of the BNCF and, since 2009, the partnership of Italian universities along with many other prestigious public and private cultural and research centers, archives,

and museums (<http://thes.bncf.firenze.sbn.it/enti.htm>). Figure 2 shows the growth of terminological heritage from 2007 to the end of 2016.

Ten years have passed since the first stage prototype, where the thesaurus resembled a tree of signifiers and signifieds, a key repository of 13,000 terms (February 2007) to be used in subject indexing. In December 2007, it had 16,000 terms while today it has 59,000 terms. Between 2007 and 2016 the thesaurus increased by 344.56% and in the last five years (2012-2016) by 22.6%. A year-to-year comparison shows a gradual and steady growth. It should be considered that in this tool there are increasing and decreasing processes (e.g., some terms are deleted), even if in the *Nuovo soggettario* thesaurus, the total number of cancellations is limited, because unaccepted forms are maintained with links of historical variants. It would be interesting to subtract the terms with proper name characteristics (e.g., the names of wars, historical events, etc.) from this number.

In the early years, it had an exponential expansion due to the physiological growth of the thesaurus. Indeed, as a result of the proposals of the indexers of Italian National Bibliography (Bibliografia Nazionale Italiana, BNI, <http://www.bncf.firenze.sbn.it/pagina.php?id=187>) and the other partners working on the project, the thesaurus has been enriched with general and specialist terminology used in the indexing procedures. Figure 3 shows that 21%

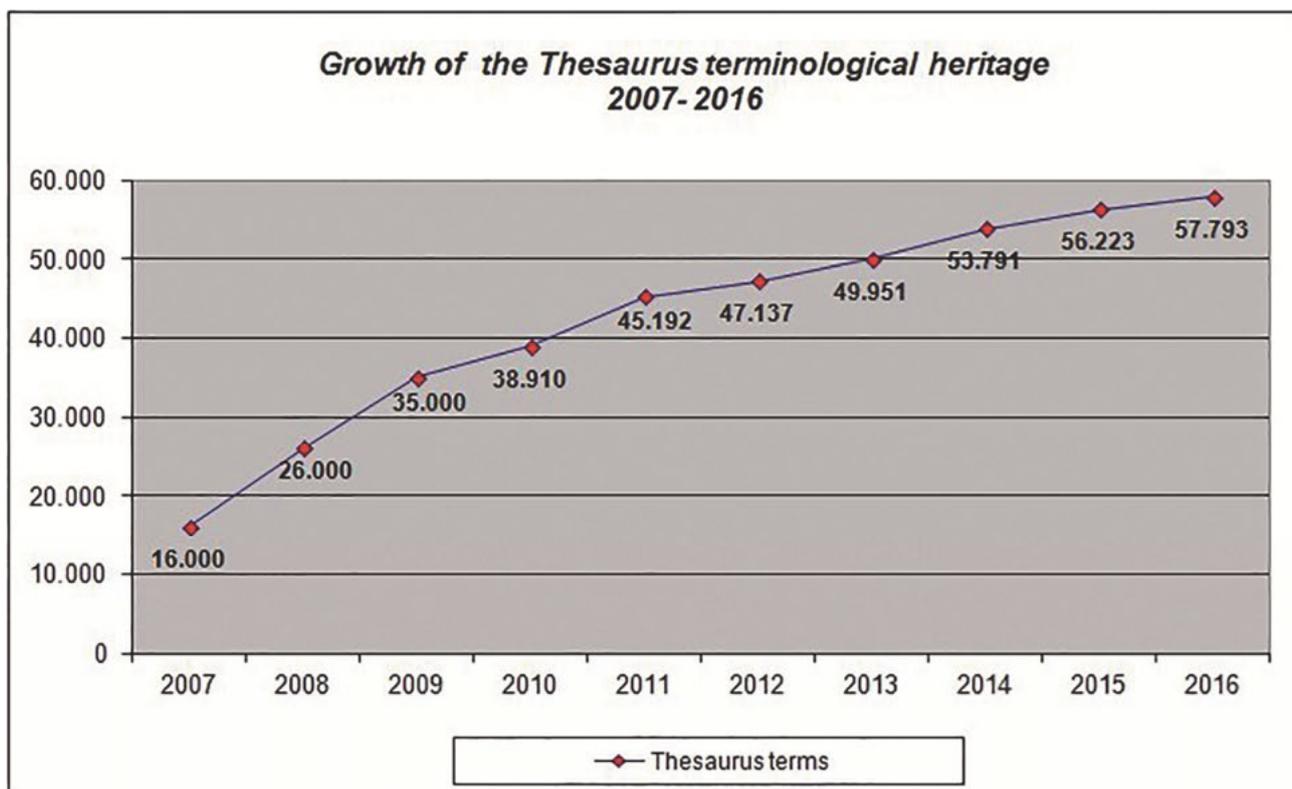


Figure 2. The growth of the thesaurus terminological heritage.

of the included terms (derived from previous indexing tools, such as the *Soggettario*) are new proposals of BNI and other institutions.

Reflecting on the disciplinary areas of the partners, it is interesting to note that there are more proposals in economics, law, and engineering than in literature, art, history, etc. The universities are the most active entities to propose new terms, because the scientific research and editorial production are more alive.

Over time, the need to acquire new terms for expressing the conceptual content of documents is going to decrease, because the thesaurus heritage will meet more and more informative needs of its users. It is interesting to look at the percentage of the increase of both analogous indexing tools, such as *Library of Congress Subject Headings* (*LCSH*, <http://id.loc.gov/authorities/subjects.html>), and *Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié* (*RAMEAU*, <http://catalogue.bnf.fr/recherche-autorite.do?pageRe>) and specialized controlled vocabularies, such as *AGROVOC* and *ThESs: Thesaurus di Economia e Scienze Sociali LIUC* of Biblioteca Mario Rostoni (Università LIUC Carlo Cattaneo, <http://www.biblio.liuc.it/pagineita.asp?codice=210>). Both the Library of Congress (LoC) and the Bibliothèque National de France (BNF) provide official information about the tools and their data. They manage and update numerical

data periodically, making them known and available to the public. However, this analysis does not have comparative assessment claims and must also take into account that there are some aspects of differentiation between the *Nuovo soggettario* thesaurus, *LCSH*, and *RAMEAU*. Three of the most significant are:

- Chronological features. These tools were created in different periods and today have different numerical sizes and consequently different growth.
- Structural characteristics. While the *Nuovo soggettario* thesaurus has only terms and not possible combinations and collocations (accessibility limited to the OPACs of the libraries that employ it). *LCSH* and *RAMEAU* include authority headings that also consist of the combinations of multiple terms.
- Characteristics of terminology. The *Nuovo soggettario* thesaurus has no proper names except some exceptional cases.

