



# Europeana DSI 2– Access to Digital Resources of European Heritage

**DELIVERABLE**

D6.4: PILOT FOR TIME-AND-PLACE DISCOVERY

<b>Revision</b>	DRAFT
<b>Date of submission</b>	
<b>Author(s)</b>	R. Simon, S. Gordea
<b>Dissemination Level</b>	Public



Co-financed by the European Union  
Connecting Europe Facility

## REVISION HISTORY AND STATEMENT OF ORIGINALITY

### Revision History

Revision No.	Date	Author	Organisation	Description
0.1	26.06.2017	Rainer Simon	AIT	First Draft
0.2	27.06.2017	Sergiu Gordea	AIT	Updated Draft
0.3	28.06.2017	David Haskiya, Timothy Hill, Hugo Manguinhas	EF	Review
1.0	30.06.2017	Rainer Simon, Sergiu Gordea	AIT	Final version

### Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

The sole responsibility of this publication lies with the author. The European Union is not responsible for any use that may be made of the information contained therein.

## Table of Contents

<b>Executive Summary</b>	<b>4</b>
<b>Introduction and Background</b>	<b>5</b>
<b>How Peripleo Works</b>	<b>6</b>
A Guided Tour of the User Interface	7
Technical Architecture	8
Enrichments & Conflation: Integrating Heterogeneous Sources	9
<b>EDSI2 Demo</b>	<b>11</b>
Integrating Places	13
Integrating Agents	14
Integrating Object Records	14
External Data Sources	15
<b>Exemplary Search Scenarios</b>	<b>16</b>
<b>Conclusion</b>	<b>22</b>
<b>References</b>	<b>24</b>

## 1. Executive Summary

This document reports on the results of the **Pilot for Time-and-Place Discovery**, a technology demonstration implemented on the basis of **Peripleo**, a prototype spatiotemporal search and visualization platform by **Pelagios Commons**.<sup>1</sup>

The task of this deliverable is to leverage Peripleo in order to: i) establish a **pilot for spatio-temporal discovery** using Europeana data; ii) test how existing Europeana enrichments - places and people - can be exploited to establish **contextual links to other specialist authorities**; iii) evaluate the setup for **potential further re-use and development**; and iv) determine how well the results can be used as an **innovative exploration environment** that could be deployed as a product.

Sections 2 and 3 of this deliverable provide an introduction to Peripleo's background, its functionality and technical architecture, and describe how Peripleo handles *conflation*, an approach to integrating data from different sources. Section 4 presents an overview of the data that has been obtained from Europeana in order to implement the pilot, as well as the external data (coming from partners of Pelagios Commons) that has been used to demonstrate cross-collection discovery and search scenarios; and describes the process by which it was integrated. Section 5 illustrates exemplary search scenarios. Section 6 concludes with a summary and an outlook on future work that may be of potential benefit to the Europeana service infrastructure.

---

<sup>1</sup> <http://commons.pelagios.org>

## 2. Introduction and Background

This document reports on the results of the **Pilot for Time-and-Place Discovery**, a technology demonstration that has been implemented on the basis of **Peripleo**, a prototype spatiotemporal search and visualization platform. Peripleo [2] is Open Source software,<sup>2</sup> licensed under the terms of the Apache 2 license. It is being maintained and developed further by Pelagios Commons,<sup>3</sup> an international Digital Humanities initiative aiming to foster better linkages between online resources documenting the past [3]; and was awarded the 1st prize in the “Best DH Data Visualization” category of the DH Awards 2016.<sup>4</sup> The work on D6.4 has been conducted primarily by Rainer Simon, Technical Director of Pelagios Commons, and lead developer of Peripleo.

Peripleo enables users to explore the geographic, temporal and thematic composition of distributed digital collections in their entirety, and to progressively filter and drill down to explore individual records. At the time of writing, Peripleo is undergoing a major re-development, which coincides with the timeline of DSI 2. The task of this deliverable is to leverage the current development of the new release, in order to:

- establish a pilot for spatio-temporal discovery
- test how existing Europeana enrichments - places and people - can be exploited to establish contextual links to other specialist authorities (such as the Pleiades Gazetteer of the Ancient World<sup>5</sup>)
- evaluate the setup for potential further re-use and development
- determine how well the results can be used as an innovative exploration environment for Europeana data, and whether it could be deployed as a product

In this document, we provide an overview of Peripleo’s features and user interface; discuss its underlying technical architecture; and present how the entity conflation and linking mechanism - which is at the heart of Peripleo’s architecture - can provide a suitable environment for interactive exploration of Europeana data. Furthermore, we describe the process of integrating Europeana authority and object record data with Peripleo. We present the results in the form of exemplary search and discovery scenarios, and conclude with a discussion of potential future work that may be of benefit to the Europeana service infrastructure.

---

<sup>2</sup> <http://github.com/pelagios/peripleo2>

<sup>3</sup> <http://commons.pelagios.org/>

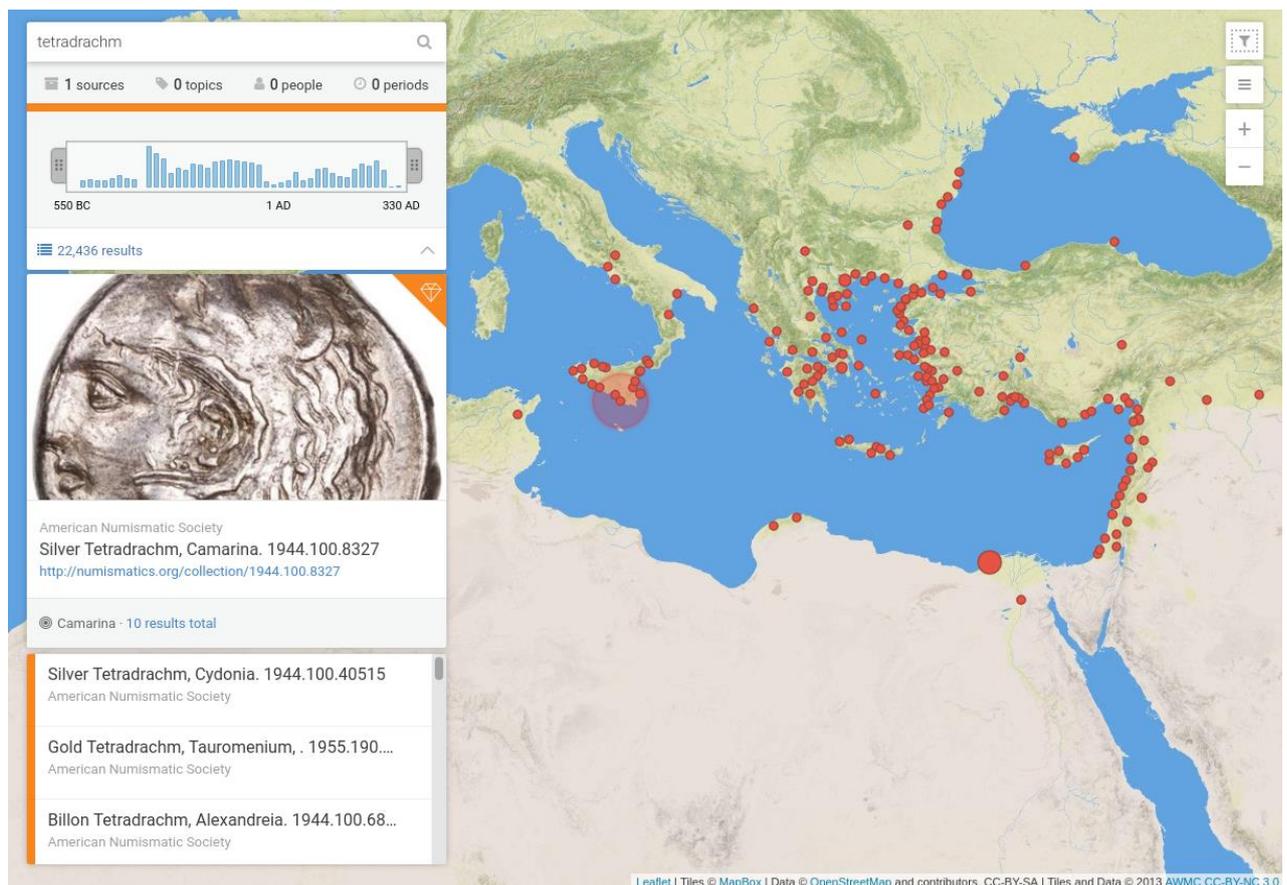
<sup>4</sup> <http://dhawards.org/dhawards2016/results/>

<sup>5</sup> <https://pleiades.stoa.org/>

### 3. How Peripleo Works

Peripleo is ancient Greek for “to sail (or swim) around” in the sense of exploration or discovery. The notion of being able to freely journey across a “sea of data” brought together from heterogeneous collections was the metaphor that had guided the original vision behind the tool. A key design goal was to provide, on the one hand, a familiar *Google-Maps*<sup>6</sup>-like interface, with full-text search and auto-completion; while allowing a more free-form mode of exploration, on the other hand [2]. Peripleo conveys a sense of the scope, breadth and structure of the data as a whole, by representing geographic coverage on the map; by showing temporal spread as a histogram; and by providing statistics on the distribution of the data across different thematic facets (such as data provider or language). This way, users can easily gain an overview first, and then filter and drill down according to their own interests.

