Leveraging Linked Data to develop rich discovery services for big heterogenous cultural data; the case of the National Cultural Heritage Aggregator SearchCulture.gr

Georgia Angelaki*†, Harris Georgiadis†, Agathi Papanoti† and Elena Lagoudi†

[1] *National Documentation Centre, Greece*

Abstract

Metadata heterogeneity is a cultural aggregator's biggest challenge. In this paper, we will present the ten-year development of SearchCulture.gr, the Greek National Cultural Heritage Aggregator, in establishing a robust and scalable aggregation infrastructure, and a public portal, providing access to nearly 1 million objects. SearchCulture.gr's success lies in its enrichment strategy that applies state-of-the-art semantic technologies and the power of Linked Data to provide both fine-grained search capabilities and a multitude of browsing options, such as displaying objects on a map and using advanced queries to create engaging thematic exhibitions that showcase the breadth and depth of Greek cultural heritage. Thanks to this innovative semantic enrichment strategy with regards to persons, types of objects, themes, geolocations and timespans/historic periods, SearchCulture.gr effectively addresses fundamental user search questions—"who," "when," "what," and "where"—resulting in very high precision in search results and overall offering the most advanced discovery services among Europeana's national and thematic aggregators.

Keywords

semantic enrichment, linked data, SearchCulture.gr, aggregation, Europeana

1. Introduction

SearchCulture.gr (https://www.searchculture.gr) is the Greek National Aggregator for Cultural Data and National Provider for Europeana. It is being developed by the National Documentation Center of Greece (EKT), a public sector organization supervised by the Ministry of Digital Governance. Since its launch in 2015, it has kept growing in numbers and expanding its functionalities. Today, it has amassed more than 860,000 records from 94 providers such as galleries, libraries, archives and museums and any type of institution that is custodian of cultural collections. Data ingested represent diverse fields, such as archeology, history, arts and crafts, folk and intangible heritage.

To deliver data to SearchCulture.gr, data providers need to comply with a set of Basic Interoperability Guidelines [1] which include submitting a set of minimum data using the OAI-PMH protocol in one of the following models: EDM, ESE or OAI_DC/OAI_DCTERMS combined with METS. The specifications also explain and promote the use of semantic vocabularies, the required standardised licenses and the minimum resolution of digital files. SearchCulture.gr's requirements are totally aligned with Europeana's specifications (Publishing Guide[2] and

* Corresponding author.
† These authors contributed equally.
searchculture@ekt.gr; angelaki@ekt.gr (G. Angelaki)
0000-0001-6360-2664 (G. Angelaki); 0000-0003-1137-6583 (H. Georgiadis); 0000-0002-3564-6739 (A. Papanoti) ; 0000-0002-3431-4513 (E. Lagoudi)

[2] https://pro.europeana.eu/post/publication-policy

Licensing Framework[3]) therefore the whole process of delivering data to Europeana is streamlined for the providers. In addition, the whole process is free for them.

SearchCulture.gr's Interoperability Guidelines are gradually being adopted in the national calls for funding of digitisation projects and by the Ministry of Culture. This initiative helps establish a common standard for interoperability and quality of digitization at the national level.

However, the Interoperability Guidelines alone are insufficient to address the semantic heterogeneity of the diverse collections aggregated in SearchCulture.gr. Achieving homogenization and semantic interoperability across all collections is essential for an aggregator to offer targeted search and browsing functionalities in large-scale datasets. To address this, we developed a state-of-the-art semi-automatic semantic enrichment strategy and infrastructure, that we use to enrich the aggregated content, along with a set of targeted Linked Data Vocabularies. EKT Vocabularies which are presented later in the text, extend, translate and link to established LOD vocabularies such as the Virtual International Authority File, UNESCO Thesaurus, Getty AAT and Geonames.

As a result of this process, every item record ingested is enriched with the new fields "EKT type", "EKT subject", "EKT person (creator/referred)", "EKT Place" and "EKT Historical period" . These fields answer the questions: "What is it", "What does it refer to", "Who", "Where" and "When", respectively.

In this paper we present the challenges we encountered, the methodology followed, the tools deployed and the new search and browsing functionalities and map-based discovery services that were gradually developed by leveraging the power of the semantic web.

2. Challenges

The source metadata ingested from a large number of providers is heterogeneous and varies in quality. Common challenges include typos, differences in syntax, the use of synonymous terms, variations in language, use of both broader and narrower concepts, and inconsistent use of singular and plural forms. For instance, dates may be formatted as "first half of the 19th century," "1800-1850," or "early 19th century." Similarly, a person's name might appear as "Rigas Feraios," "Feraios Rigas," or "Rigas Velestinlis," all referring to the same individual. Additionally, a place name like Tripoli could refer to either Tripoli in Arcadia or Tripoli in Libya, among other examples.

3. The Semantic Enrichment Scheme in SearchCulture.gr

SearchCulture.gr transforms harvested data into the Europeana Data Model (EDM) which is natively supported in the aggregator. To leverage the power of the semantic web, EDM features a number of classes devoted to the representation of "contextual" entities such as for persons and places.

In accordance with EDM that supports a Proxy mechanism to hold different views of the same Cultural Heritage Object (CHO) [2], the enrichment scheme in SearchCulture.gr is based on adding links (URI Refs) stored in separate 'EKT' fields in CHOs' metadata to terms from Linked Open Data (LOD) Vocabularies. These links are produced from curated mappings between source metadata values and terms from target vocabularies.

The implementation of the scheme is done in Semantics.gr, a platform developed in-house by EKT that serves the development, curation and interlinking of vocabularies, thesauri and authority files and their publication as LOD supporting any Data Model that can be expressed as an OWL ontology, besides SKOS.

