A Maritime Heritage Thesaurus based on a Greek project documentation case

Eleni Georgaki[1]

[1]*Department of Shipping Trade and Transport, University of the Aegean, Chios, Greece*

Abstract

This paper concerns the creation of a thesaurus in the Maritime Heritage (MH) field. The suggested controlled vocabulary could improve methods for archiving oral sources, written accounts, imagery, 3D archeology, and other multimedia objects related to people and periods from ancient times to recent history. It has been created to cover the indexing needs for resources from Greek Maritime Galleries, Libraries, Archives and Museums (GLAM) institutions hosted in a MH platform of the research infrastructure *Intelligent Research Infrastructure for Shipping, Supply chain, Transport and Logistics* (ENIRISST). Its terms are derived from the collections pertinent to maritime activities. Although the construction of this vocabulary took into consideration specific data and their documentation, the goal is to go beyond ENIRISST to assist MH documentation projects and professionals/researchers in organising and archiving MH data. The long-term ambition is to leave an imprint in the improvement of MH data communication in general, and to contribute in the promotion of MH as a separate Cultural Heritage (CH) branch. The methodological approach was sensitive to the history of the artefacts, the shipping history and terminology, CH and general vocabularies, as well as the pre-existing practices. Terms and subject headings already existing in the vocabularies of reference were leveraged.

Key words
Maritime Heritage, thesauri, visual documentation

1. Introduction

The current paper presents the methodology and the results of the construction of a thesaurus in the Maritime Heritage (MH) related to the creation of a digital Greek maritime heritage platform; building the thesaurus was part of the process and an innovative element as there has been no precedent in the area of Maritime Heritage. The suggested controlled vocabulary could definitely improve methods for archiving oral sources, written accounts, imagery, 3D archeology, and other multimedia objects related to people and periods from ancient times to recent history. It has been created to cover the particular indexing needs for museum and archive resources from Greek Maritime GLAM institutions hosted in a MH platform of the research infrastructure EN.I.R.I.S.S.T. It is comprised of terms 'derived from' the collections which include scans of models and designs that are related to emblematic categories /types of ships-boats historically, digital files of sailing boats images related to emblematic categories of sailing boats, scans of navigation instruments and related items, other museum objects of maritime museums and organizations and material related to the history of shipping.

ISKO UK conference, July 24–25, 2023, Glasgow, Scotland
EMAIL: egeorgaki@aegean.gr
ORCID: 0000-0001-8996-4807

The research is primarily focused on the construction of a thesaurus which describes topical subjects visually represented in MH documents, photographs, and artifacts, but also caters for the completion of other Dublin Core fields in the cataloguing process, such as the 'Coverage' and the 'Type' fields. It can be regarded as a 'shell' and is definitely not a complete work since the addition of new digitized and interconnected items will lead to a subsequent expansion of the thesaurus with this process estimated to be continuous. The procedure followed for the construction of a strictly controlled vocabulary was the use of a streamlined multi-tiered hierarchical arrangement and the placement of specific concepts within the hierarchical structure of terminology that function as key access points to MH collections. The basis of development of this project is the documentation of the digitilised cultural and archival material provided by Greek maritime institutions with the provided data falling into the previously mentioned data categories of the platform.

The typology of the objects includes objects/heirlooms that incorporates personal items, collection of scrimshaws collection of archaeological items, archives of shipping companies, personal and official correspondence, periodicals, photographic material from coaling stations and lighthouses [3].

2. Maritime Heritage Thesaurus (MHThesaurus)

2.1. Purpose

Although the construction of the MHThesaurus took into consideration very specific data and their documentation, the intention was not to establish a static, country-specific, vocabulary. Beyond EN.I.R.I.S.S.T., in the long run it may also aid projects of documentation of MH or the professionals in the maritime institutions in their daily work or simply be a starting point for researchers interested in organizing and archiving cultural data related to MH. While the ambition is not all projects, digital or otherwise, to adapt this vocabulary, but a goal is, with the adoption of this vocabulary, to encourage reflection on how data are described and organized.

2.2. Goal

The development of this thesaurus and an associated hierarchical arrangement of its component terms were primarily to aid the process of describing the subject content of MH data within the collection database of the platform, but above all, to have an imprint on its longevity and communication as cultural heritage and to contribute to establishing MH as a separate Cultural Heritage (CH) branch. Apart from accessibility for the researcher and the wider audience, it is also estimated that it will have immediate benefits to the cataloguing staff of the maritime museums or related research projects in the future.

2.3. Method

Developing a MH controlled vocabulary requires a methodological approach that must be sensitive to the history of the artefacts, the shipping history and terminology, cultural heritage vocabularies and general vocabularies, as well as the pre-existing practices. The procedure involved two stages: first the monolingual thesaurus was produced, and then the translation of the reduced thesaurus was carried out, sometimes, taking into account terms and subject headings already existing in the vocabularies of reference.

For the construction of a strictly controlled vocabulary the use of a streamlined multi-tiered hierarchical arrangement and the placement of specific concepts within the hierarchical structure of terminology that function as key access points to MH collections was followed as a procedure. The basis of development of this project is the documentation of the digitized cultural and archival material provided by the maritime institutions falling into the previously mentioned typology and samples of it can be seen in Figures 1 and 2. The specifications of the corresponding ELOT (the Background of Standardization in Greece) and ISO standards were followed as well as the international developments in the matters of construction and management of treasures, but also the Greek particularities in the issues of language and terminology. For instance, the National Library of Greece subject headings, as well as definitions from the Centre for the Greek Language in order to guide the cataloguer or the user in the ‘Notes’ section were adopted or seriously taken into consideration.

Figure 1*:* Sample of the digitilised material provided from the Piraeus Bank Group Cultural Foundation- Sourced from [4]

Figure 2: https://enirisst-plus.di.ionio.gr/pages/116.html

To create the thesaurus, the rules and standards of thematic indexing, creation of monolingual and multilingual thesauruses, as well as terminology standards are followed. These standards are the following:

• ELOT 1312: Documentation - Thematic analysis of documents: Examination methods documents, identifying their subjects and choosing the indexing terms (ISO 5963).

