JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

JLIS.it

Thesauri in the Digital Ecosystem

Anna Lucarelli[(a)]

a) Biblioteca nazionale centrale di Firenze

Contact: Anna Lucarelli, anna.lucarelli@beniculturali.it
Received: 5 May 2021; **Accepted:** 21 May 2021; **First Published:** 15 January 2022

ABSTRACT

In recent years, thesauri have taken on new roles, new functions, and have shown some advantages over other knowledge organitazion systems (KOS). They are increasingly important in the linked data environment of the semantic web. The *Nuovo soggettario*, created and maintained by the National Central Library of Florence, is an example of the changing uses of controlled subject systems, like thesauri and subject heading lists. Thesauri are shown to be dynamic tools, essential components for the integration of data on the web, especially for mapping and to assist with interoperability among heterogeneous resources. With the adoption of formats of the semantic web, such as RDF/SKOS, and following international standards, thesauri have evolved and have proven to be increasingly useful with free reuse and across various frameworks. To varying degrees, they have enabled increased multilingualism and conceptual equivalences, connecting information and metadata produced by institutions of different countries. As authority control systems, they interact with Wikidata and help build 'bridges' between worlds that were too far apart until not long ago, namely libraries, archives, and museums. Will the challenge of search engines, machine learning and artificial intelligence override the thesauri or will it make them even more involved?

KEYWORDS

Bibliographic control; Linked open data; Nuovo soggettario; Thesauri.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Thesauri and bibliographic control

Since the beginning of IFLA's UBC Programme, universal bibliographic control has been primarily focused on the sharing and standardization of descriptive cataloguing. Talking about thesauri gives rise to the following questions:

- the current state and the possible new future of subject indexing of which thesauri are essential components, along with subject heading lists (Petrucciani 2019, 163-173);
- the path followed by subject indexing in recent years; a path that is considered 'autonomous' compared to other cataloguing processes; a path strongly connected to the procedures and to the languages used in various countries, in several cultural, geographical and, above all, linguistic contexts;
- the relationship of thesauri with other knowledge organization systems (KOS), such as ontologies, classifications, taxonomies, and so on (Gnoli 2020);
- the role they play in current bibliographic control, considering that the concept of bibliographic control in the digital ecosystem is changing and is evolving *Dalla catalogazione alla metadazione* (from cataloguing to creating metadata), just to use the title of a recent book (Guerrini 2020) and "transitioning to the next generation of metadata" (Smith-Yoshimura 2020). This is a period when cataloguing tools, data management, and infrastructures are more than ever crossing transitional borders, tied to strategies to make bibliographic data more visible on the web.[1]

We now have an opportunity for a better integration of subject data in universal bibliographic control.
As we will see, thesauri have taken on new roles, new functions and shown some advantages over other knowledge organization systems (KOS). Yet, there are many arguments and confrontations on this issue.
In their multiple types (general or specialized domains; polyhierarchical or monohierarchical, monolingual or multilingual, etc.), thesauri continue to prove their effectiveness compared to simple lexicons or *flat lists*; they have proved to be versatile, usable, both in the framework of the post-coordinated and pre-coordinated languages, in which the rules for the citation order in the subject strings are added to vocabulary control.

Controlled vocabularies are studied by the Subject Analysis and Access Section of IFLA[2], but even by the International Society for Knowledge Organization (ISKO) with its regional chapters.[3] Thesauri are also handled by terminology associations,[4] a transversal discipline. Unfortunately, the communities that are involved are not always interactive among one another. The relationship between terminology experts and librarians engaged in subject indexing still continues to be weak and not as creative as it could be.

[1] For a selective bibliography on the current state of the subject indexing, specifically with regard to the French reality, see: *L'indexation matière en transition: de la réforme de Rameau à l'indexation automatique* 2020.
[2] https://www.ifla.org/subject-analysis-and-access.
[3] https://www.isko.org/.
[4] E.g., Associazione italiana per la terminologia (Ass.I.Term): http://www.assiterm91.it/.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Even thanks to their formalized structures, thesauri have been significantly supported by standardization. Subject indexing is one of the rare library tasks specifically regulated by the International Organization for Standardization (ISO). Let's mention the ISO 5963:1985 standard (*Methods for examining documents, determining their subjects, and selecting indexing terms*) on the conceptual analysis, recently validated in 2020, and ISO 25964:2011-2013 (*Thesauri and interoperability with other vocabularies*), just concerning the thesauri themselves, and renewing ISO-5964:1985 and ISO-2788:1986, established before the digital universe existed. However, these are not the only standards about both documentation and terminology. Within the Italian framework, groups and technical committees of Commissione UNI CT/014[5] also deal with them.

Rise or fall of thesauri?

The national libraries have tried to make their vocabularies, used for subject indexing, more 'visible' and usable, through various modes of integration with their own OPACs or with the open data hubs.

Subject indexing in libraries has however suffered a slowdown, not only because the data referred to semantic contents is considered to be less necessary, but also because indexing is a labor-intensive and hence expensive process. The lack of human resources is not only an Italian problem, though.

Nevertheless, in recent years, the development of thesauri has spread everywhere.

This spread is studied by BARTOC, a database developed and maintained at the University of Basel, since 2013. It inventories various systems for the organization of knowledge, including web applications and mapping, so far totalling 3,393 (data as of November 2021).[6]

In particular for thesauri, BARTOC monitors the establishment of thesauri across the world, by describing them by their main features, identifying 781 thesauri to date. Almost all of them are in digital format and freely accessible on the web by open licenses, many of them being multilingual.

Thesauri are inventoried and assessed by librarians and institutes around the world. Assessment includes for their performance, for their compliance with the standards, and for their level of semantic coverage.

They are assessed according to their ability to handle their own terminology expansion, starting with a particular *corpora*, and for their ability to be representative with regard to specialized domains (Folino and Parisi 2020). Such issues in Italy are examined not only by librarians, but also by CNR Institutes and by centres of excellence such as the Laboratorio di Documentazione dell'Università della Calabria.[7]

Thesauri are further assessed for the possibility of being integrated with algorithms employed for the automated indexing.

Ultimately, they are assessed according to the quality of their data (e.g., ability to be re-used) and according to the sustainability of the resulting costs, also considering that the personnel involved in creating and maintaining thesauri are required an indispensable professional development.

[5] https://www.uni.com/index.php?option=com_uniot&view=struct&id=853557&Itemid=2447.
[6] http://bartoc.org/.
[7] https://www.labdoc.it/.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

We can think that the development of these tools may depend on the fact that the terminology has acquired more and more importance within "metadata-ing". Yet, it is not only a matter of this. No longer limited to the library and documentation world, these tools have actually gone beyond the context of subject indexing and of information retrieval (IR); they have been involved in other 'universes'.

Following the standardization established by ISO 25964 and by RDF/SKOS,[8] thesauri organize both the concepts and the terms by which they are represented. The central role of the concept (that is a unit of thought, rather than a lexical element of a specific language), has ensured that the borders which separate thesauri from other knowledge organization systems have become more fluid. There are two reasons for this: for the possibility of the correct *reconciliation* of expressions in different languages, and for the opportunity to compare different systems starting from the conceptual cores on which they are established. It is no coincidence that "metathesauri" have also been set up (e.g., UMLS by the National Library of Medicine in the United States[9]).

An approaching process has been activated among classifications, schemes based on subject headings and ontologies; in the latter, the relationships among concepts are less standardized. The very relational structure of thesauri, when rigorous, encourages their evolution towards the ontologies (Biagetti 2018; Biagetti 2020).

Vanda Broughton, Leonard Will and Stella Dextre Clarke have faced such interesting issues in a recent series of virtual classes organized in 2020-2021 by ISKO UK.[10]

One characteristic of thesauri is that of being dynamic tools, obviously linked to the linguistic fabric of the context in which they are established, yet, often, through multilingual functionalities. In addition, thesauri have shown their capabilities, not only as 'tools of the trade' for librarians but even as tools for users. Yet, we know that this happens if they are provided in the right way, if they are 'well integrated' in OPACs, and if librarians employ them also as a support to reference service and to information literacy (Ballestra 2011, 395-401).

The reason why they are so costly is due to the constant maintenance work they require, along with a careful supervision of the increase mechanisms in relation to the 'literary warrant'.

They also require a continuous assessment of their structural coherence, a monitoring of the semantic relationships, particularly for synonymy and lexical variants on the one side, polysemy and new meanings on the other side.

Languages (of works, of users, of catalogues) quickly evolve, so the work to be carried out on neologisms is continuous. To give a current example, let's think about the importance to 'control' concepts connected to the pandemic we are living in and which are the subjects of works already published. SARS-CoV-2; COVID-19; Social distancing; Confinement; Lockdown; Contact tracing… (see Figures 1-5). These terms were added promptly and captured some of these new concepts from the first months of 2020. Not all vocabularies acquire new concepts with the same timeliness.[11]

[8] https://www.w3.org/2004/02/skos/.

[9] https://www.nlm.nih.gov/research/umls/index.html.

