SCIRES it
SCIentific RESearch and Information Technology
Ricerca Scientifica e Tecnologie dell'Informazione
Vol 13, Issue 2 (2023), 171-189
e-ISSN 2239-4303, DOI 10.2423/i22394303v13n2p171
Open access article licensed under CC-BY-NC-ND
CASPUR-CIBER Publishing, http://www.sciresit.it

PROPOSAL FOR THE STANDARDIZATION OF CONTROLLED VOCABULARIES FOR TELEVISION ARCHIVES: CASE STUDY AT RTVE

Virginia Bazán-Gil, Juan-Antonio Pastor-Sánchez***

*Universidad Carlos III de Madrid. RTVE, Spain
**Universidad de Murcia, Spain

Abstract

Controlled vocabularies play a crucial role in indexing and retrieving content in audiovisual archives. The integration of SKOS and ontologies can enhance search processes and metadata generation. This work demonstrates how to integrate SKOS within the ARCA system, used by RTVE for audiovisual management. The proposal focuses on adapting the relational schema structure of ARCA to unify the thesauri using the SKOS model. The process involves identifying concepts, labels, semantic relationships, and collections to create a single controlled vocabulary from the different thesauri, represented through a relational database schema. The results of the unified thesauri and the mapping of vocabulary concepts to Wikidata items are shown to reinforce integration in the realm of Linked Data.

Keywords

SKOS, TV audiovisual archives, thesauri integration, controlled vocabularies

1. *Introduction*

Controlled vocabularies or free tags are frequently used for indexing and retrieval in audiovisual archives. They have traditionally been considered an essential tool to highlight the truly significant content elements of a document, avoiding inaccurate searches in textual fields (López de Quintana, 2010).

The European Broadcasting Union (EBU) has developed various technical specifications to enable the exchange of news and archive material between its members. These specifications are based to a greater or lesser extent on use of shared controlled vocabularies to facilitate interoperability. Such vocabularies can be technical (e.g. video formats) or descriptive (e.g. television genres). The EBU – Tech 3336: EBU Reference Data & Classification Schemes (EBU, 2011) document defines the metadata schema used by EBU to publish reference data in the form of XML classification schemes, although its reference data can also be represented in alternative formats, such as Simple Knowledge Organization System (SKOS) vocabularies. The document states that the EBU has identified sets of reference data that can be replaced by similar datasets or customized and can be mapped with other classification systems. It also highlights a set of good practices for defining and maintaining vocabularies. These reference vocabularies are available for download as Linked Open Data (EBU, 2020)

In the more general context of the media, the International Press Telecommunications Council (IPTC) has developed several classification systems for identifying news content (Quinn & Parrucci, 2021). The Subject Reference System (SRS), created in 1995 in collaboration with the Newspaper Association of America (NAA), was divided into 17 subject categories, each represented by an alphanumeric code available in a number of different languages, allowing the content of a news item to be described in three levels of detail (IPTC 2022a). However, the lack of flexibility in the hierarchical structure of the Subject Reference System led IPTC to develop Media Topics, a controlled vocabulary of approximately 1,200 terms available in 12 languages (IPTC 2022b). IPTC also developed other technical (audio compression, colour, etc.) and descriptive (news provider, roles, etc.) vocabularies, all of which were represented as SKOS vocabularies in RDF/XML format (IPTC 2022c).

SKOS controlled vocabularies, available as Linked Data, are a great opportunity to establish relationships between audiovisual archives. A clear example of this is the proposal by de Boer, Priem, Hildebrand et al. (2016) to explore the connections between the archive collections of the Flemish Institute for Archiving (VIAA, Vlaams Instituut voor Audiovisuele Archivering) – now meemoo (2021) – and the Netherlands Institute for Sound and Vision (NISV). The project involved converting the VIAA thesaurus (in XML format) into SKOS, aligning the thesauri of the two institutions using the CultuurLINK tool (2021) and developing an application that allowed cross-searching of the two collections, exploiting the connections between them to the full. At the BBC, the focus is on providing a single reference for relevant entities (people, organizations, events and places) of interest to its audiences through BBC Things (2021). Following the principles of Linked Open Data, the concepts contained in BBC Things are related to other, more complete and authoritative open data sources such as Wikidata and DBpedia.

Manders and Wigham (2021) have stressed how speech to text (S2T) and facial recognition technologies have contributed to the automatic generation of metadata. They also highlight the potential of integrating thesauri, in this case onomastic thesauri, with external data sources such as Wikidata through Linked Open Data systems to facilitate the use of audiovisual archives in innovative ways, thereby also improving the search processes for users. To this end, it is necessary to adopt semantic web standards and use tools for more efficient management of thesauri, as explained by Bus and Huis in't Veld (2021).

In 2003, and in the context of Spain's television archives, del Valle and the Telecinco archive team proposed a system for the channel's news and program documentation centers. The goal was to provide Telecinco's Documentation area with an instrument that would be capable of representing the concepts and ideas that could appear on television. This involved working with a very broad thematic and contextual area with wide semantic dispersion. To this end, they defined a structured system based on two axes: a vertical axis that organized the thesaurus around macro-categories referencing the IPTC ontology and a horizontal axis formed by the thematic areas covered by television. Harnessing that framework, they distributed the 2,000 descriptors in use up until that point in the areas of News and Programs, while also detecting different connotations that were clarified through scope notes. After an evaluation process, the thesaurus (consisting of 3,000 descriptors and 500 non-descriptors) was integrated into the documentation management system for use (Valle & Jiménez, 2002; Valle, 2003).

There are also other, more theoretical approaches such as the contributions made by Caldera and Sánchez (2008; 2009), who proposed using ontologies for the control and retrieval of onomastic information (natural and legal persons seen on screen or referred to) in entertainment programs. They suggested that ontologies could be used to represent the complex social and human relationships of television personalities. The development of ontologies is not disassociated from necessary terminological control, but, indeed, also enhances retrieval by defining classes and relationships associated with a particular person and improves precision in retrieval when combined with searches in textual fields.

Eugenio López de Quintana (2010) acknowledges the dual descriptive approach (the description of what is visualized and of what is referenced) that is applied to television content analysis and proposes an alternative to manual indexing: filtering in text searches by automatically applying morpho-linguistic rules, automatic indexing of texts, and creating and maintaining ontologies that both allow for more versatile relationships than traditional thesauri and can be applied in filtering and recommendation systems for searches. According to López de Quintana, this should be a central task in the archive management process, even though the entity extraction process would be performed dynamically during the search, optimizing the relationships previously established in the ontology. It is an innovative approach that involves abandoning manual indexing but which highlights the importance of ontologies in television archives.

Finally, it is interesting to note that manual indexing in the field under discussion in this paper is not sustainable in the long term due to the constant increase in content ingested and the need to provide ever greater access to it, not only within organizations but also externally (de Boer, Ordelman & Schuurman, 2016).

Furthermore, content searching and linking are moving from a global perspective to a segment perspective where labelling each of the elements that make up the media content is considered important. Against this backdrop, Artificial Intelligence-based analysis solutions that allow entity recognition or keyword recognition from subtitles or audio-to-text transcription are gaining more and more ground. This is evidenced by the initiatives and projects that a number of different television networks are carrying out around the world.

The aim of this paper is to propose a methodology for the unification and representation in SKOS of the thesauri used by the archive of Spain's public television and radio corporation, Radio Televisión Española (RTVE). In pursuit of this goal, we provide an international overview of the use of controlled vocabularies in media archives, including those managed by radio and television companies, institutions responsible for preserving audiovisual heritage and initiatives carried out by international organizations such as the European Broadcasting Union (EBU) and the International Press Telecommunications Council (IPTC). We have conducted a comprehensive analysis of sources, primarily case studies presented at international conferences such as those of the International Federation of Television Archives (IFTA) and the committees of the EBU. These sources are only available to professionals in the sector since, in cases such as the EBU, seminars and documentation are only accessible to members of these organizations. This vision has been complemented by papers published in specialized journals and by a survey, specially designed for this research, which was distributed internationally through professional networks.

We have also conducted an analysis and description of the RTVE thesauri and their use. This has allowed us to identify opportunities and weaknesses and, in consequence, to propose the conversion, standardization and unification process presented in this article. Finally, we have analysed the main conclusions obtained from this study.