---

[3] https://pro.europeana.eu/page/europeana-licensing-framework

Semantics.gr also contains a Mapping Tool used to set *Enrichment Mapping Rules* (EMRs) in order to perform bulk data enrichment in aggregator databases and repositories. The GUI environment includes advanced automated functionalities that help the curator easily define EMRs from source datasets (resources/terms from other vocabularies, metadata records or aggregated metadata values or phrases) to terms from a target vocabulary (see fig.1 for a validated mapping in Semantics.gr).

For each dataset and target vocabulary a dedicated *Mapping Form* is created. The enrichment tool supports automatic suggestion of EMRs which is based on string similarity matching between metadata field values and indexed labels of vocabulary entries (e.g. skos:prefLabel and skos:altLabel). Besides the primary source field (ie dc:creator, dc:contributor), other fields can also be used (ie dc:subject, dc:description) as filters in order to set more refined EMRs.

The curator can create complex logical expressions using the logical operators AND, OR and NOT on the filters to avoid false positives. For instance, an EMR may assign items with dc:type “image” to the vocabulary term “vase” if they have a dc:subject value “vase” or “oenochoe” but NOT a dc:subject value “drawing representation”. Another EMR could map items with dc:type “image” and dc:subject “drawing representation” to the vocabulary term “drawing”. When the automatic suggestion function fails to produce correct rules, the curator can set EMRs manually.

The Mapping Form incorporates a self-improving automatic suggestion mechanism. The manual mappings from metadata values to usually similar or broader vocabulary terms constitute valuable knowledge that enhances the effectiveness of autosuggestions in future enrichments, thereby reducing the need for manual assignments. The curator decides whether a manual EMR should be remembered by bookmarking it.

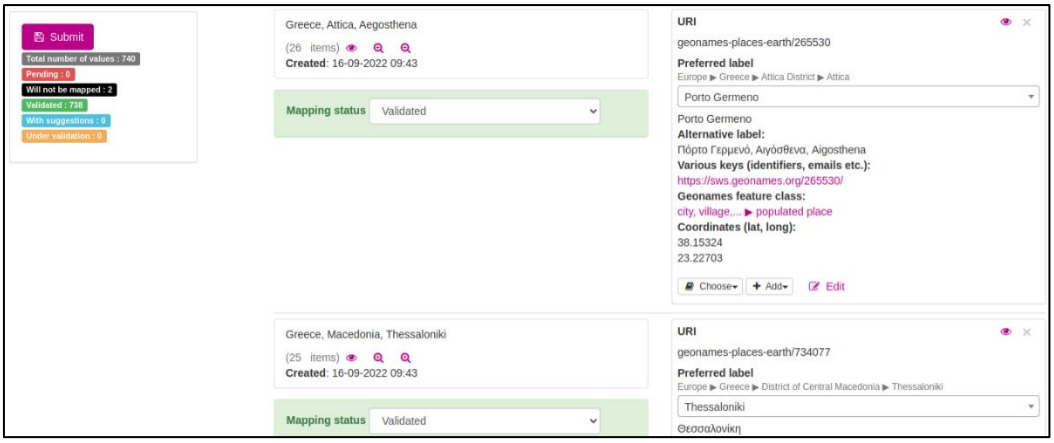


Fig. 1. Two validated mappings in a dcterms:spatial Mapping Form for a specific collection

For every different type of semantic enrichment, various extensions were made to the main enrichment scheme to provide additional functionality that would further support and automate the process as described in [3], [4] and [6].

Validated mappings are served on request via a RESTful API in JSON format which can be used by the aggregator or repository to enrich the collection easily and en masse. The tool is thoroughly described in [5].

In summary, the semantic enrichment strategy contains the following steps:

- Enrichments are performed per collection and per field in a Mapping Form
- Source metadata stay intact and enrichment values are clearly marked as “EKT” fields
- All mappings produced, either automatic or manual, are always validated by an expert curator.

- Development of target vocabularies is done in Semantics.gr. These are all bilingual and hierarchical if applicable. They are often adaptations, extensions and translations of popular Linked Data vocabularies such as Geonames, UNESCO thesaurus, etc.
- All vocabularies are openly licensed and published in Semantics via open APIs to allow for re-use

Last but not least, all related work is published and communicated via webinars, articles, etc, following EKT's open science approach.

4. Retrospective mass-scale enrichments

We started developing vocabularies and applying semantic enrichments to our collections in 2016. Every semantic enrichment work initially run as a dedicated retrospective enrichment project covering all datasets ingested in SearchCulture.gr up to that point and was then incorporated in the standard ingestion process for new collections.

Cultural heritage item types[4]**.** First, we created a SKOS-based LOD original vocabulary consisting of 193 terms that cover different types of cultural artifacts and is linked to Getty AAT. Metadata records were enriched with a separate field "EKT type" that holds references to the vocabulary's terms.

Greek Time Periods[5]. Next, we set out to homogenize and normalize chronologies and historical periods. The vocabulary Greek Time Periods was developed in 2017, and it is constructed according to the semantic class edm:Timespan of Europeana's EDM. It contains 169 terms that cover Greek history from 8000 BC to today. It is hierarchical and bilingual. Depending on whether the original temporal documentation is based on period labels or chronologies, we adopted two fundamentally different enrichment strategies, historical period-driven enrichment and chronology-driven enrichment, respectively. In the chronology-driven enrichment, chronological values are being homogenized into years or year ranges and then, based on the results, the items are enhanced with the corresponding terms from the historical periods vocabulary. Our time normalization method is fully extensible and parametric, takes into consideration language descriptors and covers four types of temporal expressions, centuries, range of centuries, years/dates and year ranges. As a result, original metadata records were enriched with two distinct fields, "EKT chronology" and "EKT historical period". A detailed presentation of the item types and chronology/period-related enrichments is provided in [3].