[10] https://www.iskouk.org/KOED.

[11] For a first survey on the terms tied to COVID-19 pandemic, inserted into Thesaurus of *Nuovo soggettario* already since March 2020: Francioni and Lucarelli 2020. On the Italian words about pandemic also Accademia della Crusca: https://accademiadellacrusca.it/it/contenuti/lacruscaacasa-le-parole-della-pandemia/7945.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Examples of new concepts related to the current world health crisis:

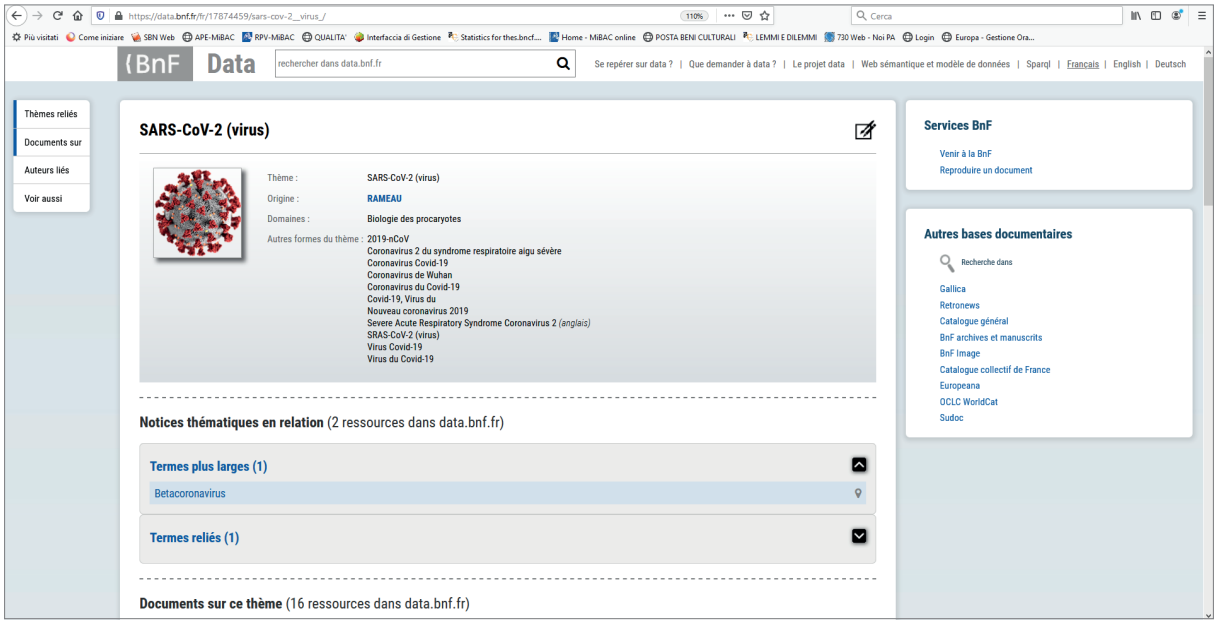

Fig. 1. The concept *SARS-CoV-2 (virus)* in RAMEAU

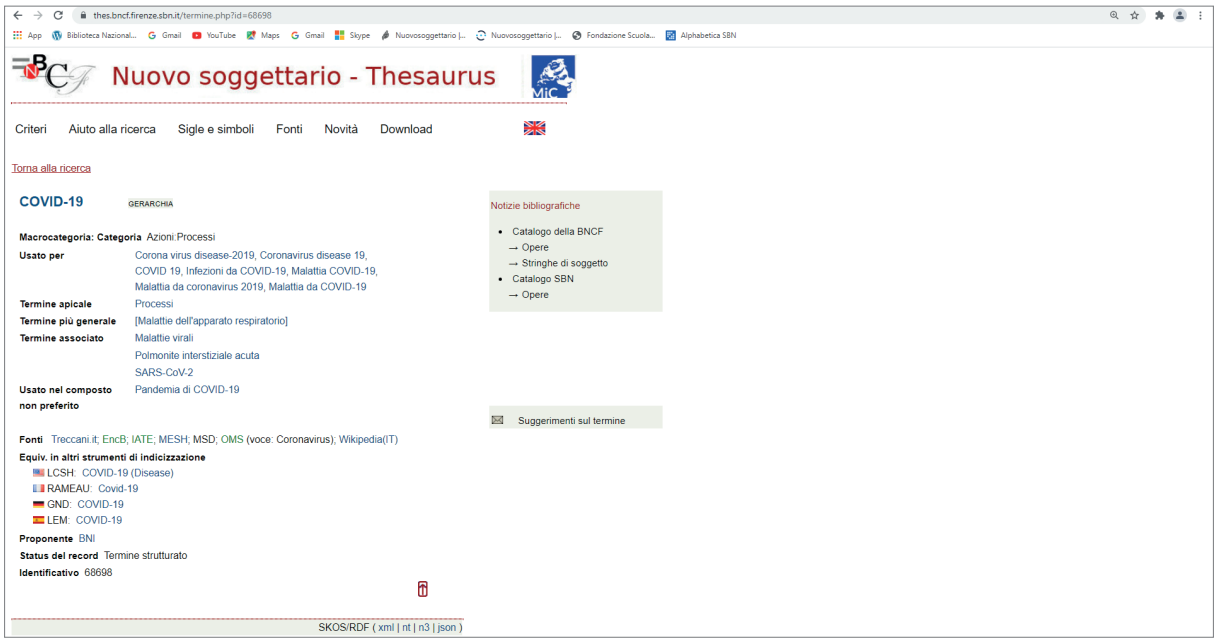

Fig. 2. The concept *COVID-19* in *Nuovo soggettario*

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

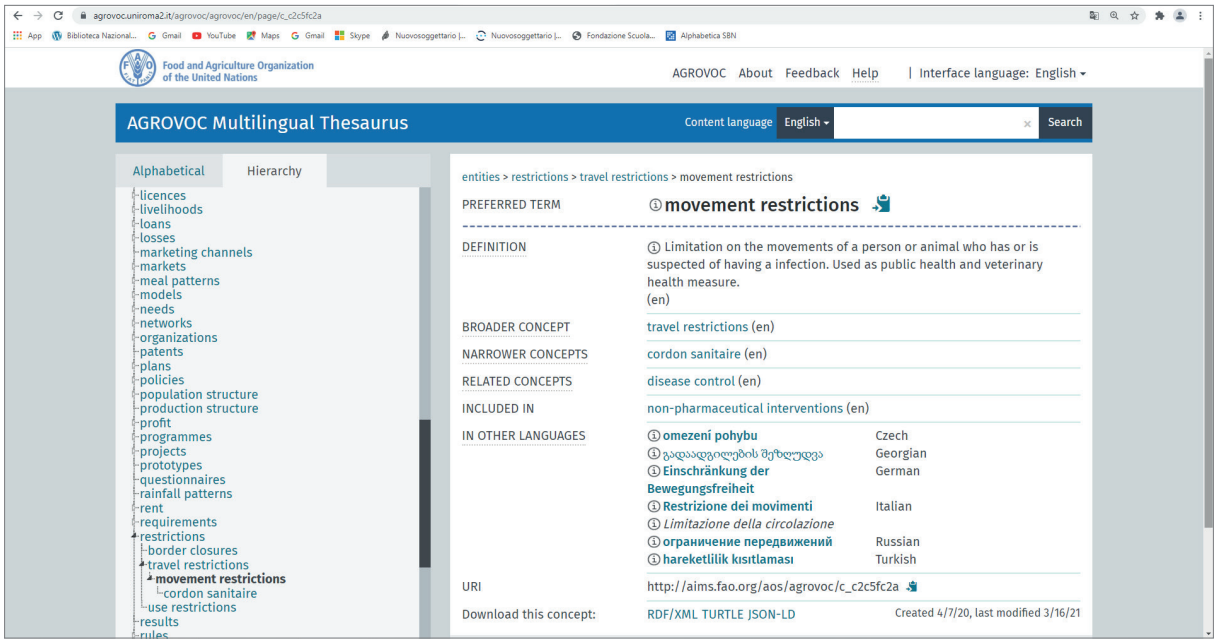

Fig. 3. The concept *Movement restrictions* in AGROVOC

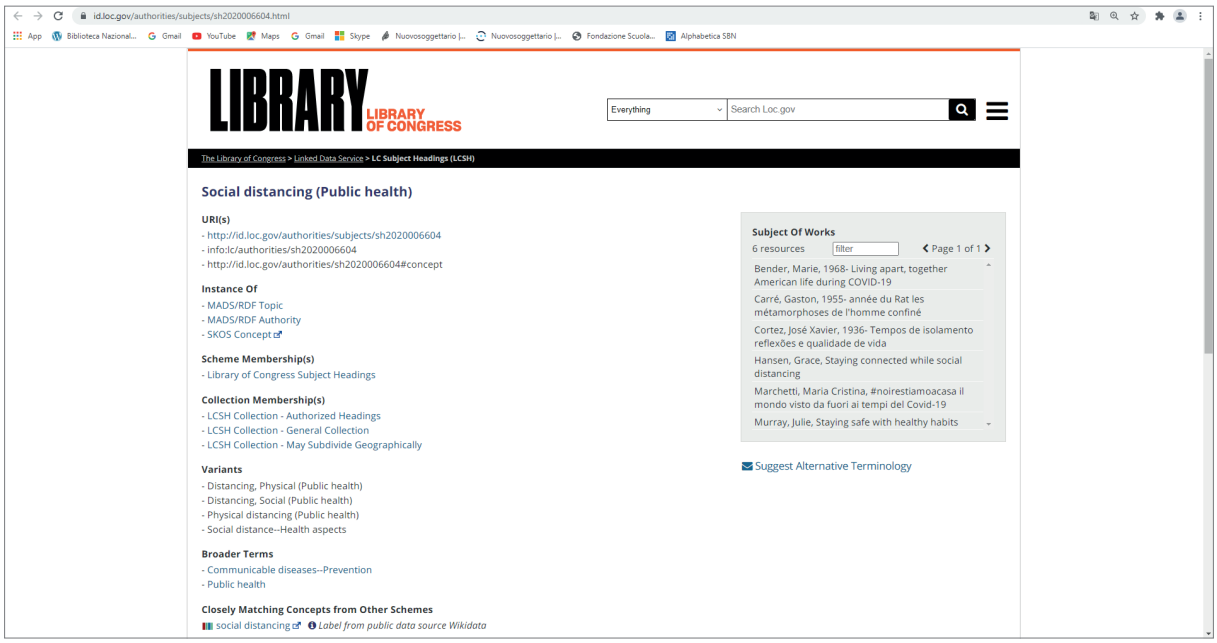

Fig. 4. The concept *Social distancing (Public health)* in LCSH

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

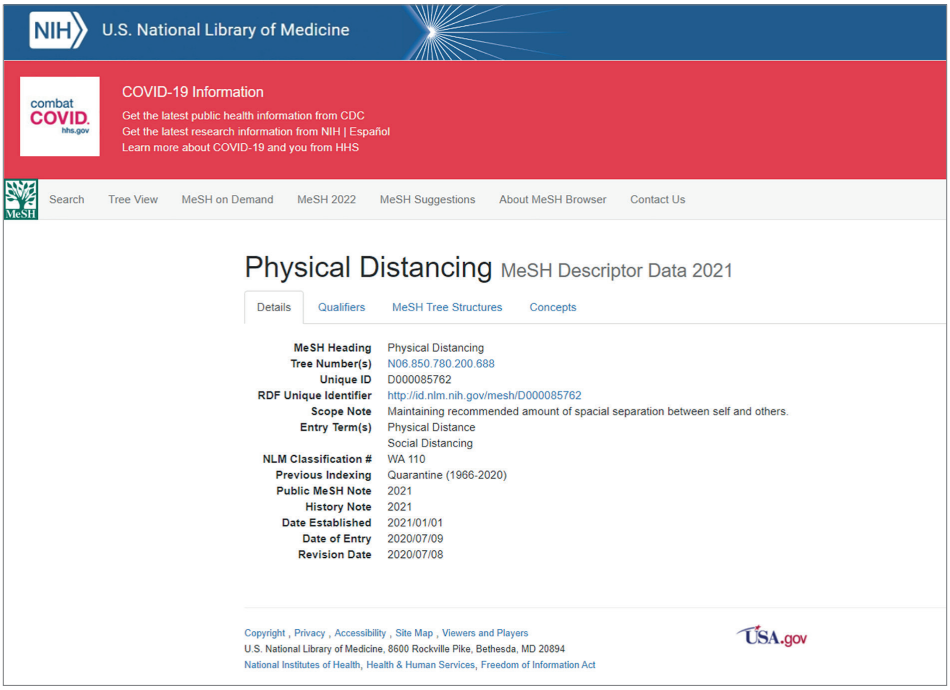

Fig. 5. The concept *Physical Distancing* in MESH

Integration of data on the web

What is important is that thesauri have proved to be the essential components for the integration of data on the web and thus fundamental elements for the affirmation of the semantic web.
We are dealing with the role of the thesauri within the semantic web at various levels and in various contexts and there are many studies on this topic (e.g., Martinez-González and Alvite Díez 2019). We could say that they are among 'the best friends' of the semantic web, for their capability to provide metadata in RDF, that is to say in open formats which allow their re-use in the most varied contexts (not necessarily library ones), because they encourage the development of mapping as well as the interoperability between heterogeneous resources (Zeng 2019, 122-146).
When we wonder which is the most re-used data among those processed by libraries, thesauri are a good example.
Many of them are connected with DbPedia.[12] Tens of other thesauri have recently connected to Wikidata.[13] The Italian Thesaurus of *Nuovo soggettario*[14] – created and maintained by the National Central Library of Florence (BNCF) – has had links with Wikipedia since 2007. Since