2. *International overview of the use of controlled vocabularies in media archives*

To further understand the use of controlled vocabularies in television archives, a brief survey was designed consisting of 10 questions. The questionnaire was published in English and Spanish on LinkedIn (Bazán-Gil, 2021a, 2021b) and distributed via Twitter on 12 July 2021, remaining available until 25 September 2021. In addition, TV archive managers in Europe, Asia and Latin America were invited to participate through personalized email invitations.

Tab. 1: Questionnaire on the use of controlled vocabularies in television archives. Source: own elaboration.

Question	Possible answers		%
1. Does the archive where you work use controlled vocabularies or thesaurus to tag content?	a.	Yes, we use controlled vocabularies.	83
	b.	No, but we are planning on using controlled vocabulary / thesaurus	3
	c.	No, we use free tags	15
2. Which kind of controlled vocabularies/thesaurus does your archive use?	a.	Persons and organizations	73
	b.	Locations	85
	c.	Topics	88
	d.	Series	10
	e.	TV genres	43
	f.	Others (please, indicate)	20
3. Who creates new items in these vocabularies?	a.	Most of them come from the production systems and are created/added by archivists.	25
	b.	Most of them come from the production systems and are created/added by producers / other staff.	10
	c.	Most of them are created/added at the archive by archivists	73
4. Are these vocabularies accurate and actively maintained?	a.	Yes, we have specific staff for this purpose.	49
	b.	Not as accurate as we would like.	33
	c.	Not accurate at all.	10
	a.	Yes	65

5. Are your archive collections or part of them available on your institution web site?	b. No	35
6. If your archive collections or part of them are available on your institution web site: Is metadata from the archive (such as summary, titles, series, etc.) reused to make contents more visible?	a. Yes, we reuse metadata from the archive system on the website	45
	b. No, new metadata is created specifically for these contents on the website	39
	c. No, archive content on the website has no metadata	18
7. If your archive collections or part of them are available on your institution web site, are these contents tagged?	a. Yes, with the same controlled vocabularies/thesaurus that are used in the archive.	21
	b. Yes, but with free tags provided specifically for these contents on the website.	39
	c. No, not at all.	21
8. Is your archive using or considering using a standard to represent vocabularies / thesaurus (e.g. SKOS, ISO or BS)?	a. Yes, we are using a standard.	20
	b. No, we are not using a standard but we are considering the idea.	33
	c. No, we are not using or considering the use of a standard.	45
9. In case you are using a standard to represent vocabularies / thesaurus, which one are you using?	Free text: SKOS (27%), the ISO standard (36%) and others which were not specified (9%).	
10. Would you like to add any comments?	Free text	
Organization	Free text	
Country	Free text	
Would you be available for further research?	Yes No	

A total of 40 responses were received from 20 different countries. Of these responses, 57% came from Europe, 30% from Latin America, 5% from Africa, 5% from Asia and 3% from North America.

The questions were structured around two main topics: internal use of controlled vocabularies in the archive and the use of controlled vocabularies on the archive's website. According to the results obtained, 83% of the archives surveyed use controlled vocabularies, compared to 15% that use free keywords. Only 3% of the archives indicated that they plan to use thesauri in the future although they do not currently use them. The most commonly used thesauri cover topics (88%), places (85%), people and organizations (73%), TV genres (43%), series (10%) and others (20%), with the functions indicated in the credits (director, producer, writer, etc.) being the most notable included in the latter thesauri.

In general, archivists are responsible for creating these terms in the archive system (73%). Less frequently, the terms are created by the archivists in the production systems (25%) or assigned by other professionals such as producers or directors (10%), although a combination of all of these may occur within the same organization.

At national preservation organizations such as the Netherlands Institute for Sound and Vision (Netherlands), INA (France) and Meemoo (Belgium), much of the metadata comes from the television channels which produce the content. These vocabularies are highly accurate and up-to-date. In 49% of cases, there are people specifically dedicated to managing thesauri. Additionally, 33% state that they are not as up-to-date as they would like, compared to 10% who claim that they are not actively managed. Regarding the use of standards to represent information (ISO, SKOS, etc.), 45% report not using or considering their use, 33% are considering their use in the future and 20% are already using them. The most used standards are SKOS (27%), the ISO standard (36%) and others which were not specified (9%).

The second part of the survey focused on the archive on the television network's website.

Among the respondents, 65% of the archives that participated in the survey have part of their collections accessible on their website, as opposed to 35% which do not. When the archive is available on the website, the most common situation (45%) is that the metadata which describes the content (title, summary, etc.) comes from the archive itself. In 39% of the cases, the metadata is generated especially for the website and just 18% of the time the content has no descriptive metadata of any type. More specifically, 39% use specific tags for the website, 36% do not tag the content posted on the website and 21% utilize the same controlled vocabularies or thesauri as those used by the archive.

3. Thesauri at RTVE

The use of thematic, onomastic and geographical indexes to standardize content indexing in RTVE's archive dates back to the creation of the Documentation Department for News in.

The development and maintenance of documentary languages "suitable for the analysis and information retrieval work necessary in the Documentation Center" (Hidalgo, 2017) became one of the functions assumed by the RTVE Documentation Center, created in 1981 to manage the content produced and/or broadcast by the Non-News Programs Area.

In 1983 the News Archive management was automated, and in 1986 the Mistral Archive Management System was implemented at TVE. This tool integrated two databases to meet the needs of two clearly differentiated production areas, Programs and News, each with its own structure, thesauri and analysis rules (Hidalgo, 2017).

In 1993, Mistral was replaced by SIRTEX, although the separate structures of the different databases, as well as the different thesauri which had been developed, were maintained. In 2005, after the digitization of the RNE (Radio Nacional de España, the radio area of RTVE) archive and the digitization project of the TVE archive that began in 2010 (de Prada, 2021), it became evident that this tool needed to be replaced by another that could manage not only the data but also the multimedia content, a Media Asset Management (MAM) system. This is when efforts were resumed to design a common data model that would facilitate the management and retrieval of the different collections that make up the RTVE archive: more than 40 documentary databases and 30 thesauri. In September 2006, after extensive work, the new data model was published, and in November of the same year, the criteria for the migration were established.

In 2011, the proposed adoption of a single data model for managing the contents of the RTVE archive became a reality. However, the number of thesauri in use, their diversity, the volume of terms they contained and the enormous differences between vocabularies with identical topics but different origins made it advisable to postpone unification for the geographical, onomastic and chronological thesauri and, as far as possible, for the thematic thesauri.

In 2021, ARCA was updated to a new, more efficient version, but the vocabulary management model and the existing thesauri remain intact, which means working with different thesauri for video, audio, text and photo databases which have been developed to differing degrees. These vocabularies are growing in parallel with the growth of documentary databases. The terms and their relationships have been managed based on the needs of each database through direct action by the documentalists and by the people responsible for the management and maintenance of these vocabularies, and more recently through data entry from production and broadcast systems, business applications with which ARCA is fully integrated.

The main thesauri applied in various contexts and information objects are shown in Table 2. These vocabularies refer to terms associated with places, persons (people and organizations), series and topics. ARCA also encompasses vocabularies of dates and forms, but these are not relevant to the process of representation and unification since they consist of simple lists of strings representing individual years or date intervals, or partial selections of complete thesauri.

The onomastic thesauri (or thesauri of persons) gather the names of individuals and organizations. These vocabularies are used to identify voices or people featured in the audio, video, photo and text databases, and allow for the indexing of production fields, the technical and artistic staff of a program, people and shots of people in video databases, and voices in audio databases, to name a few examples. These thesauri have the highest number of terms (an average of 148,161 terms) and a high growth rate that can be around 700 terms per month depending on the database. Overall, they present the highest degree of ambiguity with terms that, despite being the same, do not represent the same entity or with terms that are different but refer to the same entity.

The geographical thesauri (or thesauri of places) include entities of political geography, such as nations, cities, towns, etc., and physical geography, such as mountains, peaks, oceans, rivers, etc. They also include terms related to outer space or locations of historical civilizations that no longer exist. These terms are used to index different fields in the audio, video, text and photography databases. In the video databases, for example, these thesauri feed the controlled fields of production location, news location, geographic scope referenced in a documentary or reusable images of a specific location. Currently, ARCA manages eight geographical thesauri with an average of 13,114 terms and steady monthly growth.