The first edition of *LCSH*, development of *Subject Headings Used in the Dictionary Catalogues* of the Library of Congress was published between 1909 and 1914. In 1989, editions began to be published annually and from 1986 (the 10th ed. of *LCSH*), the introduction to each edition included statistical information (<https://www.loc.gov/>

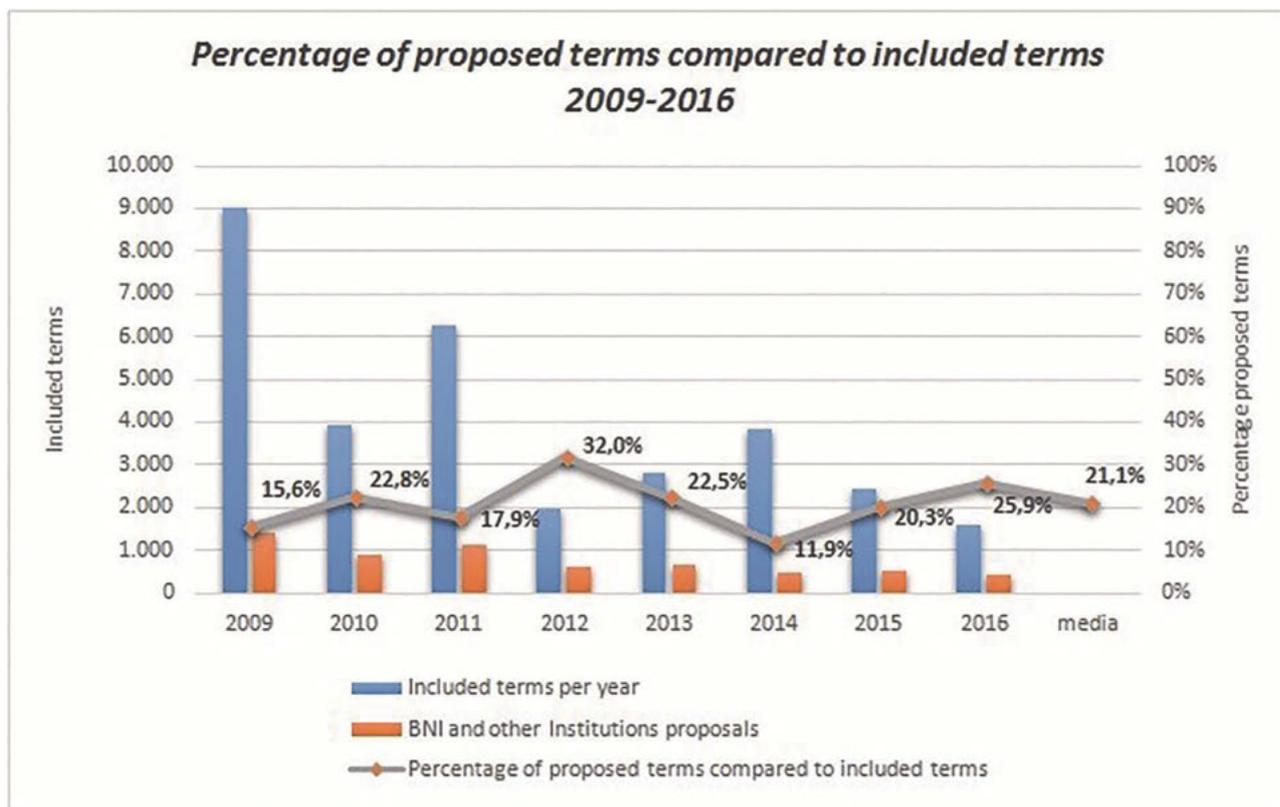


Figure 3. Percentage of proposed terms compared to included terms (2009-2016).

aba/publications/Archived-LCSH38/lcshintro.pdf). As is widely known, the subject authority database (subject headings plus subdivision records) contains personal name headings (family names, corporate headings, and meeting or conference headings), uniform titles, and topical and geographical subject headings. In 1986 the number of the authority records was 145,000, and today, it is 342,107 (24,306 personal name headings, 241,726 topical subject headings, and 61,438 geographic subject headings). From 1986 to 2016, the size of the vocabulary increased by 135.93% and in the period 2012-2016 by 2.14%. These statistics show a gradual and steady increase even if over the last five years the pace of *LCSH* growth has slowed. This is due to the fact that in 2013, the Library of Congress staff decided to cancel some headings for individual fictitious characters, mythological beings, deities, legendary characters, and names of animals; in 2014, the authority records for the Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT) became a stand-alone vocabulary and thus were removed from the statistics of *LCSH*. Although the increase in the number of deletions masks the growth of *LCSH*, approximately 5,000 new proposals (new headings, including headings with subdivisions) per year are submitted by cataloguers at the Library of Congress and by participants in the Subject Authority Cooperative Program (<https://www.loc.gov/aba/pcc/saco/about.html>).

It is also interesting to also pay attention to the French indexing system. The Bibliothèque National de France (BNF) has been using RAMEAU, the national authority list for analytical indexing since 1980. On the BNF website, a table can be found showing the growth of RAMEAU authority data from 2003 to 2016 (Statistiques d'accroissement du référentiel RAMEAU, <http://rameau.bnf.fr/informations/chiffres.htm>). The statistics refer only to the RAMEAU subject authority records whose headings include common names associated with certain proper names (names of people, battles, ships, etc.; names of communities, conventional titles or uniforms, followed by a subdivision), geographical names, titles of serial publications (periodicals and collections of monographs), fictional, mythological or legendary characters, divinities, etc.

The numerical data, found on the site, are divided into common names, geographic names, chronological subdivisions, and proper names, and at the end of the summary, there is the total of authority records broken down by years. For the comparison, only the percentage of the total and the percentage of the common names were taken into consideration. In the period 2003-2016, RAMEAU's size increased by 28.21%, ranging from 141,244 to 181,095 authority records, and in the same years, the common names

increased by 29.87%, ranging from 86,192 to 111,944. In the last five years, RAMEAU has grown by 8.08% for authority records and by 8.3 % for common names. These results show a decrease in the growth of the numerical data.

It is also interesting to make the same analysis about specialized thesauri, such as AGROVOC and ThESs thesaurus. The numerical data has been obtained thanks to an exchange of correspondence with the teams of AGROVOC and Biblioteca Mario Rostoni (Università LIUC Carlo Cattaneo, Castellanza). AGROVOC consists of 33,461 concepts in the area of interest of Food and Agriculture Organization (FAO), available in twenty eight languages (Arabic, Chinese, Czech, English, French, German, Hindi, Hungarian, Italian, Japanese, Korean, Lao, Malay, Persian, Polish, Portuguese, Russian, Slovak, Spanish, Telugu, Thai, Turkish, Ukrainian, etc.), for a total amount of 652,767 terms. It was first published in 1982 in a printed version with 8,660 concepts in three languages, and since 2000, it has been on the web. From the beginning to 2016, it has been increased with concepts by 277.09% and in the period 2012-2016, by 1.07%, ranging from 32,310 concepts (600,000 terms) expressed in twenty different languages to 32,656 concepts (652,767 terms) expressed in twenty seven different languages.

The Italian ThESs thesaurus of the Biblioteca Mario Rostoni is a controlled vocabulary built by drawing the terms primarily based on literary warrant of the economic-business, social science, and legal areas. ThESs was born in 1990, based on the British Library's experience with PRECIS, and it has grown in relation to the acquisitions of the Library and the need to express the content of the books in order to encourage research by topic. The number of terms has increased regularly, reaching the threshold of 5,000 terms already by 2000. Approximately two hundred new terms are added per year and the size of the ThESs has grown by 5%. Since 2007, it has been linked to the *Nuovo soggettario* thesaurus. This short overview shows that all thesauri have had an exponent expansion from the beginning to date. However, in the last five years, their growth (in percentage) has decreased, because the heritages of the vocabularies are now rich and respond satisfactorily to the indexing needs of cataloguers.