Peripleo is Open Source software, available through the GitHub space of Pelagios Commons.<sup>7</sup> All development carried out as part of this deliverable has been done on the upcoming new version (Peripleo 2), which is currently under active development and will be released officially in December 2017.



**Fig.1. Peripleo 2 user interface.**

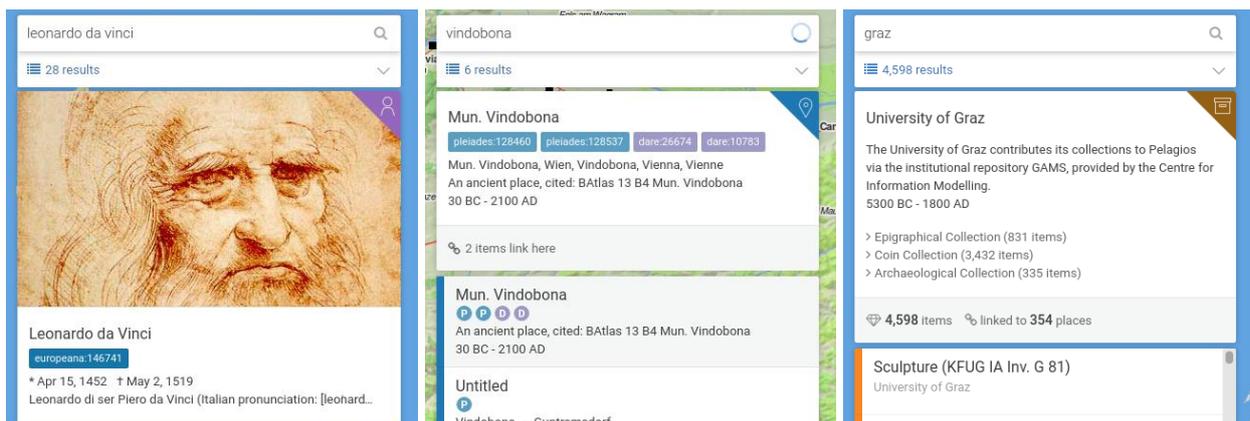
<sup>6</sup> <http://maps.google.com>

<sup>7</sup> <http://github.com/pelagios/peripleo2>

### 3.1 A Guided Tour of the User Interface

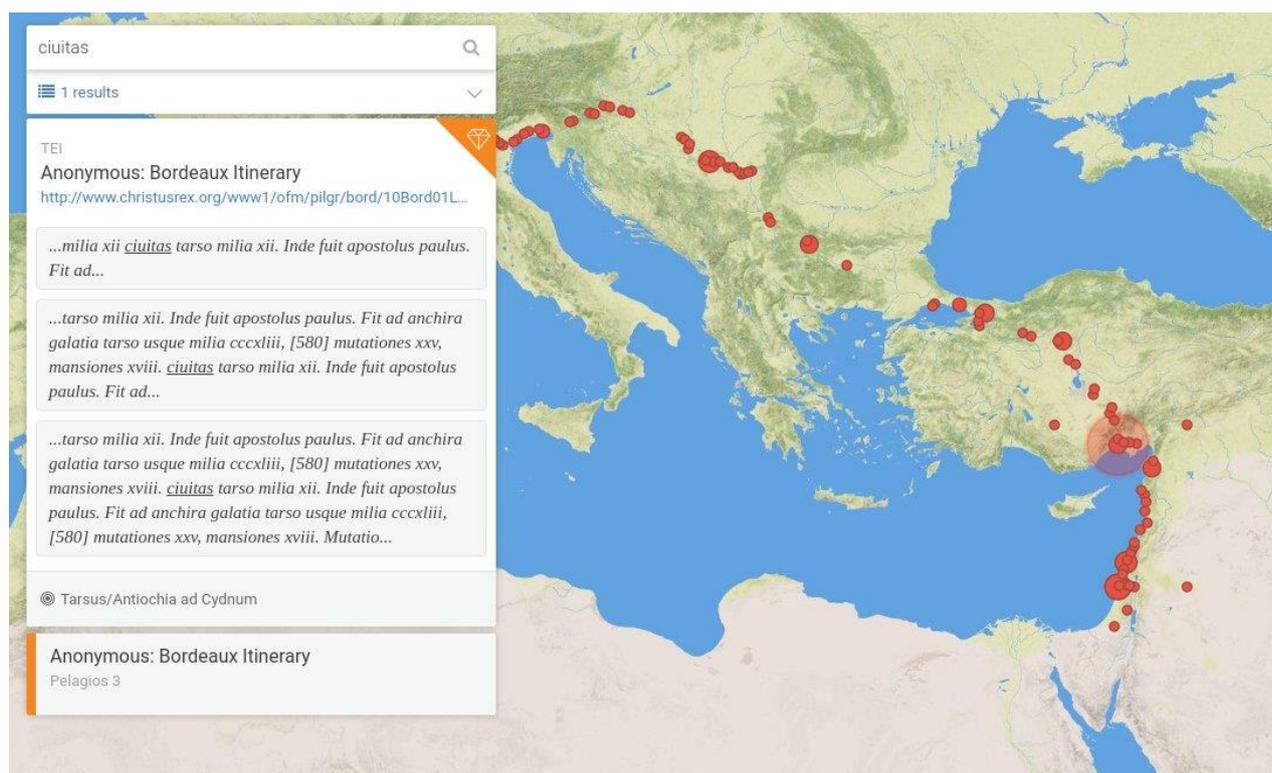
The user interface (UI) of Peripleo consists of a full-screen map, and a column of controls and information panels on the left side of the screen (Fig. 1). The top component in this column is a search entry field providing autosuggest functionality. I.e. as soon as the user starts typing, a dropdown list suggests possible search phrases - but also concrete results (places, people, objects) - to choose from. The search panel is collapsible: unfolding it reveals additional information about the structure of the current result set, e.g. the temporal distribution or facet information such as the number of sources results are from. (Once Peripleo 2 is completed, clicking on these facet counters will provide corresponding filter options. However at the time of writing, this is still under development.)

Underneath the search entry field, there are panels showing the currently selected result (if any) and full list of results. The current selection is represented with an image (if a depiction URL exists for it), metadata (title, description, date information, Web link), an icon indicating the type of the result (i.e. whether it is an object, a place, a person or a dataset), and overview information of the links that exist between the selected item and entities, such as places or people. The selection panel is specific to the type of item; hence there are different layouts for presenting information about objects (cf. Fig. 1), places, people or datasets (Fig. 2).



**Fig.2. Selection box layouts for people (left), places (centre), datasets (right).**

The map shows red markers for places where the results are located. It is important to note that a single dot in Peripleo does **not** generally represent one search result, as is usually the case in search interfaces of a similar graphical design. Instead, there is an n-to-m relationship between markers and results: on the one hand, one marker will usually represent many results located at a place. Peripleo indicates the number of results by the size of the marker. Clicking the marker will select the first search result, and provide additional information, including the number of records at this place and the option to filter the result list. On the other hand, one object may be related to multiple places - even hundreds of places, e.g. in the case of a geo-annotated text (see Fig. 3 for an example).



**Fig.3.** A single result item linking to many places (in this case, a TEI document where mentions of places were annotated with gazetteer references).

Finally, a small group of icons to at the top-right allows the user to:

- zoom the map
- choose between different base layers (presently an “empty” map - cf. Fig.1; a basemap of places and road network from the Roman Empire;<sup>8</sup> OpenStreetMap; and a Mapbox<sup>9</sup> aerial imagery layer)
- toggle filtering by viewport (if selected, results as well as facet information and temporal distribution will be shown for just the results in the currently viewed map area, not all results in the system as a whole)

### 3.2 Technical Architecture

In terms of the technical architecture, Peripleo is a Web application based on a standard 3-tier application model. It is implemented on a Java Virtual Machine (JVM) technology stack, using the *Play* Web application framework,<sup>10</sup> and the Scala programming language as a basis for the “middle tier”. This tier encompasses the application server components that implement core services, e.g. those needed for data import, indexing, querying, etc., and provides the necessary APIs to drive the browser user interface. In terms of technology, Play is based on state-of-the-art architecture concepts including RESTful APIs and rigorous decoupling of components through dependency injection. Furthermore, Play is built from the ground up as a framework for reactive

<sup>8</sup> <http://dare.ht.lu.se/>

<sup>9</sup> <https://www.mapbox.com/>

<sup>10</sup> <https://www.playframework.com/>

applications [4], a new Web application paradigm that makes more efficient use of computing resources and, in combination with other architectural measures, generally leads to more scalable and maintainable applications. Scala, in addition, is a language that complements traditional object-oriented paradigms found in languages like Java with elements of functional programming, a modern programming paradigm that favours, among other things, immutable data structures and stateless design, and is thus an excellent match for modern, reactive Web applications (cf. [1]).