Subjects. The following iteration of enrichments was thematic, adding a new SKOS-based field "EKT Subject" that includes references to terms of a bilingual and hierarchical vocabulary of subjects that is interlinked to the UNESCO Thesaurus (EKT version[6]- 1,391 terms) and a complementary vocabulary of Thematic Tags[7] that covers more specific topics (607 terms).

Persons (creators/referred persons). Next, we extended our enrichment scheme to persons, distinguishing between creators and referred persons (i.e. a person depicted in a photograph, cast of a film, a recipient of a letter or the subject of a biography). Similarly to the other enrichments, this would involve identifying person entities in metadata and mapping them to entries from a structured vocabulary, the Notable Persons in Greek History and Culture[8]. The

[4] https://www.semantics.gr/authorities/vocabularies/ekt-item-types
[5] https://www.semantics.gr/authorities/vocabularies/time-periods/vocabulary-entries
[6] https://www.semantics.gr/authorities/vocabularies/ekt-unesco
[7] https://www.semanThematic tagsics.gr/authorities/vocabularies/thematic_tags
[8] https://www.semantics.gr/authorities/vocabularies/searchculture-persons

Vocabulary has reached 9540 terms modelled according to the edm:Agent class and linking to VIAF, Wikidata and other online LD resources. Each entry was enriched, when possible, with metadata regarding place of birth and death, date of birth and death, sex, occupation, bibliographic references and links to established resources, such as the Virtual International Authority File (VIAF), Wikipedia and IMDB.

During the process, a complementary vocabulary was introduced, that of "Professions/Occupations[9]". It is an LOD vocabulary conforming to the skos:Concept semantic class and consists of 371 terms. It is hierarchical and bilingual, and its terms refer to occupations such as merchants, doctors and military officers, clergy positions, noble titles, different social movements affiliates like feminists or socialists or types of artists and literary creators. The terms of this vocabulary were used to classify entries in the Vocabulary of Persons. A detailed presentation of the person-related enrichments is provided in [4]

Places. The last retrospective semantic enrichment process regarded geographical information. Utilizing the GeoNames API, a "starter set" of ~6K terms was selected comprising entities belonging to the first level of each country's administrative hierarchy and cities with population over 100k globally. For Greece, the threshold was set to three levels of administrative divisions and all the settlements of more than 1K inhabitants. The resulting "Vocabulary of geographical names GeoNames (EKT version[10])" is hierarchical and conforms to the edm:Place contextual class of EDM. At the end of our enrichments, it reached ~12K terms.

In addition to the main vocabulary, a supplementary EKT vocabulary[11] was developed to include features that didn't fall under the strict administrative hierarchy described above, adding, for example, historical areas (e.g., Soviet Union), placenames that include many different states (e.g., the Balkans) or geomorphological elements that may transcend different states, such as rivers etc. Those two vocabularies are interconnected using two custom fields ekt:isPartOfMatch and ekt:hasPartMatch (similar to the skos:broaderMatch and skos:narrowerMatch, respectively from the SKOS data model) in order to express the hierarchical "Has-Part" relationships between the two vocabularies.A detailed presentation of the spatial enrichments can be found in [6].

Finally, it is worth noting that the EKT Vocabularies, controlled, standardised, dynamic, extensible and open, constitute particularly valuable resources that can be also used by cultural institutions in their primary documentation, enhancing the quality, interconnectivity, integration and multilinguality of the original metadata. The reuse of such resources improves accessibility to cultural heritage at national level and achieves economies of scale in documentation processes.

5. Limitations

There are inherent limitations in the enrichment process which derive from Dublin Core (DC)-based metadata models, including EDM, regarding spatial and person information representation.

In DC and in EDM, geographical information is represented in the properties dcterms:spatial or dc:coverage in an equivocal way: a toponym can either indicate the place where an item was created, where it is being kept, or its subject. Regardless, therefore, of whether more nuanced place-based information is included in the source metadata, at aggregation level this is often

[9] https://www.semantics.gr/authorities/vocabularies/professions-occupations
[10] https://www.semantics.gr/authorities/vocabularies/geonames-places-earth/vocabulary-entries?language=en
[11] https://www.semantics.gr/authorities/vocabularies/geonames-supplementary-places/vocabulary-entries

compromised due to the limited expressivity inherent in EDM and other DC-based schemata. This is why the new EKT place field produced by our enrichments may point to any place the item relates to in an indistinguishable way.

Similarly, “dc:contributor” could be the author or the subject of a book, the director, the screenwriter or the actor of a play, the sculptor, the model or the photographer of an antiquity, the sender, the receiver or the subject of a letter, etc. Moreover, while the creator in most cases is given in “dc:creator”, persons that appear in photographs or are the subject of a book, sometimes appear in “dc:subject”, some other times in “dc:contributor” or “dc:title” or “dc:description”. Ultimately, we decided that the enrichment on persons should improve SearchCulture.gr by allowing users to easily find all works of a creator (regardless of its specific role in the creation) and all the CHOs referring a person (regardless of the kind of reference). We opted for creating two separate fields, “EKT creator” and “EKT referred person” thus conducting those two kinds of person enrichments. These are the fields used for person-driven search, browsing and faceting.

6. Search and browsing functionalities

The semantic enrichment workflow aims at enhancing the user experience by providing more detailed and meaningful connections between data points in SearchCulture.gr. The array of Discovery Services built on the back of the semantic enrichments offers various gateways into the aggregated content. This approach not only improves the usability of the platform but also supports deeper scholarly research and public engagement with Greece's cultural heritage.

The controlled vocabularies developed are integrated in a flexible way, throughout the portal as search, faceted filtering and browsing options. The fact that they are bilingual, allows also english-speaking users to navigate the –mostly-greek content, up to a certain extent.

First of all, every controlled vocabulary has its own presentation page which offers dedicated search and hierarchical (if applicable) browsing options. Fig 2 illustrates the page for Persons. A subset of terms also appears on the homepage as tag clouds acting as an entry point for the user.