In a ten year span, the *Nuovo soggettario* thesaurus has evolved into a much more complex, stable, corpulent, modular, and interoperable tool which overcomes language, function, and shape barriers and thanks to qualifying links, sets up conceptual bridges, which act as an interlinking hub with other data systems and/or objects of cultural heritage institutions.

3.0 Link between *Nuovo soggettario* thesaurus and other KOSs

The strengths of the thesaurus are its rich terminological heritage and a dense network of links that allow the user to navigate other knowledge organization tools and archive and museum resources. In general, there are two types of connections between the *Nuovo soggettario* thesaurus and other knowledge organization systems: simple connections and typed links. Links are triggered by quotes or abbreviations reported in the thesaurus terms record in specific fields: *fonte* (source), *equivalenti in altri strumenti di indicizzazione* (equivalents in other subject indexing tools), *risorse di archivi e musei* (archival and museum resources). Figure 4 lists the tools which have the greater links number with the thesaurus:

- Subject indexing tools: *Library of Congress Subject Headings (LCSH)* and *Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU)*;
- Specialized thesauri: *Art and Architecture Thesaurus (AAT)*, edited by the Getty Foundation (<http://www.getty.edu/research/tools/vocabularies/aat/about.html>); *AGROVOC Multilingual Agricultural Thesaurus*, edited by the Food and Agriculture Organization (FAO); *EuroVoc Multilingual Thesaurus of the European Union*, which covers the activities of the European Union and contains terms in twenty three EU languages, (<http://euro.voc.europa.eu/drupal/?q=node>); and ThESS;
- Encyclopedia: *Treccani*, the most important Italian encyclopedia of science, humanities and arts (<http://www.treccani.it/>); the Italian version *Wikipedia* (<https://it.wikipedia.org/wiki/Wikipedia>), the most popular free online encyclopedia;

- General and specialized classification: *Dewey Decimal Classification* (the Italian version *WebDewey*, <http://deweyit.pansoft.de/webdewey/login/login.html>); *DoGi*, specialized in the classification of legal matters, edited by the *Istituto di Teoria e Tecniche della Documentazione Giuridica (ITTIG)*, (<http://nir.ittig.cnr.it/dogiswish/dogiClassificazione.php>).

The more a data set is connected, the more it acquires value in terms of credibility and maintenance. Recently, the thesaurus has entered into a loose circle of ever-widening connections, extending its links to:

- other specialized thesauri, e.g., *Bia-Net*, the classified thesaurus of Roman law and classical antiquities (<http://104.236.71.119/bianet/index.php/descrizione-del-sistema-bia-il-thesaurus/>) and the *Polimoda Thesaurus* on fashion industry edited by *Centro di Documentazione Polimoda* (<http://opac.polimoda.com/ListRecord.htm?list=table&table=5>);
- *Parole nuove*, a data base of neologisms acquired by the *Accademia della Crusca*, the most prestigious Italian academy specialized in the Italian language (<http://www.accademiadellacrusca.it/it/lingua-italiana/parole-nuove.nks>); and
- A database of archives and museums (e.g., *Fondazione Scienza e Tecnica*, <http://www.fstfirenze.it/storia-della-fondazione>).

Over time the aim of the BNCF is to develop a streamlined semantic and technical interoperability, making the thesaurus a tool within the web. As a result, conceptual mapping, intended as a lexical alignment, reconciliation and consequently a technical connection between the *Nuovo*

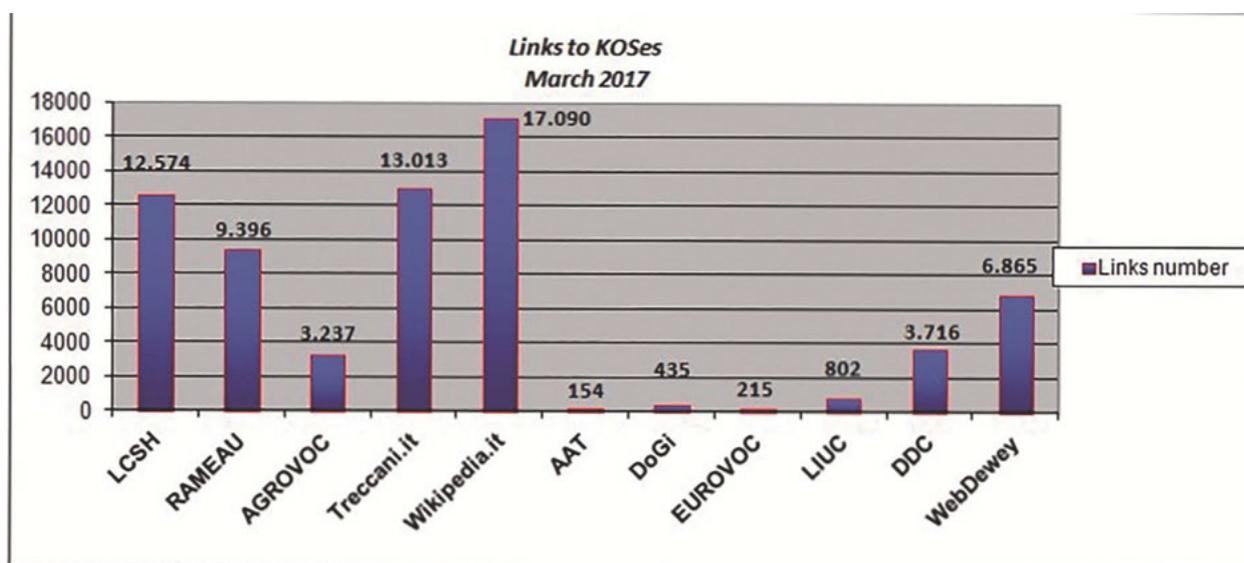


Figure 4. Links with other tools in March 2017.

soggettario terms and terms/subject strings/notations, etc., of other tools should improve the semantic enrichment and integration of heterogeneous collections and amplify the research results.

With the passing of time, the way to connect the thesaurus with other indexing tools that are different according to language, conceptual structure and domain has changed and has followed strategies that have been acquired as a result of what the semantic web developments have permitted.

Owing to the SKOS/RDF exchange format and protocols developed as a key semantic web technology (e.g., SPARQL, Simple Protocol and RDF Query Language (SPARQL 2008), used to express queries across diverse data sources, enabling integration with other RDF datasets), the data of LCSH, RAMEAU, AGROVOC, ThESS, and Wikipedia (Italian version) can be downloaded and re-used. This allows for the creation of deep links or cross-walks between the tools. Similarly to what has been done at the international level, the Italian indexing tool provides equivalents in other languages based on the American *LCSH* system and the French RAMEAU system. Its network of equivalences is comparable to what a few years ago other European national libraries experimented with the MACS-Multilingual Access to Subjects project (<http://www.nb.admin.ch>) (Landry 2001; MacEwan 2000; Olson 2003).