• ELOT 1321: Documentation-guidelines for construction and development of Monolingual Thesauri, 1993 (ISO 2788).

• ISO 5964: Documentation-guidelines for the establishment and development of multilingual thesauri, 1985 revised by ISO 25964-1: 2011/ ISO 25964-2:2013.

• ELOT 402: Terminological work – Principles and methods (ISO 704).

• ELOT 561-1 & 2: Terminological work - Vocabulary (ISO 1087-1 & 2).

More specifically these standards and rules, along with other relevant textbooks and guidelines, such as [2] and [5] were consulted before the construction of the thesaurus began in order to decide on and establish the appropriate work mode. They were also revisited during the construction process to search for examples and check equivalencies.

This work focuses on identifying the top-level-concepts (facets and hierarchies) that will become its common basis, meeting the demands for subjective and disciplinary validity, but also taking into consideration wider CH concepts. So, top-level concepts were developed by adequate abstraction from existing controlled vocabularies. The methodology followed required decisions to be made whether generic concepts subsume efficiently enough narrower terms from different thesauri and to determine whether concepts are comprehensibly enough defined in order to allow experts and cataloguers to assign the metadata of their exhibits to them.

Following this methodology, an initial set of top-level concepts was primarily designed based on other previously mentioned vocabularies but also creating a first operational draft. Thesauri cooperation and terminology integration are highly intended to ensure reliability of the endeavor and ensuring that previous cataloguing practice was integrated at least as far as the first experimental set of data is concerned, and when put in practice new entries and finer ramifications may apply.

Four facets along with their hierarchies, top terms and narrower terms' examples have been defined thus far. Facets are the most general concepts, or else, the broad categories to which terms belong to, and inherit their properties to all possible narrower terms they include. So, they are further subdivided into an open number of hierarchies (expressed by the hierarchy top terms) and at the same time they exhibit at least one specific feature which is characteristic of a certain type of terms within this hierarchy since they correspond to a broader term; thus their function, as 'structural metadata' [1] is two-fold.

The material at the disposal of the author indicated the need for further distinction and specification, the hierarchies were extended by narrower terms of classification for addressing more specific classes of categorization according to its 'guidance'. The updating and revising of the proposed classification and definitions is an ongoing process.

2.3.1. Challenges and the selection of sources

There was no established/authorised way of cataloguing from the institutions which provided the artefacts for digitisation in order to compare or base the research; this was a challenge which was addressed through leveraging other aids for the construction of the thesaurus vocabulary. The vocabulary developed is subject-specific and based on the Getty Foundation's Art & Architecture Thesaurus - AAT and Wikipedia's DBpedia semantic thesaurus, the established National Library of Greece (NLG) terms, Thesaurus for Graphic Materials (TGM) (Library of Congress), Library of Congress Subject Headings (LCSH). Excerpts from lists of thematic headings (https://southampton.spydus.co.uk/ , https://biblionet.gr/, mobius.mysticseaport.org) which are considered to contribute were also studied and adopted, or adapted accordingly, in some cases, and terminology reported in literature was also incorporated.

The current proposed version of the vocabulary was based to a large extent on the UNESCO thesaurus and its Greek adaptation through a bilingual vocabulary of thematic tags which is complementary to the UNESCO Thesaurus carried out by the National Documentation Centre (NDC). The choice is not at all arbitrary, but it serves interoperability without sacrificing precision with regard to the particular thematic area. There were cases where a decision should be reached when multiple controlled vocabularies were consulted. For instance, the term ‘transportation’ exists in the subject headings of the NLG and in TGM, but ‘transport’ was selected in order to align with UNESCO. Respectively, priority was given to the AAT over Wikidata, for example, for the same reason, and also for the granted validity these thesauri already have.

Also, research in specific trade unions sources was conducted in order to effectively grasp the occupational activities and reach specific terminology enriching related terms and describing holistically terms that should be described (e.g. Metal industry). The same objective was intended when combining terms from multiple sources, including vocabularies, thesauri, wikidata to build the terms closer to their thematic essence and building this MH controlled vocabulary based on a representative database of the field.

2.3.2. Validation

Experts, cataloguers from institutions included in the EN.I.R.I.S.S.T. project have been asked to review one or more hierarchies of their choice and comment on their structure, content, or translation.

Also, students from the department of Archives, Library science and Museology of the Ionian University were asked to evaluate the thesaurus. A questionnaire has been sent for completion in both cases and was used as a measurement instrument. The questionnaire was divided in three sections. The first section examines the competence of the introduction with criteria outlined in a measuring instrument in [9], the second one reviewed compliance with the thesaurus building criteria as derived from a thorough literature review in [3], and especially described in [6] , and the third examines particular terms, hierarchies and facets of the first operational draft of the thesaurus testing structure, content, translation and related elements, as appearing when the user browses the thesaurus online, or the Maritime Heritage platform of the EN.I.R.I.S.S.T. infrastructure selecting to view the metadata of items included in its collections.

2.3.3. Presentation of the terms

The Thesaurus is built up of descriptors and identifiers grouped into facets representing subdivisions of broad fields. As shown in Table 1 each main term indicates:

Table 1
Presentation of terms 1

Broader terms (BT) – Broader hierarchical relationship.
Narrower terms (NT) – Narrower hierarchical relationship.
Related terms (RT) – Associative relationship (non-hierarchical).
Subject category (MT) – The facet number and heading.
Historical note (HN) – The date of creation or history of the term’s use.

It may also indicate (see Table 2):

Table 2
Presentation of terms 2

Scope note (SN) – Scope note explaining the meaning or usage of the term. Where the note is taken from another thesaurus, the source is given.
Used for (UF) – Non-preferred terms or synonyms.
Use – Linking the non-preferred term to the preferred one.

