

Integrating Terminological Tools and Semantic Archaeological Information: the ICCD RA Schema and Thesaurus

Achille Felicetti¹, Ilenia Galluccio¹, Cinzia Luddi¹, Maria Letizia Mancinelli²,
Tiziana Scarselli³, and Antonio Davide Madonna³

¹PIN, VAST-LAB, Prato, Italy

²MiBACT-ICCD, Istituto Centrale per il Catalogo e la Documentazione, Rome, Italy

³MiBACT-ICCU, Istituto Centrale per il Catalogo Unico, Rome, Italy
{achille.felicetti, ilenia.galluccio, cinzia.luddi}@pin.unifi.it
{marialetizia.mancinelli, tiziana.scarselli,
antoniodavide.madonna}@beniculturali.it

Abstract. This paper describes the process of mapping, translation and publication in SKOS format of the RA Thesaurus, a terminological tool developed by the Italian Ministry of Cultural Heritage (MiBACT) as a part of the official documentation used for the recording of archaeological finds. In particular, the RA Thesaurus is intended to provide unified and meaningful terminology for the description of archaeological objects according to the MiBACT official cataloguing standards. After describing the thesaurus, the logic with which it was developed and its internal structure, we report the various phases of the conversion, both from a theoretical and implementation point of view, and the various technologies used for the publication of the thesaurus on the web. This work is a collaborative effort between PIN and MiBACT carried out under the ARIADNE project.

Keywords: Archaeology, Mapping, Thesauri, ICCD, CIDOC CRM, SKOS

1 Introduction

ARIADNE is a European project focusing on integration of existing archaeological research data infrastructures to enable the use of distributed datasets and services by means of new and powerful technologies as an integral component of the archaeological research methodology. Among other activities, ARIADNE is also actively working on building a coordinated system of multilingual terminology tools able to meet the many needs of the international community of archaeologists. As part of these integration activities, the valuable work of mapping national catalogue schemas on international standards is a critical step; at the same time integration of terminology resources is necessary to overcome linguistic barriers that frequently slow down the integration processes. We have extensively described the process of CIDOC CRM encoding of the RA Schema,

released by ICCD for documenting archaeological artefacts in Italian archaeology, in a previous work [1]. The mapping was presented as work in progress at that time. Since then, new extensions of the CIDOC CRM (and in particular *CRM_{archaeo}*) have been released which are now able to provide more possibilities for the enrichment of the semantic archaeological information and a more archaeological oriented means of documentation. The release of new versions and the creation of new extensions of the CIDOC CRM gave us the opportunity to investigate how the mapping could be improved. This allowed us to bring the mapping to a stage very close to completion, although much work still remains to be done. The RA Schema is closely linked to the RA Thesaurus, a sophisticated vocabulary providing all the necessary terminological facilities for an efficient and well-structured recording of the objects coming from archaeological excavations. The vocabulary has been implemented by ICCD to support the encoding of two specific fields (OGTD - CLS). These two fields describe the definition of the object and its class and production. This paper will focus and propose integration between the RA Schema and its thesaurus, based on W3C recommendations and using numerous tools developed and used by several partners in the ARIADNE project.

2 ICCD and the Standards for Cultural Heritage

ICCD is the Italian Central Institute for Catalogue and Documentation, one of the seven Central Institutes of the Italian Ministry of Cultural Heritage whose main goal is to create a centralized national catalogue of Italian cultural heritage. The activity of the Institute is based on the research and development of tools, methods and standards for knowledge, protection and enhancement of the cultural and artistic heritage in Italy. It mainly provides the management of the national general catalogue of archaeological, architectural, historical, artistic and ethno-anthropological heritage, the development of cataloguing methodologies and standards, and the coordination of the technical institutions involved in the cataloguing activities on the national territory. ICCD also provides tools and best practices for implementing these standards with the clear intent of unifying and streamlining processes related with the cataloguing activities, to guarantee quality and to implement standardisation and interoperability at a national level.

To ensure that this happens efficiently, the Institute creates and releases a series of organic resources and recommendations to support the standardization process in all its aspects. These include detailed regulations describing the various tools and the way they should be used, a set of schemas and forms to collect information in a structured way according to the different asset types, authority files to guarantee homogeneity for the common transversal key concepts and entities, thesauri and terminological tools to provide uniform layers of information and a common language. Among the latter category, one of the most important tools released by ICCD is the RA Thesaurus, a tool providing standard names for the definition of archaeological artefacts described using the RA schema,

the ICCD standard schema used for the recording of movable objects. The RA Schema is the most used and well established standard for Italian archaeology so far. For this reason, ICCD has invested a lot of effort in the definition of a terminological tool able to provide standardized and unambiguous names for specific fields of the schema. The creation of the RA thesaurus is one of the best results of this effort.