The data tier is implemented with ElasticSearch, an open source document store and full text index. Peripleo uses ElasticSearch's built-in parent-child relationship feature to keep a loosely-connected model of items and the connections between them. This provides for more flexible querying functionality, especially when these connections are tied to additional metadata: e.g. a specific relationship type (an object's findspot, or current location vs. a place mentioned in literature or a waypoint on a historical itinerary), or the context the connection appears in (e.g. a text snippet in literature, or an image region from a manuscript mentioning a specific person).

The presentation tier is implemented as a client-side JavaScript/HTML5 single page application, built on top of the Peripleo API. It makes use of a range of general-purpose open source utility libraries (such as RequireJS<sup>11</sup> and jQuery<sup>12</sup>) and JavaScript user interface component frameworks for specific purposes (such as Leaflet<sup>13</sup> for mapping, Typeahead.js<sup>14</sup> for autosuggest rendering, Velocity.js<sup>15</sup> for smooth animations, etc.)

### 3.3 Enrichments & Conflation: Integrating Heterogeneous Sources

Another unique feature of Peripleo not yet discussed in Section 3.2 relates the distinction between records and items, and how Peripleo deals with overlapping data coming from different sources. Let's take, as an example, the city of Rome. Peripleo harvests information about places from *gazetteers* - such as aforementioned Pleiades, GeoNames,<sup>16</sup> and the Digital Atlas of the Roman Empire<sup>17</sup> (DARE) or, in the case of this Deliverable, from Europeana. For obvious reasons, the city of Rome is listed in each of these gazetteers; and many gazetteers also include correspondence information: for example, the DARE record for Rome includes a skos:exactMatch relation to the Pleiades record for Rome, as well as a skos:closeMatch relation to the GeoNames record. The Europeana record for Rome, on the other hand, carries an owl:sameAs relation to the GeoNames record. Using this information, Peripleo can conflate all these records into one object, representing the *Place* in Peripleo terminology. Fig. 4 illustrates this example, depicting the various connections between the records for Rome in different gazetteers.

---

<sup>11</sup> <http://requirejs.org/>

<sup>12</sup> <https://jquery.com/>

<sup>13</sup> <http://leafletjs.com/>

<sup>14</sup> <https://twitter.github.io/typeahead.js/>

<sup>15</sup> <http://velocityjs.org/>

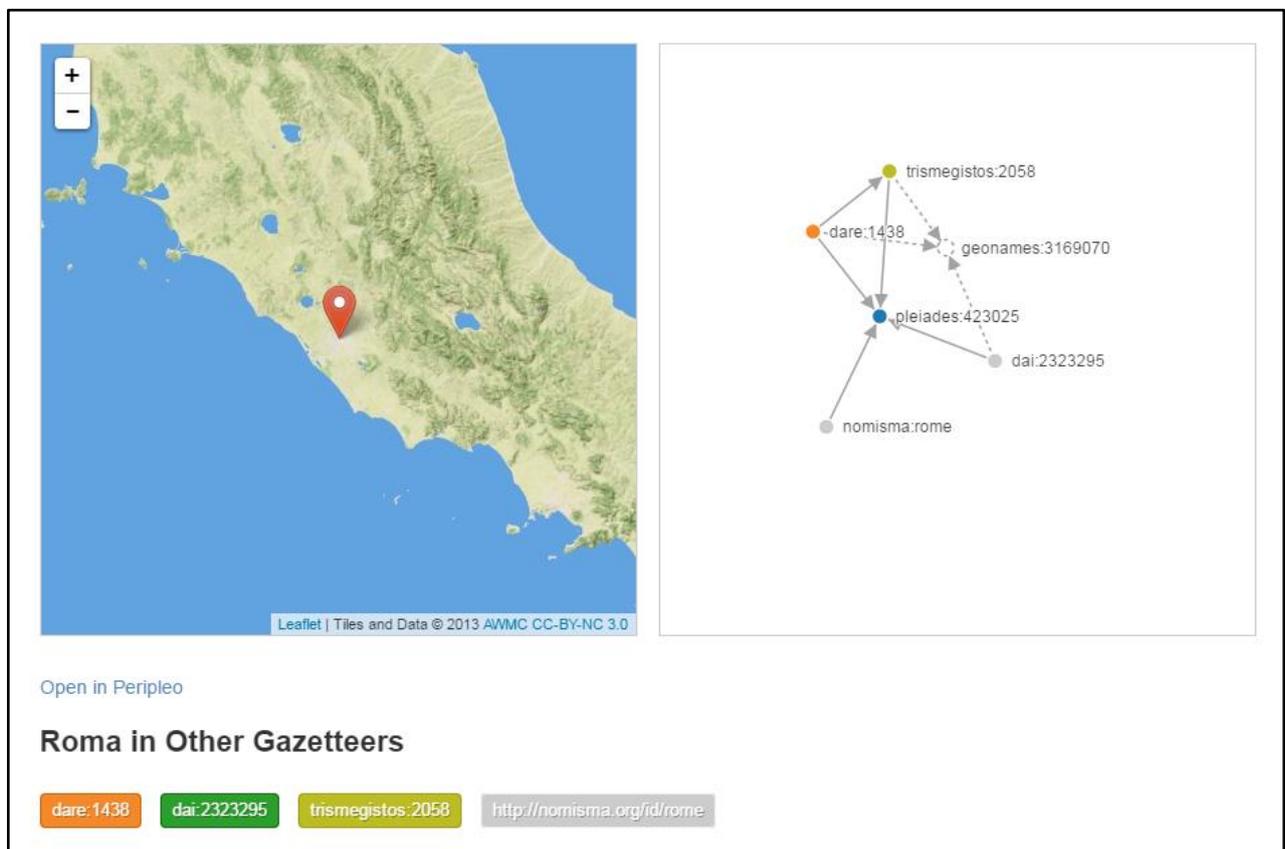
<sup>16</sup> <http://www.geonames.org/>

<sup>17</sup> <http://dare.ht.lu.se/>

It is worth noting that the term “conflation” should not be interpreted in the sense that Peripleo creates a hybrid record, mixing data from different sources. Peripleo will maintain the information from each record in its original form. The “Item” is merely a relation-aware grouping mechanism for multiple records that are linked to each other via `skos:closeMatch`, `skos:exactMatch` or `owl:sameAs` relations.

Over the course of this deliverable, Peripleo was extended to apply this conflation mechanism not just to Places, but to all types of Items. I.e. records for people, time periods and even objects are now handled the same way. The ingest process that has been implemented takes care that:

- items are properly updated when ingesting a new set of records
- their relation networks are properly merged - or split - when a new set of records is imported or an existing record is updated with new data.
- references that other items hold to the updated item (e.g. a references to a place, held by a metadata record) are properly re-written in case of merges or splits.



**Fig.4.** Records for *Rome* in different gazetteers (Pleiades, DARE, iDAI.gazetteer,<sup>18</sup> Trismegistos<sup>19</sup> and nomisma.org<sup>20</sup>) and how they relate to each other via skos relations. Solid lines are connections between gazetteers in Peripleo. Dashed lines are indirect connections, based on shared skos relations (i.e. both records carry an identical co-reference).

<sup>18</sup> <https://gazetteer.dainst.org>

<sup>19</sup> <http://www.trismegistos.org/>

<sup>20</sup> <http://nomisma.org/>



## 4. EDSI2 Demo

For demonstration purposes, a dedicated Peripleo instance was set up on a Virtual Machine hosted at AIT.<sup>21</sup> The purpose of this demonstrator is to showcase the integration of Europeana resources (i.e. records and entities) with related datasets from Pelagios Commons.

A technical and administrative decision had to be taken for the Pilot implementation, with regard to the scope of the Demo application. In discussions, two possible scenarios of using Peripleo were debated:

1. **Dedicated Europeana instance.** A dedicated deployment of a Peripleo instance, using Europeana data and open Datasets, maintained and operated by EDSI partners.
2. **Integration with Pelagios instance.** Alternatively, it was discussed whether Europeana data could be imported into the “official” Peripleo instance of Pelagios Commons (scheduled to go operational in December).

We concluded that the second solution is associated with higher risks for three reasons:

- lack of consistent coverage with structured identifiers in the Europeana sample data meant that navigation in Peripleo - which rests entirely on the existence of curated links - cannot work to its full potential;
- the sheer quantity of data in Europeana (52 million records) risks eclipsing the whole data provided by the entire Pelagios Commons network at the moment (~1 million objects). This situation is likely not to be accepted by Pelagios stakeholders;
- in terms of technology, Peripleo has not been tested with a dataset of this size either. The foundational technologies used in the system (ElasticSearch, reactive programming techniques) provide all reason to expect that Peripleo will scale to this dimension. However it is unclear whether extra hardware measures will be required (e.g. additional servers to run an ElasticSearch cluster, SSD hard drives rather than spinning disks), and who would provide hosting for a joint instance.