In recent years, these equivalents established real connections by enabling users to navigate between terms that express the same concept in different languages by displaying other countries' indexing tools and related bibliographic news. The project is based on criteria that are continually being clarified and refined over time, taking into consideration the differences that may exist between the heterogeneous tools and the resulting difficulties with their interconnections. The first step of interoperability started with the more recurring and consulted resources for attesting the form and meaning of the terms. A simple static citation of the name of the resources in tag *fonte* of the thesaurus terms was transformed into a real link, and for some of these resources, such as *LCSH* and RAMEAU (using the multilingual task of the thesaurus software), dedicated tags for deep linking were created. *LCSH* and RAMEAU equivalents are displayed near the Italian term in the list of the results obtained from clicking on *ricerca per termine* (term Search), and they are also directly searchable or shown in alphabetic order by clicking on the *equiv. in altri strumenti di indicizzazione* (search equivalents) button.

Since mid-December 2013, owing to the SKOS format, in the records of the *LCSH* data base, this connection has become reciprocal and so within the Closely Matching Concepts from Other Schemes, the Italian

terms have been visible along with the French and German terms. As it currently in the experimental stage, only one part has become reciprocal. We hope the LoC will soon complete the work started. Figure 5 shows the increase of the number of more significant link from 2011 to date (March 2017). Even if the connection with other tools started to proliferate in 2009, the basic reference is 2011, because the links began to multiply with the advent of SKOS conversion, implemented in 2010, and that permitted the marking up of the data thereby rendering them downloadable, reusable, and interoperable.

In the period 2011-2017 the number of the links increased by 263.98%. Although the two thesauri are quite different in terms of size to the *Nuovo soggettario* thesaurus, it is interesting to look at the percentage of the growth of linking in RAMEAU and *LCSH*. The RAMEAU records as defined by the Contours de RAMEAU authorities can be downloaded in SKOS/RDF format from the semantic web via data.bnf.fr. In this application, the name *Liste des themes* refers to the whole vocabulary RAMEAU.

RAMEAU data in RDF/SKOS format is placed under open license. Their re-use is free, subject to the provisions of this license, which implies maintaining the mention of source Bibliothèque nationale de France. The RAMEAU data in data.bnf.fr is updated several times a year. According to the statistics on the RAMEAU website, the number of the links increased by 65.09% in the period of 2003-2016, and at the end of 2016, there were 128,451 connections with *LCSH*. Unfortunately, a chronological summary of *LCSH* links is not available on the LoC website. Thanks to the BNCf's contacts with the staff of the LoC, it is known that today there are 80,000 RAMEAU links in their triplestore. The links were prepared by BNF around 2012, and since that time, there have been no regular updates. In addition, the LoC has also imported 4,400 links from *Nuovo soggettario* thesaurus, and possesses a significant set of GLIN (Global Legal Information Network, <https://www.loc.gov/item/prn-08-226/glin-database-offers-new-features/2008-12-03/>) thesaurus terms. The VIAF links (Virtual International Authority File, <http://viaf.org>) are the only ones which are regularly updated. Unfortunately, it has not been possible to find similar statistics about ThESS and AGROVOC. At the moment, the ThESS thesaurus has only one connection with the *Nuovo soggettario* thesaurus while AGROVOC has many connections with a wide range of tools.

AGROVOC is a SKOS-XL concept scheme (SKOS-XL 2009) and a linked data (LD) (<http://ontotext.com/Knowledgehub/fundamentals/linked-data-open-data>) set aligned with eighteen other multilingual knowledge organization systems related to agriculture, such as: Dbpe-

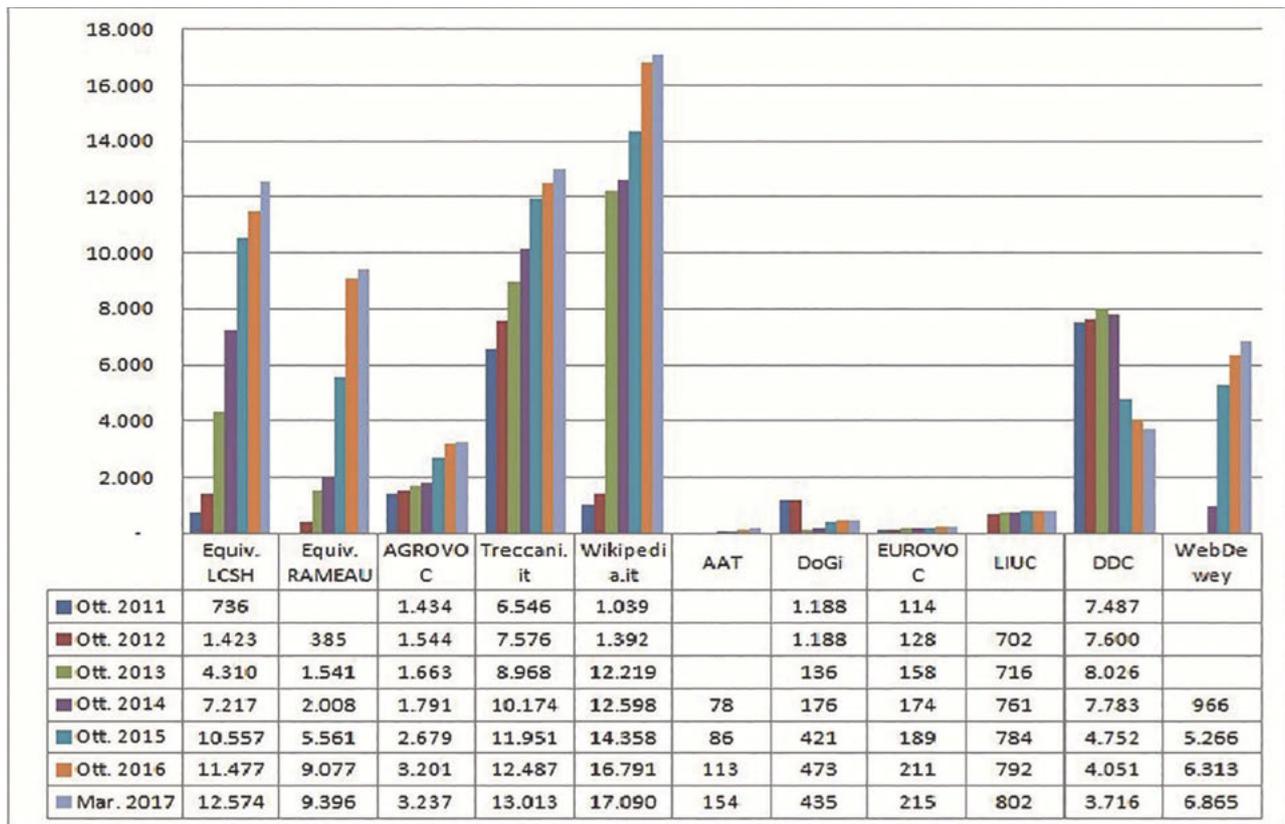


Figure 5. Links from 2011 to date.

dia (<http://wiki.dbpedia.org>), DDC (in web version), LCSH, RAMEAU, DNB (Deutsche National Bibliothek, <http://www.dnb.de>), BNCf, ASFA Thesaurus (Aquatic Sciences and Fisheries Abstracts, <http://www.4.fao.org>), EUROVOC, GEMET (General Environment Information and Observation Network, <http://www.eionet.europa.eu>), GeoNames Ontology (<http://www.geonames.org/ontology/documentation.html>), FAO Geopolitical Ontology (<http://www.fao.org/countryprofiles/geoinfo/geopolitical/resource/>), CAAS (Chinese Academy of Agricultural Sciences, <http://www.caas.cn/en/>), NALT (National Agricultural Library's Agricultural Thesaurus, <https://agclass.nal.usda.gov/>), CAB Thesaurus (essential search tool for all users of the CAB ABSTRACTS™ and Global Health databases, <http://aims.fao.org/vest-registry/vocabularies/cabi-thesaurus>), etc.