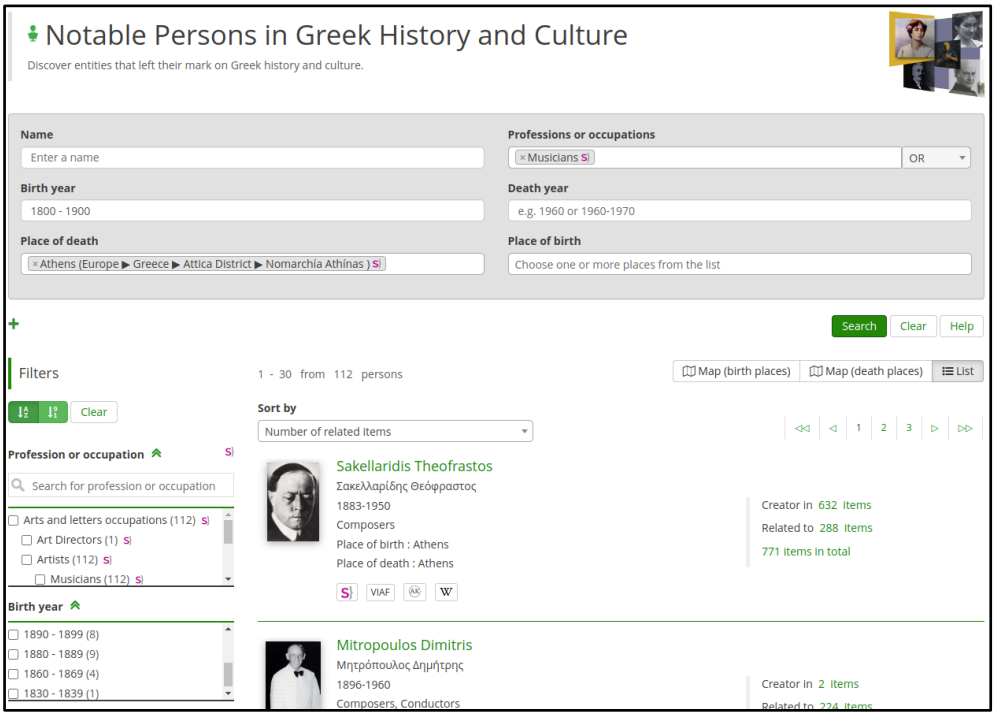
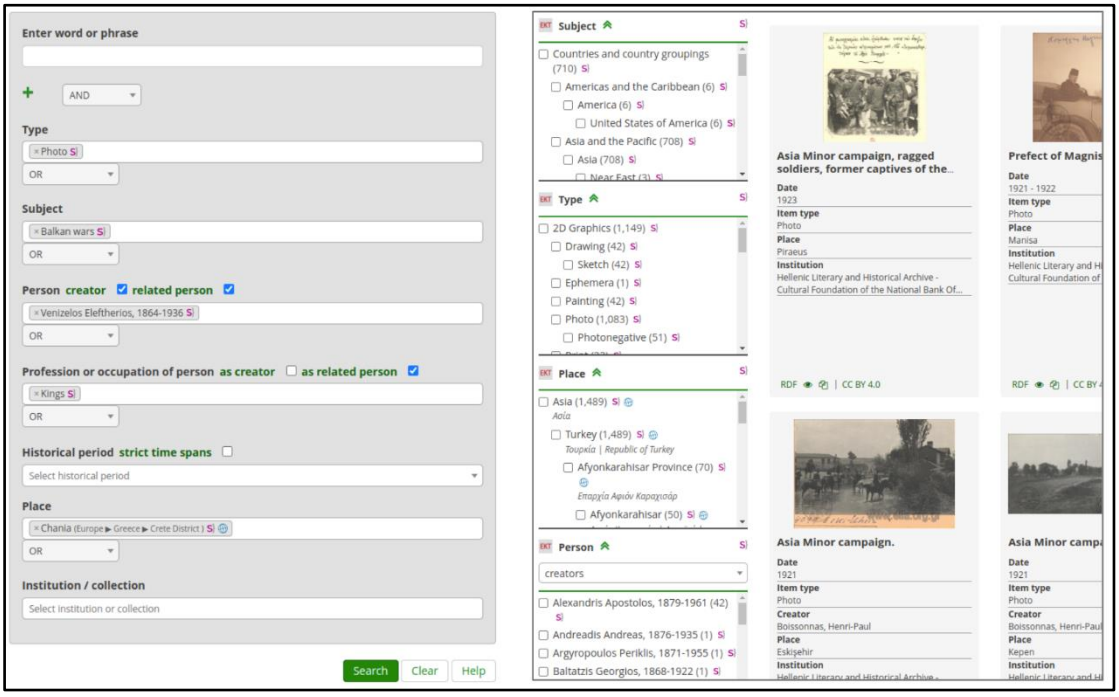


Figure 2: Searching persons by name, gender, profession, birth or death year and place

EKT vocabularies are incorporated in the respective fields in the advanced search. Every field includes autosuggestion and disambiguation functionality. The fields can be combined to enable highly targeted searches. For example, Fig. 3 (a) illustrates a query that searches for "*photos*" related to the "*Balkan Wars*", referring to "*E. Venizelos*", involving a *"King"* and focusing on "*Crete*".

The EKT vocabularies are also used as facets allowing further filtering of the results, as shown in Fig 2 (b). All resulting enrichment terms are active links pointing to all items enriched with that term. Every such term is accompanied by its semantic resource in Semantics.gr (marked with a pink "s") where one can see the complete record of the resource, such as other linked resources from other vocabularies, etc.



(a) (b)

Figure 3: The advanced search box (a) and the facets in search results (b) support all EKT fields.