Consequently, **option 1** is clearly the least risk path forward, especially given the fact that there are no resources allocated for this task beyond this deliverable. Furthermore, by maintaining a dedicated Peripleo instance, more data can be introduced gradually, adding curated datasets step by step, as methods for crosswalking and enriching Europeana data become more mature. It also means that Europeana retains full control of the application, as opposed to putting it under the stewardship of Pelagios Commons. In the future, a shared model might be conceivable as an alternative, where Pelagios Commons assists integration by maintaining two separate instances on behalf of Europeana.

Last but not least, another option we argue should be considered for the future relates to the use of Peripleo through its JSON API. Using the API, Peripleo can not only be used easily for finding alternative enrichments based on URIs. It can also retrieve statistics, at high speed, about the enrichments contained in a specific collection: e.g. how many records are in a dataset? Which distinct places or people are being referenced by records (and how often)? What is the temporal distribution of the records? etc. Our experience in Pelagios Commons has shown that Peripleo makes structural peculiarities, outliers or curation errors in a dataset visible immediately. Looking at their data through Peripleo for the first time provides at-a-glance insights into data that even collection holders themselves often did not have so far.

---

<sup>21</sup> <http://dsi-demo.ait.ac.at/>

**Table 1: Entities Overview**

Data Source	Entity Type	# Items	Co-references	Note
Europeana Entity Collection	Person	35,865	DBpedia, Wikidata, Freebase, Yago	Includes Person Entities used in Europeana Records
Europeana Entity Collection	Place	75,024	GeoNames	Includes Place Entities used in Europeana Records
Pleiades <sup>22</sup>	Place	35,041	-	A community-built gazetteer and graph of ancient places. It publishes authoritative information about ancient places and spaces.
Digital Atlas of the Roman Empire <sup>23</sup>	Place	26,423	GeoNames, Pleiades, Wikipedia, Wikidata	Coverage of co-references to other authorities varies
GeoNames (Pelagios subset <sup>24</sup> )	Place	152,249	Wikidata	Wikidata concordances are not comprehensive, dataset includes 87,555

The Semantic Entities is the core technology supporting the representation of database records in linked data formats and enabling semantic search functionality. The existing semantic vocabularies represent a good resource for generating data links between Europeana records and external sets of Cultural Heritage objects. Table 1 presents an overview of the gazetteers, and person's vocabularies available within the demo application.

In order to demonstrate the spatial-temporal discovery functionality, as the core functionality of Peripleo, several datasets that contain references to the upper Entity sets were ingested in the application. These datasets contain a subset of records describing cultural heritage objects available in Europeana and Pelagios. The overview of these datasets is presented in Table 2.

**Table 2: Records Overview**

Data source	Type of content	# Items	Note
Europeana (sculptures subset)	Rome, Sculpture	164	A set of photographs representing different sculptures from Rome
Europeana (monuments subset)	Rome, Monument	969	A set of historical buildings and monuments from Rome
Europeana (inscriptions subset)	Rome, Monument, Inscription	237	A set of inscriptions on Monuments in Rome

<sup>22</sup> <http://pleiades.stoa.org/>

<sup>23</sup> <http://dare.ht.lu.se/>

<sup>24</sup> <https://github.com/pelagios/recogito2-places-geonames>

Europeana (roman numismatics)	Roman, Coin, Münze	5,028	A numismatic data set from roman empire
Europeana (roman inscriptions)	Rome, Inscription, Monument	440	A set of records including inscriptions on monuments in Rome
University of Graz <sup>25</sup>	Coins, Inscriptions, Archaeology	4,501	Three sub-collections (epigraphy, numismatics, archaeology) of objects linked to 300+ places across the Mediterranean basin and the region of the Roman Empire, dated between 5300 BC and 1500 AD. Highest density in Rome (1.8k+ objects)
American Numismatic Society <sup>26</sup>	Coins	140,955	Coin finds from 500+ places across the Mediterranean and the region of the Roman Empire. Highest density in Rome (24k+ objects) and Alexandria (17k+ objects)
Epigraphic Database Heidelberg <sup>27</sup>	Inscriptions	55,736	Inscriptions from 500+ places in, predominantly, the region of the Roman Empire, dated between 530 BC and 740 AD.

## 4.1 Integrating Places

Europeana place entities were integrated with several gazetteers from the Pelagios network (see Table 1). Every Europeana place entity in the sample set included an owl:sameAs correspondence to GeoNames.

- **Pleiades** was chosen for this pilot since it is currently the most interconnected gazetteer in Pelagios Commons, i.e. many data providers link to places in Pleiades, and a large number of records exists that can be used for search examples in the pilot.
- A subset of **GeoNames** was integrated as a general-purpose gazetteer with global coverage, and to act as a “bridge” between Europeana place entities and those common in the Pelagios partner network.
- The **Digital Atlas of the Roman Empire** was included due to its high coverage of Pleiades-to-GeoNames correspondences. The connections between Pleiades and GeoNames that are facilitated by DARE are crucial in order to connect records linked to Europeana place IDs to external object records in Pelagios.

It was necessary to implement a custom crosswalk in order to be able to integrate Europeana place entities. This crosswalk was implemented in two stages: first, an external script - outside of Peripleo - was implemented to read individual Europeana records [5] from JSON files, normalize URIs, field names and field values, and then produce a single streamable import file (in newline-

<sup>25</sup> <http://gams.uni-graz.at/>

<sup>26</sup> <http://numismatics.org/search/>

<sup>27</sup> <http://edh-www.adw.uni-heidelberg.de/>

delimited JSON format). Second, Peripleo's built-in importer was extended to stream the contents of this import file into the backend, performing conflation as part of the process.

## 4.2 Integrating Agents

The Europeana Agents dataset (see Table 1) was the first non-gazetteer authority dataset ever to be integrated into Peripleo. Before any data could be integrated, it was therefore necessary to first extend Peripleo itself, so it can deal with types of authority records other than Places. Furthermore, UI components needed to be adapted and extended in order to present person records appropriately. The results of this work have been integrated into the official Peripleo source code repository and it will be available in the following releases. It is therefore expected that the wider community of Pelagios Commons will be able to make use of this extension in the future.

Following the development of necessary Peripleo extensions, the data ingest for the pilot was implemented in the same way as for Place data: individual agent records were downloaded from Europeana Entity Collection, and a crosswalk was developed in order to extract relevant fields (title, description, names, depiction, birth and death date) and perform minor data normalization (URIs, field values). The resulting streamable import file was then ingested into Peripleo demo.

## 4.3 Integrating Object Records

Integrating Europeana object records proved slightly more challenging. The subsets of Europeana records introduced in Table 2 were fetched by using Europeana search API<sup>28</sup>. The main criterion for selecting these subsets is the probability of having a thematic overlap with Pelagios collections (e.g. records related to same places, same historical periods, describing same or related types of content).

While some metadata records that were chosen for the pilot contained proper references to Places (in the form of URIs to Europeana Place entities), many didn't have place references, or didn't present the expected level of granularity (i.e. a large number of Place references are pointing only to the country of origin, but not to a specific location, like a town or city). However, the locations indicated as free text are clearly identifying a specific place (e.g. *Latium et Campania cum insulis (Regio I)*, *Ostia, Ostia antica (Roma)*, *Italia* would be ideally referenced by the Ostia Antica<sup>29</sup> place entity). In addition, the most metadata records did not contain structured references to people using Europeana agent URIs. In order to set up the pilot, it was therefore necessary to carry out some degree of manual enrichment. Within the follow up DSIs it is expected that both the size of the Entity Collection and the number of Entity references in Europeana records will be increased, improving so the discoverability and filtering functionality in Peripleo.

In terms of methodology, specific records were downloaded using specific Europeana search queries. Where extra enrichment was needed, queries were restricted by place search in Europeana, and a structured enrichment was added as part of the Peripleo crosswalk.

Except for the extra enrichment step, the integration procedure was identical to the procedure followed for Places and Agents, i.e. a two-stage import with external normalization following ingest from a streamable dump file.

---

<sup>28</sup> <http://labs.europeana.eu/api/search#>

<sup>29</sup> ["http://data.europeana.eu/place/base/143577"](http://data.europeana.eu/place/base/143577)

## 4.4 External Data Sources

External data sources, used within the demo, are openly licensed data from members of the Pelagios partner network. Pelagios' standard exchange format for object record data is RDF-based<sup>30</sup> and covers three major types of information: i) basic metadata properties (title, descriptions, identifiers, depictions, date range); ii) entity links (e.g. links to gazetteers); iii) correspondences (e.g. `skos:closeMatch` or `skos:exactMatch` relations).

