

# Pelagios - Connecting histories of place.

## Part I: Methods and Tools.

**Valeria Vitale, Pau de Soto Cañamares, Rainer Simon, Elton Barker, Leif Isaksen and  
Rebecca Kahn**

### **Autobiographical note**

The authors are members of the investigative teams of seven funded projects which collectively comprised the Pelagios initiative prior to its establishment as the Pelagios Network Association in 2019.

### **Abstract**

This paper provides a short history of the methods and tools developed by the Pelagios initiative: a series of seven projects dedicated to linking digital historical resources based on the geographic places to which they relate and refer. The first section of the paper situates the work within the wider field of semantic and geospatial technologies. It then describes the initial development of the Pelagios methodology through an open consortium of ancient Mediterranean data resource providers, in association with the Pleiades Gazetteer of the Ancient World. The second section details how this approach was expanded for application in other historical and archaeological periods and domains, with a special focus on tool development. In particular, we reflect on the

development and application of *Recogito*, an online semantic geoannotation platform that allows anyone to identify place references within historical texts, imagery or datasets.

### ***Prologue: A View from 2027***

*As we settle into our chairs for the monthly meeting of the Early Islamic Interest Group, some with a second cup of tea in hand, the Chair's voice assistant brings up the 'Pelagios Network spatial feed'. An expansive view of the Mediterranean appears on our screens, with current weather patterns, the latest research publications and colleagues' avatars briefly appearing above their respective locations—Egypt, Tunisia, Italy and Cyprus among them. These are, however, quickly replaced by a layer of content reflecting the main business of the day: a regional museum in Libya has digitally published their numismatic collections with associated spatial data. Reem Fathi, the museum's curator, introduces her team's work and begins by displaying the distribution of the early Islamic coins. Their alignment with settlements along the Cyrenaican coast, hinterland and adjoining islands is immediately apparent. Now the fun begins. Superimposing the new data upon those of fourteen other museums reveals clusters and associations, and, with Reem leading the discussion, we investigate those which are least expected—bringing up related imagery, metadata, scholarship and excavation reports from other projects and institutions. Some are caused by minor issues in coordinate and gazetteer information which are quickly noted and updated,*

*but one is especially intriguing: a number of eighth-century Islamic coins cluster together with several more in two separate collections and a badly eroded and illegible inscription on the outskirts of Benghazi. Pulling up aerial imagery of the immediate area reveals linear soilmarks: when the spatial feed's recommender system suggests links to recently digitised Abassid administrative records referring to a nearby location, the discovery proves yet more exciting—the possible presence of a previously unlocated early mosque. Reem plans to contact some colleagues to arrange a site visit. As the meeting draws to a close, webcams are switched off in Rome, Cairo and Tripoli, but in our shared spatial environment the conversations continue long into the weeks that follow.*

Following in the tradition of Tom Elliot and Sean Gillies' seminal paper 'Digital Geography and Classics', the scenario above is a simple sketch of one way in which historical scholars and heritage professionals might collaborate in the not-too-distant future. While the people, museum and discovery are fictitious, in this two-part article we aim to demonstrate that such forms of digital cooperation are increasingly within reach. We will do so by setting out the theory and practice of Linked Open Data (LOD),<sup>2</sup> and specifically by discussing the contributions of the Pelagios initiative of linking data through common references to place over the previous decade. Our aim is less to provide a history of Pelagios from 2011 and 2020 as it is to identify and discuss the key challenges that needed to be addressed, as well as Pelagios's evolving series of

responses to them. Fundamental to Pelagios's ongoing success has been its pragmatic approach to the implementation of Linked Open Geodata (LOG), along with its emphasis on the social element of linking data, from the co-creation of method and tools, to the development of a community of practice.

Across these two papers we articulate what motivated our intervention, what resulted from it, and some of the things that remain to be done. In the first section of Part I we provide background to the theory and technologies which underpin LOG, and the motivations and prior work which led to the Pelagios initiative. We focus on its first two project phases which established our primary methodology, known as semantic geoannotation. For the remainder of the article, we discuss Pelagios's development of several technologies to support this methodology, and particularly *Recogito*, an open source and free-to-use online platform for identifying place references in historical texts, images and datasets. The second paper, *Pelagios - Connecting histories of place. Part II: From Community to Association* (this volume), focuses on Pelagios's growth from a loose-knit community of practitioners to a formal voluntary association. We hope that by detailing and reflecting on our experience we also contribute to a wider case for the importance of investing in community-driven integration of the rich array of online resources about the past, both as LOG and as LOD more generally.

## 1. Linked Open Geodata for the humanities: the origins of the Pelagios method

Based on the opportunities afforded by the World Wide Web, LOD is now a well-established, if still relatively specialised, method for information retrieval. In this section we focus on the challenges to its adoption in the wider scholarly community by briefly setting out the ideas behind LOD, the technology on which it is based, and the issues its implementation raises. Responding to this challenge, we discuss Pelagios, whose origins in Spatial Humanities research led to the development of a specific type of LOD—linking data based on references to place, or Linked Open *Ge*odata (LOG). The LOG method that Pelagios has developed, co-created with various online providers of resources for the study of the past, emphasises simplicity and ease of use.

### 1.1 The Challenge of Linked Open Data

The primary challenge that LOD methods aim to address is the issue of discoverability on the Web: how can the user find information online that is potentially relevant to their interests?<sup>3</sup> Unlike the ‘Web of documents’ which uses embedded hyperlinks to create unidirectional relationships between Web pages, LOD is a bidirectional graph of connections to which multiple actors can contribute without being aware of each other’s work. For example, if a content provider publishes metadata about its collections as LOD, links can be made between pieces of information that were created at different times, by different authors, and for different purposes. Consequently, users subsequently encountering that content gain awareness of, and access to, a greater volume and variety of data to browse and query. For those working in heritage and

the humanities, the appeal of LOD is thus relatively clear: the possibility of creating a network of connected digital resources about the past. Through the connections that can be traced from an object (a ceramic pot, say) to the other objects, places, people, and events it relates to, it becomes easier to form narratives that enrich and contextualise its biography: where it was created; who produced, discovered, or bought it; how it complements or differs from other finds; where it can (or can't) be seen now; and so on. Yet even though the fundamentals of LOD boil down to a small number of easily articulable rules,<sup>4</sup> its practical application continues to present challenges to many who might wish to produce or make use of it. Above all, LOD is often presented as a technical solution to a technical problem, whereas in our experience both the obstacles and rewards lie just as often within the social arena. Correct and effective use of ontologies, controlled vocabularies and modelling patterns for creating LOD requires expertise, while the complexities of specific technologies for querying the data (not least query languages such as SPARQL)<sup>5</sup> can be a significant barrier to anyone who doesn't already possess advanced technical skills. Pelagios was founded in an attempt to reduce both kinds of barriers. But before discussing the method it developed, we will first define LOD a little more explicitly and explain its mechanics.

According to three of its most notable proponents: *'[t]he term linked data refers to a set of best practices for publishing and connecting structured data on the web'*.<sup>6</sup> The

concept of LOD, then, has at its core a small number of 'best practices', but it also requires thinking about what it means to publish data on the Web. These can be summarised as a series of guidelines sometimes known as 'the four rules':

1. Create (or reuse) addressable identifiers for each concept of interest;
2. Link those identifiers together to create statements about them;
3. Support incoming links from other systems;
4. Link to relevant data elsewhere.<sup>7</sup>

LOD is thus about creating and navigating connections between concepts that have something in common across multiple information systems, *'so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.'*<sup>6</sup> As is clear from the steps articulated above, a necessary prerequisite for this to work is that data must be published in a way that allows it to be linked. While this may seem self-evident, we have found it to be the single most important hurdle that practitioners face.