[12] https://wiki.dbpedia.org/.
[13] https://www.wikidata.org/w/index.php?title=Special:WhatLinksHere/Q89560413&limit=500.
[14] https://thes.bncf.firenze.sbn.it/ricerca.php. BNCF, with an almost centuries-old tradition for subject indexing (started in 1925), has the institutional task to curate the Italian subject indexing tools. *Nuovo soggettario* contains the concepts/terms employed in the framework of a pre-coordinated language that contemplates also the rules on the construction of the subject strings. Yet, thesaurus is obviously usable also for the post-coordinated indexing. It is employed by the Italian National Bibliography (BNI) and by most libraries of the Italy's National Library Service (SBN). It was also presented during IFLA General Conference 2009 (Cheti, Alberto, Anna Lucarelli, and Federica Paradisi. 2009).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

2013, reverse mapping has been implemented with a mutual browsing mechanism as well as with a synchronization realized through the field P508 (BNCF Thesaurus ID) of Wikidata (Lucarelli 2014).[15]

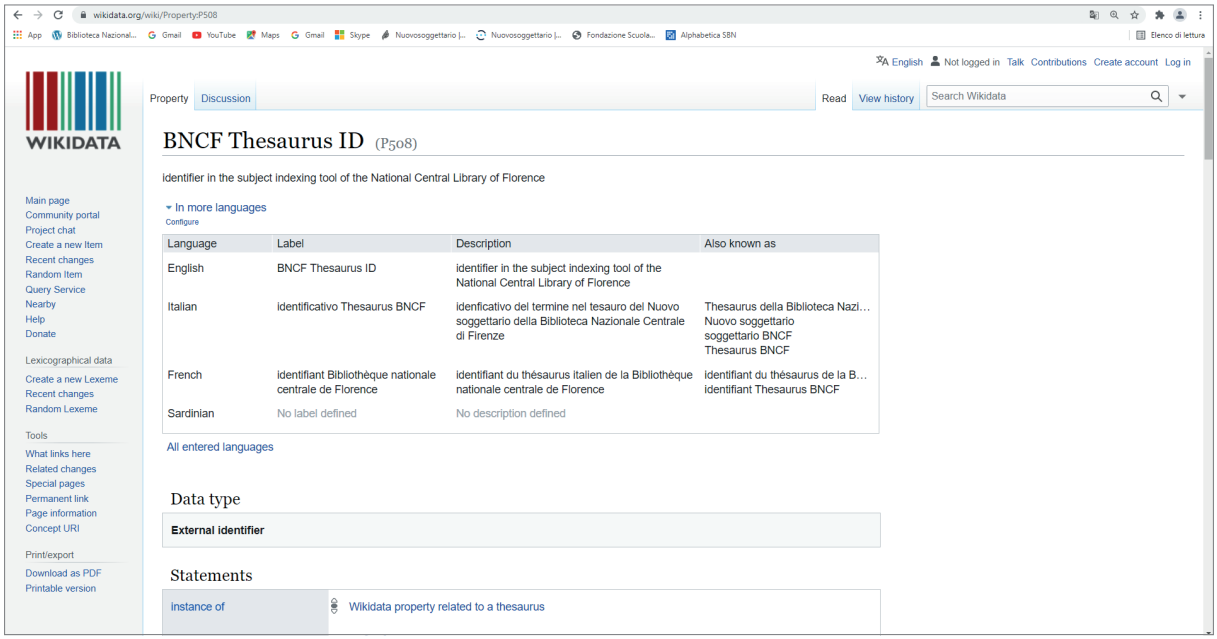

Fig. 6. Thesaurus of the National Central Library of Florence and Wikidata

Since thesauri are among the 'main actors', the interpreters of the semantic web, we must evaluate their costs on the basis of the benefits they bring to the linked open data and on the possibility of creating mapping, as Stella Dextre Clarke has recently reminded us in the above-mentioned virtual classes[16].
The opportunities offered by open data and by mapping, in both the research world and public administrations, are unquestionable. Some examples?
A few years ago, the City of Florence made use of the open data of BNCF's Thesaurus in order to organize the City's open data.
When, in May 2013, the reverse mapping from the Wikipedia entries to the corresponding terms of *Nuovo soggettario* was implemented, the number of visitors to the OPAC of BNCF has increased 28% in only one month.