The thematic thesauri (or thesauri of topics) bring together descriptors that make it possible to represent the topics addressed in a program as well as the images used to describe these topics. The average number of terms per thesaurus was 45,113. The degree of development by thematic area is uneven and depends on the type of program.

The series thesauri bring together the titles of the TV and radio series produced by RTVE or third-party productions that have been broadcast on the different TV channels or radio stations of the corporation. Unlike the previously mentioned thesauri, a series thesaurus is a multilingual thesaurus that contains the idiomatic variations of a title as well as the variants of the same title in different business applications since sometimes series can change their opening credits from the moment they are conceived as a project until they are finally broadcast. The average number of terms per thesaurus is 3,926, and they experience constant monthly growth, caused by the production and broadcasting of new programs. These thesauri feed the fields of series title, original series title and provisional series title.

Thesaurus terms have clearly differentiated origins: human cataloguing in the production system and human cataloguing in ARCA. In both cases, the candidate descriptors generated must be validated by the people in charge of thesauri management at the different cataloguing departments. These people are also responsible for establishing relationships between terms and creating scope notes when necessary. Another way of creating terms is through the massive migration of documents from other document databases. These processes only occur on an exceptional basis but generate many descriptors that need to be checked and validated.

The ARCA system is integrated with other business applications such as the Integrated Production System and the Recording and Broadcasting Management System. In these systems connected with production and broadcasting, metadata is also generated which feeds into the lexicons and thesauri in ARCA and requires constant validation processes performed by the cataloguing units.

As we mentioned earlier, vocabulary management has been uneven across the different analysis units for radio (Radio Nacional de España, RNE) and television (Televisión Española, TVE) at RTVE.

In the News department (Muñoz-de-la-Peña-Costero, Meana-Alonso & Sáez-Carreras, 2014), there are two units responsible for content cataloguing. The User and Exchange Unit is responsible for analysing original recordings of news programs, while the Cataloguing Unit is responsible for institutional broadcasts, broadcasts from territorial centers and sports events, as well as agency content. Both units work with content from the moment it is ingested into the Production System and ensure that it is retrievable both in the Production Temporary Archive (AMI) and in ARCA (where only content that will be permanently preserved is archived). The Cataloguing Unit also carries out the management and updating of controlled vocabularies.

In the Programs archive, the work is organized in a similar fashion. The Production Archive is responsible for analysing the daily magazine programs and their original recordings, as well as other recordings and agency content. In the Analysis Unit, which deals with non-daily programs and is responsible for indexing all the contents of the program database in ARCA, documentalists are responsible for creating new terms. However, management of the thesauri is centralized in a small group of people who also review candidate terms received by ARCA through other systems such as SIP.

Finally, in the archive based in Sant Cugat (Barcelona), at the RTVE Production Center in

Catalonia, pre-cataloguing is carried out in the production system, which includes metadata related to the program series and production, as well as other technical metadata. When this content is transferred to the archive, its analysis is completed with indexing. In the case of video databases, there are people dedicated to active management of the vocabularies. However, not all documentalists in all units of analysis can generate new terms directly for all thesauri; some thesauri, such as thematic ones, are closed, and terms are created in a consensual manner after a proposal.

As for RNE, content analysis is only carried out in ARCA, without any integration with other applications that could allow reuse of metadata from production or broadcasting. All cataloguing users are authorized to create terms from the data entry screen and also have permission to administer the controlled vocabularies, so they can modify and delete terms or create, modify or delete relationships or scope notes. In radio databases (word and music), it is common to have massive data loading processes, and therefore checks are carried out on the thesauri on a systematic basis, especially on the onomastic thesauri.

In 2017, a proof of concept (PoC) was carried out with the company VSN, and the results were presented at the FIAT/IFTA World Conference held in Mexico City (Bazán-Gil & Escribano, 2017). The test once again highlighted the complexity of the process and the need for appropriate resources to tackle it. In 2020, efforts to design a project focusing on unification of the thesauri were resumed, this time under the leadership of the Innovation Subdirectorate. The aim was to develop a project capable of improving the management and retrieval processes for content in the archive by introducing innovative elements that could be extrapolated to other areas of the corporation, such as the digital area. From the point of view of the archive, not only does the unification of thesauri allow for more efficient resource management, by avoiding the repetition of tasks in different units, but also the use of a single thesaurus and its integration with external data sources would represent a significant breakthrough in the way non-professional users access archive content. Additionally, the use of a standard such as SKOS lays the foundation for both a content recommendation system and the generation of ontologies, which in themselves constitute a valuable data source as has been demonstrated by Datos BNE (Biblioteca Nacional de España, 2021) and the Spanish Government's Open Data Initiative.

Tab. 2: RTVE thesauri of places, people, series and topics with their typology and datasets of application. Source: own elaboration.

Type	Scope	Domains
Video	Programs and Sports (Barcelona) News (Madrid) News (Barcelona) Programs (Madrid)	PL PE SE TO
Audio	Cell 2	PL PE SE TO
Text	Agencies Newspapers	PL PE SE TO
	Library	PL PE TO
Photos	Pictures	PL PE SE TO
	Frazen	PL PE TO

Thesaurus Domains Legends:
PL:Places, **PE**:People/Organisations, **SE**: Series, **TO**:Topics

4. *Adapting ARCA to integrate the SKOS Model*

The SKOS Data Model, defined as an OWL Full ontology, provides a very flexible way to define controlled vocabularies. Due to the nature of any controlled vocabulary, the use of SKOS places greater emphasis on organization and reuse aspects than on its application in processes with a high degree of logical formalization (W3C, 2009).

SKOS is designed for contexts in which intensive use is made of RDF. However, it is also possible to transpose its data model in the context of a relational database. Considering that ARCA uses a relational database management system and considering the scope of this work, this is without a doubt the most viable option.

The representation of SKOS using the relational model facilitates its adaptation and makes it possible to expand the structure of ARCA to represent the thesauri in a more interoperable way. Consequently, not only would it be possible to apply the functionalities of SKOS in ARCA, but the data stored according to this relational structure could easily be transformed into an RDF dataset for subsequent open publication.

SKOS is based on a conceptual organization of vocabulary. In this way, each object, idea, person, place or topic is identified with a concept. Concepts can be labelled in different languages. There are three types of labels: preferred labels, alternative labels and hidden labels. A concept can only have one preferred label in each language. In this sense, preferred labels are equivalent to traditional descriptors in thesauri. Alternative labels are used for defining synonyms (equivalent to non-descriptors), and each language can have a variable number of them associated with it. Hidden labels are used to retrieve concepts by certain search terms that editors prefer not to be visible to users when browsing concept elements.

SKOS allows concepts to be related to each other through both hierarchical and associative semantic relationships. Hierarchical relationships allow for the definition of semantically broader or narrower concepts for a given concept. Associative relationships allow for the linking of semantically related concepts outside a hierarchical relationship. Concepts that do not have any broader concepts are defined as top concepts. In cases where it is necessary to expand the typology of semantic relationships, it is possible to define new properties based on the existing ones already natively available in SKOS.

The SKOS model makes it possible to index information resources (such as audiovisual objects) by defining links with concepts. This avoids an approach where indexing is done by terms, which would mean that any change in the lexical structure of the vocabulary would require a forced alteration of previously created indexing records.

Each controlled vocabulary is identified with a concept scheme. The different concepts that make up a vocabulary are associated with that scheme. An institution like RTVE requires the definition of multiple concept schemes depending on the existing thesauri. It is also possible to define and group together sets of semantically close concepts or those that are commonly used together. These groupings are called concept collections and are useful for enriching indexing and search processes beyond existing semantic relationships between concepts.

One of the most interesting features of SKOS, from the point of view of interoperability, is the ability to define equivalences between concepts from different schemes. In this way, different vocabularies can be mapped by establishing correspondences between their elements. This enables the integration of different thesauri and the reuse of external thesauri. It is also especially useful in the processes of transforming corporate vocabulary structures. The unification or transition of controlled vocabularies within a corporation can be done gradually using the mapping possibilities between old vocabularies and those designed to replace them.