The above-mentioned indicators are given in Table 3 Greek language as well in accordance with their appearance in the Thesaurus management system (THEMAS) as shown in Figures 2, 3, 4, 5:

Table 3

Greek presentation of terms

Όρος Κορυφής (ΟΚ)- Top term
Μετ.- Translation
ΔΣ- Scope Note
ΔΣ (Μετ.)- Scope Note (Translation)
ΠΟ- Broader term
ΕΟ- Narrower term
ΣΟ- Related terms
ΧΑ- Used for (UF)
ΧΑ (Μετ.)- UF (Translation)
Πηγή -Source

The vocabulary developed is subject-specific and based on pre-existing controlled vocabularies in the CH domain or general, which include terms of the particular field. Excerpts from lists of thematic headings which are considered to contribute were also studied and adopted, or adapted accordingly, in some cases, as well as incorporation of terminology reported in literature.

Following this methodology, an initial set of top-level concepts was primarily designed based on other previously mentioned vocabularies but also creating a first operational draft. Thesauri cooperation and terminology integration are highly intended to ensure reliability of the endeavor and ensuring that previous cataloguing practice was integrated at least as far as the first experimental set of data is concerned, and when put in practice new entries and finer nuances across entries may be inserted or apply.

The research is primarily a practical design of a list of subject headings which describes topical subjects visually represented in MH documents, photographs and artifacts, but also has predicted the completion of other Dublin Core fields in the cataloguing process, such as the Coverage and the Format.

The list created is not restricted to the description of the CH area covered, i.e. Maritime Heritage although this remains predominant. This was not originally intended but it came up as a need because of the existence of unique in the MH field cultural objects, such as scrimshaws. As such, the list does not replace the role of other fields that need to be catalogued such as: type, format, coverage, historical period, creator, contributor, title, or language but is elaborated to supplement existing vocabularies, or simply assist the process of documenting the previously mentioned fields. The controlled vocabulary is restricted as much as possible to visually discernable subjects, and avoids the use of concepts and subjects related to the context of the visual item/photograph. The suggested hierarchical structure minimizes ambiguous or synonymous concepts. Four (4) facets and eleven (11) hierarchies have been singled out as access points to MH collections and are placed in the first tier of the list to successfully narrow down the scope of searching. As shown in Figure 3, there are the following facets: Concepts, Material objects, Activities, Geopolitical units, and descriptors/identifiers: Economy, Science, Culture, Information and Communication, Item types, Countries and country groupings. The "orphans" facet and hierarchy include terms that have been linked to them descriptors with an association relationship but not yet organized hierarchically.

Αποτελέσματα Αναζήτησης Θεμάτων	**Κριτήρια Αναζήτησης Θεμάτων**

Στατιστικά: βρέθηκαν 5 αποτελέσματα. **Αποτελέσματα:** 1 - 5 **σελ.:** 1 / 1

Θέμα	**Ιεραρχίες**	**Ενέργειες**
Γεωπολιτικές ενότητες (Geopolitical units)	Χώρες και ομαδοποίηση χωρών	
Δραστηριότητες (Activities)	Υπαίθριες δραστηριότητες	
Έννοιες (Concepts)	Δίκαιο, Επιστήμη, Οικονομία, Πληροφορία και επικοινωνία, Πολιτισμός, Τύποι τεκμηρίων	
ΟΡΦΑΝΟΙ ΟΡΟΙ	Ορφανοί όροι	
Υλικά αντικείμενα (Material objects)	Προσωπικά αντικείμενα, Φορτίο	

Figure 3: Facet search analysis

Translation Note

Facet	**Hierarchies**
Geopolitical units	Countries and country groupings
Activities	Outdoor life
Concepts	Law, Science, Economy Information and communication, Culture. Item types
Orphan terms	Orphan terms
Material objects	Personal accessories Freight

As regards the construction of the facets, the process of their building was inspired by the Art and Architecture Thesaurus Facet code and the BackBone Thesaurus, a meta-thesaurus, which proposes a common model of thesaurus building.

As mentioned, this thesaurus is designed to be used at the MH platform ΝΑΥΚΛΗΡΟΣ of the research infrastructure EN.I.R.R.I.S.T., but definitely not limited. It is comprised of subject headings derived from the collections of SHIPMARK which includes scans of models and designs that are related to emblematic categories / types of ships-boats historically, SHIPSAIL which includes data digital files of sailing boats images related to emblematic categories of sailing boats, SHIPEQUIP which includes scans of navigation instruments and related items, RELATED MUSEUM ITEMS, which concerns other museum objects of maritime museums and organizations and SHIPSTORIES with other material related to the history of shipping (testimonies, etc.).

It can be regarded as a 'shell' and definitely not a complete work since the addition of new digitized and interconnected items will lead to a subsequent expansion of the thesaurus and this process is estimated to be continuous. So, the development of this thesaurus and an associated hierarchical arrangement of its component terms was primarily to aid the process of describing the subject content of MH data within the collection database of the platform, but above all, to have an imprint on its longevity and communication as cultural heritage and to contribute to establishing MH as a separate Cultural Heritage branch. Apart from accessibility for the researcher and the wider audience, it is also estimated that it will have immediate benefits to the cataloguing staff of the maritime museums or related research projects in the future.

For the construction of a strictly controlled vocabulary the use of a streamlined multi-tiered hierarchical arrangement and the placement of specific concepts within the hierarchical structure of terminology that function as key access points to MH collections was followed as a procedure. The basis of development of this project is the documentation of the digitised and digitilised cultural and archival material provided by the maritime institutions-providers. These are: The Hellenic Maritime Museum, the Maritime Museum of Oinousses, the Naval History Service, the Piraeus Bank Group Cultural Foundation, the Aikaterini Laskaridis Foundation, the Photography Archive of the municipality of Kea Island of Cyclades. These data fall into the previously mentioned data categories of the platform.

The typology of the objects includes objects/heirlooms that incorporates personal items (Hellenic Maritime Museum), collection of scrimshaws (Hellenic Maritime Museum, Laskaridis Foundation, the Maritime museum of Oinousses), collection of archaeological items (the Maritime museum of Oinousses), archives of shipping companies (the Piraeus Bank Group Cultural Foundation), personal and official correspondence, periodicals (the Naval History Service), photographic material form coaling stations and lighthouses (the Photography Archive of the municipality of Kea island of Cyclades).

2.4. Synthesis

The construction of the Maritime Heritage Thesaurus (MHThesaurus) has been bottom-up, starting with more than one language simultaneously, and non-symmetrical, since it is also possible not to have the same number of descriptors in each language.