3 The ICCD RA - CIDOC CRM Mapping

In the previous work carried out together with ICCD, a detailed analysis of the RA Schema was made to map the most significant model of ICCD archaeological cataloguing system to CIDOC CRM. The RA Schema is used to record movable objects. It is one of the most used for Italian archaeology because of the huge and ever increasing amount of artefacts found during excavations. The RA Schema contains a large number of descriptive information and “cross-sections” allowing cross references with other ICCD resources. The RA Schema, together with the RA Thesaurus, features one of the best tools of this kind in the international panorama of cataloguing systems. The previous mapping work was carried out on CIDOC CRM and took advantage of version 5 of the model, released in 2013. However, in the last two years, a new version and numerous extensions of CIDOC CRM have been released. Version 6 and the *CRM_{archaeo}* [2] and *CRM_{sci}* [3] extensions, much more suitable for the description of archaeological phenomena, have strongly enhanced the representation and mapping of excavation entities. *CRM_{archaeo}*, in particular, is being developed by the ARIADNE project to facilitate the encoding of archaeological entities. Given this, we decided to update the previous mapping in order to provide a stronger archaeology-oriented logic to the various concepts and relationships that the RA Schema presents.

One of the most difficult problems to solve during the previous mapping was the representation of the “finding” event, intended as the excavation activity during which objects are found. This event is of paramount importance in archaeology because it is fundamental to trace the object’s provenance and to reconstruct its history. Following the CIDOC-CRM model, we represented the archaeological objects by using the *E22 Man-Made Object* class. However, to describe their relationships with the two important activities of “survey” (corresponding to the “RE” field of the RA Schema) and “excavation” (specified in the “DSC” field), CIDOC CRM core only provided a “change of ownership” relationship that hardly fits here but we decided to use it anyway. Our previous mapping appeared as shown in Figure 1.

Thanks to the release of the new extensions and a deep analysis of the cross-section relating the RA Schema [4], the new mapping now shows a more accurate rendering of these concepts. To express the “object found during an excavation” relationship, *CRM_{archaeo}* provides the *O19i was object found by property*, through which it is possible to link the artefact with the new *S19 Encounter Event* class, expressly designed to render the concept of “finding” as an event which occurred (*P7 took place at*) at a given Site (*E7*) identified by

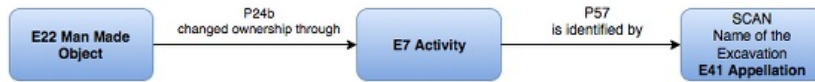


Fig. 1. ICCD-RA/CIDOC-CRM mapping

a given appellation (*P57 is identified by - E44 Place Appellation*), as shown in Figure 2. This constitutes a more accurate representation of these concepts.



Fig. 2. ICCD-RA/CIDOC CRM/CRMarchaeo mapping

3.1 The ICCD RA Thesaurus

The RA Thesaurus was developed expressly to provide standardised values for some of the “OG-OGGETTO” (*Object*) fields of the RA Schema. The content of the thesaurus is organized in a tabular structure with five columns arranged according to the hierarchical levels provided by the thesaurus. The first three columns, used to fill the CLS field of the RA schema, present the categories’ three levels of hierarchy, to which any concept can belong; column four lists the main terms for the definition of the objects; column five provides specifications of the main terms in accordance with morphological, functional or partitive criteria. Both columns four and five are meant to provide standard terms for the OGTD field of the RA Schema. Additional columns, reporting further attributes and specifications for each term and subterm, such as descriptive notes and sample images, are also present (see Fig. 3). Images are an added value of this tool for their ability to visually show what words are not always able to say. We have already investigated some of the possibilities to encode figures in our mapping, but unfortunately the tools at our disposal do not always allow a clear definition of these entities. For sure it will be important, in future versions of the thesaurus, to define a standard mechanism for associating concepts with their images even in the SKOS version of the thesaurus.

The RA Thesaurus differs from the other terminological tools created by ICCD in the very sophisticated structuring criteria it follows, made more complicated by the large amount of information deriving from Italian archaeology and the huge number of classifications and nomenclatures it provides. In particular, the thesaurus is structured according to a multilevel schema based on concept coordination, a typical KOS activity in which concepts are combined with each other in order to produce meaningful “sentences” that define complex