5. Representation of the SKOS model using a relational schema

ARCA is currently based on a relational scheme to manage the different thesauri (see Figure 1). Each thesaurus is defined on the basis of certain descriptive and configuration attributes such as title, author or language (among others). The central elements of the different thesauri are the terms that can be related to each other through different types of relationships (hierarchical, associative, synonymy).

It is also possible to define topics that allow terms to be grouped according to a typology adapted to the needs of each thesaurus. The schema also contemplates the definition of a typology of notes that are linked to a term. Finally, the system includes a change control feature, which stores the user identifier and the date on which a term, topic, relationship or note was created or modified.

The main limitation of the relational schema described above lies in its terminological approach rather than its conceptual approach. This implies a certain lack of flexibility in defining relationships and in the indexing process. It can also be observed that the relational schema does not allow storing the language information of the terms. Therefore, it is impossible to know the language of each term, which makes linguistic interoperability unfeasible. On the other hand, the change control system makes it possible to store the creation data and most recent modification data (user and date) of the terms, notes, relationships and topics. However, it is not possible to carry out an audit process that covers the entire thesaurus life cycle, as it only stores the last modification record.

The application of the SKOS model would allow greater flexibility in both the thesaurus structure and editing tasks. This line of work would allow for compliance with the ISO 25964 standard (ISO, 2011, 2013), with the consequent advantages of increased interoperability and efficiency in the application of thesauri in indexing and information retrieval processes. This would also allow greater compatibility and interoperability with other systems.

It is necessary to highlight the diversity of the 51 vocabularies used in ARCA, not only because of their number but also because of the heterogeneity of aspects such as coverage, organiation, updating, etc. In addition, several problems can be encountered when carrying out a project of this type:

- Technological limitations: initially, the direct application of the SKOS model implies the adoption of RDF-compatible technological solutions. The adoption and integration of these technologies is not straightforward, especially since the systems and application infrastructure for corporate data warehousing is often based on relational database management systems.
- The migration of the thesaurus to a new representation model requires a careful study of the controlled vocabulary ecosystem. Familiarity with the thesauri, taxonomies and keyword lists used in the description processes is essential to understand the nature of the synergies between vocabularies. It is also necessary take into consideration that the products resulting from description and indexing must continue to be maintained with the new vocabularies.
- The adoption of greater integration of corporate vocabularies requires the adaptation of editing and maintenance policies. The introduction of new vocabularies entails the implementation of new work dynamics that must be applied on a gradual basis. For this reason, adapting the different vocabularies to specific usage environments and defining adequate mapping between the new and previous vocabularies are critical factors for the deployment of such a model.

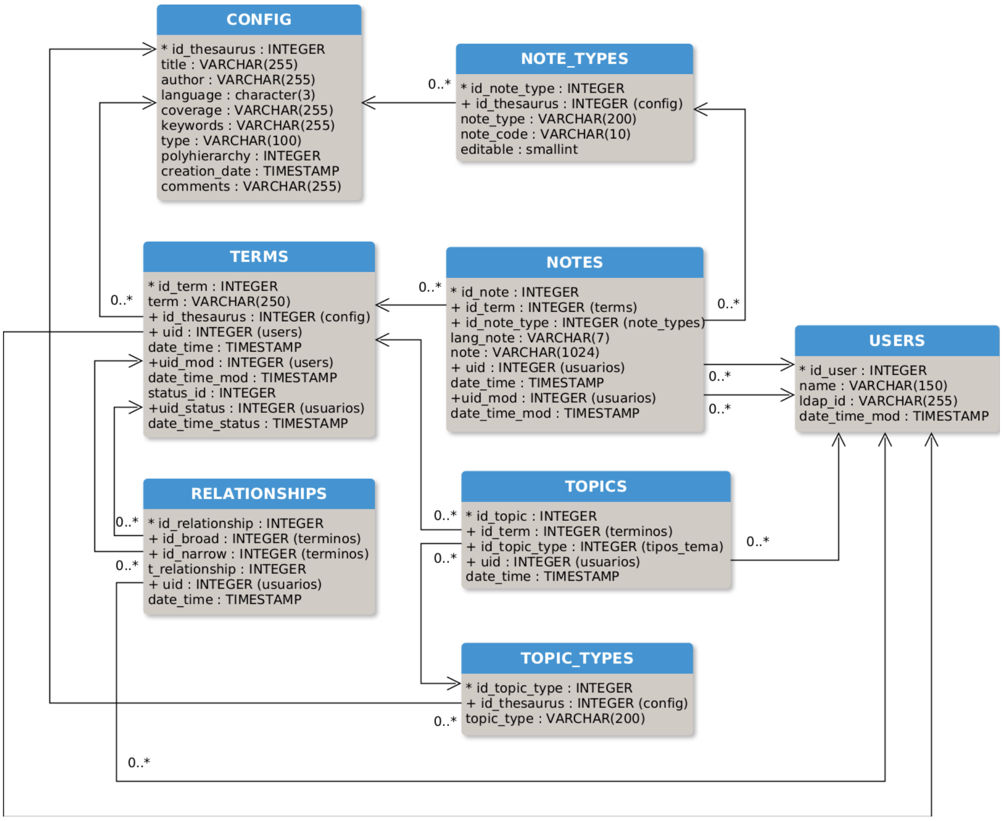


Fig. 1: Relational schema of the ARCA thesaurus manager. Source: own elaboration.

A pragmatic approach for the adoption of the SKOS model is its implementation in a relational database environment. Thus, the schema proposed above could be represented to fit the SKOS model as shown in Figure 2. In this proposal the table "nodes" stores resources of the three types of SKOS classes that are distinguished by assigning, in the attribute "node_type", the values "scheme", "concept" or "collection". The schemes correspond to each of the vocabularies. Concepts enable the definition of entities that are independent of terminological aspects and are used in the description and indexing processes. Collections are sets of concepts related to a subject. For example, it is possible to define a collection called "Environment" in which concepts such as "Here the earth", "Agrosphere" and "Man and the earth" can be included, but without these concepts belonging exclusively to the collection.

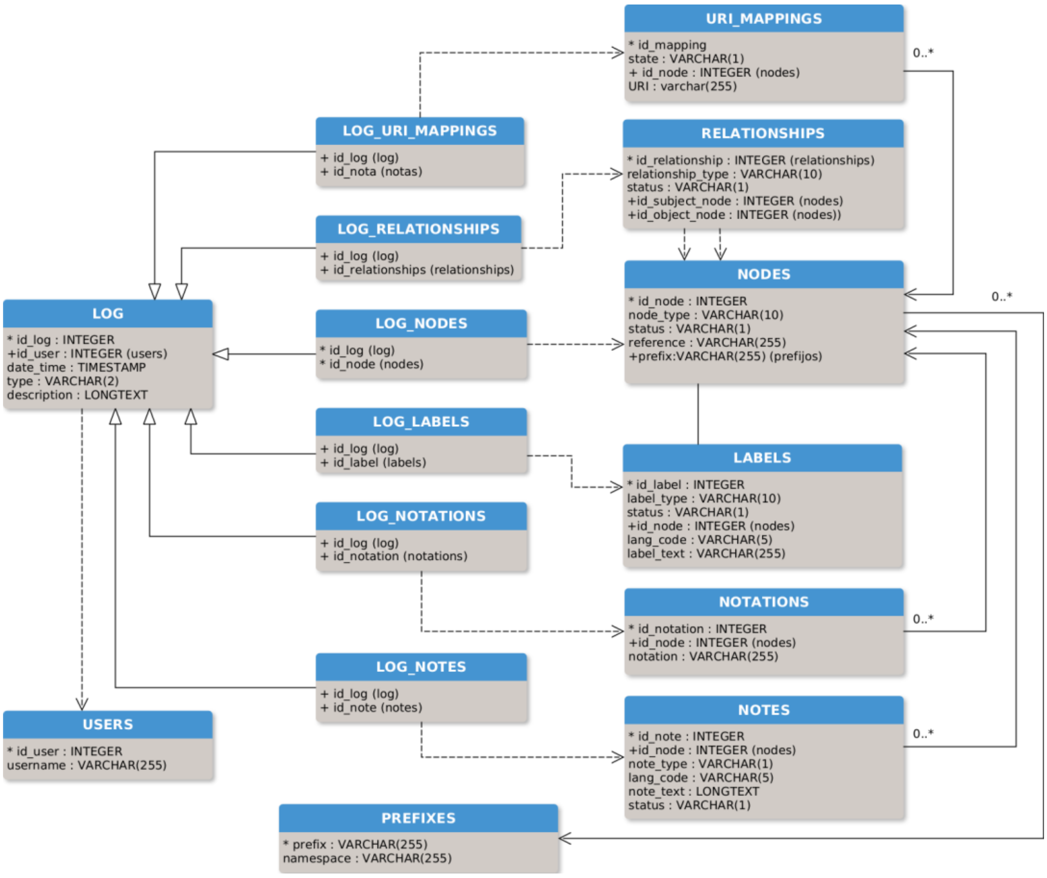


Fig. 2: Proposed relational schema for representing thesauri in ARCA compatible with the SKOS data model. Source: own elaboration.