The primary issue is addressability. In order to be able to create links, we first require an addressing scheme that encodes where those links can point to. On the Web, that scheme is provided by HTTP URIs, *Uniform Resource Identifiers*, and most commonly a specific subtype of the URI, the *Uniform Resource Locator* (usually referred to as a URL or 'web address'). We require this kind of addressability for *everything* we want to connect as LOD, including both high level concepts (such as a particular place or

person), properties (such as 'born in' or 'authored by'), and the information-carrying documents such as texts, images and data records, that refer to them. Consequently, it is a requirement: (i) that there must be URIs to which we can link; (ii) that we can rely on these URIs to remain stable for the foreseeable future, so that our links won't break; and (iii) that there exists sufficient community acceptance for whoever hosts those URIs that others will want to link to them too.<sup>9</sup>

This third requirement, which highlights the role of community, is particularly important. Introductory literature on LOD tends to stress its formal structures—the URIs, the Hypertext Transfer Protocol (HTTP)<sup>10</sup> through which browsers access data over the World Wide Web; the Resource Description Framework (RDF),<sup>11</sup> which serves as a 'language' to encode information in machine-readable terms; and so forth. All these elements are essential components, but what actually makes LOD work are the shared conventions around them that determine which URIs are used, and how. If a community agrees on a set of authority files for identifying places and people; if they use common terms when expressing relations between actors; and if they establish mapping rules between their vocabularies and those of others, then the outcome of LOD is a set of new digital pathways between those ideas. With URIs acting as the intermediaries, these pathways lead beyond the individual work, database, or website, even beyond a single scholarly domain; and as data grows and evolves over time, so do the threads that weave it together. If no agreement emerges, then the capacity to



forge wider connections is reduced, and we are essentially left little further forwards than making annotations in the margins of books.

Central to the process of linking data, therefore, is a tension between the grand vision of a freely navigable graph of richly detailed knowledge and scholarship, and the amount of community consensus and curatorial effort required to realise it.<sup>12</sup> Shared data models need sufficient granularity in order to support a particular set of information analysis or discovery tasks. 'Finding the right depth' is a major obstacle to any LOD initiative involving many contributors and consumers, summed up by the 'Goldilocks problem': what is perfectly sufficient for one person may be too complex, too simple or too inaccurate for someone else.<sup>13</sup> For example, large formal ontologies like the CIDOC-CRM<sup>14</sup> provide powerful tools for modelling entities and the relations between them in great depth, but their complexity also makes them correspondingly harder to implement and interpret. An alternative approach<sup>15</sup> is to start at the easier end of the scale, with the primary referents or entities (the people, places, periods, and so forth) mentioned within a web document (whether a text, image or dataset). The process of identifying the primary referent and attributing additional information to it is called *semantic annotation*. Building upon the traditional idea of annotation, semantic annotation encodes a link between: 1) a reference within an information-carrying resource (an online document such as a record in a numismatic database, or a toponym in a text) known as the 'target'; and 2) entities defined by URI in an authority

file, such as a list of notable individuals or places in a geographic gazetteer, known as the 'body'. What it does not do is create links between those entities that describe a relationship between them, such as 'spouse of' or 'capital of'.

Thus, if the knowledge graph of a 'full-blown' Semantic Web comprises links between many individual and abstract concepts, semantic annotation occupies a more limited territory: it focuses on document-entity relationships, rather than those between entities. While undeniably limiting its utility for making logical inferences about the things annotations refer to, we will see that it nevertheless offers several wide-ranging practical benefits. These include facilitating discovery and 'baking-in' provenance, while simultaneously making adoption easier and bootstrapping more complex semantics. It is in this space that Pelagios has aimed to make its primary contribution, by developing a set of methods, tools and community around a specialised subset of LOD: Linked Open Geodata. Of course, by doing so we do not by implication reject approaches that use more complex ontologies, but rather seek to widen the field of applications and domains in which semantic technologies may play a fruitful role.

## 1.2 The Origins of Pelagios

Pelagios originated in the coming together of a number of prior initiatives which addressed topics related to the identification of historical places. The *Hestia* project,<sup>16</sup> led by Elton Barker, explored the potential for identifying and exploring named place

entities in the *Histories* of Herodotus.<sup>17</sup> As part of its experimentation with digital mapping, *Hestia* reused the TEI-encoded<sup>18</sup> digital text of the *Histories* provided by the Perseus Digital Library,<sup>19</sup> which incorporated coordinate information from the Getty Thesaurus of Geographic Names (TGN) for many of the locations cited by its author. This facilitated a range of visualisation and network analysis techniques for exploring its narrative of events around the eastern Mediterranean and beyond. Concurrently, Leif Isaksen's doctoral research on the use of semantic technologies in archaeology concluded that social factors—in particular, the need to balance implementation costs and concrete benefits for those conducting the work—were an essential consideration for a process that required a diverse community of voluntary contributors.<sup>20</sup> A joint collaboration, *Google Ancient Places (GAP)*,<sup>21</sup> extended *Hestia's* approach by identifying place names in a corpus of multiple out-of-copyright classical texts from Google Books. This required a gazetteer that could offer publicly available and stable URIs for places in the Greek and Roman world, something that the TGN did not provide at that time. Fortuitously, *GAP* began just as the Pleiades Gazetteer of the Ancient World<sup>22</sup> was launched. In a key development which proved critical for Pelagios, Pleiades treated places as first-order concepts, distinct from location and toponymy, and assigned URIs to each.<sup>23</sup> Using Pleiades as the means to identify places across more than one text, *GAP* succeeded in demonstrating how textual collections could be cross-navigated spatially and the geographic flow of any narrative visualised, though it remained restricted to a limited set of documents provided by Google for this

purpose.<sup>24</sup> Independently, Rainer Simon was working on a tool to annotate historical maps with geospatially relevant content from Wikipedia for the EuropeanaConnect project.<sup>25</sup> Moreover, throughout the period over which these precursor projects took place, an increasing volume of openly accessible content about the classical world was becoming available online.

Launched in 2011 as part of the Jisc Geospatial Engagement and Community Outreach programme, the *PELAGIOS* project<sup>26</sup> was conceived of as a proof-of concept to address a very particular challenge: could the identification of place entities in online documents be used as a mechanism for linking classical Greek and Roman content together? Bringing together an open consortium of major ancient Mediterranean content providers,<sup>27</sup> *PELAGIOS* had the goal of defining a mechanism for recording place identifications across a broad range of web-based resources that could be collectively aggregated for visualisation or cross-browsing. In doing so it established a set of principles for producing and exploiting LOG about the ancient Mediterranean, focused exclusively on references to place, and building wherever possible on tools and specifications that were already available. The methodology it developed was based on the Open Annotation model<sup>28</sup> and the use of gazetteer URIs provided by Pleiades. Producers of LOG were encouraged to annotate place references that occurred in their content with Pleiades URIs. These so-called 'semantic geoannotations' were represented using stand-off markup: rather than requiring a

change to the content itself (such as the introduction of additional fields in an existing database or extra markup in existing data files), they would be stored in a separate dataset which could then be published independently as LOG.

As a proof-of-concept, *PELAGIOS* confirmed not only that linking data based on place references could work in practice; it also demonstrated multiple benefits of doing so, including improved discoverability, contextualisation and reuse. An example of this was the ability of consortium partners to create automated links to spatially related content in one another's collections. Importantly, it also showed how LOD could be complemented by an entirely different set of technologies: Geographic Information Systems (GIS). Semantic geoannotation creates machine-readable statements about place references by connecting them to an entry in a gazetteer (Fig. 1). This entry may include not only geographic coordinates but also complex geometries (such as lines and polygons). Such information in turn allows for the spatial footprint of different kinds of documents to be presented together on a map and for spatially adjacent content to be discovered, even when the specific locations referred to differ. Significantly, it also transcends traditional GIS capabilities by permitting the association of places even when they are unlocated, since the crucial ingredient needed for the annotation is the URI rather than a set of coordinates. When considering historical data, the case of places that are known through literary, visual or material sources, but which have no definitive location, is far from rare. An LOG approach enables

researchers to make useful statements about those places, their relevance and their role in the source, so long as there is a relevant URI in one of the available gazetteers. To take an example, it is possible to geoannotate mentions of Ogygia, the island of Calypso in the *Odyssey*, by matching them with the corresponding Pleiades URI,<sup>29</sup> even though it may only ever have existed in the cultural imagination of Homer's audience.