4.0 Links with archival and museum resources

Since 2014, the BNCf has been successfully collaborating with a working group called MAB (Italian acronym of Museum, Archives and Libraries, <http://www.aib.it/attivita/mab-italia/>), a national coordination promoted in 2011 by AIB (Italian Libraries Association, <http://www.aib.it/>), ANAI (National Association of Italian Archivists, <http://www.anai.org/anai-cms/>), and ICOM-Italia (International

Council of Museums, Italian committee, <http://www.icom-italia.org/>). The group has explored the possibility of testing how archives, museums, and libraries could cooperate and share common languages for the description of heterogeneous resources favouring cross linking and usability (Bruni, et al. 2016, 7).

The process of indexing museum objects and archival resources has some similarities to that of indexing printed publications, but it presents some interesting and challenging problems regarding the nature of the object/document and its description. It is unusual to find a museum or an archive that makes its catalogs available to visitors in the same manner as libraries. Over the past few years, much work has been done nationally and internationally to draw up some guidelines for describing objects, structuring museum data, and building a thesaurus of museum object names (Will 1993, 18; Holm 1993).

Continuing along this line, MAB has dedicated its resources to two research fields: exploring the possibility of using the new RDA (Resource Description and Access) guidelines, originally established for the description and creation of access to all types of resources, for the description of archival collections and exploring the possibility of using the *Nuovo soggettario* thesaurus for subject indexing archival resources and for describing museum collections. *Nuovo soggettario's* terms were used for the subject

indexing of some archival units of the Italian writer Giovanni Pascoli (<http://www.pascoli.archivi.beniculturali.it/index.php?id=66>) and of the Italian politician, journalist, and anti-fascist activist Ernesto Rossi (<http://siusa.archivi.beniculturali.it/cgi-bin/pagina.pl?TipoPag=compare&Chiave=316359&RicProgetto=personalita>) and, in addition, for the description of some scientific tools of the *Fondazione Scienza e Tecnica* (<http://www.fstfirenze.it/storia-della-fondazione>). At the end of the experiment, the *Nuovo soggettario* thesaurus implemented links with various data. Furthermore, by using the dedicated tag *risorse di archivi e musei* of the *Nuovo soggettario* record terms, it is now possible to navigate the resources of the Museo Galileo (<http://www.museogalileo.it/>), the Fondazione Scienza e Tecnica (<http://www.frame-test.it/>), and the SIUSA-Archivi di personalità (Archivi di personalità, census of the personality archives of culture in Tuscany between 1800 and 1900, <http://siusa.archivi.beniculturali.it/cgi-bin/pagina.pl?RicProgetto=personalita>).

5.0 Interoperability between *Nuovo soggettario* thesaurus and *Wikipedia*

Another important aspect of the *Nuovo soggettario* thesaurus opening and linking is the collaboration with Wikipedia encyclopedia, edited in more than two hundred eighty languages. In the optics of the Wikipedia universe, the primary purpose of cultural institutions is the free access to knowledge for everyone. If the blending of different cultural environments is well managed, it would mutually benefit everyone involved. In recent years in Italy and around the world, many projects have been carried out for linking libraries, archives, and museums with Wikipedia, Wikidata, and Wikisource. One such example took place at the British Library in which a session was held to collectively edit Wikipedia entries containing information about the library's collection; another example, the State Library of Queensland, which, with the collaboration of local libraries and Wikimedia-Australia, organized a project for editing its webpage on Wikipedia.

In the sphere of the *GLAM-Wiki Initiative* (<https://en.wikipedia.org/wiki/Wikipedia:GLAM/About>), the project that promotes the sharing of resources of cultural institutions with the Wikipedia universe, it has developed the promotional campaign *Wikipedia Loves Libraries* to link the encyclopedia with open access libraries and archives. GLAM was born from the awareness that these institutions, caretakers, and managers of documents, cataloguing resources, images, photographs, etc., could make a fundamental contribution to the world of knowledge and, in particular, to the *Wikipedia* community. In this project, libraries play a central role due to the linked data development, contributing towards giving

new, open, and more productive value to the content and data that they produce (Lucarelli 2014, 54).

It is always curious whenever librarians and Wikipedians cross paths; two very different workflows and cultures meet, clash, and gradually merge into one. The librarians are professionals who work in a pyramidal organization with formal methods and encoded best practices in order to manage structured data compliant with standards. However, Wikipedians are volunteers who work in a horizontal organization without a hierarchy control, with great team spirit and liveliness, collaborating in open and online projects with likeminded souls who come from the world of the internet. Many librarians find a stimulating environment and enthusiastic comprehensive approach on Wikipedia. The cultural melting pot of librarians and Wikipedians produces the best for everyone. Both love collaboration, integration, and sharing; therefore, their partnership could move forward in the future if new initiatives and projects are identified that recognize common goals, workflows, etc. (Ayers 2016; Ayers and Zanni 2017, 57).

Since 2012, many initiatives have been born in Italy and disseminated throughout the country due to the work and collaboration of librarians and Wikipedians. Numerous meetings, conferences, and training courses have been organized and agreements have been signed for projects related to cataloguing and authority control. The ICCU (Istituto Centrale per il Catalogo Unico, <http://www.iccu.sbn.it/opencms/opencms/it/>) began work to enhance the quality of authority records edited by SBN (Servizio Bibliotecario Nazionale), increase Italian participation in VIAF, and established links with Wikipedia entries.

In January 2013, the BNCf began its first contacts with Wikimedia Italia given that the *Nuovo soggettario* thesaurus for some years had welcomed Wikipedia (Italian version: https://it.wikipedia.org/wiki/Pagina_principale) into the sources to check the form and the meaning of the terms. After a few months, an agreement was settled to collaborate further development of the links between the terms of the thesaurus and the *Wikipedia* equivalents. In May of the same year, due to the BNCf activity, reciprocal links were implemented, thus allowing for crosswalks. This has been an example for other national versions of Wikipedia that have been linked to national indexing tools through the Authority control field. Connections of Wikipedia with a multidisciplinary subject indexing tool generated the first prototype of interoperability ever produced in the world of libraries.