Integration of data from Pelagios Commons is generally much more straightforward, since the community:

- takes care of properly curating place references in their data. As Pelagios generally deals with specialist scholarly content, curation is often a manual and highly-quality controlled process. Unattended automatic enrichment from unstructured metadata is the exception rather than the norm.
- actively maintains dump files which can be (re-)harvested into Peripleo easily, and updated periodically if desired. Support for importing this format is natively built into Peripleo. Therefore no extra conversion steps (except for minor sanity checks and URI normalization) are required.

---

<sup>30</sup> <https://github.com/pelagios/pelagios-cookbook/wiki/Joining-Pelagios>

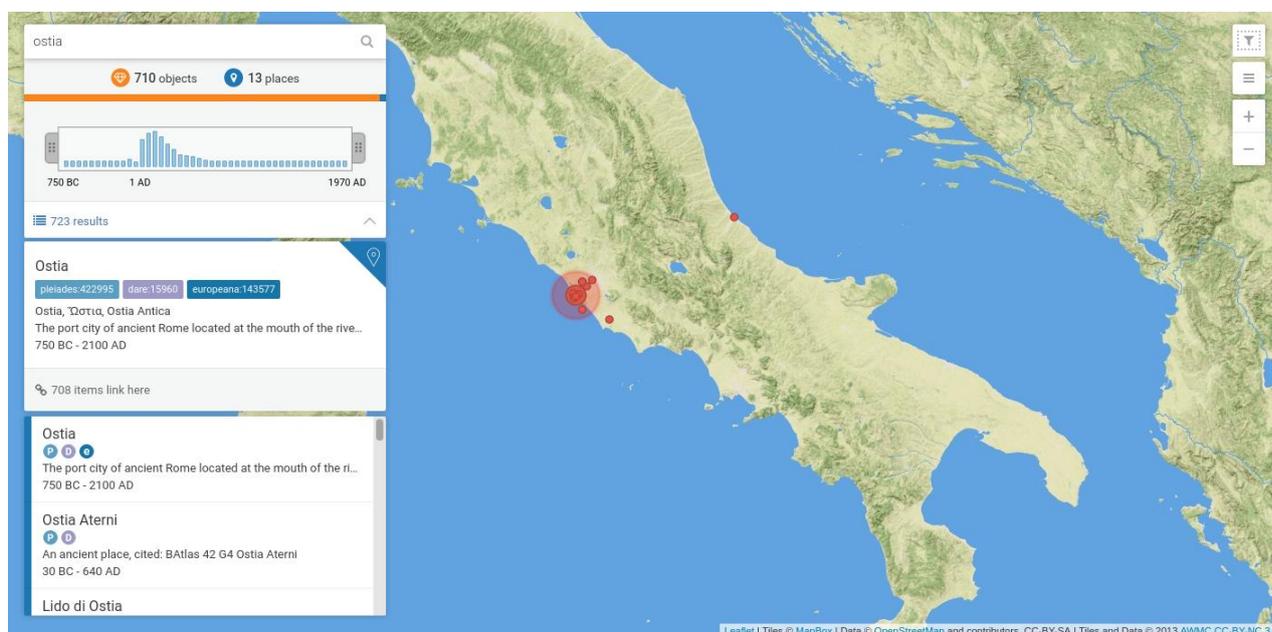
## 5. Exemplary Search Scenarios

In this section, we present a number of exemplary search scenarios that emphasize the key benefit of integrating Europeana data with Peripleo: the ability to link non-Europeana data with resources from other sources; and thus the possibility to create bridges to different domains.

### 5.1 Searching items by place

In this scenario, the user is interested in finding objects that relate to the ancient place of *Ostia*, Italy. In this case, the user starts by typing the phrase *ostia* into the search box. Peripleo returns a list of results (places and objects), with the place *ostia* as the highest-ranked result (see Fig. 5).

Selecting the place from the result list brings up additional information. Most importantly, Peripleo lists the gazetteers it holds information from. In this case, the user interface shows us that there are records for Ostia in Pleiades<sup>31</sup>, the Digital Atlas of the Roman Empire<sup>32</sup> - and Europeana.<sup>33</sup> Thanks to the enrichments, these records have become connected inside Peripleo; and any object linked to one record (e.g. Pleiades) is now also linked to all other records (DARE and Europeana).



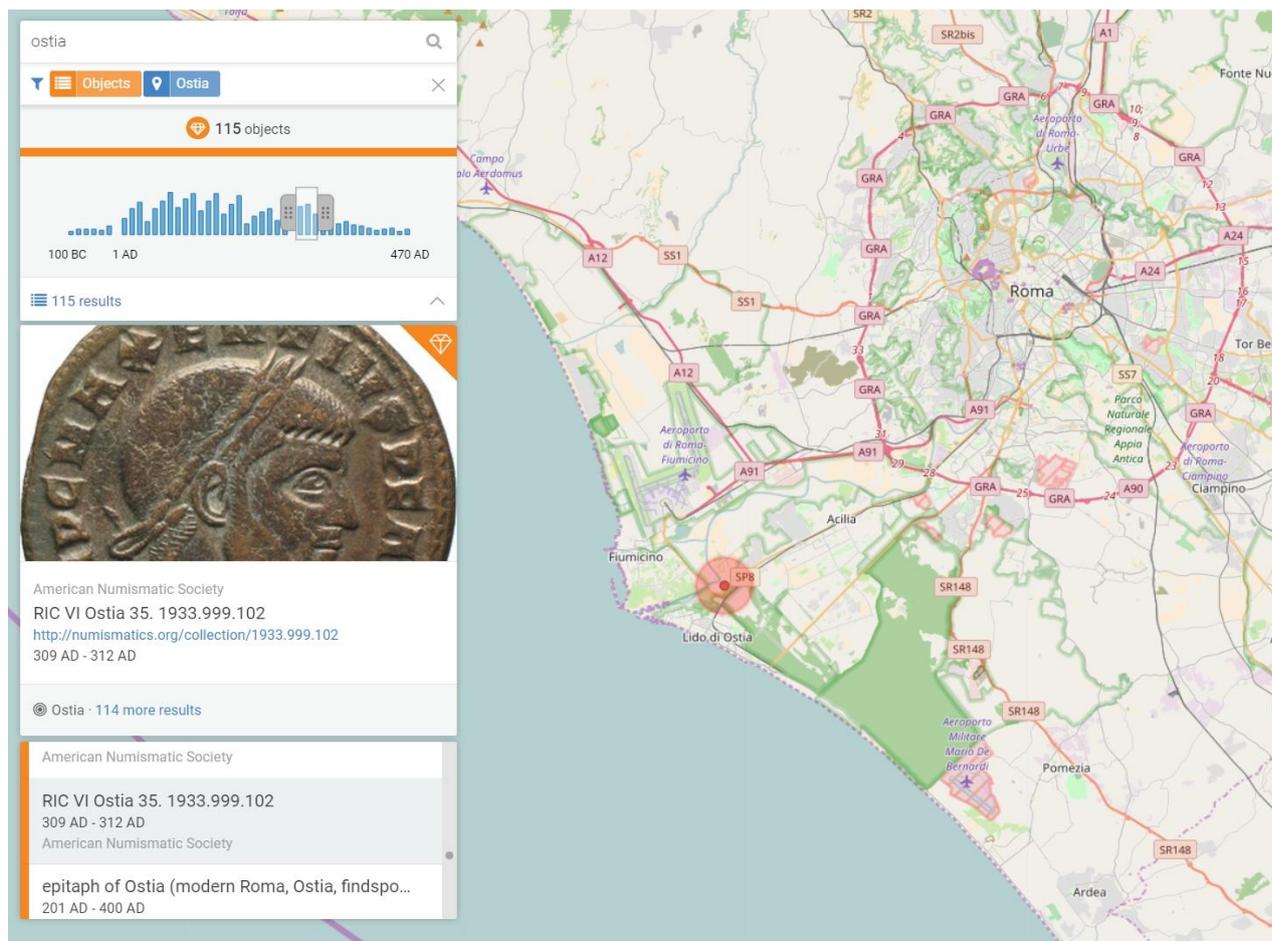
**Fig.5. Searching for ‘Ostia’ returns the combined place record: the Europeana entity is connected to its corresponding counterparts in other gazetteers: Pleiades, and the Digital Atlas of the Roman Empire.**

The interface also shows the user that Peripleo has 708 items linked Ostia (see Fig. 5). By clicking the link, the user can move into a filtered search, showing these 708 objects. A key point to note is that, due to the connections across gazetteers in Peripleo, the results include content from Europeana, as well as from non-Europeana sources, such as the American Numismatic Society (see Fig. 6).