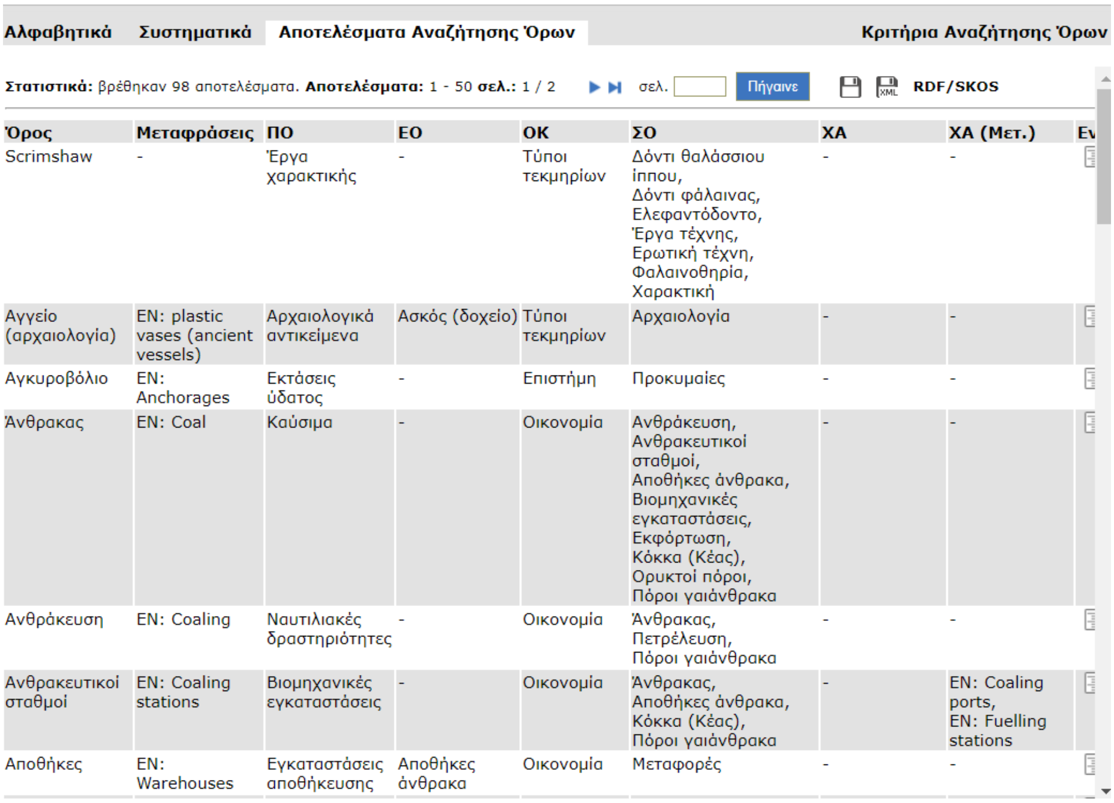
Single descriptors can be combined to express compound concepts since post-coordination was followed (in contrast to subject heading building procedures). Nevertheless, pre-combinations, such as adding qualifiers, to provide clarity, or handle translation issues have been applied.

2.5. Searches

The thesaurus was implemented with the Web-TMS (Thesaurus Management System) software supported by the Cultural Informatics research group of the Institute of Computer Science – FORTH, University of Crete. This system works in a graphical internet environment and provides for the construction and management of bilingual multi-thematic thesauruses, in this case with Greek as the dominant language and English as the reference language.

It follows the specifications of the corresponding ELOT and ISO standards and the international developments in the matters of construction and management of thesauri, but definitely also the Greek peculiarities in matters of language and terminology.

For the terms search quick search and search by selecting criteria is provided. Users have the ability to search for a term in general as well as specifically by general field or combination of fields using logical operands (see Figure 4).



Αλφαβητικά | Συστηματικά | Αποτελέσματα Αναζήτησης Όρων | Κριτήρια Αναζήτησης Όρων

Στατιστικά: βρέθηκαν 98 αποτελέσματα. **Αποτελέσματα:** 1 - 50 **σελ.:** 1 / 2 σελ. Πήγαινε RDF/SKOS

Όρος	Μεταφράσεις	ΠΟ	ΕΟ	ΟΚ	ΣΟ	ΧΑ	ΧΑ (Μετ.)	Εν
Scrimshaw	-	Έργα χαρακτικής	-	Τύποι τεκμηρίων	Δόντι θαλάσσιου ίππου, Δόντι φάλαινας, Ελεφαντόδοντο, Έργα τέχνης, Ερωτική τέχνη, Φαλαινοθηρία, Χαρακτική	-	-	
Αγγείο (αρχαιολογία)	EN: plastic vases (ancient vessels)	Αρχαιολογικά αντικείμενα	Ασκός (δοχείο)	Τύποι τεκμηρίων	Αρχαιολογία	-	-	
Αγκυροβόλιο	EN: Anchorages	Εκτάσεις ύδατος	-	Επιστήμη	Προκυμαίες	-	-	
Άνθρακας	EN: Coal	Καύσιμα	-	Οικονομία	Ανθράκευση, Ανθρακευτικοί σταθμοί, Αποθήκες άνθρακα, Βιομηχανικές εγκαταστάσεις, Εκφόρτωση, Κόκκα (Κέας), Ορυκτοί πόροι, Πόροι γαιάνθρακα	-	-	
Ανθράκευση	EN: Coaling	Ναυτιλιακές δραστηριότητες	-	Οικονομία	Άνθρακας, Πετρέλευση, Πόροι γαιάνθρακα	-	-	
Ανθρακευτικοί σταθμοί	EN: Coaling stations	Βιομηχανικές εγκαταστάσεις	-	Οικονομία	Άνθρακας, Αποθήκες άνθρακα, Κόκκα (Κέας), Πόροι γαιάνθρακα	-	EN: Coaling ports, EN: Fuelling stations	
Αποθήκες	EN: Warehouses	Εγκαταστάσεις αποθήκευσης	Αποθήκες άνθρακα	Οικονομία	Μεταφορές	-	-	

Figure 4: Term search results

From the results screen, users can view the results, save them to a file or print them, or, by clicking on a term, view that term's card (Figure 5).