Drawing upon substantive technical expertise across its consortium, *PELAGIOS* did not include tooling within its scope. Instead, depending on the skills and interests of partners, annotations were produced in a variety of ways, including hand-encoding and automated scripting, and integrated into online resources through mechanisms ranging from pop-up dialog boxes to simple hyperlinks. It was nevertheless clear that for the semantic annotation method to gain wider acceptance, we would need to make it much easier for others to adopt. A follow-on project, *Pelagios 2*,<sup>30</sup> sought ways to make the core model more accessible and more useful to a wider body of stakeholders by creating a basic cataloguing service for annotations, embeddable widgets for websites that displayed related content, and a cookbook containing 'recipes' for creating annotations across a range of different scenarios. A further addition was the development of a prototype viewer for visualising the graph of connections between data providers. This basic visualisation tool, the *Pelagios Graph Explorer*, was developed to demonstrate how individual annotations could collectively form a combined network of documents and places, which in turn could help bring to light

new relationships or spatial correlations (Fig. 2).<sup>31</sup> A significant challenge presented by the need to represent collections as a graph was the heterogeneity of aggregated data, which might derive from a single text, or a large collection of individual museum records. To address this, we made use of the Void RDF vocabulary,<sup>32</sup> allowing contributors to define a collection as a particular set of annotations which in practice could link to multiple sources, or a subset of annotations from a single source.

While *Pelagios 2* remained focused on the ancient Mediterranean world, it had become evident that there were instances in which the use of alternative gazetteers might be needed, such as references in secondary literature to modern cities or countries. This in turn presented an issue known as the ‘co-referencing problem’:<sup>33</sup> as the content of Pleiades and a contemporary gazetteer such as GeoNames overlap, how can we allow for the use of different URIs referring to the same place, while still maintaining connectivity across the graph? This required us to start addressing the issue of cross-linking place definitions across gazetteers—a process called alignment—using vocabularies such as SKOS.<sup>34</sup> As with annotation however, solutions to this challenge are not solely technical, and raise important considerations relating to who is responsible for conducting and maintaining such alignments, and whether specialist gazetteers should align to a single ‘spinal’ gazetteer or, alternatively, to as many other gazetteers as possible. These early considerations would in time lead to the Pelagios Gazetteer Interconnection Format, discussed in section 2.1 below.

By mid-2012, the *PELAGIOS* and *Pelagios 2* projects had established the basic principles, methodology and a nascent community for semantic geoannotation of place references within ancient Mediterranean resources. Pelagios enabled data providers from a range of institutions, each with their own particular form of data (text, image or database), to interlink their content. Key to its success was the prior existence of a relatively mature digital ecosystem of resources, technical expertise to implement the new method, and, above all, a gazetteer (Pleiades) which provided stable and persistent URIs for places. The next challenge was to take this proof of concept, which had worked to interlink important ancient Mediterranean resources, and develop tools that would make it accessible to a much wider audience including those in other humanities subjects and disciplines.

## **2. From Method to Tools: Widening Participation in Linked Open Geodata**

Supported by the Andrew W. Mellon Foundation's Scholarly Communications programme, *Pelagios 3: Early Geospatial Documents*<sup>35</sup> explored whether Pelagios could contribute to the study of pre-Cartesian modes of geographic representation more generally, where the ability to identify and compare different spatial representations—either texts or maps—holds great potential for rethinking non-modern cultures of knowledge. To demonstrate the applicability of our approach



beyond the ancient Mediterranean, we expanded our temporal and geographic scope to encompass early Christian, Islamic and Chinese traditions with a focus on their earliest geographic texts and maps. By addressing the issues that this expansion presented, our aim was to make some headway in—if not necessarily solve—some of the practical challenges of working beyond the ancient Mediterranean. For instance, many of these documents are unique manuscripts and difficult to access due to being offline, behind paywalls, or simply hard to find in the first place. Scholarship is often undertaken by independent researchers with little technical support, and the need to use historical gazetteers in addition to Pleiades would obviously be essential. In order to address challenges such as these, and once again drawing on the expertise of its community, Pelagios developed three novel technologies: *Peripleo*, the Gazetteer Interconnection Format, and *Recogito*.

## 2.1 Three Pelagios technologies: The Gazetteer Interconnection Format, *Peripleo* and *Recogito*

When moving beyond the online ancient Mediterranean, we were quickly confronted by an obvious difficulty: the lack of specialist historical gazetteers akin to Pleiades that could serve research communities beyond those studying classical antiquity. An ostensible solution would be to develop a new ‘global’ gazetteer—either by extending Pleiades or else using a pre-existing one, such as GeoNames—which would provide an authority record for all places in all periods. Yet such a gazetteer risked being both

prescriptive and unworkable. Gazetteers, like ontologies, also suffer from the Goldilocks problem: the ideal level of spatial resolution (say street-level locations) for one situation may be completely unsuited and unsuitable to another. Consequently, any global gazetteer faces the twin risks of either being poorly optimised for most cases, or highly bespoke and heterogeneous, and thereby opaque and arbitrary in their level of coverage. Our solution returned to the principle of supporting a devolved ecosystem of online resources. Instead of attempting to build a single 'supergazetteer', we developed the Pelagios *Gazetteer Interconnection Format*.<sup>36</sup> a minimal specification for enabling different URI-based gazetteers (each serving the needs of their particular community) to be aligned to each other. The interchange format allows data from divergent traditions to be interlinked, since the documents are linked to entries in separate gazetteers, while at the same time the gazetteer entries cross refer to one another. An Early Modernist, for example, might reference the 'Rome' defined by TGN<sup>37</sup> for example, while a Classicist might use one provided by Pleiades,<sup>38</sup> but alignment allows both scholars to ascertain that they are effectively referring to the same place. That being said, the need for historical URI-based gazetteers remains a central and ongoing challenge for the Pelagios community and those working with historical LOD in general, and one we will return to in the Part II of this paper.<sup>39</sup>

The second challenge was to visually demonstrate the value of producing LOG, given the comparative lack of available software that could make use of it. Our first

visualisation tool, the *Graph Explorer*, displayed the connections between collection content topologically, but was highly abstract and not easy to navigate. Although it displayed conceptual relationships between places and documents, it did not possess a geospatial interface that could provide meaningful context or allow users to find content related to their areas of interest. In order to demonstrate better the potential of connecting historical data through place references, we developed *Peripleo*, a map-based LOG visualisation prototype with both a Graphical User Interface (GUI) and an Application Programming Interface (API).<sup>40</sup> Combining a variety of ancient Mediterranean-related annotation sets—many created during *Pelagios 1* and *2*—*Peripleo* displayed them as dynamic and searchable data on top of a bespoke, high-resolution map of the Roman empire contributed to the project by Johan Åhlfeldt.<sup>41</sup>

As a pilot demonstrator, *Peripleo* made the data queryable as a whole, facilitating the exploration of spatial relationships that would otherwise be hard, or even impossible, to discover in isolation (so called 'co-contextualisation'). Items and places from multiple data sources can be projected onto the map, allowing the user to follow the LOG connections between them (Fig. 3). In this respect, it offered the potential to generate new lines of inquiry by enabling the discovery of otherwise unseen connections between data of different nature and provenance. Depending on the availability of individual dataset attributes, features also included full-text and adjacency search, dynamic filtering by collection and time, and the display of IIF

images. Additionally, the API provided developers with programmatic access to the data to incorporate within their own search or visualisation tools. Although limited to a specific set of 48 annotation datasets and 19 gazetteers, *Peripleo* was highly effective as a showcase of the potential of LOG, and ignited enthusiasm in the ancient Mediterranean scholarly community and the digital humanities more generally.<sup>42</sup> Archives and Museums in particular were quick to realise the value of *Peripleo*, since it allowed them to provide access and showcase collections which had already been converted to LOG.