A node can have multiple labels in different languages. In the "labels" table (Figure 2) there is an attribute called "label_type" which indicates whether it is a preferred, alternative or hidden label (skos:prefLabel, skos:altLabel, skos:hiddenLabel). Multilingual labelling is possible by assigning the corresponding language code in the "lang_code" attribute. Therefore, synonymy relations become an attribute of the labels: concept nodes can be labelled with a single preferred label in each language and multiple alternative labels to represent synonymous terms.

The relationships between elements are stored in the "relationships" table, where the links between concepts, collections and concept schemes are represented. The value of the "relationship_type" attribute makes it possible to identify the following relationship cases:

- broader: Broad hierarchical relationship between concepts.
- "narrower": Narrow hierarchical relationship between concepts.
- "related": Associative relationship between concepts.
- "inScheme": Relationship representing the membership of a concept in a concept scheme.

- "member": Relationship representing the concepts or nested collections contained in a collection.
- "hasTopConcept": Relationships of a concept scheme to the top-level concepts.
- "topConceptOf": Relationship representing when a concept is a top-level concept of a concept scheme.

The values above coincide with the naming of the equivalent properties of the SKOS recommendation. The table "relationships" would also allow the storing of mapping correspondences between concepts belonging to different schemas using the field "relationship_type" with the values "exactMatch", "closeMatch", "broadMatch", "narrowMatch" and "relatedMatch".

The table "uri_mappings" defines equivalence relations (owl:sameAs property) between vocabulary concepts and other resources such as Wikidata, DBpedia, Geonames, etc. This feature enables the integration of RTVE's thesauri into the Linked Open Data ecosystem.

This scheme also contemplates assigning different types of notes and notations to any type of node. Finally, it also includes a change log for all elements (node, label, relation, notation or note), identifying the user who made the change, the date it took place and the inclusion of descriptive comments.

Figure 3 shows an example using two concepts from the series thesaurus, between which two skos: related semantic relationships are defined. Both concepts are linked to a collection and to the concept scheme representing the thesaurus. The definition of preferred labels can be observed, which includes both the text strings and the language used. The owl:sameAs property defines the mapping to Wikidata entities.

Figure 4 shows the most relevant aspects of the representation of the previous example according to the proposed relational schema.

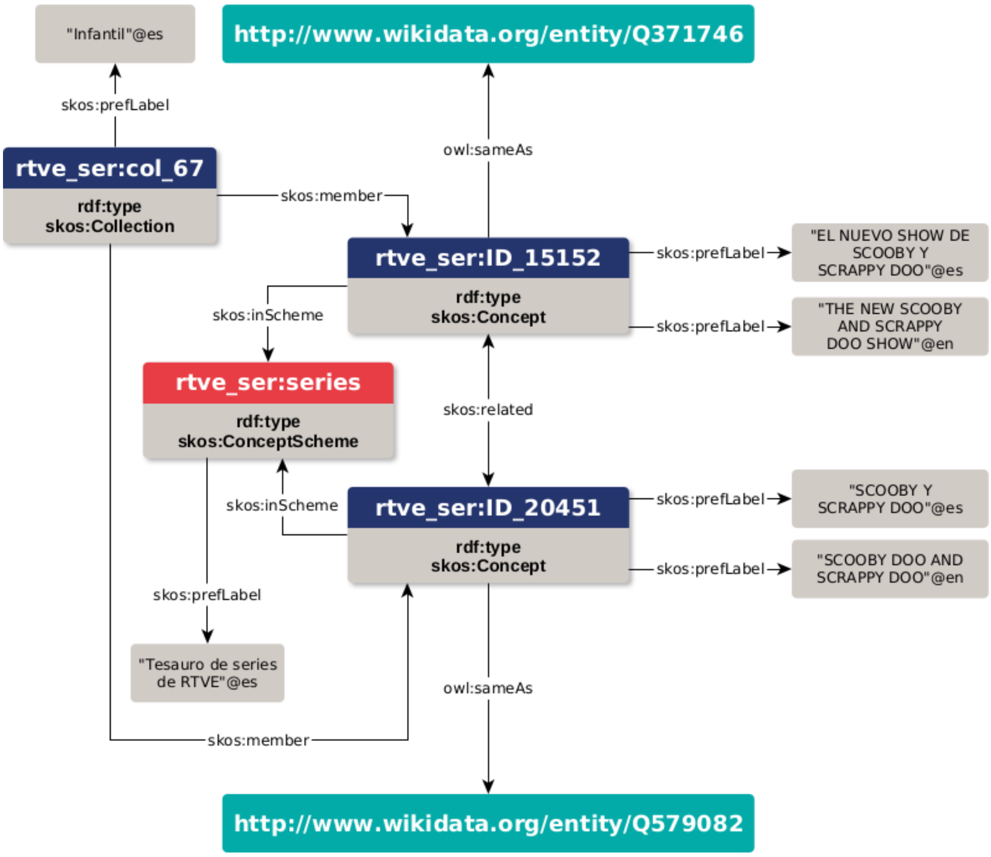
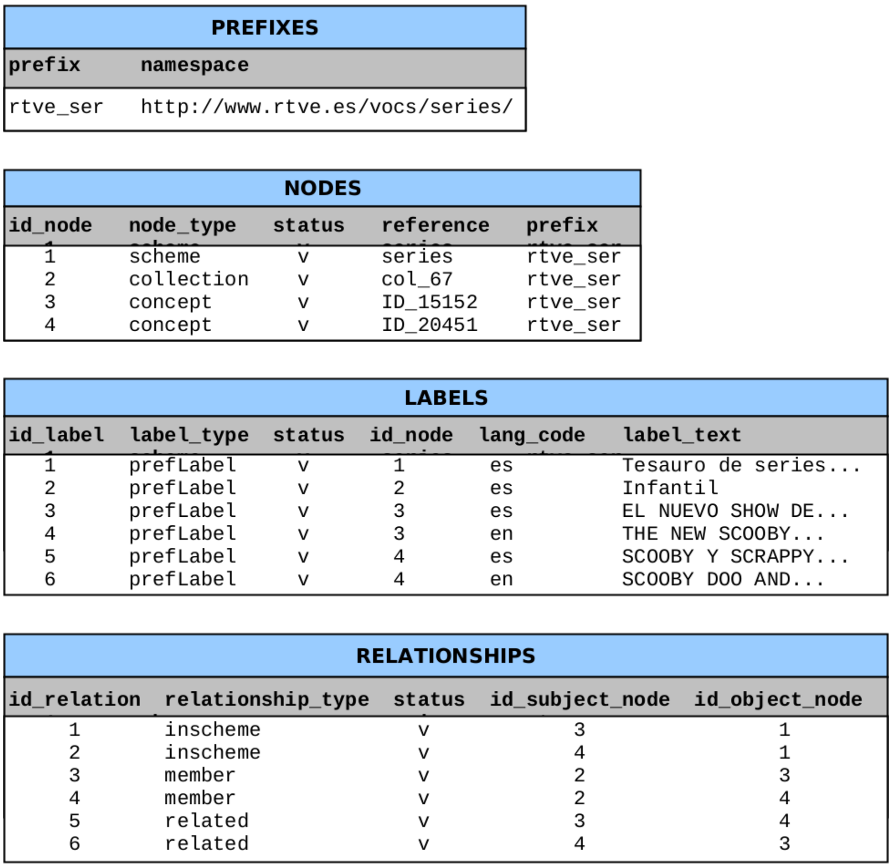


Fig. 3: Basic example that shows two concepts from the series thesaurus, and the equivalence with the representation according to the proposed relational schema. Source: own elaboration.