As we can see from this example, it is no longer possible to talk about thesauri without talking about interoperability. The international standard ISO 25964 dedicates the second of its two parts to the methods for the realization of this interoperability. Furthermore, the interoperability ac-

[15] The *Nuovo soggettario* was the first general thesaurus to activate a form of interlinking with a version of Wikipedia in a specific language, preceded, at an international level, by experiences in specialized sectors as in the case of Thesaurus for Economics of Leibniz-Informationszentrum Wirtschaft (http://zbw.eu/stw/version/latest/about).
[16] About the costs resulted from the procedures of the bibliographic control, also Bergamin 2020, p. 167.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

tivated by thesauri has made them fundamental 'hubs' as well as 'bridges' for the connection between data from different institutions. Mapping also has been realized with particular success in the context of multilingualism. Not only multilingual vocabularies (such as the notable AGROVOC, AAT, EUROVOC, IATE, that are often cited, in a crossed mode, interconnecting one another), but also monolingual vocabularies with equivalences in other languages in the form authorized by those other vocabularies or subject heading schemes. At the same time, Pat Riva explained the importance of multilingualism and of the internationalization of the bibliographic description in order to facilitate access (Riva 2021).

In the revolution of open data, thesauri are thus on the 'front line'. Many of them have implemented new formats for the publications and the exchange of metadata (i.e., SKOS) by exceeding the previous ones (i.e., Zthes). They have become structures "of" the web.
In the linked open data cloud, many controlled vocabularies are represented, including those created and maintained by the national libraries.[17]
The Thesaurus of *Nuovo soggettario* has been in SKOS since 2010 and has achieved the 'five stars' of Tim Berners-Lee.[18] It can also be found in the hub *dati.beniculturali* of the Ministero della Cultura.[19]

The initiatives of national libraries and national bibliographies

Since the publication of IFLA's *Guidelines for subject access in National Bibliographies* ten years have passed, but many indicated best practices are still valid.[20] Following these guidelines, both national libraries and national bibliographies that are assigned to the bibliographic control of our countries, have implemented important choices in the field of subject indexing.
Many national libraries have updated their bibliographic tools to follow the latest standards and entered the world wide web of data, following new 'conceptual models'.
For some of these institutions it has been a period of reforms, like for the Bibliothèque Nationale de France which, in 2019, made public its *Réforme de Rameau.*[21]
Regardless of the subject indexing language used, the national libraries continue to benefit from the controlled vocabularies even when indexing graphic resources, audio resources, ancient works and, in certain countries, works of fiction as well. They even use controlled vocabularies when providing Genre/Form descriptions, a practice that is also supported by IFLA.[22] In some cases, they use expressly dedicated thesauri, for the indexing of particular types of resources, for instance, the *Library of Congress Genre/Form Terms for Library and Archival Materials* (LC-GFT).[23]

[17] https://lod-cloud.net/clouds/publications-lod.svg.
[18] https://lod-cloud.net/dataset/bncf-ns.
[19] https://dati.beniculturali.it/altri-linked-open-data-del-mibact/.
[20] https://www.ifla.org/publications/ifla-series-on-bibliographic-control-45.
[21] https://rameau.bnf.fr/syntaxe.
[22] https://www.ifla.org/node/8526.
[23] https://id.loc.gov/authorities/genreForms.html.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

National libraries generally use these vocabularies for projects of automated indexing or semi-automated indexing of online resources, by having them interact with implemented algorithms. For example, this is part of the subject cataloguing policies of the Deutsche Nationalbibliothek, as explained by Ulrike Junger since the beginning (Junger 2018), and also more recently described by Mödden and Suominen (Mödden 2021; Suominen 2021).
In the name of the data quality, the use of vocabularies continues to rely on uncontrolled keywords. Thanks to mapping to RDF and to open data's hubs, the national libraries' vocabularies encourage a connection among different OPACs, which hopefully is a prelude to additional future forms of connections; some of these connections were originated from the project named MACS (Multilingual Access to Subjects) which was exceptionally innovative and whose operational phase started in 2005.[24]

As we can see in the figures below, starting a search with the subject term employed by the Deutsche Nationalbibliothek, one sees the connected publications, but it is also possible to move to the French equivalence of data.bnf, where resources on that topic can be explored. Through the correspondent RAMEAU page, it is possible to browse towards *Library of Congress Subject Headings* (LCSH) where works about the same topic can be explored in the catalogue of the Library of Congress. The equivalences are generally ensured by the form of the closely matching concepts from other schemes, as well evidenced by LCSH.

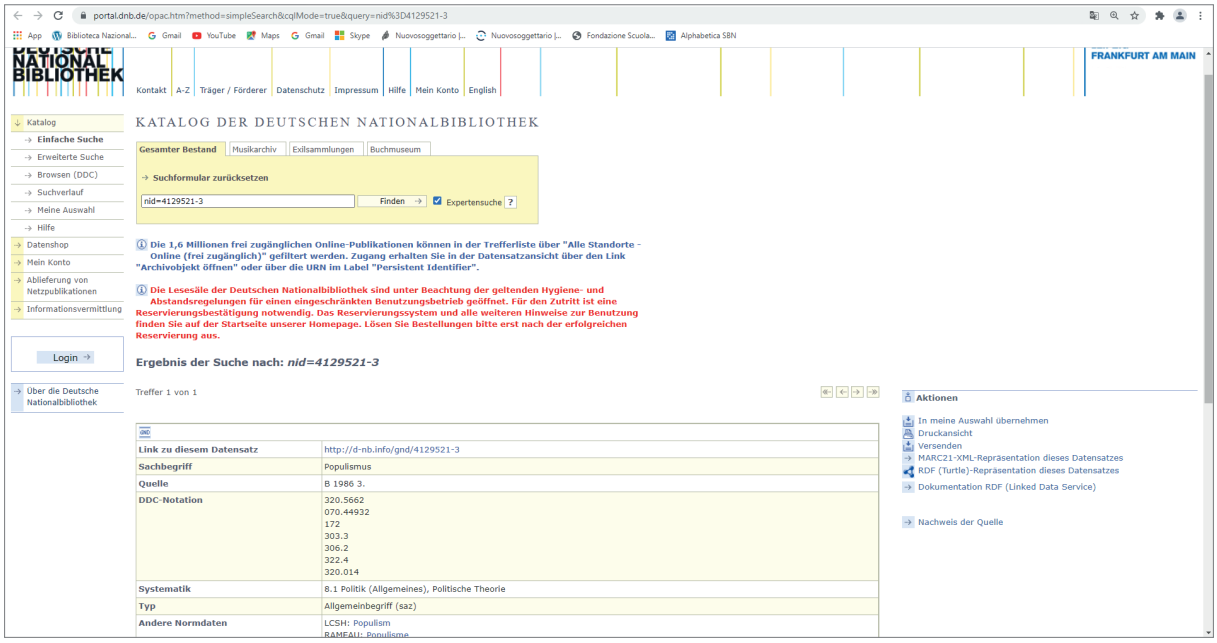

Fig. 7. Concept with equivalences in Gemeinsame Normdatei (GND) and links

[24] https://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/9041.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

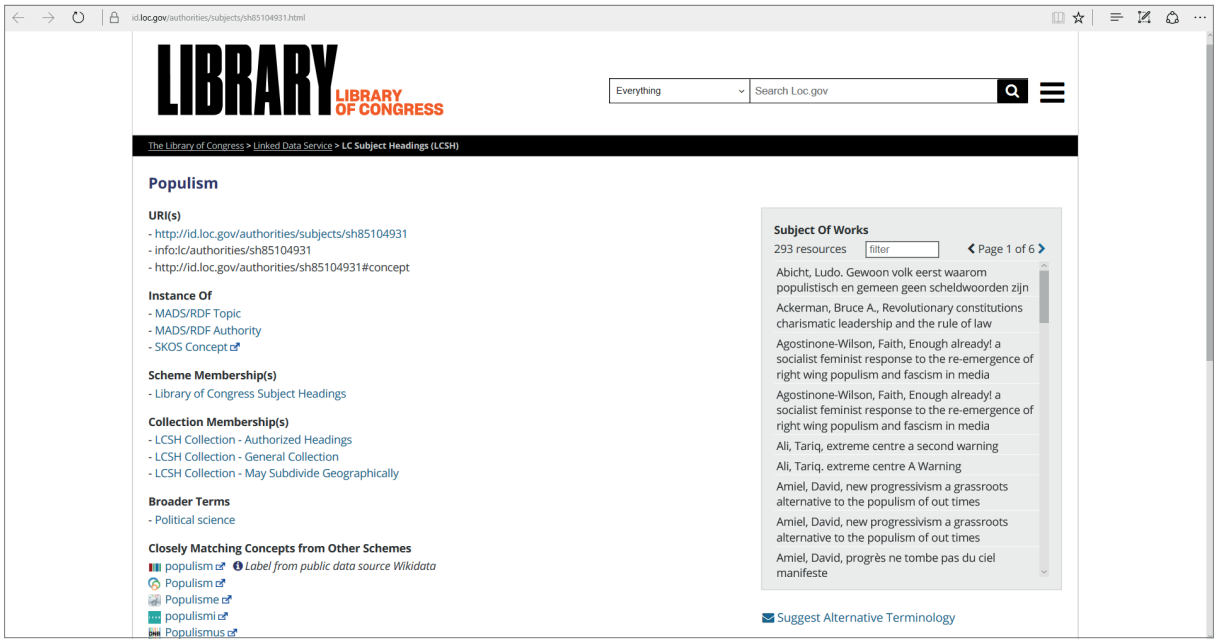

Fig. 8. Concept with equivalences in *Library of Congress Subject Headings* (LCSH) and links