Integrating Terminological Tools and Semantic Archaeological Information

| LIVELLI GERARCHICI PREVISTI NEL THESAURUS | | | | | | | | | |
|--|-------------------|-------------------|---|------------------------|------------|---|-----------------------|--------------|--------------------------|
| SONO UTILIZZATI PER VALORIZZARE CAMPI DIVERSI DEL TRACCIATO DELLA SCHEDA RA 3.00, paragrafo OG-OGGETTO (vedere di seguito le istruzioni specifiche) | | | | | | | | | |
| LIVELLI DA UTILIZZARE PER LA COMPILAZIONE DEL CAMPO <i>CLS Categoria - Classe e produzione</i> <small>Per la compilazione del campo CLS sono selezionate le definizioni gerarchicamente relazioni al termine e alle sue eventuali specifiche scelse dai successivi livelli 4 e 5 del thesaurus (in linea in proposito alle istruzioni per l'uso del vocabolario aperto per il campo CLS della scheda RA, pubblicate sul sito ICCD)</small> | | | LIVELLI DA UTILIZZARE PER LA COMPILAZIONE DEL SOTTOCAMPO <i>OGTD - Definizioni</i> | | | ATTRIBUTI DEL TERMINE INSERITO IN UNO DEI LIVELLI 1-6 | | | |
| LIVELLO 1 | LIVELLO 2 | LIVELLO 3 | LIVELLO 4 | LIVELLO 5 | | | | | |
| CATEGORIA LIVELLO | CATEGORIA LIVELLO | CATEGORIA LIVELLO | TERMINI | TERMINI PIU' SPECIFICI | | | TERMINI PREFERENZIALI | NOTA D'ABITO | INDICAZIONE SEMPLIFICATA |
| | | | | FUNZIONE | MORFOLOGIA | PARTE | | | |
| col. 1 | col. 2 | col. 3 | col. 4 | col. 5 | | | | | |

Fig. 3. The ICCD RA Thesaurus model

concepts. Generally speaking, there can be two types of concept coordination: pre-coordination and post-coordination. The key distinction between the two relies on when the actual coordination occurs in relation to an information retrieval event. Pre-coordination is decided and implemented before the information retrieval time, by a KOS maintainer or by an indexer who is using the KOS itself. This occurs, for instance, when an indexer takes two existing concepts from a concept scheme, such as “Coins” and “Mintage”, and explicitly combines them with a given syntax, such as “Coins-Mintage”, to index a particular document. Post-coordination, on the other hand, is performed as part of an information retrieval task, for instance through a SPARQL query able to retrieve all documents indexed using both “Coins” and “Mintage” concepts [5]. The RA Thesaurus follows the post-coordination approach to create ad hoc concepts by using the elements of a given schema. Each concept is in fact provided with all the necessary subterms depending on it, which can belong to three specific semantic areas according to the specification provided: either functional (i.e. relative to the specific function of the object), partitive (i.e. relative to a specified part of the object) or morphological (i.e. linked to the different forms that from time to time an object may present). The structure of the thesaurus is obviously functional to the specific cataloguing activities. Each concept is thus created on the fly by combining the main terms with all the related subterms required to render the specific name that a concept should show in a given context. Figure 4 provides an example of how the thesaurus is structured by reporting the various facets of the term “cintura” (belt) and its related functional and morphological subterms:

It is evident from the example above that the thesaurus itself does not offer a closed and exhaustive list of all possible terms that can be used during the compilation of the schema. Instead, it is a reference tool that, after a general term is fixed, assists the user in proceeding to further specifications by the addition of suitable subterms to gradually approximate the precise semantic meaning of the object to be described.

The flexibility of this structure allows it to achieve a significant depth of semantics, where required, and to build specific definitions of several types of objects,

| Reference Term | Specific Terms | Scope notes |
|----------------|-------------------------------|-------------|
| cintura | | |
| | a fascia | MORPHOLOGY |
| | a losanga | MORPHOLOGY |
| | a placche | MORPHOLOGY |
| | affibbiaglio | PART |
| | borchia | PART |
| | multipla | MORPHOLOGY |
| | per la sospensione delle armi | FUNCTION |
| | puntale secondario | PART |

Fig. 4. Example of the thesaurus structure

including those in fragmentary conditions (for instance by means of the partitive subconcepts).

This will overcome the necessity to define in advance the entire terminological apparatus suitable to describe the infinite variety of situations the archaeologists may face.

Just to remain with the example above, from a logical point of view, if an archaeologist finds a stud (*borchia*) pertaining to an ancient belt (*cintura*) intended for the suspension of a sword or other similar weapons (*per la sospensione delle armi*).

A valid definition would be composed as follows:

Cintura (main term) +
per la sospensione delle armi (morphological aspect of the main term) +
borchia (part of the object that was found)

in order to have an entry like this:

Cintura per la sospensione delle armi, borchia (Belt for weapons suspension, stud)

representing an exhaustive explanation of the fragmentary object itself and of the bigger object which is part of, and also as a valid entry from the terminolog-

ical point of view following the formal recommendations provided by the ICCD guidelines and validation systems.