Despite this enthusiasm, many of the researchers and heritage professionals with whom we interacted were prevented from joining the LOG ecosystem by the technological barriers they faced in creating their own annotations. While the *Gazetteer Interconnection Format* and *Peripleo* were important for enhancing and expressing the value of LOG created by our project partners, a third major challenge remained: how to enable non-specialists to produce LOG themselves. The initial phases of *Pelagios* had largely focused on restructuring and/or repointing data that had already been derived or extracted from historical documents and collections. *Pelagios 3* forced us to work directly with original sources. In order to make this process both efficient and enjoyable we developed *Recogito*,<sup>43</sup> an online tool for semantic geoannotation of digital texts and images, supported by a straightforward map visualisation of the place references. Although other annotation platforms were

available,<sup>44</sup> *Recogito* provided a variety of unique functions to support geoannotation, including automated geo-coding suggestions from a selection of historical gazetteers, an interface to facilitate place disambiguation, and automatically generating map-based visualisations. In line with Pelagios's underlying goals, it was designed with fairly simple semantic relationships in mind, prioritising usability over complexity and richness. *Recogito* started off only able to geoannotate texts, but as the project's focus expanded to include medieval sources such as portolan charts and early Islamic maps image annotation was added to allow inclusion of these non-literary documents. Options for different kinds of annotations (places, events, and relations) were also gradually incorporated in subsequent evolutions of the platform.<sup>45</sup>

Though originally developed as an 'in-house' tool for the *Pelagios 3* Investigative Team, it quickly became evident that *Recogito* might not only be used by others, but also held significant potential for research and heritage management. In subsequent phases of Pelagios, it has been redeveloped from the ground up as *Recogito 2.0*: a complete online platform for which anyone can register, providing them with a private workspace that can be shared publicly where copyright licencing permits. Development has increasingly been in response to community interaction and feedback managed through workshops, outreach activities, mailing lists and issue tracking, as well as emphasising sustainability and interoperability.<sup>46</sup> Given *Recogito's* growing popularity within the digital humanities,<sup>47</sup> the remainder of this section

focuses on its current features, real-world applications, and increasing use as a collaborative and teaching platform.

## 2.2 *Recogito 2.0*

*Recogito* helps users with varying degrees of digital proficiency to produce annotations that highlight spatial information in historical documents.<sup>48</sup> At its basis is the ability to render more easily the identification and disambiguation of place references in a range of source documents (text, image, tabular data) over multiple formats.<sup>49</sup> This practice of geoannotation works by way of a two-step process. First a user highlights a character string in a text or selects part of an image and makes an assertion that it (the text or image at this point) refers to a place; second, using *Recogito's* suite of gazetteers and map visualisation, the user then selects the most appropriate gazetteer record, which assigns the assertion an authoritative URI (Fig. 4). Because of this critical reliance on gazetteers, annotating in *Recogito* becomes more challenging where such a global authority is unavailable (for Early Modern England, say, or smaller regions such as provinces or cities). However, *Recogito's* very demonstration of the practical utility of gazetteers for geoannotating historical resources has directly led to communities developing them for their own fields, sometimes even using *Recogito* in the process.<sup>50</sup>

As well as its easy-to-use annotation screen for identifying and resolving place references against gazetteers, *Recogito* comes with an in-built map view in which annotations associated with a gazetteer URI are visualised in their spatial context (Fig. 5). Even in its most basic form, a map of place references can be useful in stimulating research questions about the document, the nature of the historical information provided, and even the socio-political and cultural narratives that surround such places. Allowing us to see the places referred to in a text or rendered on a map means being able to unfold and make visible the world—actual, imagined or remembered—that is shaped by the historical source. A place is represented on the map proportional to the number of references associated with it in the document: the bigger the marker, the more references it has. Additionally, by clicking on a marker, users can browse the corresponding text passages, one by one. As one researcher has put it, '*Recogito allows us to see and prove things that our intuition as readers tells us, but that are hard to demonstrate*'.<sup>51</sup>

Thus, while it shares some of the characteristics of more traditional location-based and GIS-oriented approaches, *Recogito's* URI-based annotation approach differs substantially. Principally, by aligning place references in a document to gazetteer URIs, *Recogito* enables the creation of LOG by domain specialists rather than data technicians: a browser-based digital workspace, *Recogito* does not require any installation or specialist knowledge in semantic technologies and web-mapping.

*Recogito's* different views also facilitate the easy and swift transition from document to map, and back again, thereby better enabling the spatial analysis of the historical source. In fact, as a platform *Recogito* offers a number of additional features that support the use of geoannotation for research or teaching purposes. By registering a *Recogito* account, users are able to upload, manage and share the digital documents they want to annotate, either in text, image or tabular format.<sup>52</sup> While annotation is supported by a number of automated processes, the annotator is kept aware of the process throughout, and maintains full control over it. 'Bulk' operations allow users to re-apply the same annotation to all identical occurrences of a place reference in a text, considerably speeding up the process.

Arguably, one of the most useful additional features when dealing with text-based documents is Named Entity Recognition (NER). As a tool that enables the automatic identification of place names in multiple languages<sup>53</sup> its application within the *Recogito* annotation environment demonstrates a core principle of this platform. Using NER, a *Recogito* user can quickly and easily produce visualisations of the main place references that appear in their document, which can be of particular benefit when analysing larger corpora. Yet, even advanced NER algorithms are unlikely to recognise that phrases such as 'the capital of the Empire', or 'the City of the Seven Hills' (let alone pronouns such as 'it' or 'there') may—but do not necessarily—refer to the place entity 'ancient Rome'. That is to say, it is only by close reading of the source



document that one may identify less explicit references, where places are not mentioned by proper names, but indirectly via adjectives or periphrasis.<sup>54</sup> While *Recogito* makes the manual confirmation and correction of algorithmically generated annotations as intuitive and efficient as possible, it initially presents NER annotations in grey—both the highlighted word in the annotated document and the corresponding place icon on the map. The grey colouring indicates that a human reader has yet to verify them as a place reference or to align them to a particular gazetteer record; once they have been verified and aligned, the annotation and map icon turn green. Thus *Recogito's* interface aims to make clear that semantic annotations are not 'facts': they are assertions of meaning by a reader, where context remains king.<sup>55</sup> Reflecting their interpreted nature, each annotation also records its author and is given a timestamp, thereby rendering the process as transparent as possible and open to scrutiny. The simple structure of the annotations, the commitment to openness, and the emphasis on human intervention in automated processes have been strategic decisions to give annotators control of their workflow, and viewers the ability to assess it, so as not to be reliant upon a technological black box.<sup>56</sup> While these are among the core capabilities of *Recogito*, other features include image and dataset annotation, use of tags to complement annotation, aggregated statistics for each document, and recording non-place-entities and relationships, as well as tools for sharing and collaboration in research and educational settings. In our next two sections we will illustrate some of these features by way of some specific applications.

### 2.3 Recogito in Use

Our first example is the semantic annotation of the *Lives of the Most Excellent Painters, Sculptors, and Architects*.<sup>57</sup> Written by the sixteenth-century Italian artist Giorgio Vasari, the *Lives* was one of the most influential books in the study and perception of Italian art at the time of its publication and over subsequent centuries. In it Vasari recounts the lives of those whom he deems to be the most notable artists in the Middle Ages and Renaissance, interweaving their biographies with information about their works. Geoannotating this seminal book helps not only to confirm basic assumptions about the places where these artists flourished, but, through the use of *Recogito's* map view, to highlight less obvious spatial patterns within Vasari's work. For example, the visualisations foreground the predominance of Tuscany (and the cities of Florence, Pisa, Siena, Arezzo) that was not only Vasari's birthplace but also the area under the patronage of the Medici family (Fig. 5, above). At the same time, the use of geoannotation helps to surface other locales that played a role in the evolution of Italian art according to Vasari, while the capacity in which they appear can also be identified: as provenance of source material (such as marble or pigments), houses of famous workshops, centres of valuable commissions, and so on.