Likewise the Thesaurus of *Nuovo soggettario* has been connected to the works described in the online catalogues of the National Central Library of Florence and Italy's Servizio Bibliotecario Nazionale (SBN), as shown in Figure 9, it has also been possible to navigate to Datos. BNE, that is, to the controlled equivalents of the Biblioteca Nacional de España, and, from there, it has been possible to explore the *Obras* on the same subject in the BNE catalogue.[25]

[25] https://datos.bne.es/tema/XX525409.html.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

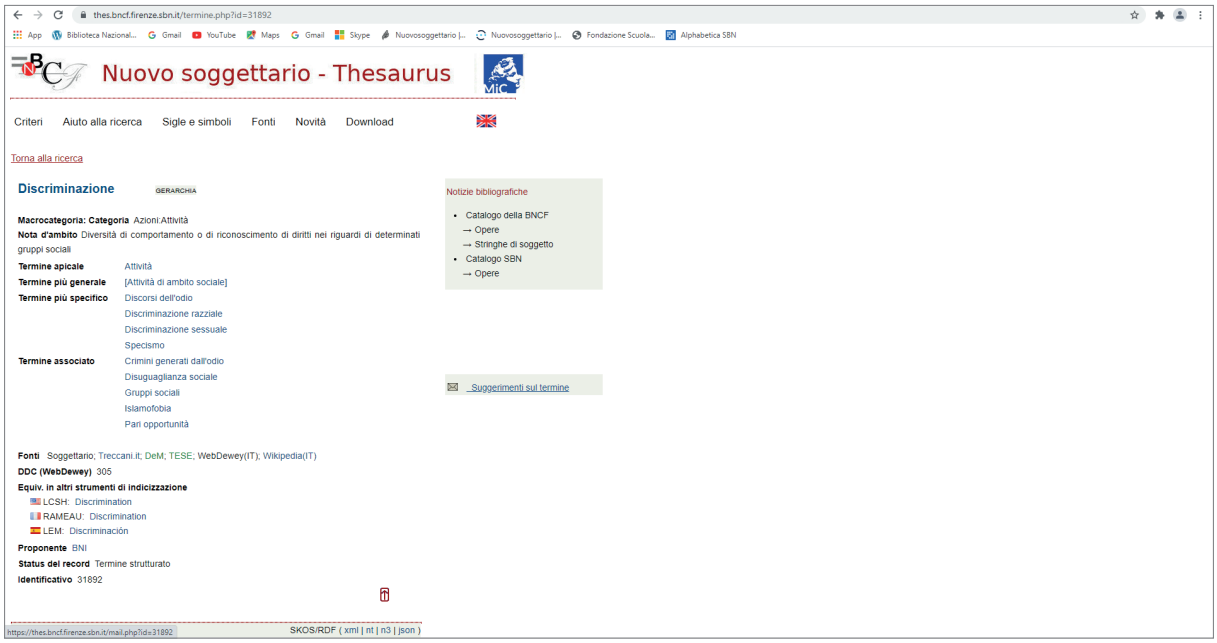

Fig. 9. Concept with equivalences in *Nuovo soggettario* and links[26]

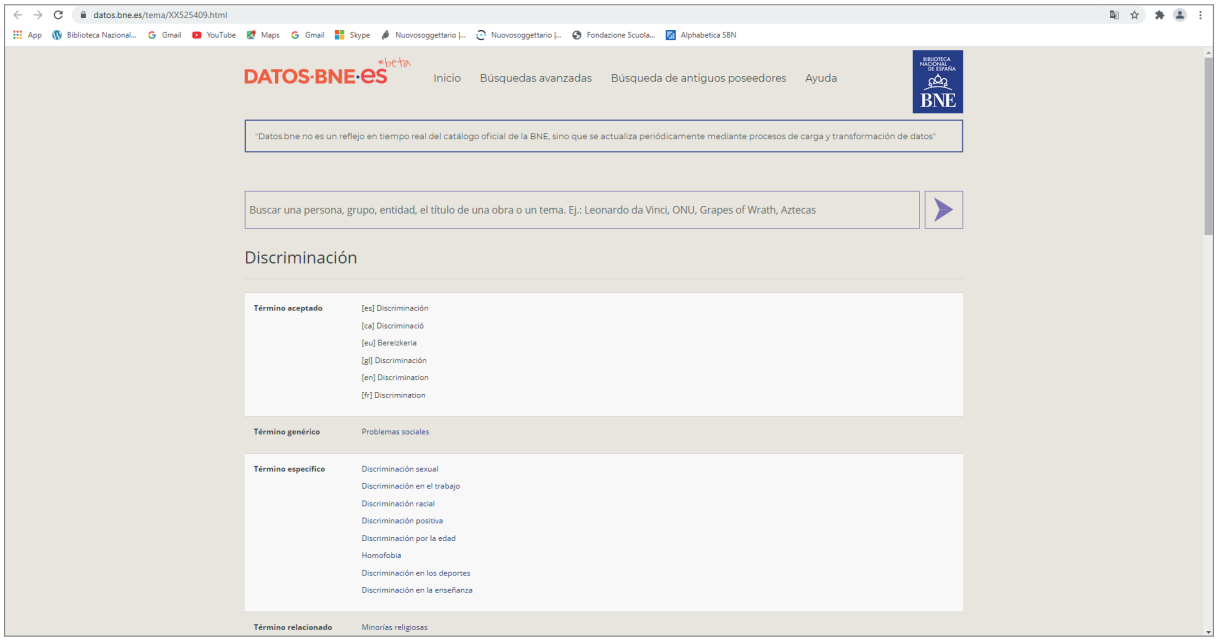

Fig. 10. Concept with equivalences in EMBNE and links

[26] The figure refers to the result of the research carried out on the Thesaurus at the time of the Conference (8th-12th February 2021) . For current results, see: https://thes.bncf.firenze.sbn.it/termine.php?id=31892

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

In fact, over the years, the Italian Thesaurus of *Nuovo soggettario* has considerably increased the mapping with other KOS and with equivalents of other vocabularies and continues to link to more equivalences (Viti 2017, 624-637). The number of links with LCSH has increased from 390 in 2011 to the current 14,970; with the French terms of RAMEAU from 380 in 2012 to the current 13,380 links; with the German terms from 130 in 2018 to the current 2,200; with the Spanish terms from 300 in 2019 to the current 2,270.[27]

Making such links is challenging work, requiring careful mapping and not without problems. For instance, there are challenges about the level of equivalences among concepts, especially across languages. This was explained by Pino Buizza in one of his latest papers on mapping between the Thesaurus of *Nuovo soggettario*, in Italian, and the two subject heading lists produced by national bibliographic agencies in the United States and in France: the *Library of Congress Subject Headings*, in English, and the *Repertoire d'autorité-matière encyclopédique et alphabétique unifié*, in French (Buizza 2020, [59]-68).

The equivalences found in *Nuovo soggettario*, when downloaded through SKOS, can activate mutual connections or give rise to the indication of the variant in Italian, as data.bnf shows in Figure 11 under the term *Épidémies*, among the *Autres forms du thème*.

Such initiatives demonstrate the importance that policies be activated among national libraries.