<sup>31</sup> <http://pleiades.stoa.org/places/422995>

<sup>32</sup> <http://dare.ht.lu.se/places/15960>

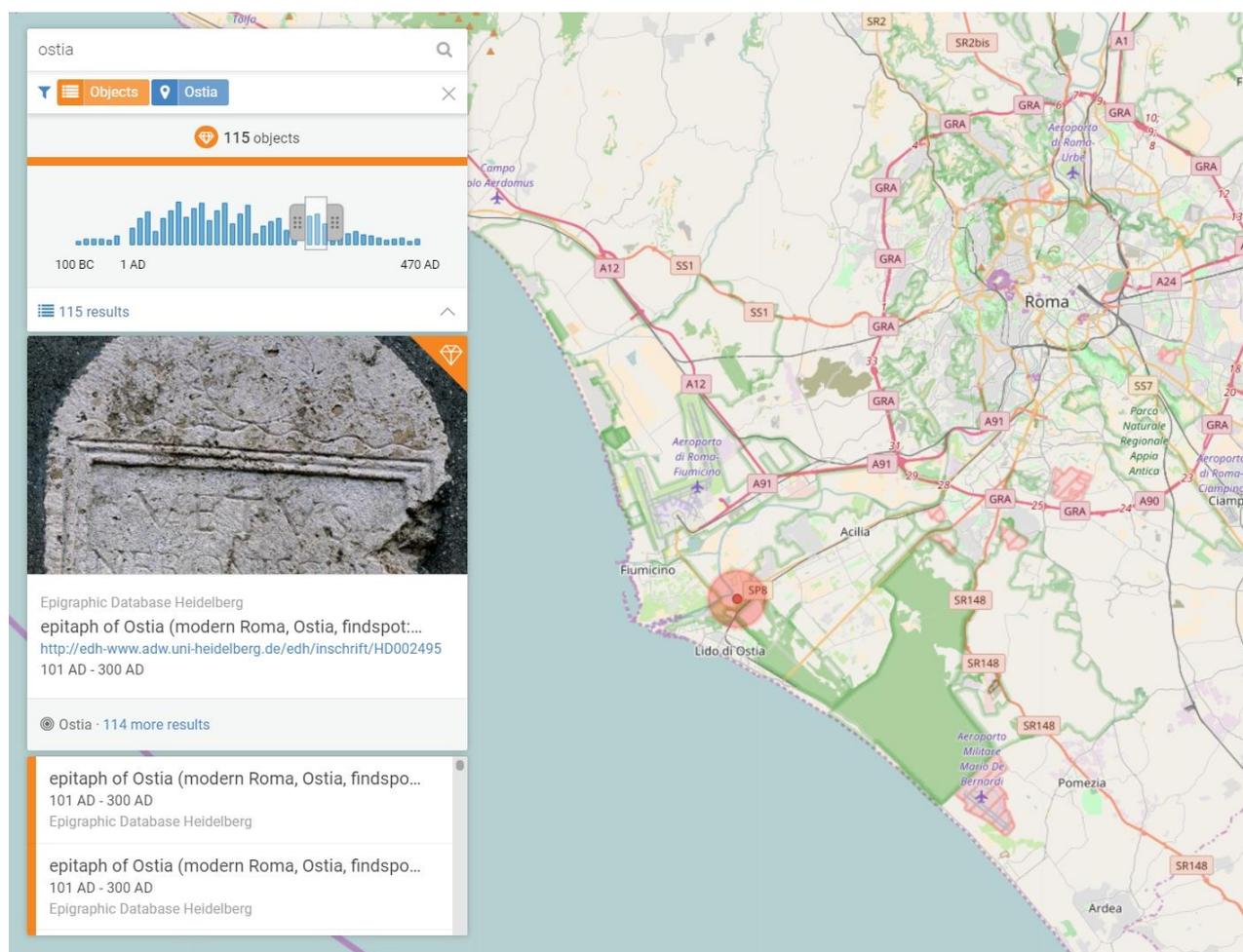
<sup>33</sup> <http://data.europeana.eu/place/base/143577>



**Fig.6. Filtering the search to Ostia enables the user to see all content linked to this place. Note that non-European objects - such as a coin record from the American Numismatic Society collection - are now connected to the Europeana place record for Ostia. Additional filter options can be applied, such as by type ('Objects' only) or time range.**

It is also possible to apply additional filters to narrow down the search further (e.g. by setting a time range with the sliders on the time histogram, see Fig. 6); or to switch between different basemaps.

Fig. 7 shows another example result from a filtered search on Ostia, in this case the record for an inscription from the Epigraphic Database Heidelberg.



**Fig.7. Alternative example for filtered search (type *Object*, place *Ostia*, restricted time range), with a selected inscription record from the Epigraphic Database Heidelberg.**

## 5.2 Exploring content across places

Scenario 5.1 showed one important “entry path” through which users can start exploring content in Peripleo: by searching for a place. Another typical scenario is that the user will be interested in a specific topic, object or, generally, search phrase rather than a specific place. In this scenario, the user searches for the phrase *antoninian*<sup>34</sup>, a particular type of Roman coin.

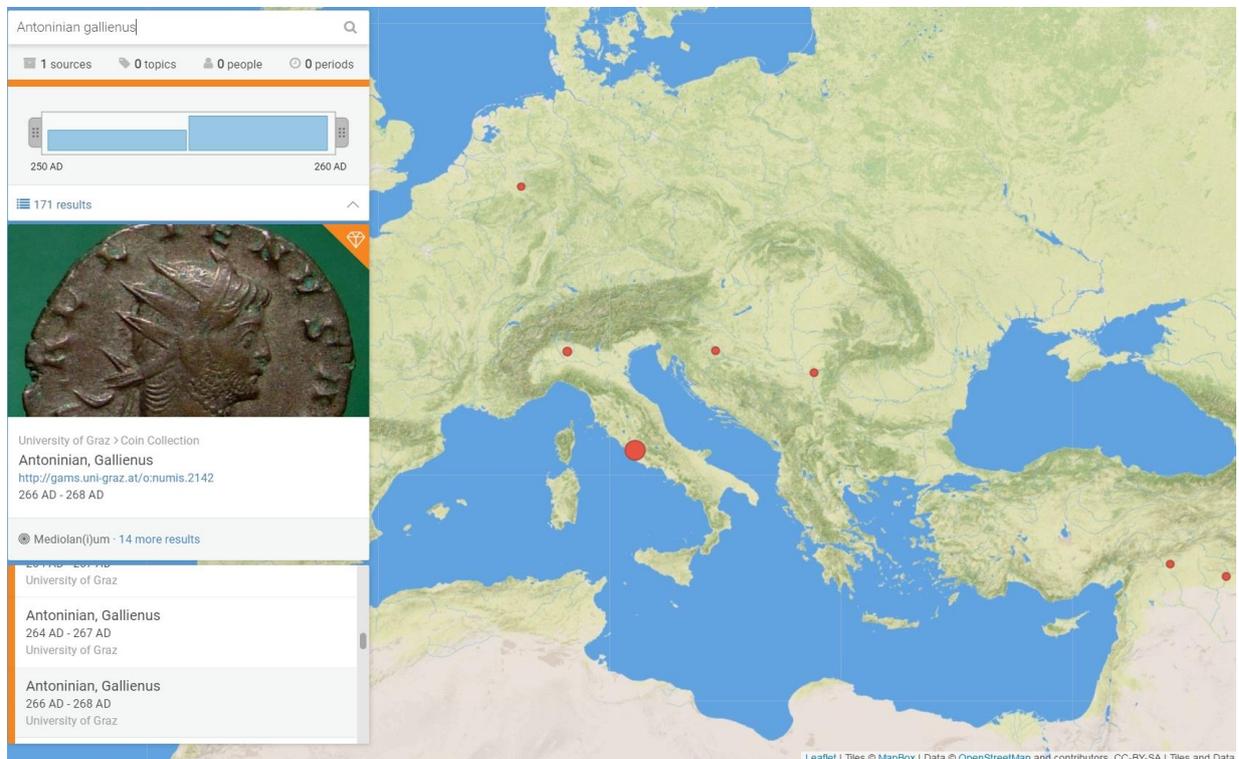
Fig. 8 shows the corresponding search result in Peripleo: 712 results are returned, distributed across a number of places (Rome being the most prominent location with the highest number of results); and across a time period between 230 AD and 290 AD.

As in the previous case, the scenario shows how the search result combines content from both Europeana and non-Europeana sources.