The identification of different roles is made possible by the ability to provide tags that accompany the geoannotation. Tagging is a powerful feature of *Recogito*: this free text

option enables users to enrich information for each place reference, and provide structured data about the text's spatial organisation as a whole. Users are free to develop their own typology for defining the spatial information, which *Recogito* 'remembers', so as to ensure consistency of tagging throughout the text. In the most straightforward case, we might reuse the text's own spatial description, as shown in Figure 6. This visualisation (taken from the *Recogito* map view) shows the Catalogue of Ships in Book 2 of the *Iliad*, according to Homer's own description of those who fought at Troy by Greek contingent (e.g. the Boeotians) or Trojan ally. The simple tag effectively highlights which places belong to which contingents, and help provide a sense of Homer's spatial patterning.<sup>58</sup>

*Recogito's* automatically generated maps can be particularly helpful when interpreting historical itineraries or travelogues. An ancient journey can be gradually, and virtually, rebuilt on the map, helping transform the words in a text into places that have their own spatiality and are experienced sequentially.<sup>59</sup> For example, mapping the itinerary listed on the Vicarello Cups<sup>60</sup> between Gades (modern-day Cádiz) and Rome, shows a clear preference for avoiding mountainous areas, even where they would reduce the overall length of the journey. Again, by using different tags, *Recogito* can highlight the nature of different stages in the journey through colour-coding, while symbol size helps to distinguish between places that appear only once and those that reoccur. This approach can help bring together both the spatial dimension of the travel, grounded

in its physical locations and in the connections between cities, and the narratives that surround those experiences.

Each 'document' in *Recogito* in fact consists of one or more individual files, a feature which can be used either to split a single text into sections (such as chapters), or to compare multiple but divergent versions. While each source file remains distinct and individually accessible through a table of contents in the interface, the annotations produced are counted and represented together within the map view; colour-coding can again be used to distinguish between different versions or different parts of the text. The same approach can also be used to compare entirely different documents relating to the same historical event. For example, a comparative visualisation of geographical references related to the second Punic War in the accounts of ancient historians such as Polybius (II century BCE, from Megalopolis in Arcadia), Livy (I century BCE, from Padua) or Appian (I century CE, from Alexandria of Egypt) might stimulate research questions about the places—and therefore the moments of the wars—that they choose to highlight, as well as about their interpretations of Hannibal's military strategy and Rome's response.

So far our examples have been drawn from texts, but *Recogito* also enables the geoannotation of images, which is particularly effective for enabling the visualisation and analysis of non-Cartesian maps. Annotation works in the same way as texts, with

the exception being that a user records the place name on the image in a transcription box, before aligning that transcribed reference to a gazetteer. While this is done out of necessity—*Recogito* cannot automatically read place names in images—it also has the virtue of enabling the capture of an image’s specific toponymy, thereby facilitating comparison, say, of maps of the same area from different traditions or in different languages. A good example is the geoannotation of a portolano by Grazioso Benincasa (Fig. 7).<sup>61</sup> The individual folios, when uploaded together in *Recogito*, produce a metadocument that collects all the geoannotations, and presents them together, providing an at-a-glance representation of the areas covered by the Atlas. By colouring the annotation symbols in the map view according to the manuscript page it relates to, users can immediately see how different maritime zones are divided between the folios, as well as areas of overlap (Fig. 8).

In fact, the flexibility of LOG allows documents in image-based media to be compared and contrasted with texts, and vice versa. For instance, as a popular international visitor attraction since the eighteenth century, Pompeii has generated a large corpus of both visual and literary documents specifically targeting tourists. By geoannotating early maps of the partially buried city, as well as its first travel guides, it is possible to pinpoint recommended attractions, and to estimate the movements of visitors in the first ever archaeological site open to the public.<sup>62</sup> Seeing what places appear in *Recogito's* map view helps rebuild a progressive visual account of the areas of the site

as they gradually became accessible to the public (Fig. 9). Such a visualisation can facilitate comparison to actual excavation data, and help determine which monuments and buildings, among those uncovered, became part of the standard 'bucket list' of attractions during the years of the Grand Tour. This work, in turn, can contribute to the investigation into the practical and ideological variables that influenced such choices, as well as their subsequent crystallisation.<sup>63</sup> A close examination of the early recommendations for tourists in Pompeii is also enhanced by the availability of alternative base-maps in *Recogito*. One of these is OpenStreetMap,<sup>64</sup> which depicts the present-day excavation site in great detail. Exploiting this open resource, as well as the coverage of Pompeii in the Pleiades gazetteer, it is possible not only to display annotations on a detailed contemporary map, but also to generate hypotheses on the connectivity of, and spatial relations between, the Pompeian attractions.

By giving these brief examples of *Recogito* in use, our aim is not to dwell on the topics or documents that underlie them, but rather to demonstrate *Recogito's* broad range of possible applications, formats and historical domains. In the final section we discuss some of the ways in which *Recogito* can also be utilised as part of longer pedagogical, technical and collaborative workflows.

## 2.4 Teaching, Collaboration and Workflows

In addition to the research scenarios described above, *Recogito* has also proven to be a useful teaching and assessment tool in various pedagogical contexts, across many

historical disciplines, and at different levels of Higher Education.<sup>65</sup> Annotating place names in primary sources helps students familiarise and engage with unfamiliar texts and their background context, thereby facilitating the learning process. *Recogito* can be an effective way to learn about the physical and cultural roles that particular places and landscapes have played in shaping historical events. Thanks to the features that enable collaborative annotation, it is well suited to group assignments where students can collectively define annotation and tagging criteria and determine their own research questions. A rich set of sharing options allows users to choose the level of access to others they deem suitable for their work. In a classroom environment the annotated documents might be made visible only to the student and their teacher/instructor. As each annotation is tied to a specific annotator, assessment can be conducted both at group and individual level.<sup>66</sup>

By default, documents uploaded in *Recogito* are visible only to the user who has created them, ensuring both privacy to researchers who are still experimenting and not ready to present their results to a wider audience, and legal protection for documents which are under copyright and cannot be shared online. On the other hand, when both copyright and privacy considerations allow it, *Recogito* users can share their (annotated) local copy of the document within *Recogito* itself, inviting other users to contribute, or simply browse the annotations and the map-based visualisations.

*Recogito* associates a stable URI with each document and allows it to be publicly shared in order to engage a larger audience.

While *Recogito* offers a variety of opportunities for teaching and can provide students with an accessible first stepping-stone into the world of digital scholarship, larger-scale research projects will quickly outgrow *Recogito's* built-in visualisation, analysis and publication features. For this reason, *Recogito* provides a range of upload and download options that allow users to import and export data in different formats and commence or continue their analyses in other platforms and tools.<sup>67</sup> An important case is support for the Text Encoding Initiative (TEI). If an uploaded document conforms to TEI Guidelines, *Recogito* merges new annotations into the original TEI text. For plain text documents, it can optionally apply a basic TEI-XML template, to allow for full text export of literary documents with embedded annotations. Additional formats may be available depending on the document type and the kinds of annotations that the user has created, such as formats specifically for use with network analysis tools (see below), or for using text annotated in *Recogito* as ground truth for training natural language processing algorithms. Support for these commonly used formats as input and output allow *Recogito* to be used in sophisticated digital workflows that start with physical manuscripts and end with complex analyses and compelling or informative visualisations of their contents.