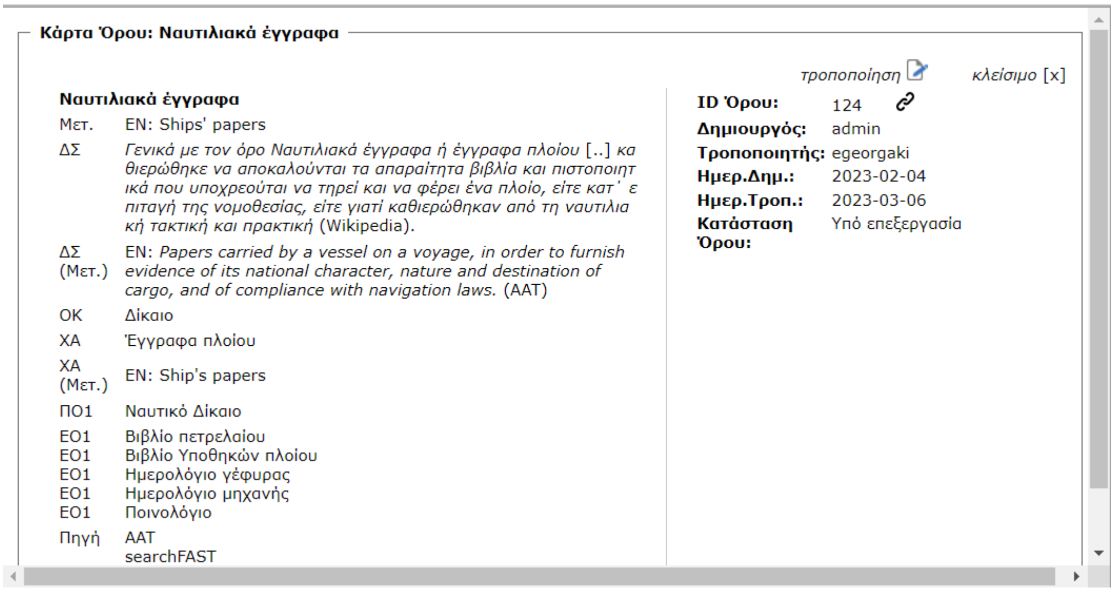


Figure 5: Term card

In the present application, alphabetical (Figure 4), hierarchical (Figure 5) and graphical (Figure 6) presentation of thesaurus terms, is possible. When viewing terms, users can also jump to one or more term positions at the hierarchies to which it belongs and return to the top of the page.

2.5.1. Alphabetical presentation

In the alphabetical presentation (Figure 6), all data relating to a descriptor are displayed as well as cross-references from non-descriptors (non-preferred terms) to descriptors (preferred terms). Details of the displayed data can be found in the legend.

Αλφαβητικά | **Συστηματικά** | **Αποτελέσματα Αναζήτησης Όρων** | **Κριτήρια Αναζήτησης Όρων**

ΣΟ Φαλαινοθηρία

Άνθρακας

Μετ.	EN: Coal
ΟΚ	Οικονομία
ΠΟ1	Καύσιμα
ΣΟ	Ανθράκευση Ανθρακευτικοί σταθμοί Αποθήκες άνθρακα Βιομηχανικές εγκαταστάσεις Εκφόρτωση Κόκκα (Κέας) Ορυκτοί πόροι Πόροι γαιάνθρακα

Ανθράκευση

Μετ.	EN: Coaling
ΔΣ	η προμήθεια, ο εφοδιασμός πλοίων ή ατμομηχανών με άνθρακα, που τον χρησιμοποιούν για την κίνησή τους [Λεξικό Τριανταφυλλίδη, όπως αναφέρεται στο www.greeklanguage.gr]
ΟΚ	Οικονομία
ΠΟ1	Ναυτιλιακές δραστηριότητες
ΣΟ	Άνθρακας Πετρέλευση Πόροι γαιάνθρακα

Ανθρακευτικοί σταθμοί

Μετ.	EN: Coaling stations
ΟΚ	Οικονομία
ΧΑ (Μετ.)	EN: Coaling ports EN: Fuelling stations
ΠΟ1	Βιομηχανικές εγκαταστάσεις
ΠΟ2	Εξοπλισμός και εγκαταστάσεις
ΣΟ	Άνθρακας Αποθήκες άνθρακα Κόκκα (Κέας)

Figure 6: Alphabetical presentation of terms

2.5.2. Hierarchical presentation

In hierarchical presentation, the position of a term in the hierarchy or hierarchies to which it belongs is shown. The different levels of belonging to the hierarchy are represented by the symbol "--" (Figure 7). As mentioned, this presentation can also be saved and printed.

Κριτήρια αναζήτησης:
Ιεραρχική παρουσίαση όρου: **Άνθρακας**.

Αποθήκευση αρχείου ως.. *Εκτύπωση αρχείου..*

Βρέθηκαν οι παρακάτω αναφορές:
1 αναφορά στην ιεραρχία: Οικονομία. *Μετάβαση στην αναφορά:* *1*

Οικονομία
```
--  Βιομηχανία
--  --  Βαριά βιομηχανία
--  --  --  Βιομηχανία μετάλλου
--  --  Ναυπηγοεπισκευαστική βιομηχανία
--  --  Παραγωγή
--  Γεωργία
--  --  Αλιεία
--  Δραστηριότητες
--  --  Εκφόρτωση
--  --  Φαλαινοθηρία
--  Εξοπλισμός και εγκαταστάσεις
--  --  Βιομηχανικές εγκαταστάσεις
--  --  --  Ανθρακευτικοί σταθμοί
--  --  --  Ναυπηγεία
--  --  Εγκαταστάσεις αποθήκευσης
--  --  --  Αποθήκες
--  --  --  --  Αποθήκες άνθρακα
--  --  Οχήματα
--  --  --  Πλοία
```

Figure 7: Hierarchical presentation of terms

2.5.3. Graphical presentation

The graphical presentation shows the entire tree of a hierarchy, including the terms and their position in the hierarchy, as well as the corresponding English. Different colors help to represent the relationships within it (Figure 8).