<sup>34</sup> see <https://en.wikipedia.org/wiki/Antoninianus>



**Fig.8. Distribution across space, time and data sources for the search term *antoninian*.**

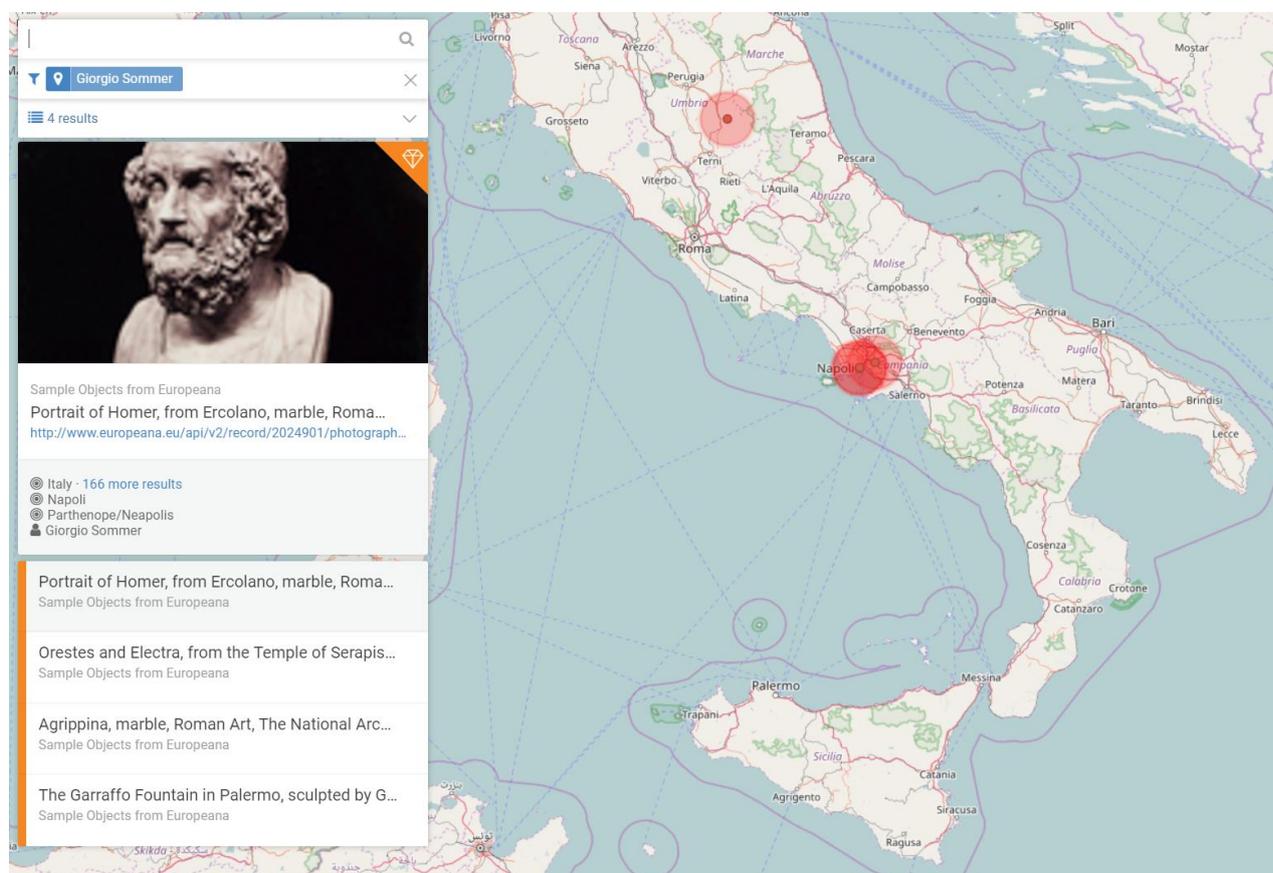


**Fig.9. Alternative search example using a more refined search term (*Antoninian Gallienus*).**

### 5.3 Searching by person

A third major navigation path for users of Peripleo is to explore via connections to *people*. For example, users could search for the name of a person directly, or search for a specific object (e.g. a work of art) and then explore further the connections to people linked to it.

In Fig. 10, the user might, for example, have searched for the term *portrait* and stumbled upon the Portrait of Homer, available in Europeana through the photography taken by Giorgio Sommer<sup>35</sup>. The object is linked to the authority record of Giorgio Sommer via the corresponding entity URI.<sup>36</sup> Entering a filtered search by a single click of a button, the user can explore all other items by Giorgio Sommer, as well as gain an instant overview of the relevant places and time period of his work.



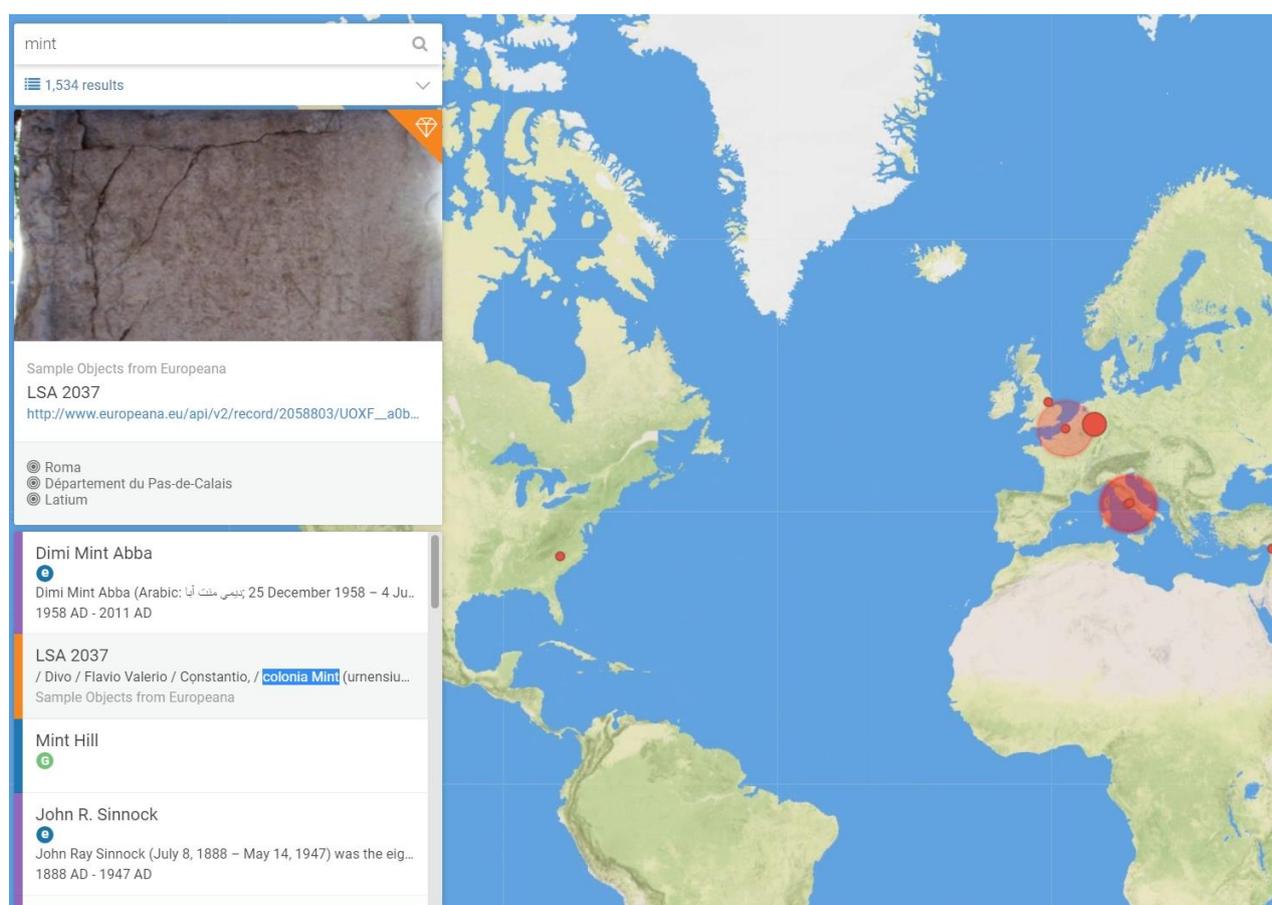
**Fig.10. Navigating by person: content linked to the artist *Giorgio Sommer*.**

<sup>35</sup> [https://en.wikipedia.org/wiki/Giorgio\\_Sommer](https://en.wikipedia.org/wiki/Giorgio_Sommer)

<sup>36</sup> <http://data.europeana.eu/agent/base/5486>

## 5.4 Exploratory browsing, discover loose relations or serendipitous content

Another common way of user interaction with a large repository is based on exploratory browsing and discovery pattern. This demonstrates the *“the design of systems that support users to explore and experience, to learn and be amazed, and engage with the material from many different angles”*<sup>37</sup>. A sample scenario illustrated in the following Figures starts from the common word *mint*. While the most common use of the word refers to the mint plant, the other common use refers to coin manufacturing<sup>38</sup>. When using it as search term in the demo application, one can discover different entities and items that include the word in their descriptions (see Figure 11) like: places that contain the word in their official names (e.g. Mint Colony, Mint Hill, Mint Wall), coins (i.e. as the term relates to mint state in numismatics) or persons that are related to the term through their occupation (see Figure 12).