As much as annotating with *Recogito* is useful in its own right as a process for studying and close reading source material, its primary function is to enable non-IT experts to create LOG annotations. *Recogito* is not a LOG hosting platform per se (it does not provide a SPARQL query endpoint, for instance), but it does make it possible to share annotation projects online. This includes supported data export formats, such as the annotations in RDF format, as well as the annotated documents and map visualisations themselves. It also generates stable URIs which can be shared online for each single annotation, for the entire annotated document, and for the map visualisation. Thanks to its flexible sharing options, *Recogito* can be a straightforward way to share and communicate complete or work in progress research, and to invite other users to give feedback or become contributors. Yet it is only when users make their annotations available online in RDF format that the potential of LOG may be more fully realised: it enables those annotations to be connected, unpredictably, to various other, and often heterogeneous, RDF statements. The relationships created in *Recogito* are designed to be simple, following the principle that in the *social* Semantic Web it is as important to make it easy to create a connection as it is to describe that connection in detail.

The benefits of working collaboratively have also led to direct improvements in *Recogito* itself in the form of feature requests and suggestions for improvement.<sup>68</sup> For example, as its increasingly diverse community grew, many users were interested in annotating not only entities, but also the relationships between them. While not

semantic annotation in the strictest sense, the frequent demand for this functionality led us to develop a new 'relations' mode, that enables previously annotated entities in a text to be related to one another and their relationship to be defined with a tag (Fig. 10). Data created in this way can then be downloaded and processed in commonly used Network Analysis Software such as *Gephi* or *Visone*.

\*\*\*\*\*

Notwithstanding all these features, *Recogito* is clearly not intended as a panacea for geospatial teaching and research in the humanities or indexing collections. The principal limitation, which goes beyond Pelagios, pertains to the inevitable bias that comes with the selection of available data. Any amount of LOG that we can capture, access or query (via a SPARQL endpoint or some more user-friendly interface like *Peripleo*) is only ever a subset of the existing information about a given topic, and likely just a fragment of what is available digitally, let alone in print or manuscript. As with any resource, analyses and visualisations of LOG datasets must always be considered within their limits, bearing in mind that while the results may provoke interesting lines of research, they are seldom if ever complete or definitive.

This 'Open World' principle<sup>69</sup> has always lain at the heart of the Semantic Web, acting as both a cautionary dictum (that there will always be 'unknown unknowns') and as a call-to-arms to continue linking related content to improve the scholarly environment

for future researchers, learners and digital explorers. In responding to this call, the personal use of applications like *Recogito* only get us so far. In the end, increasing the availability of linked and digitally actionable data about the past needs to come from a large number of individuals and organisations working together. Promoting and facilitating such collaboration, and the establishment and growth of a community of practice dedicated to this activity is a subject we return to at the end of this special issue in Part II of our discussion of Pelagios.<sup>70</sup>

---

<sup>1</sup> With acknowledgement, and apologies, to Tom Elliott and Sean Gillies. See: T. Elliott, and S. Gillies, 'Digital geography and classics', *Digital Humanities Quarterly*, 3,1 (2009).

<sup>2</sup> There have been extensive debates in the computer science literature about whether Linked Data is necessarily, or ought to be, 'open'. We take the position that many proprietary knowledge graphs, such as those used by Google and Microsoft for their search engines, might reasonably be considered Linked Data. However, the inability to reference their underlying concepts makes them unsuitable for linking to third-party content. As this is the fundamental goal of Pelagios, we therefore focus exclusively on Linked *Open* Data for the purposes of this paper.

<sup>3</sup> We are not here concerned with some of the processes that semantic technologies more broadly aim to support, such as inferencing or complex domain modelling.

<sup>4</sup> Most famously articulated in T. Berners-Lee, *Linked Data - Design Issues*, (2006, revised 2009, 2010), accessed 11 January 2021. <https://www.w3.org/DesignIssues/LinkedData.html>.

<sup>5</sup> See: <https://www.w3.org/TR/sparql11-query>.

<sup>6</sup> C. Bizer, T. Heath and T. Berners-Lee, 'Linked Data - the story so far', *International Journal on Semantic Web and Information Systems*, 5, 3, (2009) 1-22.

---

<sup>7</sup> Originally formulated as: '1. Use URIs as names for things; 2. Use HTTP URIs so that people can look up those names; 3. When someone looks up a URI, provide useful information, using the standards (RDF\*, SPARQL); 4. include links to other URIs so that they can discover more things.' Berners-Lee, *Linked Data - Design Issues*.

<sup>8</sup> Berners-Lee, *Linked Data - Design Issues*. See, also, the 2010 addendum, *Is your Linked Open Data 5 Star?*

<sup>9</sup> It is for this reason that Wikipedia and its LOD derivatives Wikidata and DBpedia have often played a central role in the linked data ecosystem: they offer stable URIs for things and ideas about the world that operate as intermediary hubs for connecting related content. The information they provide at the end of those links (however reliable or unreliable it may be) is of only secondary importance from an LOD connectivity perspective.

<sup>10</sup> The HTTP/1.1 specification is described at <https://tools.ietf.org/html/rfc2616>.

<sup>11</sup> A formal Primer on RDF is available at: <https://www.w3.org/TR/rdf11-concepts>.

<sup>12</sup> "Instead of asking machines to understand people's language, it involves asking people to make the extra effort". T. Berners-Lee, *What the Semantic Web is Not*, (1998c).

<https://www.w3.org/DesignIssues/RDFnot.html>.

<sup>13</sup> L. Isaksen, 'Archaeology and the Semantic Web', (unpublished Ph.D. thesis, University of Southampton, 2011), <https://eprints.soton.ac.uk/206421/>

<sup>14</sup> CIDOC Conceptual Reference Model. See <http://www.cidoc-crm.org>. The CRM has a primary focus on the relationships between entities and leaves domain communities to work out appropriate taxonomic vocabularies by subclassing its `E55 : Type` class.

---

<sup>15</sup> We emphasise that the approach is only one along a continuum balancing complexity with descriptive power. Since there is no universally optimal level of complexity for data models, ontology selection should always take into account the specific needs and practical realities of a given project.

<sup>16</sup> AHRC, 2008-2010; 2014-2015. See <https://hestia.open.ac.uk>.

<sup>17</sup> The project was co-led by Stefan Bouzarovski (Birmingham) and Christopher Pelling (Oxford), with technical support from Leif Isaksen (Southampton), see <https://hestia.open.ac.uk/publications> for publications.

<sup>18</sup> Text Encoding Initiative. See <https://tei-c.org>.

<sup>19</sup> <http://www.perseus.tufts.edu>.

<sup>20</sup> L. Isaksen, 'Archaeology and the Semantic Web'.

<sup>21</sup> A collaboration between Barker, Isaksen, Eric Kansa (Alexandria Archive), Kate Byrne (Edinburgh), Nick Rabinowitz, and Enrico Daga (Open University); Google Digital Humanities Awards, 2010-2011; 2012.

<sup>22</sup> Pleiades started as a digitisation of the *Barrington Atlas of the Greek and Roman World*, (ed., R. Talbert, Princeton, 2000) but grew quickly beyond that, becoming a highly collaborative project inviting expert contributions from users. See <http://pleiades.stoa.org>.

<sup>23</sup> See S. Gillies and T. Elliott, 'What's an un-GIS?' (2010, revised 2013). Accessed 11 January 2021, <https://pleiades.stoa.org/docs/papers-and-presentations/whats-an-un-gis>.

<sup>24</sup> Google, 2010-2011; 2012, see <https://googleancientplaces.wordpress.com>.

---

<sup>25</sup> See R. Simon, B. Haslhofer, J. Jung, 'Annotations, Tags and Linked Data. Metadata Enrichment in Online Map Collections Through Volunteer-Contributed Information'. *e-Perimtron*, 6,3 (2011), 129-137.

<sup>26</sup> The recursively-titled *PELAGIOS: Enabling Linked Ancient Geodata In Open Systems*. The Joint Information Systems Committee (Jisc) was at that time a primary funder of digital resources across the UK Higher Education sector.