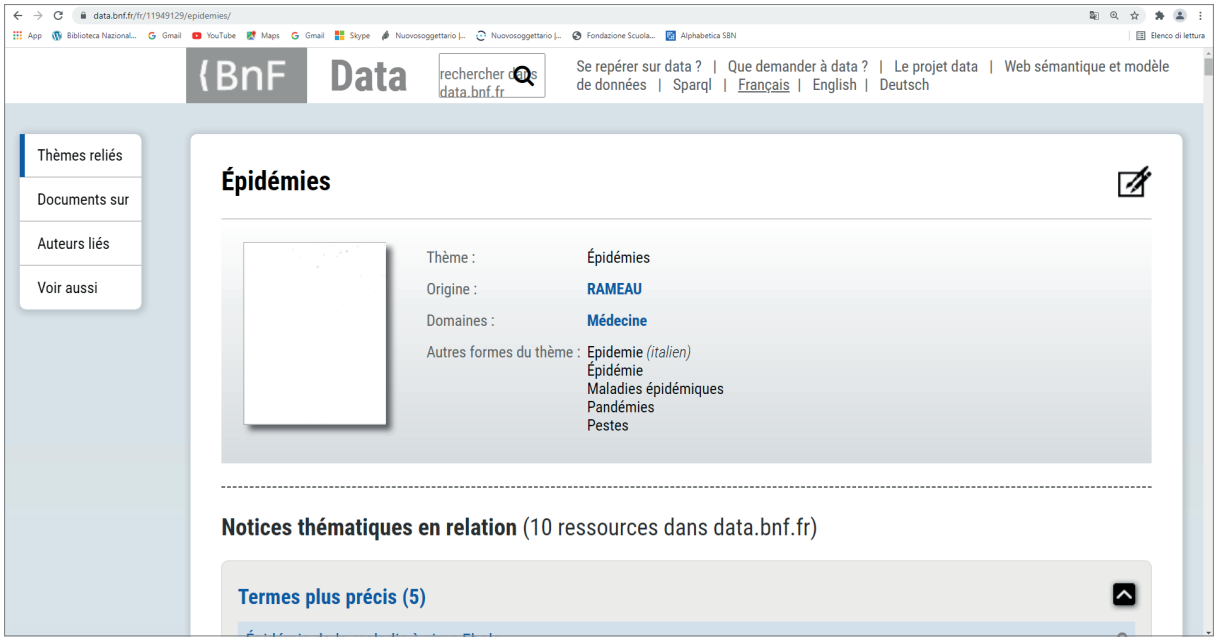

Fig. 11. Mutual connections or the indexing of the variant in Italian

[27] The comprehensive data on the trend of the equivalences in other languages are visible in: https://thes.bncf.firenze.sbn.it/stat.php.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Thesauri and Authority control: connection with other interlocutors

Collaborating on policies does not mean that the different indexing languages used by national libraries and connected through the respective vocabularies must have the same characteristics, the same syntactic rules. Not all tools have the same compliance with the standards, the same structure or functionality. Not all of them are polyhierarchical. Not all of them have comprehensive hierarchies up to the top term. Not all receive both common and proper names. Not all have the same integration with an OPAC and open data.

What brings them together is the progressive alignment among one another, the fact that they achieve common features, for example, to be integrated within Wikidata, so they are all visible on Wikipedia.

Who would have imagined that an encyclopedia would connect its own entries with the most important controlled vocabularies created by national libraries for the purpose of the bibliographic control? At the bottom of the Wikipedia page, you can find the box 'Authority control' with the relevant links.[28]

We know that libraries are not the only producers of bibliographic data, and that other operators are involved in universal bibliographic control. Yet, the data produced by 'certain' major libraries keep reflecting the highest level of quality.

Thesauri and recent features in today's context

Other issues related to thesauri within the digital ecosystem might be added to the above-described panorama. I take a cue from the *Nuovo soggettario* to outline some particularly interesting ones:

1. It has grown in size.
   Nuovo soggettario, in compliance with ISO 25964, has so far had a remarkable quantifiable increase: from 13,000 terms of the prototype to the current 67,000 terms.

2. It has a new interface.
   Since 2020, it has had a new, more user-friendly interface, implemented during the development of BNCF's new web site.

3. It interfaces with classification systems.
   Beyond the above-mentioned multilingualism, *Nuovo soggettario* maps with the Italian WebDewey (Crociani, Giunti, and Viti 2016), etc.

4. It has increased coverage in various subject domains.
   Thanks to the institutions that collaborate with BNCF,[29] it has largely enhanced its general coverage and expanded coverage in specific domains.

[28] See, for instance, the connections to the main thesauri at the footnotes of *Arredo urbano* of the Wikipedia in Italian language through the "Controllo di autorità": https://it.wikipedia.org/wiki/Arredo_urbano.
[29] https://thes.bncf.firenze.sbn.it/enti.htm.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

5. It can be employed for the Genre/Form indexing.
   This will be possible once our OPACs implement the tag MARC 655.

Also, the Thesaurus of *Nuovo soggettario* is employed apart from BNCF for the subject indexing of specialized resources:

- for audio and audiovisual resources, as for instance, in projects on oral sources of the Istituto centrale per i beni sonori e audiovisivi (ICBSA) (Magrini 2021);
- for graphic resources, for instance for photographs, also in BNCF but additionally in photographic libraries, for example, in the Fototeca - Biblioteca Panizzi;[30] for iconographic resources and maps, for instance in the Museo Galileo (Pocci 2020);[31]
- for archival resources, for example, for the documents indexed in projects of BNCF in collaboration with both Soprintendenza archivistica e bibliografica della Toscana,[32] and Historical Archives of the European Union.[33]

Integration of the *Nuovo soggettario* with databases of archives and museums

This connection of the Thesaurus of *Nuovo soggettario* with databases of archives and museums is quite interesting.

Let's look first at the Gallerie degli Uffizi, one of the most important museums in the world.[34] In 2019, BNCF started a partnership, a "Research pact," with the Uffizi.[35]

From *Violini* of *Nuovo soggettario* [36] it is possible to browse through the records of the Gallerie degli Uffizi catalogue thanks to the connection with *Violino* of the "Scheda OA" (Opere/oggetti d'arte) for the object's definition.[37] A reverse connection can also be seen from the record of the Museum. When the concept from the *Nuovo soggettario* indicates an iconographic subject (for instance *Albero della vita* [tree of life]), the link is to the Uffizi works that represent that subject, as shown in Figure 12.

From some terms, for example *Sestanti* [Sextant] (as shown in Figure 13), it is possible to view the resources of both the Gallerie degli Uffizi and the Museo Galileo.[38]

[30] http://panizzi.comune.re.it/Sezione.jsp?titolo=Fototeca&idSezione=233.

[31] https://www.museogalileo.it/it/biblioteca-e-istituto-di-ricerca/biblioteca-digitale/collezioni-tematiche/747-bibliotheca-perspectivae.html.

[32] http://sa-toscana.beniculturali.it/index.php?id=2.

[33] https://www.eui.eu/en/academic-units/historical-archives-of-the-european-union.

[34] https://www.uffizi.it/.

[35] https://www.beniculturali.it/comunicato/uffizi-e-biblioteca-nazionale-di-firenze-patto-per-la-ricerca.

[36] https://thes.bncf.firenze.sbn.it/termine.php?id=17664.

[37] http://www.iccd.beniculturali.it/it/ricercanormative/29/oa-opere-oggetti-d-arte-3_00.

[38] https://thes.bncf.firenze.sbn.it/termine.php?id=30976; http://catalogo.uffizi.it/it/29/ricerca/iccd/?search=*&fromRA=true&filter_OGTD-words=%3D&filter_OGTD=Sestante; https://catalogo.museogalileo.it/oggetto/Sestante.html.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