Interoperability between the *Nuovo soggettario* thesaurus and *Wikipedia* has evolved in a variety of ways and remains an activity that is constantly subject to intellectual analysis and review. After an initial phase of association

[Criteri](#)
[Aiuto alla ricerca](#)
[Sigle e simboli](#)
[Fonti](#)
[Novità](#)
[Download](#)

RDF/XML

+ Ricerca per termine del Nuovo soggettario

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

Sistema Nuovo soggettario

- o home page
- o manuale applicativo

Anemometri

Gestuchin Notizie bibliografiche

✉

Categoria/Faccetta: Cose:Oggetti

Nota d'ambito: Strumenti per la misura dell'intensità del vento o per l'indicazione per mezzo di indici dei valori momentanei della direzione e dell'intensità del vento

TT [Oggetti](#)

BT [Strumenti meteorologici](#)

NT [Anemometri a filo caldo](#)
[Anemometri a laser Doppler](#)

RT [Venti](#)

Fonte: [Soggettario](#); [Treccani.it](#); [VLI](#); [DeM](#); [PT](#); [Zin](#); [Brenni](#); [DTS](#); [ESG](#); [Lueger](#); [AGROVOC](#); [Wikipedia\(IT\)](#)

Equiv. in altri strumenti di indicizzazione

[LCSH: Anemometer](#)

[RAMEAU: Anémomètres](#)

Risorse di Archivi e Musei

[Museo della Fondazione Scienza e Tecnica](#)

[Museo Galileo](#)

Done



Museo Virtuale

Anemometro

Home > Indice delle Sale > Sala XIV. L'industria degli strumenti di precisione > La meteorologia >

Collocazione: Sala XIV

Costruttore: sconosciuto

Data: prima metà sec. XIX

Materiali: marmo, ottone

Dimensioni: altezza 415 mm, diametro 268 mm

Inventario: 803

Tipologia

- Anemometro

Contesto

- Banderuole e anemometri
- Meteorologia

Dizionario

- Banderuola
- Livella
- Rosa dei venti

Approfondimenti

- Bussola
- Mercurio

Anemometro costituito da una banderuola di ottone montata su un recipiente cilindrico e inserita in un pozzetto praticato in una base di marmo. Il mercurio, che in origine riempiva il pozzetto, permetteva alla banderuola di galleggiare e di muoversi liberamente. Sotto la banderuola sono impennate orizzontalmente due coppie di ruote a pale e un conoide filettato. Su di esso originariamente era fissato un filo con un contrappeso. L'azione del vento, facendo ruotare il conoide, faceva avvolgere il filo su circonferenze sempre maggiori. Ad un certo punto, la forza del vento che agiva sulle pale controbilanciava esattamente la forza del contrappeso: le palette cessavano di ruotare e venivano bloccate nella loro posizione da un nottolino. Osservando la posizione del contrappeso si poteva così stimare la velocità massima raggiunta dal vento. Nella base è presente una corona metallica che reca una bussola ed una livella a bolla; su di essa è incisa una rosa dei venti.

Figure 6. An example of link between *Nuovo soggettario thesaurus* and Fondazione scienza e tecnica and Museo Galileo.

of conceptual equivalents through manual procedures, automatic association procedures were also tested based on the exact equivalence of the terms' form. From the evaluation of the results, the BNCf realized that the automatic matches (about 9,000 terms) needed the intellectual/manual control over forms and meanings. Thus, the hypothesis of developing interoperability only with automatic procedures has been abandoned for the time being.

The BNCf continues to work with the Wikipedia community to understand how Wikidata and Dbpedia, as multilingual databases, could be employed as hub structures to link different subject indexing tools, built in different languages and in different countries, by automatic or semi-automatic methods and to optimize reciprocal links. The BNCf looked to take advantage of pre-existent links between the *Nuovo soggettario* thesaurus, Italian Wikipedia, and English Wikipedia to automatically match LCSH equivalents and to connect them to non-linked *Nuovo soggettario* terms. For example, BNCf has exploited the work of the librarian John Mark Ockerbloom who established some crosswalks between LCSH and Dbpedia, which were then matched to the results of the BNCf works, thereby automatically acquiring further six hundred new links for *Nuovo soggettario* (Lucarelli and Viti 2015, 53; Viti 2014).

Today, after several experiments, the attribution of Wikipedia equivalents is based on specific activities:

1. Identification of the semantic correspondence and citing the Wikipedia in the *fonte* field of the thesaurus term;
2. Verification. Upon the download to the user display, our programmer verifies the formal and semantic correspondence between the terms of the *Nuovo soggettario* thesaurus and the Wikipedia entries. In this verification step, the links on the disambiguation pages, proper names, titles of works, films, etc. are deleted;
3. Creation of links. SPARQL creates direct links and thus obtains a triple list which includes the identifier of the *Nuovo soggettario* term, the link SKOS close-match and the identifier of the Wikidata entry;
4. Creation of crosswalks. BNCf sends this list of triples to the Wikidata referent who creates reciprocal links; and
5. Final check. BNCf checks the accuracy of the inverse links from *Wikipedia* to the *Nuovo soggettario thesaurus*.

This procedure is very dynamic and always requires a lot of maintenance. When structurally different tools are connected, logically there might be some problems in getting them to communicate and keep alignments. The quality of interoperability is far from automatic procedures, because it is subjected to intellectual actions of analysis and con-

tinuous review of the forms and the meanings of terms. As of September 2017, the links between *Nuovo soggettario* thesaurus and Wikipedia are greater than 17,000. The *Wikipedia* readers can link to the thesaurus and the BNCf OPAC through the tag *Collegamenti esterni* on the *Wikipedia* page by clicking on the thesaurus term.

The involvement of the BNCf in Wikimedia Foundation projects has further expanded. The idea of sharing information, knowledge, and training between different professions is the basis of the project *Wikipedian in residence* (WIR). Like many prestigious cultural institutions and international libraries, in 2016 the BNCf welcomed an *Wikipedian*, for six months, to share ideas, experiences, and ways of working and to facilitate the BNCf collaboration with the *Wikipedia* community (Lucarelli 2014, 9).

5.0 Mapping from the *Nuovo soggettario* universal thesaurus to the *Dewey Decimal Classification* in the Italian web version

Since the beginning of the construction of the *Nuovo soggettario* (2007), the most closely matching classification notation has been found for each preferred term. From 2007 to July 2014, it has been assigned numbers of the 21st, 22nd and 23rd editions, and since July 2014, it has been assigning *WebDewey* (Italian version) numbers, which are displayed in a specific field in the record of the term. As of March 2017, 10,581 *DDC* notations have been mapped including 6,865 with *WebDewey*. This is an ongoing project that takes into consideration the differences between the two indexing tools that do not share the same structure, and so the criteria on which the resulting actions are continually being clarified.

The mapping criteria formulation are in compliance with ISO 25964 pt. 2, the new standard that looks at the question of multilingualism and mapping from the point of view of the semantic web and linked data, and also with EDUG's recommendations for best practice in mapping involving *Dewey Decimal Classification* (*DDC*) (<http://edug.pansoft.de/tiki-index.php?page=EDUG+workshops>) (European *DDC* Users Group, ed. 2016). It is based on an intellectual process that checks the suitability of the mapping case by case in order to identify exact or inexact concept equivalents in the *Dewey Decimal Classification*. A structural model is used in which the *Nuovo soggettario* thesaurus is the source vocabulary and the *DDC* is the target vocabulary. The mapping is in only one direction, from thesaurus concepts, expressed by the *Nuovo soggettario* terms, to the *DDC* classes, expressed by notations, and it does not distinguish the differences between the degrees of equivalence.

Only one *DDC* notation, which represents the interdisciplinary or comprehensive number, has been assigned

to the *Nuovo soggettario* concept/term. The interdisciplinary or comprehensive number has been found in the index and/or in the schedules of the *DDC*. This choice is independent from the category, the scope note, and the hierarchical structure of the *Nuovo soggettario* concept. This policy has some limited exceptions if the *DDC* does not have the interdisciplinary notation or comprehensive number. For example, if the *DDC* has a number related to a specific disciplinary field, the indexer could assign this only if there is a semantic equivalence with the meaning of the thesaurus term.