3.2 A SKOS Mapping Proposal

SKOS is the standard chosen by the ARIADNE project for the encoding of all terminological resources to be used in its integration plan, and for the undeniable advantages provided to integration and interoperability by its RDF-based format. As one can easily understand from what was previously stated, the “combinatorial” nature of the RA Thesaurus, and especially of the sections intended for the encoding of the OGTD field (column four and five), makes it very difficult to encode in a SKOS compatible format, which requires that a complete, self-consistent and self-sufficient definition exists in the thesaurus for each item or concept. The SKOS vocabulary itself does not provide any mechanism for expressing that a given concept may consist of other pre-coordinated concepts. It is, of course, possible to extend SKOS to establish a pattern for representing coordinated concepts, for instance by stating a new sub property, as in the following example:

```
iccd:coordinationOf a rdf:Property;
  rdfs:domain skos:Concept;
  rdfs:range rdf:List.
```

and then use the new property this way:

```
iccd:coinsMintage a skos:Concept;
  iccd:coordinationOf (iccd:coins iccd:mintage);
  skos:prefLabel "Coins-Mintage"@en.
```

However, patterns for pre-coordination have not yet been exploited by the SKOS community and solutions of this kind have not been explored fully enough to warrant their inclusion in the official SKOS vocabulary. Analyzing the RA Thesaurus, PIN and ICCU identified a possible solution. We tried to follow a different approach, more “pre-coordination oriented” to rearrange, where possible, the original content according to semantic criteria in order to define meaningful self-consistent concepts in the SKOS representation. After discussing the matter in depth, we proposed the following solutions:

1. The partitive specification subterms are in many cases independent terms related with the main term mostly by a part-whole relationship. Thus, it is possible to describe this relationship by using the *skos:narrowerPartitive* property to define them. This is particularly suitable if we consider that the same partitive term could occur for different main concepts: both a belt and a flag could have a *puntale* (ferrule) as partitive concept. Therefore, it is important to clearly define the hierarchy of these kinds of objects. Alternatively, it would be possible to combine main terms with their partitive terms in order to define complete and

self-consistent concepts, to be then defined as narrower terms of the main ones. In the previous example, we could define, for instance, a new *puntale di cintura* (belt ferrule) term, which would be clearly distinguished by a *puntale di insegna* (flag ferrule), the two being totally different, although very similar, objects.

2. The morphology and functional specification subterms are meaningless in themselves. They become meaningful only when combined with their main term. Creating SKOS narrower terms from these elements requires, for each morphological or functional term, the creation of a subterm obtained by combination with the super concept, in order to obtain a set of semantically consistent narrow terms. There is no semantic meaning in *multipla* itself unless this concept is used together with *cintura* in order to specify, in this case, the typology of a given belt. *Cintura multipla* is, on the other hand, a perfectly consistent concept.

Multiple combination of partitive, morphological and functional sub concepts to create specific entries, even if not impossible, would be very difficult to implement in SKOS due to the exponential growth of all possible combinations. At present, we propose not to extend the pre-coordination operations beyond the minimum requirements of semantic understandability and to use more than one SKOS concept to describe specific archaeological objects if required.

4 SKOS Encoding of RA Thesaurus

From a technological point of view, the RA Thesaurus was created starting from 2008 on the basis of the terms extracted from the database maintained by the “Sistema Informativo Generale del Catalogo” (SIGEC). Its development also went through various phases of data cleaning and strengthening. The RA Thesaurus is currently an “open vocabulary”, meaning that it is not meant to have a stable form since its content can be updated and modified by ICCD during further stages of research. Currently, the available version of the vocabulary is in textual format that is organized in a tabular structure, whose fields comply with the ISO standard norms for thesauri. In order to make the original textual information interoperable and ensure integration with semantic terminological tools, it was necessary to encode them in a structured and standard format.

The process we implemented for the SKOS encoding of the RA Thesaurus is a proposal for its *re-engineering* as a formal ontology and for making the knowledge it provides explicit in a formal sense. The whole process of encoding required a set of subsequent steps for data analysis, adjustment, conversion, publication and enrichment, in which the original textual data has been processed using both open source tools and *ad hoc* scripts.