In order to exploit the hierarchy incorporated in the respective vocabularies (places, item types, subjects) we index for each item all broader terms as well, using a separate auxiliary Solr field, thus, supporting hierarchical searching and faceting. This way, for example, when a user searches items of "Attica" (the prefecture) the results will also include items of "Athens".

Moreover, the enrichments are used to locate the items on an interactive map (Fig. 5). The implementation of the map navigation was based on leaflet.js and OpenStreetMap. To support the display of a large number of items on the map, items appear in clusters that can be further expanded as the user clicks or zooms on the map. Users can retrieve items belonging to a specific place or all items located in the current map frame (within its coordinates). The new visualization feature was also added to the Thematic Exhibitions, providing a new dimension with regards to showcasing the items included.

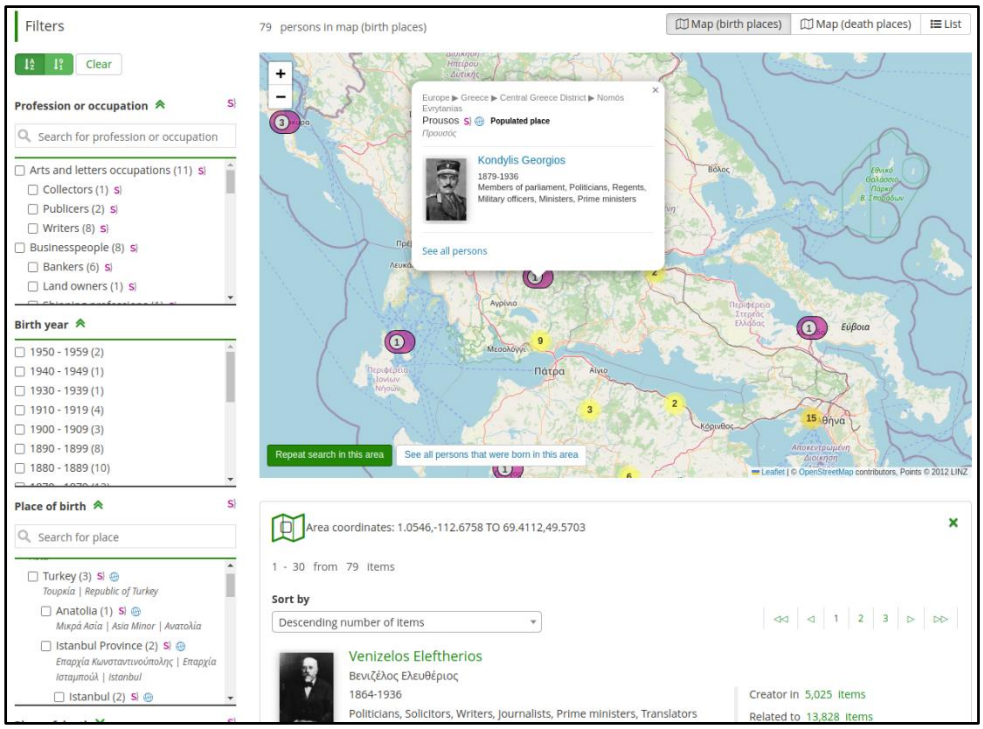


Figure 4: Exploring persons on the map by their birth or death place

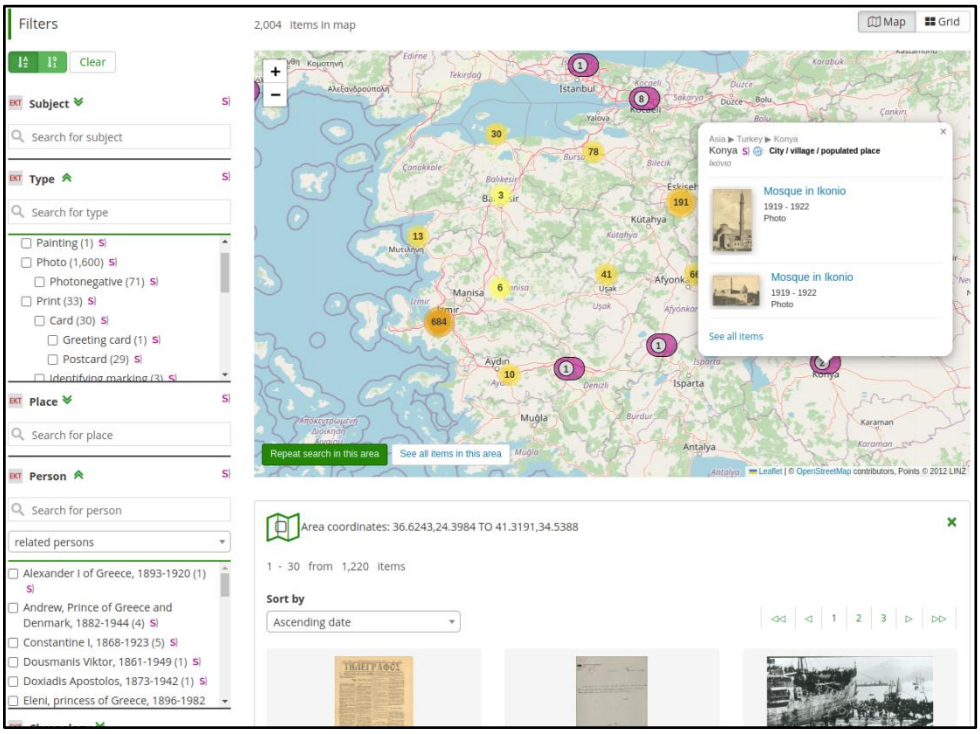


Figure 5: Exploring items on the map

By combining the enrichments with the presentation features, we are also able to offer more "niche" and playful search options, making the content more engaging and interesting overall. For example, one could search for female writers born in Crete at the beginning of the 20th century and approach this search in various ways, such as through the advanced search box, the person browsing page, or the map.