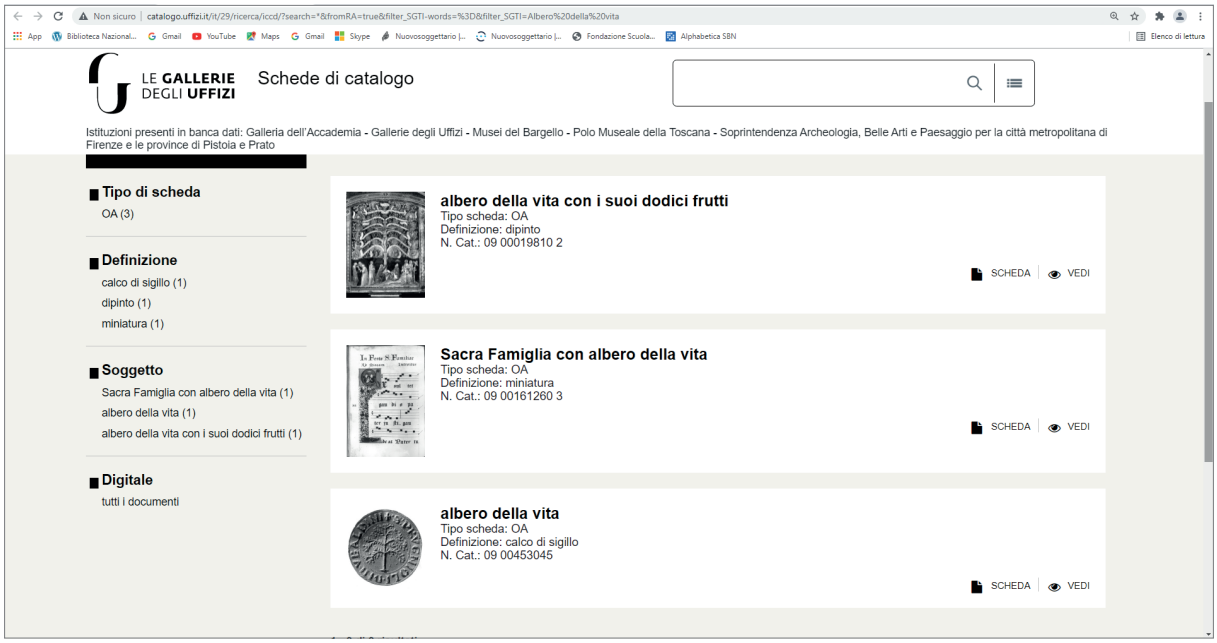

Fig. 12. The Uffizi works on *Albero della vita*

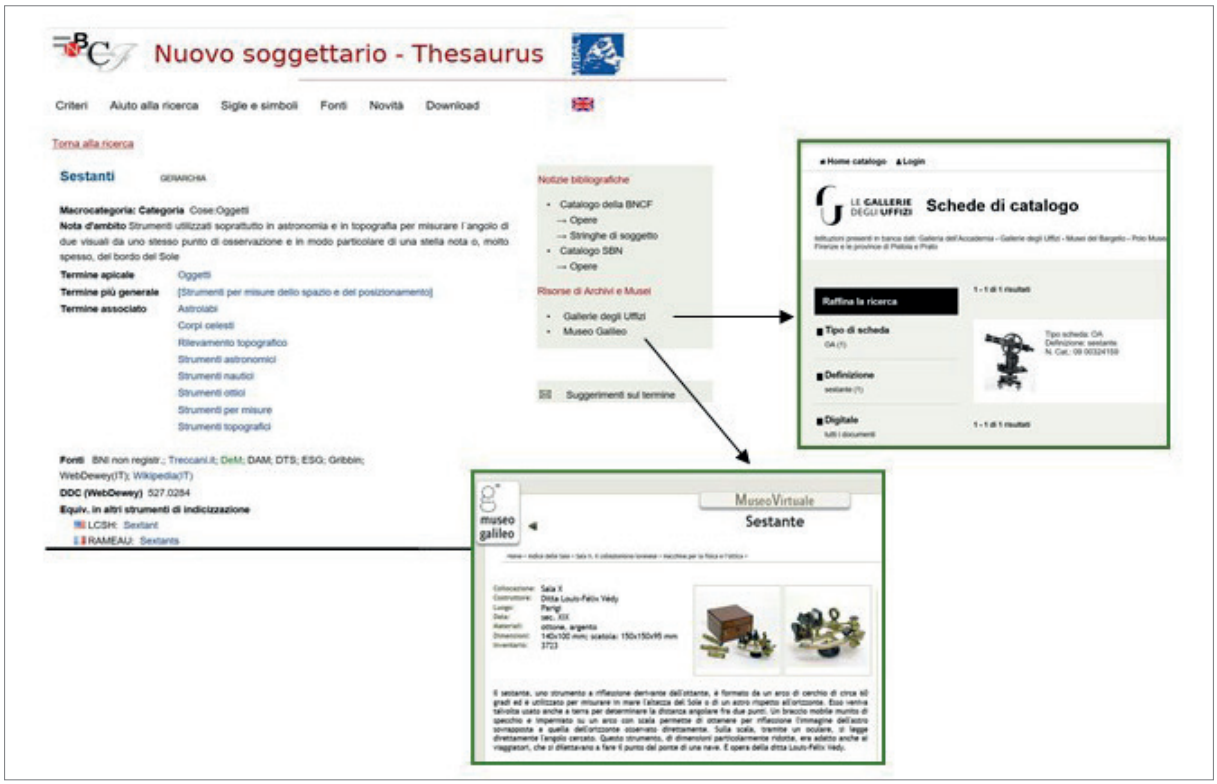

Fig. 13. The term *Sestanti* as seen in A. the *Nuovo soggettario*, B. the Gallerie degli Uffizi's Schede di catalogo, and C. the Museo Galileo's Museo virtuale

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

An example of links with archives can be seen in Figure 14, where *Federalisti europei* in the *Nuovo soggettario* is linked with the Ernesto Rossi fund of the Historical Archives of the European Union.[39]

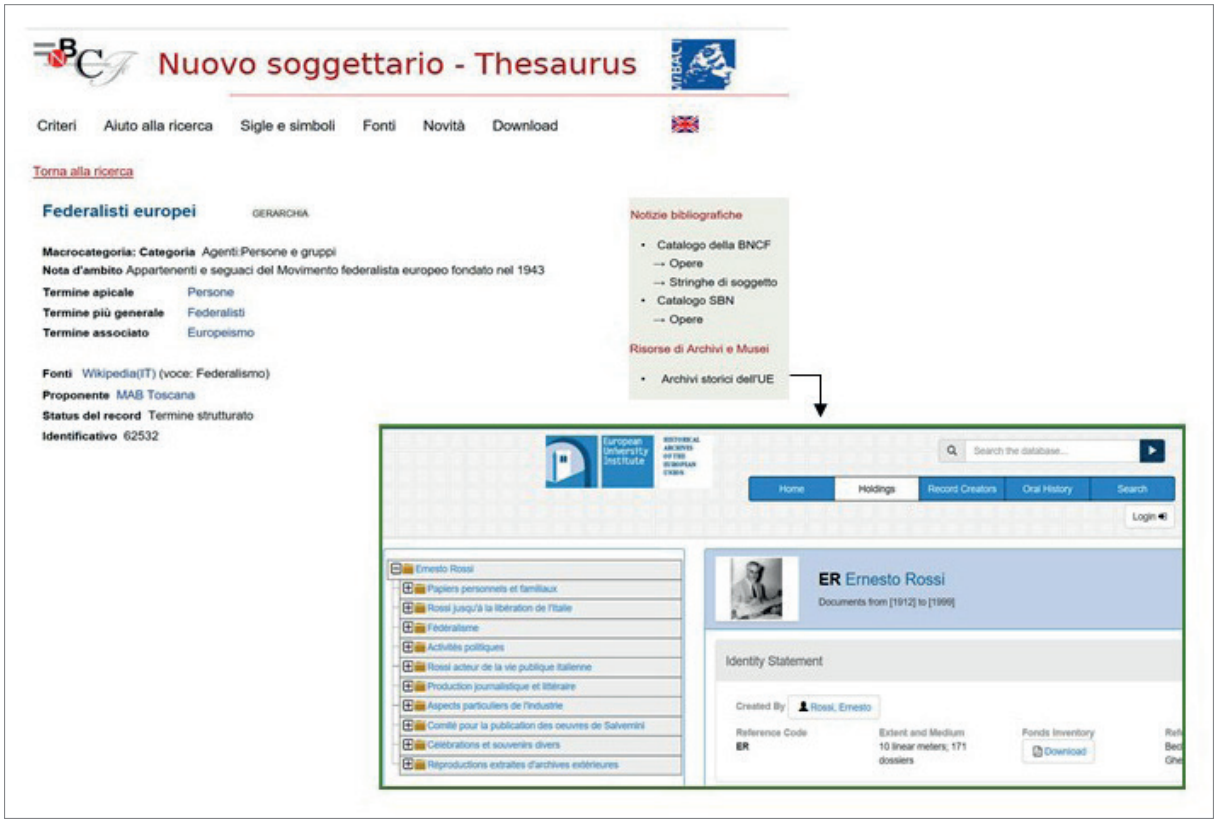

Fig. 14. Links between *Nuovo soggettario* and the Historical Archives of the European Union

These examples of the GLAM (Galleries, Libraries, Archives and Museums) perspective are also promoted by the Wikipedia universe,[40] and in Italy by the MAB (Musei Archivi Biblioteche) projects in which BNCF has participated while joining various research efforts.[41] Likewise, it is hoped there will be future possible connections, for example, with the controlled vocabularies of the Sistema Archivistico Nazionale (SAN)[42] or collaborations with institutions dealing with the standardization of the terminology employed for the cataloguing of the cultural heritage, such as the Istituto Centrale per il Catalogo e la Documentazione (ICCD) (Birrozzi et al. 2020).

[39] https://thes.bncf.firenze.sbn.it/termine.php?id=62532; https://archives.eui.eu/en/fonds/115005?item=ER. About the experimentation on subject indexing of Ernesto Rossi Fund: Becherucci et al. 2019, 24-48.

[40] For example, see the recent Gruppo Wikidata per Musei, Archivi e Biblioteche, https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche.

[41] https://www.aib.it/attivita/mab-italia/.

[42] http://san.beniculturali.it/web/san/home;jsessionid=66BD6878BF6E20807ACABB005C45C7CE.sanapp01_portal.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

The perspectives of machine learning, artificial intelligence, automated subject indexing

In which directions will the future of the *Nuovo soggettario* go? Its challenges are not that different from those of other thesauri.

Within the current context, which has much changed due to the predominant role of the Internet, subject indexing is interacting with the semantic capabilities of search engines, such as Google, with the development of both artificial intelligence and machine learning and, of course, with the dissemination of the ever increasing number of digital resources.

At the same time, we know that it is wrong to assume that sources transmitting information be only those 'hooked' by Google, just as it is wrong to confuse the functions of our catalogues with those of other tools for access to information.

However, as often pointed out by the indexing experts, such as Stella Dextre, one must also be aware that libraries and other institutions, dealing with information retrieval, have much fewer resources to be earmarked for 'manual' subject indexing, that is 'intellectual' indexing, as compared to Google's algorithms.