The process can be subdivided into two analytic phases (see Fig. 5):

1. In the first analytic phase we focused on encoding the key fields of the original thesaurus, such as concepts and classes. The result of the first phase consisted in the creation of a SKOS/RDF version of the RA Thesaurus obtained through the mapping between the main concepts and the SKOS Core Vocabulary.
2. In the second phase, we focused on the integration of all morphological, functional and partitive aspects related to thesaurus’ concepts. The analysis of this

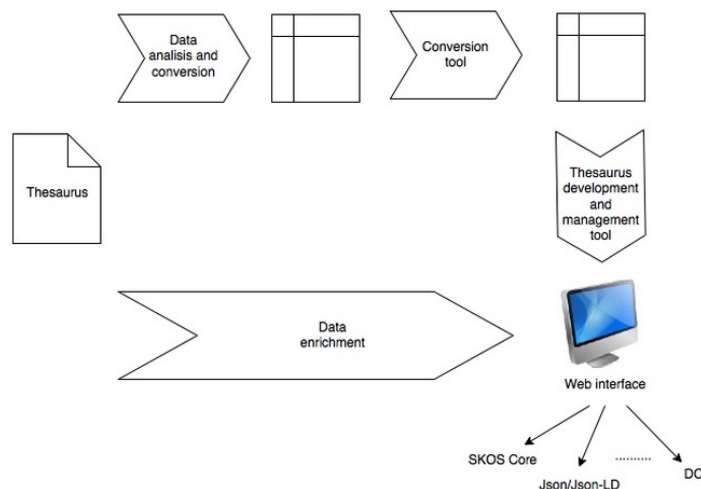


Fig. 5. SKOS encoding process

additional information required further investigation into how SKOS extensions could be used for the publication of thesauri in a semantic format.

4.1 Thesaurus Conversion Using SKOS Core Vocabulary

The conversion of the RA Thesaurus initially required a deep data analysis to define a precise mapping between its main fields and the SKOS Core Vocabulary [6] in order to use its set of properties and classes to express the conceptual content of the thesaurus as an RDF graph. The fields examined in the first analytic phase are levels one and two, containing categories and subcategories, and level four, containing the main terms for the description of the artefacts. With reference to level five, we limited our analysis to the functional facet only and we considered the descriptive notes in the attribute fields. Classes and terms were mapped using the *skos:Concept* entity, main terms were mapped as *skos:prefLabel*, non-preferential terms as *skos:altLabel*, notes were encoded using *skos:scopeNote*. The *skos:broader* and *skos:narrower* properties were used to express the hierarchical relationships between categories or concepts. The functional specification of a term was expressed through the *skos:narrower* relation with a subterm obtained by combination with the super concept.

Figure 6 shows an example of the mapping expressed by using SKOS entities. Each concept coming from the RA Thesaurus is represented by a blue circle. The central circle depicts the concept of Cintura (belt) while the red circle represents the thesaurus itself. Arrows connecting the various circles represent the SKOS relationships existing among them. The mapping definition on the SKOS Core Vocabulary was followed by the use of an *ad hoc* script and of a specific tool that allowed the conversion of a huge textual file into RDF format.

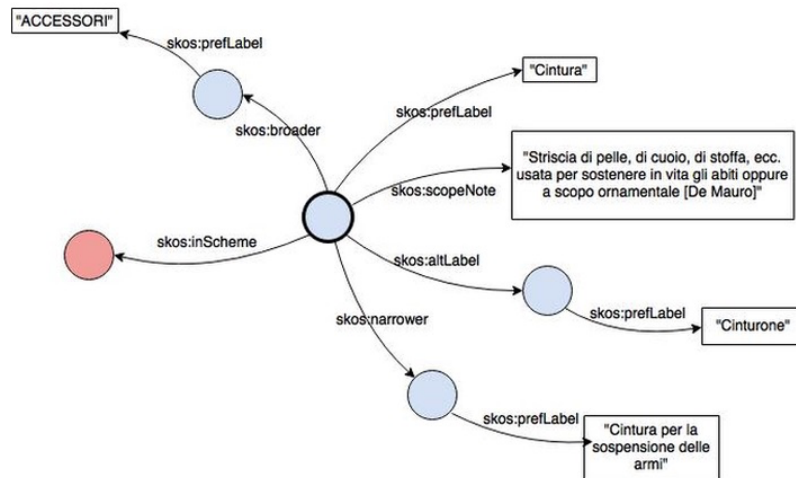


Fig. 6. Example of mapping expressed by using SKOS entities

At first, the original thesaurus was manipulated and converted in order to create a CSV file that satisfied some specific technical requirements. The script was developed in Perl language and was intended to select a specific thesaurus' subset of fields, to sort and to clean the information and to convert them into a custom CSV file. Subsequently, the Stellar Console tool was applied to further elaborate this file. Stellar Console is an open source command line utility application developed in the framework of the AHRC-funded project "Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources" (STELLAR) [7]. The Console accepts input format such as CSV in order to produce a more structured output such as SKOS/RDF or CIDOC-CRM/RDF by applying a set of customizable templates. The templates look for the presence of particular field names in the input data, and process each row in turn using the values contained in these fields. The use of the conversion feature from the custom CSV files to the SKOS/RDF of Stellar Console is the final step for the conversion of the main subset of RA Thesaurus from a textual format to a structured, semantic and interoperable format.