7. Thematic exhibitions

In response to the urgent need for digital cultural content for education and use throughout the closure of physical cultural spaces, during the COVID pandemic, the SearchCulture.gr team developed the thematic exhibitions functionality and created a series of exhibitions which found much favor with the public, increasing user numbers significantly.

The Thematic Exhibitions feature a smaller or larger set of CHOs from different digital collections that share storytelling relevance. The back end of this functionality also utilizes semantic enrichments. The curator uses a Query Form (similar to the Advanced Search Box) to select a Type, Place, Person, or Subject, and enters search parameters such as time, type sub-category, inclusion or exclusion of items, collections, or organizations. Following traditional curatorial practices, the digital curation process begins with a conceptual query. The curator refines this query through trial and error, selecting and organizing objects to create a coherent and compelling narrative by layering, juxtaposing, comparing, combining, or excluding objects. As the query is refined through this iterative process, the selected items begin to form a conceptually cohesive and narratively unified exhibition.

The interpretative phase involves using tools to clarify and narrate the exhibition's stories in both Greek and English, incorporating images, quotes, and hyperlinks to provide context, depth, and additional information. The exhibition can be presented through map-based storytelling or a grid layout, depending on the content's relevance. Each virtual exhibition includes a key image, a title, and context-providing subtitle, accompanied by an interpretative text. The Exhibitions Query Form enables editors to create thematic exhibitions by publishing the retrieved query results in a curated, organized manner.

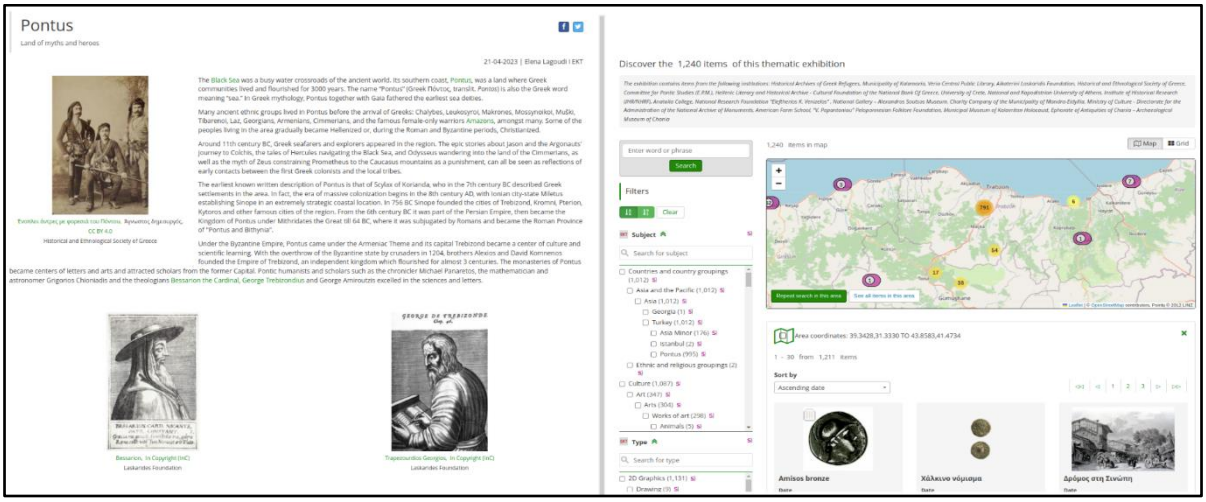


Figure 6: Exploring items on the map

With the aim of showcasing various aspects of Greek cultural heritage, the Thematic Exhibitions developed range from simple, type-based archaeological exhibitions—such as the story of the oil lamp—to more nuanced topics, like Olfactory or Industrial Heritage. From 2020 to 2024, eighty (80) Thematic Exhibitions were created, covering arts and crafts, archaeology, music and theater, religion, architecture, folklore, oral traditions, and social issues. Following the incorporation of geolocation enrichments in 2023, a series of place-based exhibitions was developed, highlighting many areas of Asia Minor where Hellenism thrived, as well as Greek islands and regions of strong local interest. For a detailed presentation of the new functionality see [7].

Recently, a new sub-category of thematic exhibitions was released highlighting people, either individuals or groups of people that share the same ideas, passions, genres of writing or painting (Fig. 7).