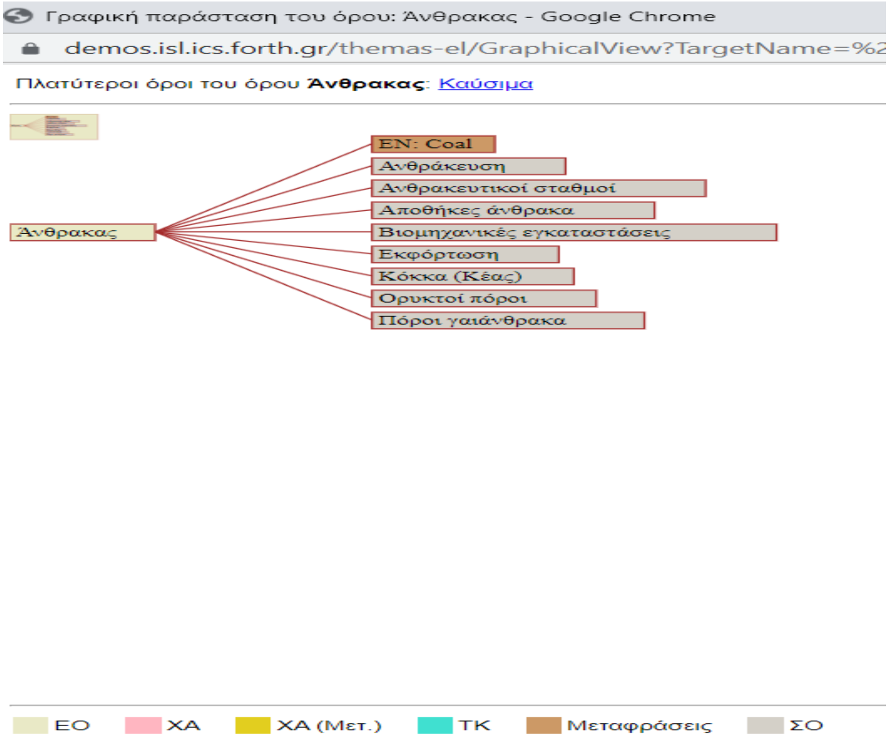


Figure 8: Graphical representation

2.6. Specific examples

As previously mentioned, the research is primarily a practical design of a thesaurus which describes topical subjects visually represented in MH data, but also has predicted the completion of other Dublin Core fields in the cataloguing process, such as the ‘Coverage’ and the ‘Type’ fields. In Figure 9 below there is an example of the usage of the described MHThesaurus in order to document the topic/subject of the related museum artefact as well as resource type. The term Αρχαιολογία/ Archaeology has been used to describe the thematic category this artefact falls in, whereas the term Αγγείο (Αρχαιολογία)/ plastic vases (ancient vessels) is selected from the MHThesaurus to describe the ‘Type’ Dublin Core field. Interoperability is also served in this particular example since the choice about the Greek term was made identical with the Wikipedia term because a qualifier was used to clarify the essence of the noun in Greek contrary to their homonyms in other fields. The English term is selected according to the AAT Getty Vocabulary; a qualifier also exists and the rationale for its adoption is the same. As the author in [2] supports *the usual method of removing ambiguities caused by homographs is to add qualifiers (printed within parentheses) after the terms, to distinguish the two or more different meanings,* and in this case the qualifier becomes an integral part of the term. The ‘Notes’ field, in the above-mentioned example, is enriched with the respective definitions derived from the respective vocabularies which are referenced. Hence, the user is able to reach the terms from the external vocabularies. Also, the possibility of using them as same or related terms providing the URIs is given through the SKOS format. The term card can be seen at Figure 10 and the thesaurus can be found at https://demos.isl.ics.forth.gr/themas-el/Index?logout=true&lang=en with username MHTanagnostis and MHTanagnosi as a password and the link to the platform is https://enirisst-plus.di.ionio.gr/en/museums with the respective documentation.

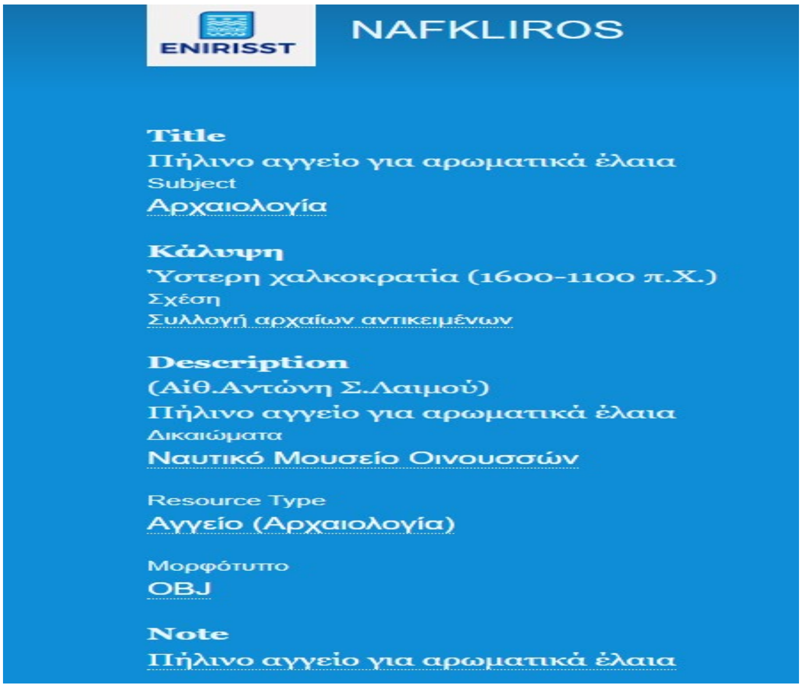


Figure 9: Documentation example with the MHThesaurus https://enirisst-plus.di.ionio.gr/el/node/89