4.2 Thesaurus Publication and Enrichment Using SKOS Extensions

In the second analytic phase, the publication of the thesaurus was analysed and tested on a vocabulary server. Possible solutions for mapping and integrating the fields that were not converted in the first analytic phase were consequently studied and tested.

In order to produce the necessary results for the RA Thesaurus publication, it was important to consider two fundamental aspects. The first was a vocabulary web server supporting international standards such as SKOS and the ISO

thesaurus norms; the second was a vocabulary web application which supports multilingualism, semantic thesauri and data enrichment. All these aspects, in our opinion, are fundamental to make the RA Thesaurus even more flexible for future study phases by expanding and integrating it with further multiple extensions.

We considered different possibilities to achieve the above-mentioned results, by choosing TemaTres as the most pragmatic solution. TemaTres is an open-source, web-based thesaurus management package [8] that supports the handling of vocabularies in accordance with the ISO standard thesaurus norms, including the last ISO-25964 [9]. The main features of TemaTres include a functional user interface for editing and browsing, good search capabilities, and the ability to export all or part of the thesaurus in a number of standardized forms (Json, Json-LD, SKOS Core, DC etc.). TemaTres easily allows data import in SKOS/RDF format and some of the more advanced features include the ability to link terms between two different vocabularies. A test version of TemaTres was installed on a local server and used to import the SKOS/RDF thesaurus version containing the main concepts in order to proceed further with the enrichment work. The TemaTres publication of the RA Thesaurus provides many editing and search facilities. One of the most important is the ability to customize and automatically generate URIs used to unambiguously identify and reach resources from any context. For generating suitable URIs we have used - by means of testing - the official ICCD namespace (<http://www.iccd.beniculturali.it>), which will be useful for the future installation of TemaTres on the ICCD server and for the creation of consistent and unambiguous URI/URL to make the RA Thesaurus available also in a Linked Open Data format. The conversion of the fields related to morphological and partitive specification of terms required further actions on the data. We used the TemaTres administration facilities for this semantic enrichment. We mapped the morphological specifications using *skos:broader* and *skos:narrower* properties. The partitive specification subterms was mapped using the last ISO standard on thesauri ISO 25964[10]. One of the innovations introduced by the current norm is the possibility to make explicit the nature of semantic relationships, in particular we focused on the changes regarding the hierarchical relationships. To extend the richness of thesauri, the SKOS Core hierarchical relationships depicted through the tags BT and NT can be further divided into generic (BTG/NTG), partitive (BTP/NTP) and instancial (NTI/BTI). ISO 25964 specifies that this relationship holds “between a pair of concepts when the scope of one of them falls completely within the scope of the other” [11]. We introduced the BTP and NTP relationships using the corresponding property in the ‘iso-thes’ namespace: *iso-thes:broaderPartitive* and *iso-thes:narrowerPartitive* [12]. The example in Figure 7 shows that a “fibbia di cintura” (belt buckle) concept stated this way, for instance, specifies that the fibbia is part of a “cintura” (belt), whereas a “fibbia” (buckle) per se could also be part of other objects, for instance, a weapon, a garment and so forth. Therefore, the BTP/NTP relationships cannot be automatically inferred by the subconcept only because it could be part of many objects.

The image field of the RA Thesaurus is also a very interesting case. As already mentioned, images increase the richness and meaningfulness of concepts, their presence being sometimes crucial, especially in cases where proper understanding of the archaeological objects may remain ambiguous. In a 2005 version of the SKOS Core Guide W3C Working Draft [13], the Working Group proposed the use of symbolic labels, as part of the labelling properties, to label a concept with an image. Symbolic labels could be used to assign preferred and alternative symbolic labels to a concept by means of the *skos:prefSymbol* and *skos:altSymbol* properties. This solution would have been the most appropriate for the mapping of the RA Thesaurus sample image, but in the subsequent W3C Recommendation [14], symbolic labelling elements were removed, although no explicit deprecation axioms were expressed in the schema. In order to achieve a publication of the thesaurus that complies as far as possible with the W3C specifications, we preferred not to use the solution proposed in the SKOS Core Guide W3C Working Draft, but to use the current W3C Recommendation only. According to the latter, sample images can be regarded as accessorial information of the SKOS concepts. The relationship can be mapped using the *skos:note* property, considering that there is no restriction on the nature of the information that the property can associate with the concept.