**Fig.11. Exploratory browsing: content retrieved by using the term *mint*.**

As the mint facilities became important institutions for modern states, several different professions or occupations got the mint term in their names. The Peripleo free text search works against the title and descriptions of all items and entities found in the database. When applying the type filter to retrieve only persons in the results list, a couple of uncommon or unexpected personalities appear in the list. However, there are good explanations of such surprising results. From the list of 19 agents that are currently available in the demo database and contain the term *mint* in their free text descriptions, three representative cases are presented in Figure 12. *Lajos*

<sup>37</sup> <http://pro.europeana.eu/page/issue-4-discovery>

<sup>38</sup> <https://en.wikipedia.org/wiki/Mint>

*Beran* was a Hungarian artist and a coin engraver, which worked as art director for the State Mint in Hungary<sup>39</sup>. Sir *Isaac Newton* is well known for his scientific work, however in the last part of his life, when he lived in London he served for more than 30 years as Warden and Master of the Royal Mint. The third example is the *Fyodor Petrovich Tolstoy*<sup>40</sup>, is a representative of the famous and powerful Tolstoy family. For 40 years he was Vice-President of the Royal Academy of Arts in Russia. But before that he worked for the Department of Mint and became the founder of the medal working in Russia.

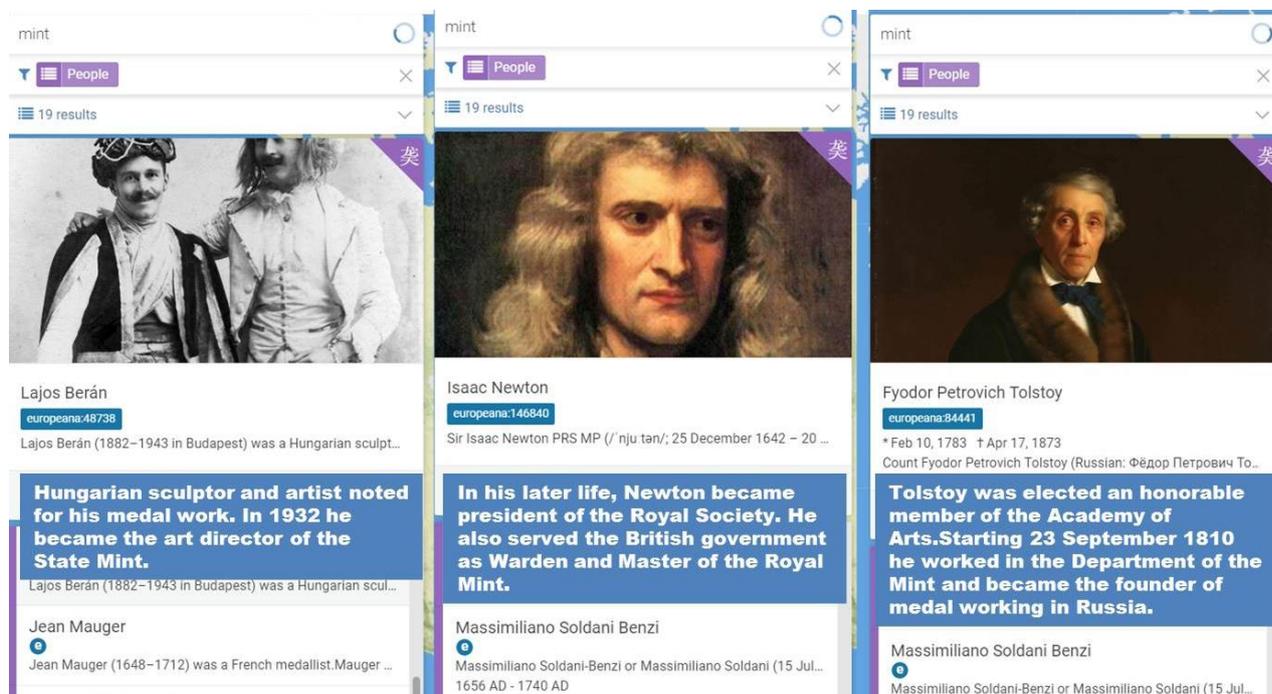


Fig.12. Exploratory browsing: unexpected occupations of famous people referring the term *mint*.

## 6. Conclusion

This Deliverable describes the pilot for time and place based discovery, which is based on Peripleo tool, a spatio-temporal search engine developed by the Pelagios project. The demonstrator integrates sample resources from Europeana and Pelagios platforms.

A set of positive aspects regarding the feasibility of implementing spatio-temporal discoverability services based using existing metadata from Europeana records were identified during the setup of the demonstrator, as follows:

- Usage of Entities.** It is fairly straightforward to integrate Europeana entity data - Places and People - into Peripleo. This is potentially a significant aspect for two reasons: on the one hand, the simple interlinking that currently exists between Europeana and GeoNames provides an entry point into a whole chain of Linked Data from other gazetteers in the Digital Humanities space as well. This extra information could be harvested into Europeana in the future as a means to provide further sources of enrichment.

<sup>39</sup> [https://en.wikipedia.org/wiki/Lajos\\_Ber%C3%A1n](https://en.wikipedia.org/wiki/Lajos_Ber%C3%A1n)

<sup>40</sup> [https://en.wikipedia.org/wiki/Fyodor\\_Petrovich\\_Tolstoy](https://en.wikipedia.org/wiki/Fyodor_Petrovich_Tolstoy)

- **Alternative ways of Exploration.** Peripleo, while still under active development, is a fairly suitable platform for enabling alternative modes of exploration. Instead of focusing on search for specific items, Peripleo supports an exploratory mode of navigating “collections of collections” as a whole and getting an idea of “what’s inside” - in terms of spatial footprint, temporal distribution, and - albeit at the time of writing to a lesser extent - thematic composition.

However, the pilot also revealed a number of caveats that made integration of Europeana difficult in some cases. We expect that the usage of Europeana Entity Collection within the Europeana records will be enhanced in the near future. A good coverage of semantic enrichments over the metadata records will facilitate in the first instance the support for semantic search queries through the search API and will also enhance the automatic discoverability of additional resources from linked data. However, for setting up a fully functional demonstrator, this challenge could only be addressed by resorting to additional manual enrichments performed on selected content. The following data processing steps required couldn’t be performed automatically as expert knowledge was required to interpret textual data and manually map them to semantic entities or concepts.

- **Disambiguation issues.** Metadata for places and time are not (always) normalized in Europeana and often represented as text only. Ambiguity in identifiers is another issue that has proven challenging: for example, GeoNames - being built as a gazetteer of administrative units - will frequently have different identifiers for a city and its administrative region (e.g. in the case of Vienna, there is both a record for the city and the “Bundesland” - the federal state) Unless a gazetteer in Peripleo includes a relation between the two (e.g. skos:closeMatch), records linked to the one will not show for the other, and vice versa. While technically correct, will usually go against user expectation.
- **User expectations/experience.** A low coverage, and interlinking of logically related entities, can significantly impact the user experience, as objects that might be relevant to a search may be excluded from the result set. Also many Europeana items do not have time references; therefore the time filter automatically eliminates the item for which the metadata field year is not populated. We expect that this challenge will be highly collection-specific. I.e. some Europeana collections will be more affected than others (due to different data curation practices and quality levels). This would significantly reduce their visibility in Peripleo. A systematic investigation into the effects on user experience is beyond the scope of this deliverable, and would have significantly exceeded the effort available. However, it could be a potential question for future work.
- **Semantic categorization.** There is currently a lack of semantic categorizations. Lists of localized concept tags, in different languages, are often used instead of URIs. Peripleo does provide built-in support for semantic category tags in principle. However this also a new development in Peripleo 2, and not currently used in live Pelagios data. In that sense, it is a challenge that confronts both Europeana and Pelagios Commons at the same time - and thus a potential area of future collaboration.

Another area of future work is time periods: Pelagios Commons has recently started to advocate the use of PeriodO,<sup>41</sup> a gazetteer of period definitions for linking and visualizing data; and first steps have been made to integrate PeriodO into Peripleo as well.<sup>42</sup> Coordinating practices for temporal tagging in Europeana could open up new ways for interlinking with the wider Digital Humanities community as well.

---

<sup>41</sup> <http://perio.do/>

<sup>42</sup> <https://twitter.com/aboutgeo/status/877423514571202561>

## 7. References

- [1] Rainer Simon, Elton Barker, Leif Isaksen, Pau de Soto Cañamares. 2017. Linked Data Annotation Without the Pointy Brackets: Introducing Recogito 2. In Journal of Map & Geography Libraries, vol. 13, issue 1, pp. 111-132.
- [2] Rainer Simon, Leif Isaksen, Elton Barker, Pau de Soto Cañamares. 2016. Peripleo: a Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time. In Code4Lib Issue 31 (2016-01-28). ISSN 1940-5758.
- [3] Rainer Simon, Elton Barker, Pau de Soto Cañamares, Leif Isaksen. 2014. Pelagios. In ISAW Papers 7.27. <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/simon-barker-%20desoto-isaksen/>
- [4] Manuel Bernhardt. 2016. Reactive Web applications. Manning Publications.
- [5] Europeana Data Model – Mapping Guidelines v2.1 (2014)  
<http://travesia.mcu.es/portalnb/jspui/bitstream/10421/5982/1/EDM%20Mapping%20Guidelines%20v2.1.pdf>