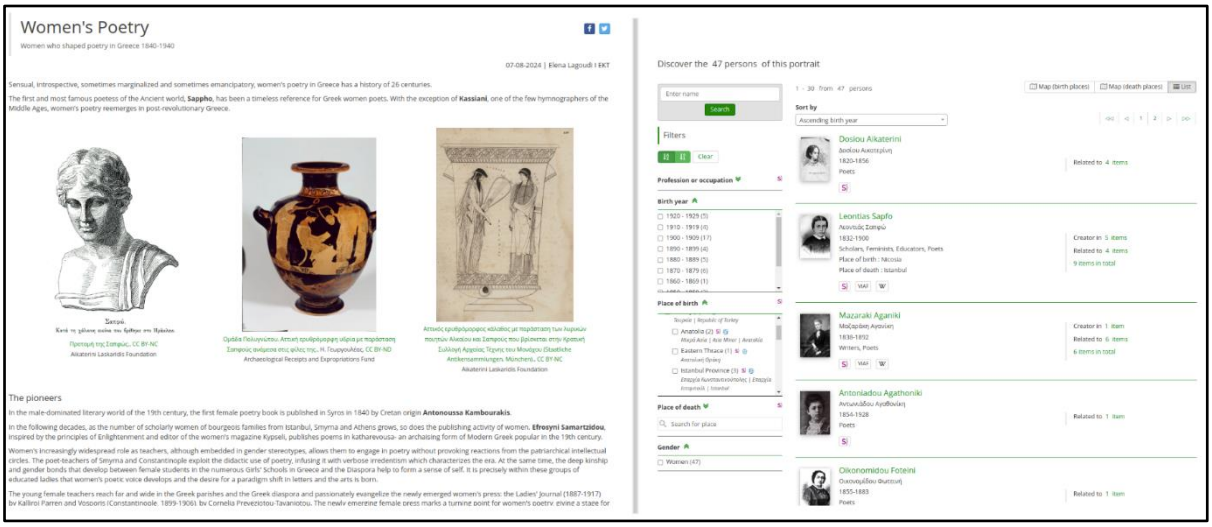


Figure 7: Person/Group of persons'-related exhibitions and functionality

The strategic editorial approach of the national aggregator SearchCulture.gr aligns closely with that of Europeana, integrating best practices outlined in the Europeana Editorial Guidelines[12] and utilizing digital storytelling techniques recommended by the Europeana Network Association. These techniques include making content personal, blending expertise with an informal tone, uncovering hidden stories, using visual and audio materials, ensuring clear narrative structure, beginning with specific details, and employing evocative imagery.

8. Expert knowledge in the enrichment and digital curation processes

The SearchCulture.gr team consists of four core members: one technical developer who also serves as the scientific supervisor of the infrastructure, one digital heritage expert responsible for network development, and two archaeologists working full-time on the semantic enrichment discussed earlier and digital curation processes described below. Since 2016, several humanities scientists have contributed on a part-time basis to the semantic enrichment process.

It is essential to emphasize that the involvement of expert knowledge throughout the process is indispensable for ensuring the quality and validity of the vast amount of EMRs, mappings, and vocabulary terms created over the years, as well as of the enrichments themselves. In addition, curators have provided invaluable feedback throughout the gradual development of the semantic enrichment infrastructure.

Furthermore, the disambiguation and identification process often requires substantial historical knowledge and thorough research using reliable online and offline resources. For instance, over 5,000 settlements in the Greek territory were renamed—sometimes as frequently as every 20 years—during the 20th century for historical and political reasons. One example is the village of Γκρόπινο (of Bulgarian origin), which was renamed Τρόπινο in 1928 as part of an effort to "Hellenize" the name. In 1940, it was renamed again to Βαλτολείβαδο (meaning "meadow with swamps," a reference to its natural surroundings), and finally, in 1961, it was given the more "elegant" name Δάφνη (Laurel). Archival research in this case was essential for

[12] https://pro.europeana.eu/page/writing-for-europeana-pro

SearchCulture.gr curators to accurately assign the correct Geoname, as related resources are not readily available.

Moreover, the deep familiarity with incoming collections gained through semantic enrichment significantly facilitates and improves the digital curation process of thematic exhibitions, as discussed in the previous section.

All the above create a virtuous cycle between data ingestion, semantic enrichment, infrastructure development, collections' curation, and presentation. The investment in expert human curation required for the semi-automatic enrichment of the collections is, therefore, justified by the tremendous added value it brings to the discovery of the collections, as presented below, that is, at least for the size of the collections that is ingested in SearchCulture.gr.

9. Related work

Discoverability and re-use can be influenced by the level of metadata heterogeneity and semantic interoperability [8]. Different semantic enrichment strategies are, therefore, adopted by large cultural heritage aggregators as a means to contextualise resources, disambiguate, add multilinguality and offer search and browsing functionalities across multiple heterogenous source datasets [9].

Among the domain and thematic aggregators that form the Europeana Aggregators' Forum, some demand the data is enriched prior to ingestion, transferring the responsibility to the providers [10], others undertake semantic enrichment post ingestion [11], while the majority just indexes string data without applying any semantic enrichment before delivering data to Europeana.

Europeana enriches aggregated CHOs by automatically linking text strings found in the metadata to controlled terms from established LOD vocabularies such as Geonames, GEMET and Dbpedia [12],[13]. However, complete automated enrichment on structured fields (such as dc:type) adopts an "enrich-if-you-can" strategy, horizontally, resulting in non-negligible percentages of mistakes [12] and in relatively low enrichment coverage - despite using extremely large target thesauri, such as DBpedia and Geonames. Additionally, Europeana will not employ sophisticated methods to extract related information from other descriptive fields or to create more nuanced matches (e.g., between synonym terms).

Automated annotation methods on more descriptive fields (such as dc:title) yield similar challenges [14]. All these techniques enhance searchability and multilingualism. However, due to the relatively low enrichment coverage, the extensive target thesauri used, and the significant percentage of enrichment errors, they cannot achieve sufficient homogenization to enable aggregators to offer advanced methods for content exploration, such as browsing and faceting on enriched fields.

In the comparative evaluation performed by EuropeanaTech Task Force [16] the necessity of human-in-the loop methodologies to complement automatically produced enrichments is implied. Many EU-funded projects deal with the complexities of fully automatic or crowdsourced enrichments such as Enrich+, St George on a Bike and Europeana XX. SAGE [17] is a semantic enrichment and validation platform that deploys state-of-the-art AI tools assisted by human-in-the-loop validation mechanisms to produce automatic mappings. However, it lacks the sophistication provided in Semantics.gr such as the use of filters to refine mappings.

Regarding the search and browsing functionality offered, the way heterogeneous data is curated—whether effectively or not—reflects in the options presented to users. For example,

place-based search is available through Deutsche Fotothek[13] and the German Digital Library[14]. CulturaItalia.it[15] provides place-based filtering of results, while the Swedish Kringla[16] offers province-based filtering and map-based search, though it geolocates only a fraction of the objects on the map and the functionality is only available in Swedish. Among other cross-domain aggregators, the German Digital Library provides advanced disambiguation and search functionality for persons, using the GND thesaurus as the source vocabulary and the OpenRefine tool to enrich person values. Each individual has a landing page that includes basic biographical information and a list of works for which they are either the author or a referenced entity.