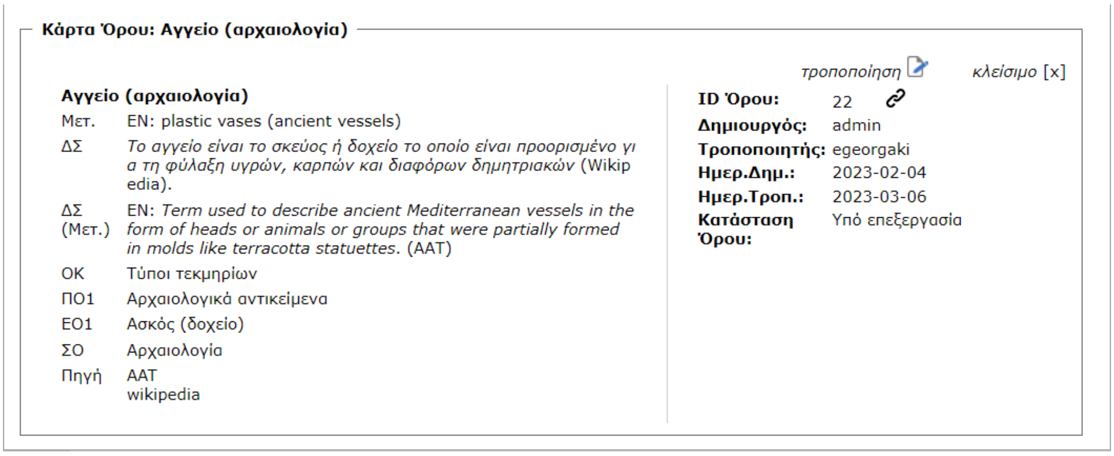


Figure 10: Example of term card

In Figure 11 below, the Subject field is filled with the MHThesaurus terms ‘Coaling stations’, ‘Barges’, ‘Warehouses’, and ‘Coal’. The current proposed version of the vocabulary was based at a large extent on the UNESCO thesaurus and its Greek adaptation that is a bilingual vocabulary of thematic tags that is complementary to the UNESCO Thesaurus carried out by the National Documentation Centre (NDC) in Greece. The choice is not at all arbitrary, but it serves interoperability without sacrificing precision in the particular thematic area. There were cases where a decision should be reached when multiple controlled vocabularies were consulted. For example, priority was given to the AAT over Wikidata, for this particular reason, and also for the granted validity this thesaurus already has. An example of this in the current vocabulary is ‘coaling station’ instead of ‘fuelling station’ in Wikidata, as demonstrated in Figure 12. Interoperability is also served by providing the AAT URI for skos:exactmatch or skos:closeMatch properties.

Figure 11: https://enirisst-plus.di.ionio.gr/el/node/68

Κάρτα Όρου: Ανθρακευτικοί σταθμοί

Ανθρακευτικοί σταθμοί

Μετ.	EN: Coaling stations
ΔΣ	Ίδιο με http://vocab.getty.edu/page/aat/300263399
ΔΣ (Μετ.)	EN: Structures used for storing and loading coal onto vehicles (AAT).
ΟΚ	Οικονομία
ΧΑ (Μετ.)	EN: Coaling ports EN: Fuelling stations
ΠΟ1	Βιομηχανικές εγκαταστάσεις
ΣΟ	Άνθρακας Αποθήκες άνθρακα Κόκκα (Κέας) Πόροι γαιάνθρακα
Πηγή	AAT

Translation note:

ΔΣ Ίδιο με	Scope note Same as
ΟΚ Οικονομία	Top Term Economy
ΠΟ1 Βιομηχανικές εγκαταστάσεις	Broader term1 Industrial facilities
ΧΑ (Μετ.)	UF (Translation)
ΣΟ Άνθρακας Αποθήκες άνθρακα Κόκκα (Κέας) Πόροι γαιάνθρακα	Related terms Coal Coal warehouses Kokka (Kea) Coal resources
Πηγή AAT	Source AAT

Figure 12: 'Coaling stations' term card

An important note to be made is that the non-preferred terms or synonyms may be given in most cases in one of the two languages in the MHThesaurus. The reasoning of this can be easily comprehended and the expected procedure is twofold. On the one hand, because of the fact that this a multilingual vocabulary and its building was almost concurrent in both languages since the Greek vocabularies were based on the UNESCO thesaurus and, on the other hand, because the English terms in some cases prevailed as the already existing general controlled vocabularies were in English (UNESCO, LCSH). This affected the choices of preferred terms, but Greek terms were proposed in the related terms to allow disambiguation and assist Greek experts and researchers. Of great importance towards this direction also were the Scope Notes (SN). In other cases the bilingual established pair from the National Documentation Center (NDC in Greek EKT) of Greece prevailed if estimated that served understanding of the term better. For instance, "Παραγωγή"-"Manufacturing Industries" where the plural number in the English term enables the enrichment of the related terms with multiple subsectors (see Figure 13).

The connection of the generated terms of the proposed vocabulary to the related terms is given providing the URIs in most of the terms indicating either relevance or loan from another vocabulary/thesaurus.

Another important note to be made was the mismatch between the Greek and the English singular and plural number of terms which has arisen in the same way opting for established terminologies (eg. Όρμος- Bays, Θάλασσα-Seas) (see Figures 14 and 15).

In addition, not every term is bilingual at the current phase. This is because there is neither equivalency in both languages nor established translations, as in the term card "Scrimshaw" demonstrated in Figure 16. This obstacle is expected to be overcome after feedback is given from the experts and the professionals of the institutions involved in the project.