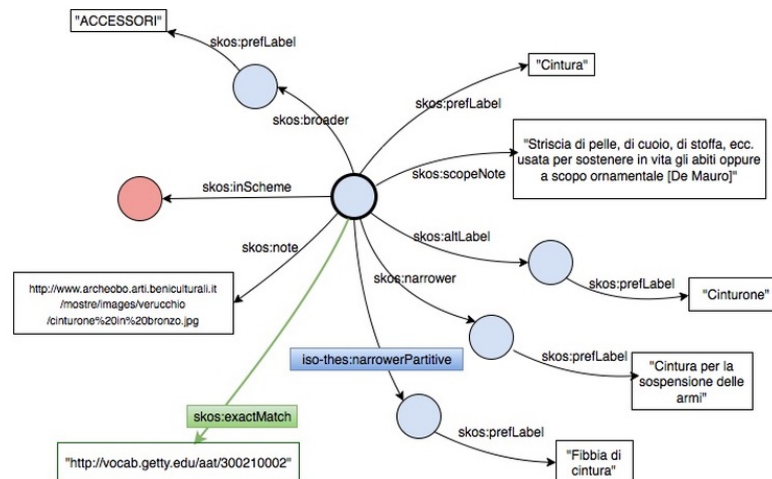


Fig. 7. Example of mapping expressed by using SKOS extensions

5 Getty AAT Mapping

The AAT thesaurus (Art Architecture Thesaurus - Getty Institute) [15] was chosen by ARIADNE to represent a common spine and to constitute a facet

allowing search and faceted browsing across all the terminological tools that the project is collecting. Integration will be based on mappings of national/local vocabularies to the AAT thesaurus. This will allow interoperability over the subject metadata in different partner languages via the common AAT spine. The issue of multilingualism is a matter that needs to be taken into account, not only because of the variety of national thesauri that are going to be integrated by the ARIADNE initiative, but also for the future creation of common and transnational terminological tools. Linguistic issues often make the direct mapping of a concept via the *skos:exactMatch* property on AAT concept difficult, but hopefully the most significant issues will be resolved by the end of the project. The conceptual mapping between the ICCD RA Thesaurus and AAT has been completed and revised; for this purpose it was decided to manually construct a mapping from the various terms and functions (if any), following in sequences the three main categories of the RA Thesaurus. The work pattern was based on an Excel representation of the thesaurus to which additional columns were added in order to specify:

- the *targetLabel* and the unique identifier (ID) of the corresponding definition/term selected in AAT;
- the SKOS schema properties (*skos:closeMatch*; *skos:exactMatch*; *skos:broadMatch* and *skos:matchURI*);
- the name of the institution in charge of the definition of each specific mapping (creator).

Only a subset of the RA Thesaurus was taken into account to demonstrate the feasibility of these operations. The subset includes one thousand, one hundred and ninety one terms related to ten major categories (highlighted in the original source as “livello_1_categoria ”) relating to:

- CLOTHING AND ACCESSORIES
- FURNISHING
- TRANSPORTATION
- CONSTRUCTION INDUSTRY
- PAINTING
- ARCHAEOBOTANICAL FINDINGS
- ARCHAEOZOOLOGICAL FINDINGS
- SCULPTURE
- INSTRUMENTS - TOOLS AND OBJECTS OF USE
- GENERAL TERMS

The analysis for finding the corresponding entries in the AAT thesaurus took into account the information provided by scope notes and images accompanying each concept; extensive web searches were performed to find the most appropriate matching term between Italian and English; and terminological researches was carried out using different resources to identify synonyms to make the associated *targetLabel* as unique and as precise as possible.

The mapping work has identified :

- 457 *broadMatch* associations
- 104 *closeMatch* associations
- 630 *exactMatch* associations

Three examples of association are provided in the following table:

| Categoria | | | | | | |
|--------------------------------------|-----------------|--------------------|---------------------------|---------------------------|-----------|-------------|
| livello1 | livello2 | livello3 | Livello4 termine | targetlabel | ID | matchlabel |
| Mezzi di trasporto | Terrestri | A trazione animale | cisium | two-wheeled carriages | 300215685 | broad match |
| Strumenti - Utensili e Oggetti d'uso | Armi e Armature | Armi da difesa | farsetto da armare | arming doublets | 300226824 | close match |
| Scultura | | | imago clipeata | clipei (portraits) | 300178246 | exact match |

Fig. 8. Example of RA/AAT associations

At the end of the mapping work we can say that the most significant activity, from the scientific-methodological point of view, has been the review of the whole process. Started as punctual control “1: 1” of correspondence between the terms of the two terminology tools (thesaurus ICCD / RA and AAT), this review has, in fact, been expanding by realizing the mapping of the terminological categories relating to individual entries with the codes referring to the facet and the hierarchy AAT. This has made possible:

1. disambiguating and correcting matches previously selected - and often lexically corrected - but decontextualized from the original domain of belonging;
2. providing the basis for a future matching job between different categories of multilingual thesauri.