However, among all the large cultural heritage data aggregators investigated, and to the best of our knowledge as members of the Europeana Aggregator Forum, SearchCulture.gr is the only national cross-domain aggregator that adopts such a systematic and meticulous semantic enrichment strategy across the majority of the EDM contextual classes. This approach enables US to build very fine-grained search and browsing options, effectively answering the "who," "what," "when," and "where" questions in a refined and sophisticated manner. The enrichment has remarkably improved the searchability of the SearchCulture.gr collections as demonstrated in[3].

10. Conclusions and future work

Semantic enrichments improve the quality and validity of data, add a layer of multilingualism to search, clarify concepts and individuals, support the conceptual interconnection between items found across various repositories, and enhance advanced search capabilities. Highlighting often unexpected correlations between people, places, themes, historical periods, and types of content opens new horizons for understanding and researching Greek culture.

Given the related efforts, the semantic enrichment scheme presented in this paper, is aligned with the Europeana TF recommendations [15] and achieves high coverage and effective disambiguation because i) it adjusts to the documentation particularities of the individual collections ii) it combines self-improving, automatic and fuzzy-based suggestions with a suite of tools that support the curation and disambiguation process, iii) uses controlled target vocabularies that are gradually expanded to cover the needs of the specific collections, and iv) employs expert knowledge for the validation of the mappings.

This model systematic curation process brings added value to aggregation and supports the development of advanced possibilities for searching and navigating the heterogeneous richness of the country's cultural heritage.

Building further on such a robust, versatile, and scalable semantic infrastructure, our future plans include exploring the representation of Events and Intangible Cultural Heritage, developing relevant targeted semantic vocabularies, opening the thematic exhibitions functionality to end users, exploring both crowdsourcing and AI technologies to improve metadata quality, and adding more engaging visualization elements to our portal.

13 https://www.deutschefotothek.de/

14 https://www.deutsche-digitale-bibliothek.de/?lang=en

15 http://culturaitalia.it/

16 https://www.kringla.nu/kringla/

References

1. EKT 2024, Basic Interoperability Guidelines https://ariadne.ekt.gr/ariadne/handle/20.500.12776/17194, 2nd edition
2. Europeana Data Model Primer, https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf
3. H. Georgiadis, A. Papanoti, M. Paschou, A. Roubani, D. Chardouveli, E. Sachini (2018). Using type and temporal semantic enrichment to boost content discoverability and multilingualism in the Greek cultural aggregator SearchCulture.gr. International Journal of Metadata, Semantics and Ontologies, 13(1), 75–92.
4. H. Georgiadis, A. Papanoti, E. Lagoudi, G. Angelaki, N. Vasilogamvrakis, A. Panagopoulou and E. Sachini, Enriching the Greek National Cultural Aggregator with Key Figures in Greek History and Culture: Challenges, Methodology, Tools and Outputs. International Conference on Theory and Practice of Digital Libraries TPDL 2022
5. H. Georgiadis, G. Angelaki, E. Lagoudi, N. Vasilogamvrakis, K. Stamatis, A. Papanoti, A. Panagopoulou, K. Bartzi, D. Charlaftis, E. Angelidi, D. Hardouveli, P. Karagianni and E. Sachini (2022). Publishing LOD Vocabularies in Any Schema with Semantics.gr. In: Garoufallou, E., Ovalle-Perandones, MA., Vlachidis, A. (eds) Metadata and Semantic Research. MTSR 2021.
6. A. Papanoti, E. Lagoudi, G. Angelaki, H. Georgiadis and Evi Sachini Greek Culture on the Map: Place-based Enrichment Scheme at the Greek National Cultural Data Aggregator. MTSR 2023
7. A. Papanoti, G. Angelaki, E. Lagoudi, H. Georgiadis, Every good search tells a story: Creating virtual exhibitions on the Greek national cultural aggregator through query-based curation. 5th CAA-GR CONFERENCE 2024
8. E. Garoufallou, and C. Papatheodorou: A critical introduction to Metadata for e-Science and e-Research, International Journal of Metadata, Semantics and Ontologies ((IJMSO), 9(1), pp.1–4. http://dx.doi.org/10.1504/IJMSO.2014.059143
9. S. Peroni, F. Tomasi, F. Vitali: The aggregation of heterogeneous metadata in Web-based cultural heritage collections. A case study. International journal of Web Engineering and Technology, Vol. 8 (4), pp 412-432 (2013)
10. M. Smith: Linked open data and aggregation infrastructure in the cultural heritage sector, A case study of SOCH, a linked data aggregator for Swedish open cultural heritage, in Information and Knowledge Organisation in Digital Humanities, Routledge (2021),
11. Linked Open Data for Libraries, Archives and Museums, EuropeanaTech Insight, Issue 7, https://pro.europeana.eu/page/issue-7-lodlam
12. H. Manguinhas, Europeana Semantic Enrichment Framework, technical documentation 2016, available at https://pro.europeana.eu/page/europeana-semantic-enrichment
13. J. Stiller, V. Petras, M. Gäde, A. Isaac.: Automatic Enrichments with Controlled Vocabularies in Europeana: Challenges and Consequences. EuroMed: 238-247 (2014)
14. E. Agirre, A. Barrena., O. Lopez de Lacalle, A. Soroa, S. Fernando, M. Stevenson: Matching Cultural Heritage tems to Wikipedia. In: Proc. LREC 2012. Istanbul, Turkey. (2012)
15. EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy: final report, https://pro.europeana.eu/project/multilingual-and-semantic-enrichment-strategy
16. Europeana Task Force on Enrichment and Evaluation Final Report https://pro.europeana.eu/project/evaluation-and-enrichments
17. E. Kaldeli, Combining AI tools with human validation to enrich cultural heritage metadata, https://pro.europeana.eu/post/combining-ai-tools-with-human-validation-to-enrich-cultural-heritage-metadata