PREFIXES

prefix	namespace
rtve_ser	http://www.rtve.es/vocs/series/

NODES

id_node	node_type	status	reference	prefix
1	scheme	v	series	rtve_ser
2	collection	v	col_67	rtve_ser
3	concept	v	ID_15152	rtve_ser
4	concept	v	ID_20451	rtve_ser

LABELS

id_label	label_type	status	id_node	lang_code	label_text
1	prefLabel	v	1	es	Tesauro de series...
2	prefLabel	v	2	es	Infantil
3	prefLabel	v	3	es	EL NUEVO SHOW DE...
4	prefLabel	v	3	en	THE NEW SCOOBY...
5	prefLabel	v	4	es	SCOOBY Y SCRAPPY...
6	prefLabel	v	4	en	SCOOBY DOO AND...

RELATIONSHIPS

id_relation	relationship_type	status	id_subject_node	id_object_node
1	inscheme	v	3	1
2	inscheme	v	4	1
3	member	v	2	3
4	member	v	2	4
5	related	v	3	4
6	related	v	4	3

Fig. 4: Representation of the example in Figure 3 using the proposed relational schema. Source: own elaboration.

6. *Conversion, standardization and unification process*

The unification of the thesauri managed through ARCA has been designed based on the study of existing vocabularies. The transformation of the data from the different thesauri into a structure compatible with the SKOS model requires a conversion and standardization procedure. The result is a single controlled vocabulary for each of the domains (persons, places, series or topics). This process entails analysing both the terms of the thesauri and the relationships defined between them. For this purpose, the following tasks have been performed:

- Identification and unification of concepts (nodes) from the analysis of the terms of the different thesauri of a domain. Identical terms make it possible to identify unique concepts.
- Identification of collections of concepts based on associating terms with thematic categories. The collections are definedby these categories.
- Extraction of preferred labels and labelling of unified concepts.
- Processing of BT (broader term), NT (narrower term) and RT (related term) relationships for the definition of hierarchical and associative relationships.
- Processing of USE and UF (used for) relationships for identification of alternative labels.
- Processing of cross-linguistic relationships to identify labels in other languages, applying techniques for automatic recognition of the language of the term.
- Processing of other types of relationships (NI, SN and FI) that identify scope notes.
- Mapping of concepts with Wikidata through an entity recognition process based on the execution of SPARQL queries and the use of the EntitySearch Wikibase API.

This process is represented in Figure 5.

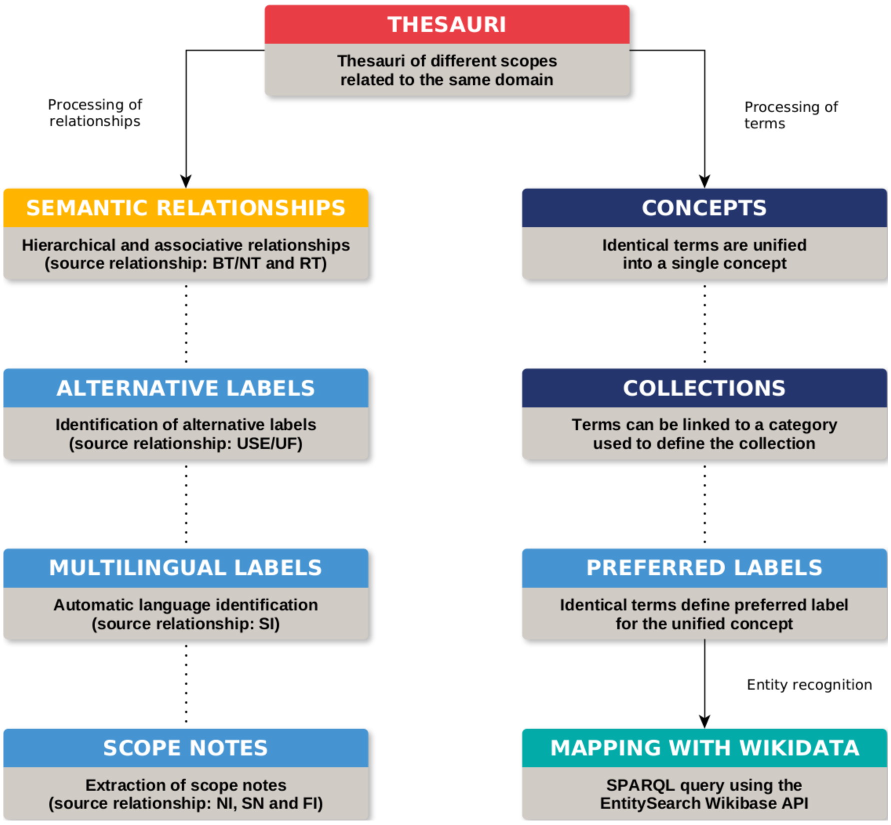

THESAURI
Thesauri of different scopes related to the same domain
Processing of relationships
Processing of terms
SEMANTIC RELATIONSHIPS
Hierarchical and associative relationships (source relationship: BT/NT and RT)
CONCEPTS
Identical terms are unified into a single concept
ALTERNATIVE LABELS
Identification of alternative labels (source relationship: USE/UF)
COLLECTIONS
Terms can be linked to a category used to define the collection
MULTILINGUAL LABELS
Automatic language identification (source relationship: SI)
PREFERRED LABELS
Identical terms define preferred label for the unified concept
Entity recognition
SCOPE NOTES
Extraction of scope notes (source relationship: NI, SN and FI)
MAPPING WITH WIKIDATA
SPARQL query using the EntitySearch Wikibase API

Fig. 5: Overview of the steps in the process of unifying the thesauri of a domain.

The identification of the nodes requires a previous analysis of the terms of those thesauri of different fields to be unified linked to the same domain. This analysis is necessary due to the duplication of terms between different thesauri of the same domain. The recognition of identical labels from different thesauri implies the creation of a single concept to represent them all. After the initial analysis, the conceptual level of the unified vocabulary is defined. The labels collected in this phase are then used to establish the preferred labels for the concepts.

Thesaurus terms are associated with one or more thematic categories. These categories have been used to define collections of concepts. In addition to defining these collections, the membership of these collections by the corresponding concepts has also been determined.

The processing of relationships between terms enables the identification of semantic relationships. Broader term (BT) and narrower term (NT) relationships have been used to represent hierarchical relationships. Related term (RT) relationships have been used to define associative relationships between concepts. USE/UF synonymy relationships have allowed the identification of alternative labels.

Relationships between equivalent terms from different languages (SI) have been used to represent preferred labels in languages other than Spanish. However, one of the main problems encountered in the conversion process to the new representation model is related to multilingual labelling. The thesauri managed in ARCA identify the equivalences between labels of different languages by means of a specific relationship at the terminological level. However, there are no cases in which the language in which any of the terms are found is identified. It is therefore necessary to apply a procedure to recognise the language of the terms in order to carry out the appropriate labelling. The notes are represented by relationships represented with FI, SN and NI codes. These relationships have been used for the definition of the corresponding scope notes.

The mapping of concepts to Wikidata entities is obtained with SPARQL queries to search for entities whose name matches the preferred label of the concept. Each SPARQL query makes use of the EntitySearch Wikibase API to find matches between the preferred thesaurus labels and the labels used in the Wikidata entity descriptions.

The process described above makes it possible to obtain a single vocabulary resulting from the unification of the different thesauri of a given domain. However, it is necessary to adapt the thesaurus management module in accordance with the following guidelines:

- It would switch to working with an approach based on managing concepts (instead of terms) that can be labelled by preferred or alternative labels.
- The software must control the assignment of a language to each label, whether in a default language or explicitly in another language when necessary.
- Semantic relationships are established between concepts and not between terms. These relationships must conform to the traditional set of hierarchical (broader/narrower) and associative relationships.
- The software should also control the constraints at the labelling level: avoid duplication of labels and control the assignment of one single preferred label in a language to each concept.
- It should also verify the consistency of the establishment of semantic relationships. This control is aimed at avoiding the creation of cycles in the hierarchical structure or the establishment of associative relationships between concepts located in the same taxonomic line.