<sup>27</sup> Pelagios has continued to use the 'open consortium' model in which a small number of guaranteed (and sometimes funded) partners guarantee a critical mass of collaboration while actively fostering engagement with a much wider range of stakeholders. Core partners of the first Pelagios consortium comprised the Open University, the University of Southampton, AIT: Austrian Institute of Technology, The Institute for the Study of the Ancient World (Pleiades), Tufts University (the Perseus Digital Library), the University of Cologne (Arachne), and King's College, London (*SPQR* project).

<sup>28</sup> <http://www.openannotation.org/spec/core>. It would later provide the basis for the W3C's Web Annotation Recommendation, which Pelagios now uses. See <https://www.w3.org/TR/annotation-model>.

<sup>29</sup> <https://pleiades.stoa.org/places/456105>.

<sup>30</sup> *Pelagios 2* (Jisc, 2011-12). Partners directly funded as part of *Pelagios 2* included the Open University, the University of Southampton, AIT: Austrian Institute of Technology; University of Oxford (CLAROS); University of Cologne (Arachne); King's College London (SPQR); American Numismatic Society (Nomisma) and the University of Reading (Ure Museum), although a much larger number of projects participated over its lifetime.

<sup>31</sup> See <http://pelagios-project.blogspot.com/2011/11/pelagios-graph-explorer-information.html>.

---

<sup>32</sup> Vocabulary of Interlinked Datasets. See <https://www.w3.org/TR/void>.

<sup>33</sup> For a broader discussion see, for example: H. Glaser, A. Jaffri, and I. Millard, 'Managing Co-reference on the Semantic Web'. *Proceedings of the WWW2009 Workshop: Linked Data on the Web (Spain, 2009)*.

<sup>34</sup> Simple Knowledge Organization System. See <https://www.w3.org/TR/2009/NOTE-skos-primer-20090818>.

<sup>35</sup> *Pelagios 3* was the first of three separate stages of Pelagios funded through the Andrew W. Mellon Foundation's Scholarly Communications programme (*Pelagios 3*, *Pelagios 6* & *Pelagios 7*). Directly funded Partners for the *Pelagios 3* project included the Open University, the University of Southampton, AIT: Austrian Institute of Technology, the British Library Map Collections, the University of Portsmouth (A Vision of Britain Through Time), Drew University (Digital Mappa), Harvard University (China Historical GIS), The University of Edinburgh (Edinburgh Geoparser), The Institute for the Study of the Ancient World (Pleiades), and the Queen Mary University, London. It also brought a new member to the Investigative Team, Pau de Soto Cañamares, an archaeologist from Catalonia with a background in geospatial technologies and the geography of the ancient Mediterranean.

<sup>36</sup> <https://github.com/pelagios/pelagios-cookbook/wiki/Pelagios-Gazetteer-Interconnection-Format>. The basis of the format was developed at a Pelagios workshop held at the Institute for the Study of the Ancient World in 2014.

<sup>37</sup> <http://vocab.getty.edu/page/tgn/7000874>.

<sup>38</sup> <https://pleiades.stoa.org/places/423025>.

<sup>39</sup> The Gazetteer Interconnection Format has subsequently evolved into the Linked Places Format (<https://github.com/LinkedPasts/linked-places>), which we discuss further in *Part II* of this article.

---

<sup>40</sup> <https://peripleo.pelagios.org>. See R. Simon, B. Haslhofer, J. Jung, 'Annotations, Tags and Linked Data. Metadata Enrichment in Online Map Collections Through Volunteer-Contributed Information'.

<sup>41</sup> This would also become the basis of the *Digital Atlas of the Roman Empire*, <https://imperium.ahlfeldt.se>, also developed by Johan Åhlfeldt.

<sup>42</sup> *Peripleo* was voted 'Best DH Visualization' in the 2016 international Digital Humanities Awards (See <http://dhawards.org/dhawards2016/results>). It is not currently supported, but a list of all the projects and institutions who previously contributed data to *Peripleo* is available at <https://peripleo.pelagios.org/about>.

<sup>43</sup> See R. Simon, E. Barker, L. Isaksen and P. de Soto Cañamares, 'Linking early geospatial documents, one place at a time: annotation of geographic documents with *Recogito*'. *e-Perimtron*, 10,2 (2015), 49–59.

<sup>44</sup> See, for instance, *Hypothes.is* (<https://web.hypothes.is>) and *Pundit* (<https://thepund.it>).

<sup>45</sup> See <https://web.archive.org/web/20191109062202/http://commons.pelagios.org/2018/03/you-can-now-do-everything-in-recogito> for an archived summary of *Recogito* updates.

<sup>46</sup> *Pelagios 4*, funded by the Arts and Humanities Research Council of the UK and led by the Open University, allowed this special edition's contributors to apply the tools and data produced by *Pelagios 3* to their own research so that their benefits for research could be evaluated critically. The SEA CHANGE project (effectively 'Pelagios 5') funded by an Open Humanities Award and led by the Austrian Institute of Technology, trialled *Recogito* in two workshops held at German universities in order to solicit feedback from large groups of first-time users.

<sup>47</sup> *Recogito* was voted 'Best DH Tool or Suite of Tools' in the 2018 international Digital Humanities Awards (See <http://dhawards.org/dhawards2018/results>).



---

<sup>48</sup> R. Simon, V. Vitale, R. Kahn, E. Barker, L. Isaksen, 'Revisiting Linking Early Geospatial Documents with Recogito', 14,3 (2019), 150–163. See also Recogito2 on Github: <https://github.com/pelagios/recogito2> and the comprehensive tutorial guide to using Recogito: <https://github.com/pelagios/pelagios.github.io/wiki>.

<sup>49</sup> These currently include .txt, TEI-XML, .tiff, .jpg, .png, IIIF, and CSV.

<sup>50</sup> The *KIMA* project, an attestation-based gazetteer of Hebrew place names, is a good example of this. The project is explored in more detail in *Part II* of this paper.

<sup>51</sup> Jesse W Torgerson, Wesleyan University, personal communication, July 2017.

<sup>52</sup> Though effectively annotating local copies, it is possible to associate the copy—and thus the annotation—with the original source URI.

<sup>53</sup> At time of writing these include English, French, German, Spanish, Hebrew and Latin, and further NER engines can be added as plug-ins.

<sup>54</sup> For an example of the semi-automated annotation mode in *Recogito* used to analyse a literary corpus, see also C. Porter, R. Milligan and K. Lilley, 'Hidden geographies and digital humanities: analysing and visualising the literary corpus of Humphrey Llwyd', *Literary Geographies*, 6,1 (2020) 96-118.

<sup>55</sup> The balance between NER- and user-generated annotations in *Recogito* is also discussed in Palladino 2018.

<sup>56</sup> See E. Barker, C. Bissell, L. Hardwick, A. Jones, M. Ridge and J. Wolffe, 'Digital technologies: help or hindrance for the humanities?', *Arts and Humanities in Higher Education*, Vol 11.1-2, 2012, 185–200.

<sup>57</sup> G. Vasari, *Le vite de' più eccellenti pittori, scultori, e architettori*, Torrentino (Firenze, 1550).

---

<sup>58</sup> For some preliminary insights into Homer's spatial patterning, see B. Jasnow, C. Evans, and J. S. Clay, 'Poetic and geographical organization in the Catalogue of Ships', *Transactions of the American Philological Association* 148 (2018) 1-44. Critical to this work was the bespoke Neatline mapping exhibit produced for this analysis: <http://staging.ships.lib.virginia.edu/home>.

<sup>59</sup> An example of the use of Recogito to support the study of historical itineraries is briefly discussed in O'Doherty 2021 (this volume).

<sup>60</sup> [https://en.wikipedia.org/wiki/Vicarellò\\_Cups](https://en.wikipedia.org/wiki/Vicarell%C3%B2_Cups).

<sup>61</sup> Egerton MS 2855, British Library. Document available at: <https://doi.org/10.21250/pel06>.