On an experimental basis, built numbers have been assigned according to *DDC* instructions, and it is indicated by a specific tag of the thesaurus *DDC commento*, which is invisible in the user interface. Thanks to the conversion of the thesaurus data in *Marc 21 Format for Classification Data* (US Library of Congress Network Development and MARC Standards Office), it is possible to visualize the *Nuovo soggettario* equivalent into the notation of *WebDewey* (with the *LCSH* equivalents). The *DDC* is actually a reference not only for mapping but also for linguistic variants. In fact, the thesaurus builders use the *DDC* to check the forms of terms, other synonyms, etc. For this reason, in some cases, the *DDC* acronym could also be displayed in the *fonte* tag of the record of the term, verifying, in this way, any formal agreement between the preferred words in the thesaurus and preferred words in the Italian translation of the *DDC*.

Today the BNCF is assessing the following possibilities:

- If it is more convenient to map only with the interdisciplinary number or the comprehensive number or with the *DDC* notations for specific points of view; the letter of is more coherent with the meaning assigned in the thesaurus;
- Whether to assign more than one number;
- Whether to enrich WebDewey index with *Nuovo soggettario* thesaurus terms; and
- Whether to expand the *DDC* classes with built numbers, which are the equivalent of complex concepts of the thesaurus terms (Crociani, Giunti and Viti 2016, 56).

6.0 Conclusion

The core of the *Nuovo soggettario* system is a universal-general thesaurus, which involves different domains and that is applicable for indexing different information resources. Detailed reports, from the prototype phase in 2007 to date, have shown a positive level of growth and stability. Owing to its open structure, it is able to acquire both general and specialized terminology. Today it has a strong structure with a wealth of notes that clarify the

meaning of the terms, document bonds with shapes in the old indexing tools, and which establishes a conceptual bridge with other data sets. Since 2009, the BNCF has had among its objectives, interoperability and mapping between the *Nuovo soggettario thesaurus* and other KOSs, establishing reciprocal equivalences by the conversion of an expression of the thesaurus into a corresponding expression in one or more other vocabularies. During the information retrieval, this should allow for the navigation among the resources of different institutions.

Criteria and procedures have been established and revised over time, because connections and interoperability involve several aspects, such as the conceptual structure of the tools, languages, specificity of the domains of belonging, the meaning of the terms (polysemy and monosemy), the conventions of grammar and syntax in regional or national use, the type of links established (exact or partial equivalence; broader or narrower equivalence), etc.

The conversion of the *Nuovo soggettario* thesaurus into mark-up formats suitable for the semantic web and linked data world, improves its function as an interlinking hub for direct searching and for organizing content by different professional communities. If the data are open and downloadable in RDF, they can be usable and reused by institutions on the web. Figure 7 shows the number of accesses within the timeframe 2011-2016. The data reported relate to the number of IP addresses that have access to the *Nuovo soggettario* website.

As can be seen, there was a sharp increase in the number of accesses by 317,835 in 2014. Probably, this is due to the fact that in that year a reciprocal interoperability with Wikipedia (Italian version), the fifth most consulted site in the world, began.

Analysing the IP addresses, it was realized that the thesaurus in SKOS version has been downloaded not only from libraries (such as the Ticino Library System or organizations that collaborate with the BNCF such as Università commerciale Luigi Bocconi) but also from other institutions such as Bartoc.org (Basel Register of Thesauri, Ontologies & Classifications is a database of knowledge organization systems and KOS related registries, developed by the Basel University Library, Switzerland, <http://bartoc.org/it/node/18385>), local authorities, the European Union, private German individuals, Japanese companies, etc. This shows the spread of the Italian language (which is the fourth most spoken language in the world) and the appreciation that the thesaurus also receives abroad.

Since the thesaurus's birth, the BNCF has believed in the development of the semantic and technical interoperability and, today, offers its users the opportunity to navigate in the documentary heritage of prestigious Italian and foreign cultural institutions and to the users of other institutions the opportunity to acquire the heritage

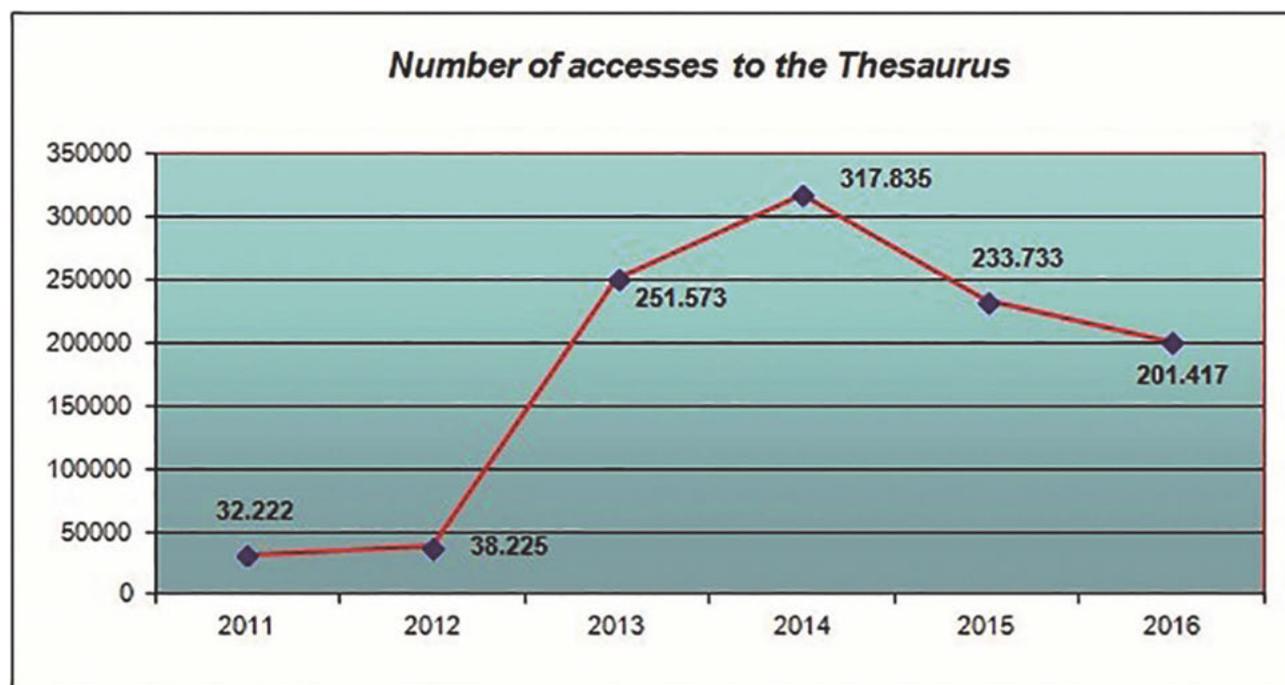


Figure 7. Number of accesses to the Thesaurus.

of the BNCF through the activation of crosswalks. The *Nuovo soggettario* is not only a web-accessible tool but has become part of the semantic web where content, objects, pointers, links, and deferrable URIs are found. A dataset is never exhaustive if it is published on the web without connections to other datasets. Tim Berners Lee, the inventor of the web and linked data, suggests a five-star deployment scheme for open data and explains the costs and benefits that come along with it. He emphasizes that interlinking adds the maximum value to the linked data “link your data to other data to provide context” (Berners-Lee 1996 and 2009). The *Nuovo soggettario* thesaurus, through structured data and SKOS format, is published on the DataHub platform (<https://datahub.csail.mit.edu/www/>) thus giving a broad visibility to the BNCF and its precious patrimony.