Κάρτα Όρου: Παραγωγή

Παραγωγή

Μετ. EN: Manufacturing Industries

ΟΚ Οικονομία

ΠΟ1 Βιομηχανία

Translation note:

Μετ.	Translation: Manufacturing industries
ΟΚ Οικονομία	Top term Economy
ΠΟ1 Βιομηχανία	Broader term1 Industry

Figure 13: 'Manufacturing industries' term card

- **Κάρτα Όρου: Όρμος**

Όρμος

Μετ.	EN: Bays
ΔΣ	Κόλπος, οποιαδήποτε εσοχή ή διείσδυση της θάλασσας στη ξηρά
ΔΣ (Μετ.)	EN: Body of water connected to an ocean or lake, formed by an indentation of the shoreline
ΟΚ	Επιστήμη
ΠΟ1	Εκτάσεις ύδατος
ΣΟ	Γεωμορφολογία Λιμάνια Προκυμαίες Φάροι

Translation note:

Μετ.	Translation: Bays
ΔΣ	Scope note
ΔΣ (Μετ.)	Scope note (Translation)
ΟΚ Επιστήμη	Top term Science
ΠΟ1 Έκτασεις ύδατος	Bodies of water
ΣΟ Γεωμορφολογία Λιμάνια Προκυμαίες Φάροι	Geomorphology Harbours Piers & wharves Lighthouses

Figure 14: 'Bays' term card

– **Κάρτα Όρου: Θάλασσα**

Θάλασσα

Μετ.	EN: Seas
ΟΚ	Επιστήμη
ΠΟ1	Εκτάσεις ύδατος
ΣΟ	Ενάλια αρχαιολογία Νερό

Translation note:

Μετ.	Translation: Seas
ΟΚ Επιστήμη	Top term Science
ΠΟ1 Έκτασεις ύδατος	Bodies of water
ΣΟ Ενάλια αρχαιολογία Νερό	Underwater archaeology Water

Figure 15: 'Seas' term card

Κάρτα Όρου: Scrimshaw

Scrimshaw

ΔΣ Objects, whether decorative or household items, carved from w
hale ivory, baleen, or similar materials from marine mammals.
Originally specifically those objects made by whalers from whal
e teeth or baleen while on board ship. (AAT)
Engravings and carvings done in bone or ivory, created by sailor
s (Wikidata)

ΟΚ Τύποι τεκμηρίων

ΠΟ1 Έργα χαρακτικής

ΣΟ Δόντι θαλάσσιου ίππου
Δόντι φάλαινας
Ελεφαντόδοντο
Έργα τέχνης
Ερωτική τέχνη
Φαλαινοθηρία
Χαρακτική

Πηγή AAT
wikipedia

Translation note:

ΟΚ Τύποι τεκμηρίων	Top term Item types
ΠΟ1 Έργα χαρακτικής	Broader Term 1 Engravings
ΣΟ Δόντι θαλάσσιου ίππου Δόντι φάλαινας Ελεφαντόδοντο Έργα τέχνης Ερωτική τέχνη Φαλαινοθηρία Χαρακτική	Walrus ivory Whale ivory Ivory Works of art Erotica Whaling Engraving process
Πηγή	Source

Figure 16: 'Scrimshaw' term card

2.7. Future steps in the development of the MH Thesaurus

The population of the thesaurus with terms from maritime legislation documents is currently designed and in progress. Authors in [7] developed a deep learning technique for text mining and discovering knowledge from textual data, focusing on legal texts for extracting concepts from the broader domain of maritime heritage. The results of their work, approximately 4000 words, extracted from an initial corpus of 80.000 included in legal texts on maritime matters in the Government Gazette, between the years 1975-1999, will be incorporated in the MHThesaurus as a further development step.

Nevertheless, the enrichment of the thesaurus followed after the evaluation process took place when new terms were proposed by Greek experts, cataloguers and professionals in the Maritime Heritage field, and added as preferred or non-preferred terms, RT, BT or NT is an endless pathway. In-house vocabularies- no matter how restricted they are- are always valuable sources of terms because they are material-oriented. The same applies with related glossaries/controlled vocabularies work which can always be input to update later versions of the thesaurus. Also, new Maritime Heritage institutions can be approached, to provide their feedback perhaps related to new kinds of material not only in Greece, but also abroad-an element missing from the first evaluation process. Last but not least, sub-thematic population of the MHThesaurus from more specialized vocabularies of the field (e.g. Maritime law, maritime fixtures etc) would be steps in that direction. As regards the connection to external resources, in future development steps, other skos labels such as skos:closeMatch, skos:broadMatch, skos:narrowMatch could also be taken advantage of for the purpose of data interoperability and integration [3].

3. Acknowledgements

This research was co-financed by the European Union and Greek national funds through the project title: “Intelligent Research Infrastructure for Shipping, Transport and Supply Chain -ENIRISST”.

4. References

[1] ANSI/NISO Z39.19-2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies ISBN: 1-880124-65-3 URL: https://ils.unc.edu/courses/2015_fall/inls151_002/Readings/NISO.pdf
[2] J. Aitchison, D. Bawden, A. Gilchrist, Thesaurus construction and use: a practical manual, 4th ed. Aslib, London, 2000.
[3] E. Georgaki, Digital visual documentation and communication of Greek maritime heritage [thesis in progress]. University of the Aegean.
[4] E. Georgaki, K. Konstantopoulou, Critical Presentation of Maritime Heritage Records Digitization and Documentation Program, Arxeion Taxis. Under publication
[5] IFLA, Guidelines for Multilingual Thesauri: working group on guidelines for multilingual thesauri IFLA classification and indexing section, 2005. URL: https://archive.ifla.org/VII/s29/pubs/Draft-multilingualthesauri.pdf.
[6] F. Lacasta, G. Falquet, G. Zarazaga-Soria, J. Nogueras-Iso, An automatic method for reporting the quality of thesauri, Data & Knowledge Engineering 104 (2016) 1–14. URL: http://dx.doi.org/10.1016/j.datak.2016.05.002.
[7] D. Mouratidis, E. Mathe, Y. Voutos, K. Stamou, K. Kermanidis, P.Mylonas and A. Kanavos, Domain-Specific Term Extraction: A Case Study on Greek Maritime Legal Texts, in: Proceedings of the 12th Hellenic Conference on Artificial Intelligence (SETN '22), Association for Computing Machinery, New York, NY, 2022. URL: https://doi.org/10.1145/3549737.3549751.
[8] NAFKLIROS, (2023). URL: https://enirisst-plus.di.ionio.gr/en
[9] L.F. Spiteri, The Essential Elements of Faceted Thesauri, Cataloging & Classification Quarterly, 28(4) (2000) 31-52. doi:10.1300/J104v28n04_05
[10] THEMAS - Thesaurus Management System, 2001. URL: https://demos.isl.ics.forth.gr/themas-el/Index?logout=true&lang=en.