Despite the presence of search engines and their powerful automatic and semi-automatic indexing, the role of thesauri does not seem to be outdated.

For instance, Birger Hjørland, professor at the Royal School of Library and Information Science of Copenhagen, has very recently questioned about the reasons why the search engines, despite they apply principles of semantic type, do not make knowledge organization (KO) and mapping of the relationships among concepts superfluous at all (Hjørland 2021).

'Human' taxonomists working for Google, support the well-known Google Knowledge Graph, which is connected with DBpedia, Wikidata and with the linked data. This is a project about which very many reservations have been expressed.[43]

The procedures for automated and semi-automated translation/indexing are dealt with within IFLA[44] but also within countless other frameworks; to give some examples, in Italy these procedures are studied at the Istituto di linguistica computazionale di Pisa, at the Universities of Padua and Udine. In 2011, BNCF took its first steps by starting a project for the semi-automated indexing of digital doctoral theses. At that time, we used MAUI and other open source software. Should we have the resources and the possibility to restart this project, we could build on the important experiences of other national libraries, such as the Deutsche Nationalbibliothek or utilize tools like those implemented by National Library of Finland.

Studies will continue on machine learning, knowledge graphs like Google's, *corpora* of terms, and the benefits that thesauri can bring to our users, because not only the artificial intelligence world will benefit from such insights, but also libraries and the national bibliograpies world in their mission for the dissemination of knowledge.

In closing, here are some Keywords for the future of thesauri and for their challenges: creativity, versatility, sharing.

A special thank goes to Barbara Tillett who sent me many comments and suggestions.

[43] https://en.wikipedia.org/wiki/Google_Knowledge_Graph.

[44] Automated subject analysis and access Working Group, https://www.ifla.org/node/92551.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

References

(Last consultation of the websites: 15 July 2021).

Ballestra, Laura. 2011. "Information literacy education in Italian libraries: evidence from an Italian University." Bibliothek Forschung und Praxis 35, no. 3 (December):395-401.

Becherucci, Andrea, Silvia Bruni, Benedetta Calonaci, Emilio Capannelli, Walter Fochesato, Anna Lucarelli, and Sonia Puccetti. 2019. "Libri per gli internati militari italiani durante la Seconda guerra mondiale: un inedito di Ernesto Rossi." Biblioteche oggi 37, (May):4-48. DOI: http://dx.doi.org/10.3302/0392-8586-201904-024-1.

Bergamin, Giovanni. 2020. "Postfazione." In Guerrini, Mauro. 2020. Dalla catalogazione alla metadazione. Tracce di un percorso, 167-168. Roma: Associazione italiana biblioteche.

Biagetti, Maria Teresa. 2018. "A comparative analysis and evaluation of bibliographic ontologies." In Challenges and opportunities for knowledge organization in the digital age. Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal, edited by Fernanda Ribeiro, Maria Elisa Cerveira, 501-510. Baden Baden: Ergon.

Biagetti, Maria Teresa. 2020. "Ontologies (as knowledge organization systems)." In ISKO Encyclopedia of Knowledge Organization, edited by Birger Hjørland and Claudio Gnoli. https://www.isko.org/cyclo/ontologies.

Birrozzi, Carlo, Barbara Barbaro, Maria Letizia Mancinelli, Antonella Negri, Elena Plances, and Chiara Veninata. 2020. "Catalogare nel 2020. La digitalizzazione del patrimonio culturale." Aedon. Rivista di arti e diritto on line no. 3. http://www.aedon.mulino.it/archivio/2020/3/birrozzi.htm.

Broughton, Vanda. 2020. "General principles underlying knowledge organization systems (KOS)." In *KO-ED Introduction to Knowledge Organization*. https://www.iskouk.org/event-4025408

Buizza, Pino. 2020. "Thesaurus and heading lists: equivalence and asymmetry." In Knowledge Organization at the Interface. Proceedings of the Sixteenth International ISKO Conference, 2020 Aalborg, Denmark, herausgegeben von International Society for Knowledge Organization (ISKO), prof. Marianne Lykke, prof. Tanja Svarre, prof. Mette Skov, Daniel Martinez Avila, [59]-68. Baden-Baden: Ergon.

Cheti, Alberto, Anna Lucarelli, and Federica Paradisi. 2009. "Subject indexing in Italy: recent advances and future perspectives.". https://www.ifla.org/past-wlic/2009/200-lucarelli-en.pdf.

Clarke, Stella Dextre. 2020 "How should today's thesaurus earn its keep?." In *KO-ED Introduction to Knowledge Organization*. https://www.iskouk.org/event-4048801

Clarke, Stella Dextre. 2020 "What is a thesaurus? How and Why so?." In *KO-ED Introduction to Knowledge Organization*. https://www.iskouk.org/event-4048800

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 (January/April):87-101. DOI: https://doi.org/10.2426/aibstudi-11408.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Folino, Antonietta and Francesca Parisi. 2020. "Rappresentatività e copertura semantica dei KOS." AIDAinformazioni 38, no. 3/4:93-112.

Francioni, Elisabetta and Anna Lucarelli. 2020. "Nuovi concetti, nuovi termini ai tempi del Coronavirus." Bibelot: notizie dalle biblioteche toscane 26, no. 1 (January/April). https://riviste.aib.it/index.php/bibelot/article/view/12038.

Gnoli, Claudio. 2020. Introduction to Knowledge Organization. London: Facet Publishing.

Guerrini, Mauro. 2020. Dalla catalogazione alla metadatazione. Tracce di un percorso; prefazione di Barbara B. Tillet; postfazione di Giovanni Bergamin. Roma: Associazione italiana biblioteche.

Hjørland, Birger. 2021. "Search engines and Knowledge Organization (or why we still need Knowledge Organization).". KO-ED Theoretical Perspectives. https://www.iskouk.org/event-4058726.

L'indexation matière en transition: de la réforme de Rameau à l'indexation automatique, sous la direction d'Etienne Cavalié. 2020. https://www.bnf.fr/sites/default/files/2020-03/biblio%20indexation%20matiere%2011mars20.pdf.

Junger, Ulrike. 2018. "Automation first – the subject cataloguing policy of the Deutsche Nationalbibliothek.". http://library.ifla.org/2213/1/115-junger-en.pdf.

Lucarelli, Anna. 2014. "«Wikipedia loves libraries»: in Italia è un amore corrisposto..." AIB studi 54, no. 2/3 (May/December). DOI: https://doi.org/10.2426/aibstudi-10108.

Magrini, Sabina. 2021. "«Ti racconto in italiano»: management, description and indexing of oral sources. A project by the ICBSA (Istituto Centrale per I Beni Sonori e Audiovisivi)." In Conference BC 2021. Video. https://www.youtube.com/embed/Yo6Vi72E1T4?start=10942&end=12772.

Martínez-González, M. Mercedes, and María Luisa Alvite Díez. 2019. "Thesauri and semantic web: discussion of the evolution of thesauri toward their integration with the semantic web." IEEE Access, 7. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8873649.

Mödden, Elisabeth. 2021. "Artificial intelligence, machine learning and DDC Short Numbers." In Conference BC 2021. Video. https://www.youtube.com/embed/Yo6Vi72E1T4?start=124&end=1523.

Petrucciani, Alberto. 2019. "C'è un futuro per l'indicizzazione?." In Viaggi a bordo di una parola. Scritti sull'indicizzazione semantica in onore di Alberto Cheti, a cura di Anna Lucarelli, Alberto Petrucciani, Elisabetta Viti; presentazione di Rosa Maiello, 163-173. Roma: Associazione italiana biblioteche.

Pocci, Adele. 2020. "Bibliotheca perspectivae: una sperimentazione del Nuovo soggettario nell'ambito specialistico dell'iconografia scientifica." Bibelot: notizie dalle biblioteche toscane 26, no. 3 (September/December). https://riviste.aib.it/index.php/bibelot/article/view/12798.

Riva, Pat. 2021. "The multilingual challenge in bibliographic description and access." In Conference BC 2021. Video. https://www.youtube.com/embed/Yo6Vi72E1T4?start=12826&end=14355.

Smith-Yoshimura, Karen. 2020. Transitioning to the Next Generation of Metadata. Dublin, OH: OCLC Research. https://doi.org/10.25333/rqgd-b343.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Suominen, Osma. 2021. "Annif and Finto AI: developing and implementing automated subject indexing." In Conference BC 2021. Video. https://www.youtube.com/embed/Yo6Vi-72E1T4?start=1892&end=2953.

Viti, Elisabetta. 2017. "My First Ten Years: Nuovo soggettario growing, development and integration with other Knowledge Organization Systems." Knowledge Organization 44, 8:624-637.

Will, Leonard. 2020. "From concepts to knowledge organization systems." In *KO-ED Introduction to Knowledge Organization*. https://www.iskouk.org/event-4043820

Zeng, Marcia Lei. 2019. "Interoperability." Knowledge Organization 46, 2:122-146. https://doi.org/10.5771/0943-7444-2019-2-122.