A vocabulary stored in a relational database structure can also be exported to an RDF dataset which would not include the information about the editorial process made by the thesaurus managers. In this way the dataset could be published as open data for reuse or downloading in an interoperable format. Another possible application would be the evaluation of the quality of the vocabulary and the analysis of its structure through a specialized tool such as Skosify, VocBench, Poolparty or Qskos.

7. *Results*

In the case of the series thesaurus, almost 20,000 concepts were detected. This thesaurus has a low number of alternative labels and semantic relationships. However, the source relationships of equivalence between languages were numerous, which has given rise to many labels (more than 4,500) in other languages, of which it has been possible to automatically translate 96%.

The thesaurus of places resulting from the unification is twice as extensive (almost 41,000 concepts). In this case, a greater number of alternative labels has been obtained, but, in any case, the ratio of alternative/preferred labels is quite low (0.081). On the other hand, there are a considerable number of semantic relationships (more than 50,000). Since this is a thesaurus of places, with multiple city names and place names, it has been decided not to perform translations. In this case the detection of the original language of the labels is more complex as many of them are in Basque, Galician and Catalan, with the consequent difficulties of the different lexical processors to recognize the language of the label with 100% accuracy.

The subject thesaurus includes more than 21,500 concepts. The percentage of alternative labels is the highest of all the thesauri analysed (22%). It is also the one that includes the highest number of semantic relationships, with a total of 28,642, of which 16% correspond to associative relationships and the rest to hierarchical relationships. This vocabulary corresponds to a structure closer to the traditional thesaurus.

The thesaurus of persons includes more than 300,800 concepts. This vocabulary includes names of individuals and organizations. There is no indicator or structure to distinguish between individuals and organizations. The percentage of alternative labels with respect to the total is almost 6%. The alternative labels apply to both organizations (20TH CENTURY FOX -> Alternative: TWENTIETH CENTURY FOX) and individuals (JEROME DAVID SALINGER -> Alternative: J D SALINGER). The labelling also includes explanatory texts (Jesse Jackson (Father) -> Alternative: JESSE JACKSON (1941-). Given the number of concepts, the number of semantic relationships is very low (3,554). Most of these relationships are associative (65%) and are applied to define links between organizations, persons or both. In reality, this thesaurus of people is an authority file.

Table 3 summarizes the quantitative results of the analysis and unification process.

Tab. 3: Summary of the results obtained after the unification process and application of the SKOS data model and Wikidata entity recognition. Source: own elaboration.

Feature	Series	Places	Topics	People/Organisations
Thesauri processed	7	8	8	4
Concepts	19,532	40,908	21,462	331,645
Preferred labels	19,532	40,908	21,462	331,645
Alternative labels	458	3,351	6,335	21,499
Hieracchical relationships (*)	196	48,002	24,064	1,280
Associative relationships (*)	348	2,222	4,578	2,540
Recognized entities	4,326 (22.15%)	21,212 (51.85%)	9,127 (42.54%)	128,672 (38.8%)
Details of the main classes of recognized entdities (**)	TV series (Q5398426): 690 (15.95%) TV program (Q15416): 58 (8.27%) Film (Q11424): 345 (8%)	municipality of Spain (Q2074737): 4092 (19.3%) city (Q515): 1900 (8.95%) big city (Q1549591): 1784 (8.41%) municipality of Catalonia (Q33146843): 1222 (5.76%)	taxon (Q16521): 403 (4.41%) academic discipline (Q11862829): 234 (2.56%) type of sport (Q31629): 166 (1.81%)	human (Q5): 106714 (82.94%)

(*) The semantic relationships between concepts are counted in both directions: Concept A→Concept B and Concept B→Concept A). (**) Only the classes that exceed 5% of the total number of recognized entities or the three most frequent ones are included.

A number of problems associated with the unification process were identified. One is the use of capital letters for the labels of the original thesauri. The solution requires prior conversion of the entire label – except the initial letter – to lower case letters and using a morphological analyser to identify proper nouns, which could solve part of this problem. When we applied this method to the series thesaurus, we obtained 68% efficiency. However, the percentage drops significantly (15% of the labels analysed) when it is applied to the thesaurus of places, since the morphological analysers have difficulty identifying proper nouns associated with places and city names. The results obtained in the case of the topic thesaurus are similar to those of the series thesaurus (73%). Furthermore, the use of this method in the thesaurus of persons is quite inconsistent due to the large number of surnames that are identical to common nouns and to the large number of acronyms and abbreviations. For this analysis we used the Stanza morphological parser created by the Stanford NLP Group and the model available for Spanish and English. In this respect, it may be advisable to apply complementary techniques based on entity recognition (NER).

Another problematic aspect is the pre-coordinated nature of many labels. It is common to find terms such as "TENNESSEE - OAK RIDGE NATIONAL LABORATORY". The solution to this problem would be to search for substrings in those labels that contain any type of separator characters (commas, semicolons, full stops, hyphens, etc.). The cases identified as positive would require the labelling to be corrected and the corresponding hierarchical relationships to be established. In line with the previous example, this would be rendered as “TENNESSEE” - narrow concept -> "NATIONAL LABORATORY OAK RIDGE". Another approach would be the use of extended hierarchical relationships as contemplated by ISO 25964 (Alexiev, 2016).

Duplication represents a further challenge. There are duplicate concepts but with different labelling. An example of this is in the onomastic thesaurus, which includes a concept with the preferred label "LOS ANGELES LAKERS" but also includes another whose preferred label is "LOS ANGELES LAKERS BASKETBALL TEAM".

There is also a problem of terms being duplicated between different thesauri. An example of this would be where it is possible to find a concept labelled "MADRID (CAPITAL)" in the thesaurus of places, while the thesaurus of topics contains the concept with the label "MADRID" and the thesaurus of persons includes two concepts labelled "MADRID" and "MADRID (CAPITAL)", respectively.

As well as the aforementioned difficulties, the fact that the translation is not entirely reliable and the fact that there are no connections with external information sources must also be added. Defining mapping relationships with Wikidata, for example, would have multiple benefits: enrichment of the semantic structure of vocabularies, retrieval of complementary alternative labels, automatic recognition of entities and the definition of multilingual labels. Finally, other low incidence and easily solved errors are the identification of some cases of hierarchical cycles in the thesaurus of topics and the absence of preferred labels in some concepts included in the geographical and thematic thesauri.

The Wikidata mapping process has been done by searching preferred labels using the EntitySearch Wikidata API. The results are conditional on the type of vocabulary, the language of the preferred label and, obviously, the Wikidata coverage. The results obtained for the thesaurus of places, where more than 50% of the vocabulary items were recognized in Wikidata, are particularly noteworthy. The results for the onomastic thesaurus reflect that more than 38% of the elements were recognized and that of these more than 80% corresponded to Wikidata class Q5 (people), which is an indicator of the coherence of the process. It is evident that elements of vocabularies with a strong local component and no presence in Wikipedia/Wikidata are not recognized in this process. However, the main advantage of this process lies in obtaining results from a relatively flexible search procedure that could be extended by also using the alternative labels.

8. Conclusions

The number of works related to the use of controlled vocabularies in audiovisual archives is limited. Nonetheless, growing interest can be observed due to the role of vocabularies in semantic searches in the context of media archives. There is also a clear trend towards automatic keyword generation through the application of technologies such as NLP. Although these keywords can be integrated with existing thesauri, establishing connections with external data sources such as Wikidata and Geonames is considered especially pertinent.

The data collected in the survey on the use of controlled vocabularies in television archives shows that there is still a long way to go to achieve full standardization and interoperability and demonstrates a clear tendency to adopt local or media-specific solutions. The survey also shows limited use of thesauri at an international level with a predominance of thesauri of persons, which constitute one of the major search elements in the media.