In conclusion, the indexing tool made available to Italian libraries is developing, on the one hand, with a wealth of open and reusable data in different contexts, and on the other, with an open, creative, and collaborative spirit. It is interesting to note that, in parallel with this quantitative and qualitative increase, there was an optimization of search and retrieval of information. It would be a valuable contribution to do an investigation to understand how a relational structure, made up of links and open data, generated by a tool such as the *Nuovo soggettario* thesaurus, enhances the search methods of users and therefore the effectiveness and pertinence of the results.

References

- Ayers, Phoebe. 2016. At Scale, At Large: Wikipedia & Libraries. *Medium* (blog). <https://medium.com/@phoebeyayers/at-scale-at-large-wikipedia-libraries-3e90a9618398>
- Ayers, Phoebe and Andrea Zanni. 2017. “Wikipedia e biblioteche: Una prospettiva globale.” *AIB studi* 57: 119-25.
- Bergamin, Giovanni and Anna Lucarelli. 2013. “The *Nuovo soggettario* as a Service for the Linked Data World.” *JLIS.it* 4, no. 1. doi:10.4403/jlis.it-5474
- Berners-Lee, Tim. 1996. The World Wide Web: Past, Present and Future. <https://www.w3.org/People/Berners-Lee/1996/ppf.html>
- Berners-Lee, Tim. 2009. Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html>
- Biblioteca Nazionale Centrale di Firenze, ed. 1956. *Soggettario per i cataloghi delle biblioteche italiane*. Florence: Stamperia il Cenacolo.
- Biblioteca Nazionale Centrale di Firenze. 2006. *Nuovo soggettario: Guida al sistema italiano di indicizzazione per soggetto; Prototipo del Thesaurus*. Milan: Editrice Bibliografica.
- Bruni, Silvia, Francesca Capetta, Anna Lucarelli, Maria Grazia Pepe, Susanna Peruginelli, and Marco Rulent. 2016. “Towards the Integration of Archives, Libraries and Museums.” *JLIS.it* 7, no.1. doi:10.4403/jlis.it-11482
- Buizza, Pino. 2011. “Subject Analysis and Indexing: An ‘Italian Version’ of the Analytico-Synthetic Model.” In

- Subject Access: Preparing for the Future*, ed. Patrice Landry, Leda Bultrini, Edward T. O'Neill and Sandra K. Roe. IFLA series on bibliographic control 42. Berlin: De Gruyter Saur, 25-36. <https://books.google.it/books?id=Fy6pqEGaSKQC&pg=PA25&dq=Subject+analysis+and+indexing:+an+%E2%80%9CItalian+version%E2%80%9D+of+the+Analytico-synthetic+model&hl=it&sa=X&ved=0ahUKEwjzhuzfgOvVAhUHxRQKHfUPDZ0Q6AEIFDAA#v=onepage&q=Subject%20analysis%20and%20indexing%3A%20an%20%E2%80%9CItalian%20version%E2%80%9D%20of%20the%20Analytico-synthetic%20model&f=false>
- Crociani, Laura, Maria Chiara Giunti, and Elisabetta Viti. 2016. "Trent'anni di Dewey in Italia: Il ruolo della Biblioteca nazionale centrale di Firenze e i nuovi sviluppi sul fronte dell'interoperabilità con altri strumenti di indicizzazione semantica." *AIB Studi* 56, no. 1. doi:10.2426/aibstudi-11408
- Dewey Decimal Classification and Relative Index*, 23rd ed. 2011. Dublin, OH: OCLC. Also available as online database, WebDewey (Italian version), at: <http://deweyit.pansoft.de/webdewey/login/login.html>
- European DDC Users Group, ed. 2016. EDUG's Recommendations for Best Practice in Mapping Involving Dewey Decimal Classification (DDC). Presentation from 2015 EDUG Workshop: Mapping to/from the Dewey Decimal Classification (Version 2016/06/21). <http://edug.pansoft.de/tiki-index.php?page=EDUG+workshops>
- Holm, Stuart A., comp. 1993. *Guidelines for Constructing a Museum Object Name Thesaurus*. MDA occasional paper 19. Cambridge: Museum Documentation Association.
- International Organization for Standardization. 2011. *ISO 25964. Thesauri for information retrieval*. Pt.1 of *ISO 25964: Information and Documentation; Thesauri and Interoperability with other vocabularies*. Geneva: ISO.
- International Organization for Standardization. 2013. *Interoperability with Other Vocabularies*. Pt.2 of *ISO 25964: Information and Documentation; Thesauri and Interoperability with Other Vocabularies*. Geneva: ISO.
- Landry, Patrice. 2001. "The MACS Project: Multilingual Access to Subjects (LCSH, RAMEAU, SWD)." *International Cataloguing and Bibliographic Control*. 30, no.3: 46-9. www.ifla.org/IV/ifla66/165-181.pdf
- Lucarelli, Anna. 2014a. "Nuove scommesse della BNCF: Wikipediani in residence, Wikisource e altro ancora." *DigItalia* 9, no. 2: 100-6.
- Lucarelli, Anna. 2014b. "Wikipedia Loves Libraries: In Italia è un amore corrisposto." *AIB studi* 54, nos. 2/3. doi:10.2426/aibstudi-10108
- Lucarelli, Anna. 2015. "Web dei dati alla Biblioteca nazionale centrale di Firenze." *DigItalia* 10: 30-9. <http://digitalia.sbn.it/article/view/1471/977>
- Lucarelli, Anna and Elisabetta Viti. 2015. "Florence-Washington Round Trip: Ways and Intersections between Semantic Indexing Tools in Different Languages." *Cataloging & Classification Quarterly* 53: 414-29. doi: 10.1080/01639374.2014.1004008
- Marc 21 Format for Classification Data*. Washington, DC: Library of Congress. Concise version available online at <http://www.loc.gov/marc/classification/>.
- MacEwan, Andrew. 2000. "Crossing Language Barriers in Europe: Linking LCSH to Other Subject Heading Languages." *Cataloging and Classification Quarterly* 29, nos. 1-2: 199-207.
- Olson, Tony, 2003. "Integrating LCSH and MeSH in Information Systems." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*, ed. I.C. McIlwaine. UBCIM Publications. 25. Munich: K.G. Saur, 21-24.
- Viti, Elisabetta. 2014. *Wikipedia/Wikidata nuove opportunità di multilinguismo nel Thesaurus del Nuovo soggettario*.
- Will, Leonard. 1993. "The Indexing of Museum Objects." *The Indexer* 18: 157-60.
- World Wide Web Consortium. 2008. SPARQL Protocol of RDF. W3C Recommendation, 15 January 2008.
- World Wide Web Consortium. 2009a. SKOS Simple Knowledge Organization System Reference. W3C Recommendation, 18 August 2009. <http://www.w3.org/TR/skos-reference>
- World Wide Web Consortium. 2009b. SKOS Simple Knowledge Organization System Primer. W3C Working Group Note, 18 August 2009. <http://www.w3.org/TR/skos-primer>
- World Wide Web Consortium. 2009c. SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL) Namespace Document-HTML Variant. W3C Recommendation, 18 August 2009. <http://www.w3.org/TR/skos-reference/skos-xl.html>