It is worth underlining that the focus of the whole work of mapping is the concept of individual terms meant as records entered in a complete hierarchical structure of related terms and notes. Among the results which have been achieved - and which are highlighted though the mapping between classes - we can state the high level of correspondence between the ICCD/RA thesaurus entries and the AAT Thesaurus record types. Out of one thousand, one hundred and ninety one basic records one thousand, one hundred and sixty four among them are linked to “concept” and only twenty seven to “guide term”. According to the AAT Thesaurus guidelines:

- Concept: Refers to records in the AAT that represent concepts; records for concepts include terms, a note, and bibliography.
- Guide term: Refers to records that serve as place savers to create a level in the hierarchy under which the AAT can collocate related concepts. Guide terms are not used for indexing or cataloguing.

6 Conclusions and Further work

The study and analysis of the RA Thesaurus allowed us to fully understand the complexity of the challenges arising from the need to define, by means of standard nomenclatures, objects of such various and multifaceted nature as archaeological objects are. The ICCD RA vocabulary, being the result of years of research by a team of experts in the field of Cultural Heritage, is definitely an irreplaceable resource that adequately meets this need. Its structure is certainly an important point of arrival on the road to standardization. From a methodological point of view, the work carried out has highlighted both conceptual and procedural challenges that arise when attempts are made to handle a complex structure in a standard tool. The results achieved so far are considered satisfactory, also in consideration of the fact that the work is at an intermediate stage and that further studies and investigations will be necessary before the conversion of the entire thesaurus can be completed. Future activities will include a clear and unambiguous definition of complex concepts, such as those arising from the combination of multiple terms and subterms; and the definition of precise criteria for the inclusion of images, which, as stated, is one of the distinctive features of this vocabulary. The choice of AAT as the common standard partially solves the multilingualism issues, providing labels in different languages for the terms already mapped. We must instead provide appropriate translations for those that have no equivalent in the thesaurus of the Getty Institute. At the end of the ARIADNE project, the RA Thesaurus will become part of the rich set of terminological tools that the project is already collecting in order to integrate them into the platform on which real interoperability will take place. The ARIADNE Portal will make this resource available and easily accessible online for external use outside of the project. The publication as Linked Open Data, also provided by the project, will guarantee its availability in other Cultural Heritage scenarios.

7 Acknowledgements

The present work has been supported by the ARIADNE project, funded by the European Commission (grant 313193) under the FP7 INFRA-2012-1.1.3 call. The authors opinion do not necessarily reflect those of the European Commission

References

1. Felicetti, A., Scarselli, T., Mancinelli, M.L., Nicolucci, F., (2013) Mapping ICCD Archaeological Data to CIDOC-CRM: the RA Schema, Vladimir Alexiev, Vladimir

- Ivanov, Maurice Grinberg (eds.): Practical Experiences with CIDOC CRM and its Extensions (CRMEX 2013) Workshop, 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), Valetta, Malta, September 26, 2013, CEUR-WS.org/Vol-1117, pp 11-22
2. CRMarchaeo http://www.ics.forth.gr/is1/index_main.php?l=e&c=711
 3. CRMsci http://www.ics.forth.gr/is1/index_main.php?l=e&c=663
 4. Ronzino, P., Amico, N., Felicetti, A., Niccolucci, F., European standards for the documentation of historic buildings and their relationship with CIDOC-CRM, International Conference on Theory and Practice of Digital Libraries (TPDL 2013), 2013.
 5. <http://www.willpowerinfo.co.uk/glossary.htm>
 6. SKOS Core Guide <http://www.w3.org/2004/02/skos/core/guide/>
 7. STELLAR project <http://hypermedia.research.glam.ac.uk/kos/stellar/>
 8. TemaTres website <http://www.vocabularyserver.com/>
 9. ISO 25964 thesaurus schemas. <http://www.niso.org/schemas/iso25964>
 10. International Standard Organization (ISO). Documentation - Thesauri and interoperability with other vocabularies: Part 1, Thesauri for information retrieval, 2011. Report ISO 25964.
 11. Alexiev, V., Isaac, A., On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI), International Journal on Digital Libraries, pp 1-10, 2015.
 12. Cardillo, E., Folino, A., Trunfio, R., Towards the reuse of standardized thesauri into ontologies, Workshop on Ontology and Semantic Web Patterns (WOP2014), co-located with the 13th International Semantic Web Conference (ISWC2014), Riva del Garda, 2014.
 13. SKOS Core Guide, W3C Working Draft 10 May 2005 <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20050510/>
 14. SKOS-Reference <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>
 15. Getty website <http://vocab.getty.edu/http://vocab.getty.edu/>