Since its creation, the RTVE archive has managed multiple thesauri whose structure, growth and use are aligned with the specific needs of each of the analysis units. The absence of a common standard, such as SKOS, has resulted in heterogeneity in the structure of the vocabularies. There is also a great deal of semantic inconsistency due to the distributed management of vocabularies and the different criteria of the editors. This fact has impacted the level of interoperability of the metadata generated in the processes of indexing and describing audiovisual content. This lack of interoperability, which may perhaps be less critical in exchanges of programs with other archives, is nevertheless essential in order to optimize processes within the organization itself. The associated standards that separate the conceptual and lexical levels of the controlled vocabularies allow greater flexibility in the search and retrieval of audiovisual objects.

RTVE's thesauri are geared towards the peculiarities of an archive with a long history. The division of production into news and non-news programs and the separation of the radio and television archives has given each of these vocabularies their own distinct character which is closely linked to the characteristics of the programs analysed and the recovery and reuse needs of the various areas and departments. This explains the differing levels of detail between the geographical and the onomastic, with a higher level of detail applied to the description of the original recordings versus that of the broadcast programs. The origin of their indexing terms is also diverse. Many of these terms are assigned in the analysis units by the documentalists (although a significant number also come directly from the production or broadcast areas), which highlights the need to reuse metadata and seek interoperable models that are useful for all business areas within the corporation.

Additionally, the multilingual nature of the audiovisual productions and the geographical structure of RTVE have a direct impact on the vocabularies. The coexistence of terms in different languages is especially evident in the thesauri of places and series. Multilingualism is not applicable to other vocabularies such as topics or people and organizations. SKOS is an optimal solution for managing multilingualism.

The flexibility of the SKOS model enables its application in the context of RTVE, where thesauri are managed within a relational database platform. The proposal contained in this work allows the use of SKOS functionalities in ARCA and would enable the exporting of RDF datasets for subsequent publication.

Nevertheless, interoperability is not enough. For integration into the linked data ecosystem it is necessary to go one step further by establishing links between vocabularies and open data sources such as Wikidata. The results shown in this research are promising and suggest that the procedure used could be refined to incorporate alternative labels into entity recognition processes, as well as other data sources beyond Wikidata.

The application of SKOS and the adaptation of its data model to a relational environment is feasible. This study focuses on controlled vocabularies in media files. However, the proposed methodology enables extrapolation to other environments in which the use of controlled vocabularies remains focused on terminology management. The management of vocabularies with two levels (conceptual and lexical) is far more flexible. All this facilitates the integration of vocabularies at a corporate level as well as the reuse of external vocabularies with the consequent optimization of resources.

REFERENCES

Alexiev, V., Isaac, A., & Lindenthal, J. (2016). On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI). *International Journal on Digital Libraries*, 17, 39-48. https://doi.org/10.1007/s00799-015-0162-2

Bazán-Gil, V. (2021a). Use of controlled vocabularies in TV Archives. In *LinkedIn*. https://www.linkedin.com/posts/virginia-baz%C3%A1n-gil_survey-on-the-use-of-controlled-vocabularies-activity-6820361793211269120-NF1Y?utm_source=share&utm_medium=member_desktop

Bazán-Gil, V. (2021b). First findings of our research on the use of controlled vocabularies on TV Archives!. In *Twitter*. https://twitter.com/VirginiaBazang/status/1428267534093012992

Bazán-Gil, V., & Escribano, M. (2017). Raiders of lost order: reordening thesaurus in a digital envioremеnt. FIAT/IFTA World Conference: Living in the Digital Age; Connecting Roots and Cultures.

BBC. (n.d.). *BBC Things*. Retrieved from https://www.bbc.co.uk/things/

Biblioteca Nacional de España. (2022, October 31). *Datos.BNE.es El portal de datos bibliográficos de la Biblioteca Nacional de España.* Retrieved from https://datos.bne.es/inicio.html

Bus, H., & Huis in't Veld, V. (2021). Thesaurus Management at Sound and Vision: the switch to a new editor. FIAT/IFTA World Conference: Advancing the Digital Dividend. Retrieved from https://fiatifta.org/index.php/world-conference-2021-online-edition/

Caldera-Serrano, J., & Sánchez-Jiménez, R. (2008a). Ontología para el control y recuperación de información onomástica en televisión. *El Profesional de La Informacion, 17*(1), 86–91. https://doi.org/10.3145/epi.2008.ene.10

Caldera-Serrano, J., & Sánchez-Jiménez, R. (2008b). Recuperación de secuencias de información audiovisual con rdf y smil. *El Profesional de La Informacion, 18*(3), 291–300. https://doi.org/10.3145/epi.2009.may.06

de Boer, V. (2017). Getting down with LOD tools at the 2nd CLARIAH Linked Data workshop. Retrieved from http://www.victordeboer.com/tag/cultuurlink/

de Boer, V., Ordelman, R. J. F., & Schuurman, J. (2016). Evaluating unsupervised thesaurus-based labeling of audiovisual content in an archive production environment. *International Journal on Digital Libraries, 17*(3), 189–201. https://doi.org/10.1007/s00799-016-0182-6

de Boer, V., Priem, M., Hildebrand, M., Verplancke, N., de Vries, A., & Oomen, J. (2016). Exploring Audiovisual Archives Through Aligned Thesauri (pp. 211–222). https://doi.org/10.1007/978-3-319-49157-8_19

de Prada, A. (2021). Archivos Audiovisuales de RTVE entre el patrimonio empresarial y la memoria. Nueva Revista de Política, Cultura y Arte.

EBU. (2011). EBU – TECH 3336: EBU Reference Data & Classification Schemes. Retrieved from https://tech.ebu.ch/docs/tech/tech3336.pdf

EBU. (2020). Index of /metadata/cs. Retrieved from https://www.ebu.ch/metadata/cs/

Hidalgo, P. (2017). Preservación del patrimonio audiovisual de televisión El archivo de Televisión Española (TVE): de los orígenes a la digitalización. https://eprints.ucm.es/id/eprint/41938/1/T38624.pdf

IPTC. (2022a). IPTC CV Server Guidelines. https://www.iptc.org/std/NewsCodes/guidelines/cv.iptc.org-guidelines.html

IPTC. (2022b). Media Topics Subject Taxonomy for the Media: the successor to the Subject Codes.

IPTC. (2022c). News Codes. Retrieved from https://iptc.org/standards/newscodes/

ISO. (2011). ISO 25964-2:2011. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval.

ISO. (2013). ISO 25964-2:2011. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies.

López de Quintana, E. (2010). Transformación y compatibilidad: el documento audiovisual en los archivos municipales. XVIII. Jornadas de Archivos Municipales. Cuadro de Clasificación de Fondos. Pilares de La e-Administración: Cuadro de Clasificación y Tesauro. Retrieved from http://www.madrid.org/archivos/images/ACTIVIDADES/PUBLICACIONES/XVIIIjarchivosmunicipalescuadro.pdf

Meemoo Flemish Institute for Archives. (2022). Meemoo: a vision for the future, and the past. Retrieved from https://meemoo.be/en/meemoo-a-vision-for-the-future-and-the-past

Muñoz-de-la-Peña-Costero, P., Meana-Alonso, S., & Sáez-Carreras, S. (2014). Cinco años de experiencia digital en los Servicios Informativos de TVE: una nueva gestión de contenidos. *El profesional de la información, 23*(1), 72-79. https://.doi.org/10.3145/epi.2014.ene.09

Quinn, B., & Parrucc, J. (2021). IPTC NewsCodes: Controlled Vocabularies for the News Media. EBU MDN Workshop 2021. Retrieved from https://tech.ebu.ch/docs/events/mdn2021/presentations/1_1620_Quinn_IPTC_Parucci_NYT_IPTC_NewsCodes.pdf

Valle Gastaminza, F. del, & García Jiménez, A. (2002). Construcción de un tesauro para el Centro de Documentación de Telecinco. *Scire, 8*(1), 103–118. Retrieved from https://www.ibersid.eu/ojs/index.php/scire/article/view/1162/1144

Valle Gastaminza, F. del. (2003). Tesauros e Información Audiovisual. Estudio de caso. *Documentación de Las Ciencias de La Información*, 26, 165–180. Retrieved from https://revistas.ucm.es/index.php/DCIN/article/view/DCIN0303110165A

W3C. (2009). SKOS Simple Knowledge Organization System Reference. *World Wide Web Consortium Recommendation.* Retrieved from http://www.w3.org/TR/2009/REC-skos-reference-20090818.