<sup>62</sup> As with texts, annotating images requires that a suitable gazetteer be available and this is still not common at city level. Pompeii is one of the few cases within the Pleiades gazetteer that has URIs for individual buildings. See also V. Vitale and S. De Beer, 'What are urban gazetteers and why do we need them?', *EuropeanaTech* 12 (2019), <https://pro.europeana.eu/page/issue-12-pelagios#what-are-urban-gazetteers-and-why-do-we-need-them>.

<sup>63</sup> For example, were these decisions mostly driven by functionality, prioritising ease of access and affordability? Or were there more agendas at play, such as the wish to portray the alleged greatness of Roman cities with a focus on grandiose buildings and striking decorative features, at the expenses of buildings with greater historical or cultural value?

<sup>64</sup> The other basemaps are aerial and satellite imagery provided by MapBox (<https://www.mapbox.com/>) and the Digital Atlas of the Roman Empire basemap produced by Johan Åhlfeldt (see above).

---

<sup>65</sup> For a more in-depth discussion of pedagogical applications of *Recogito*, see S. Dunn and V. Vitale, 'Linking text and maps: annotation as a critical tool for teaching in the Spatial Humanities', *Literary Geographies* (forthcoming).

<sup>66</sup> Since 2018, *Recogito* has been used in the teaching of Classics, Ancient History, Historical Geography, Mapping Technologies, and Digital Approaches to Cultural Heritage, and the tasks performed have ranged from the identification of ancient places, the creation of map visualisations, the comparison of documents from a spatial perspective, and the extraction of historical place-names in order to build an entirely new gazetteer.

<sup>67</sup> Import formats at time of writing include plaintext, TEI-XML, various image formats including IIIF, and CSV datasets. Annotations can be downloaded in RDF (the native format of LOD) as well as in CSV, a spreadsheet-compatible format. Users solely interested in mapping data can download GeoJSON or KML (formats widely compatible with GIS systems and online mapping toolkits). Network relationships can be downloaded in multiple formats that are supported by mainstream Network Analysis software.

<sup>68</sup> Feature requests for *Recogito* have included: enhancements to the export formats; additional features for controlling the bulk annotation mode; better support for right-to-left languages; functionality to upload a bespoke controlled vocabulary, and for re-arranging tag order in an annotation. Part II discusses at greater length the role of the community, including in the shaping of *Recogito* development.

<sup>69</sup> [https://en.wikipedia.org/wiki/Open-world\\_assumption](https://en.wikipedia.org/wiki/Open-world_assumption).

<sup>70</sup> R. Kahn, L. Isaksen, E. Barker, R. Simon, P. d. S. Cañamares and V. Vitale, 'Pelagios - Connecting histories of place. Part II: From Community to Association' (this volume).

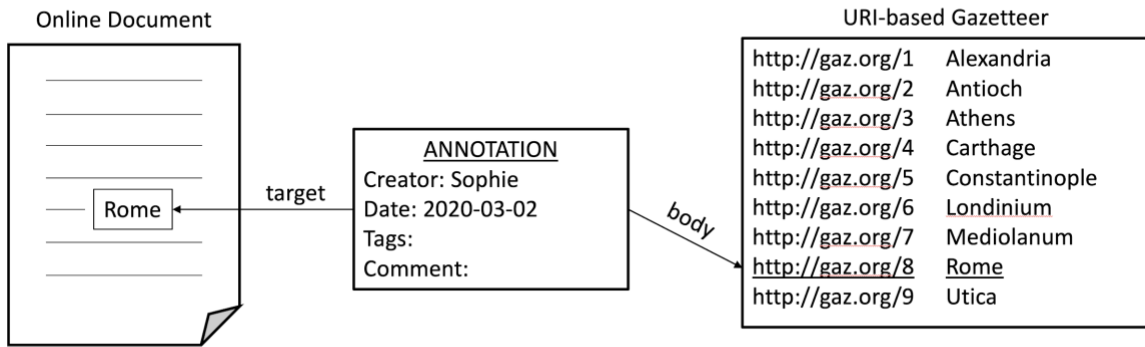


Figure 1 - The notional structure of a semantic geoannotation

Add a Place:

Places Searched:

Corsica  
Ostia  
Aegyptus

CLEAR

Data View:

Arachne

1374 references in selected places

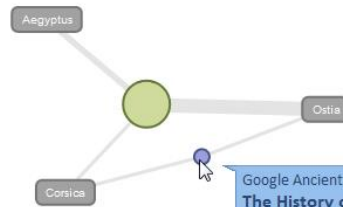
Reference to [Ostia](#) in The History of the Decline and Fall of the Roman Empire:Page 176

The History of the Decline and Fall of the Roman Empire, Edward Gibbon, 1830, Page 176

Reference to [Corsica](#) in The History of the Decline and Fall of the Roman Empire:Page 176

Datasets:

- SPQR
- nomisma
- Arachne
- Google Ancient Places
- Perseus



Google Ancient Places  
**The History of the Decline and Fall of the Roman Empire**  
Page 176

This dataset contains 29 place references in total.

- 1 references to Ostia
- 1 references to Corsica

Click to view details in the Data View!

Map



Figure 2 The Pelagios Graph Explorer

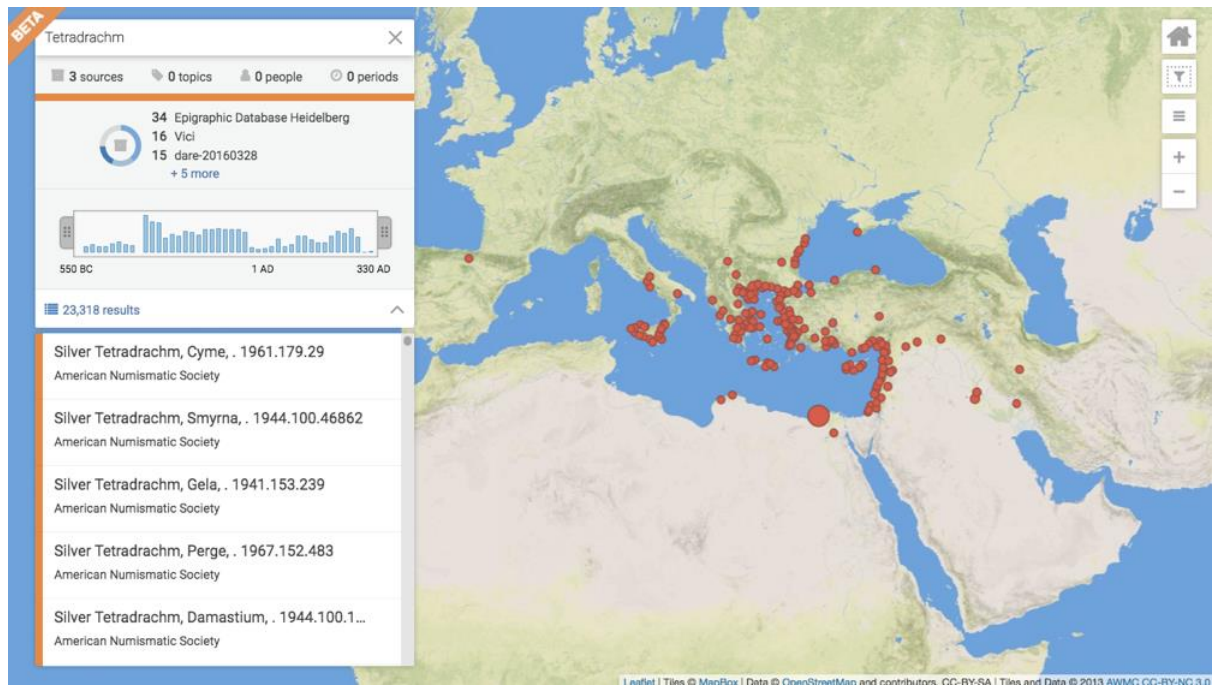


Figure 3 A screenshot from Peripleo, showing the results for the query 'Tetradrachm'. On the left, the complete list of the 23,318 results, the time slider, and the breakdown by contributing dataset

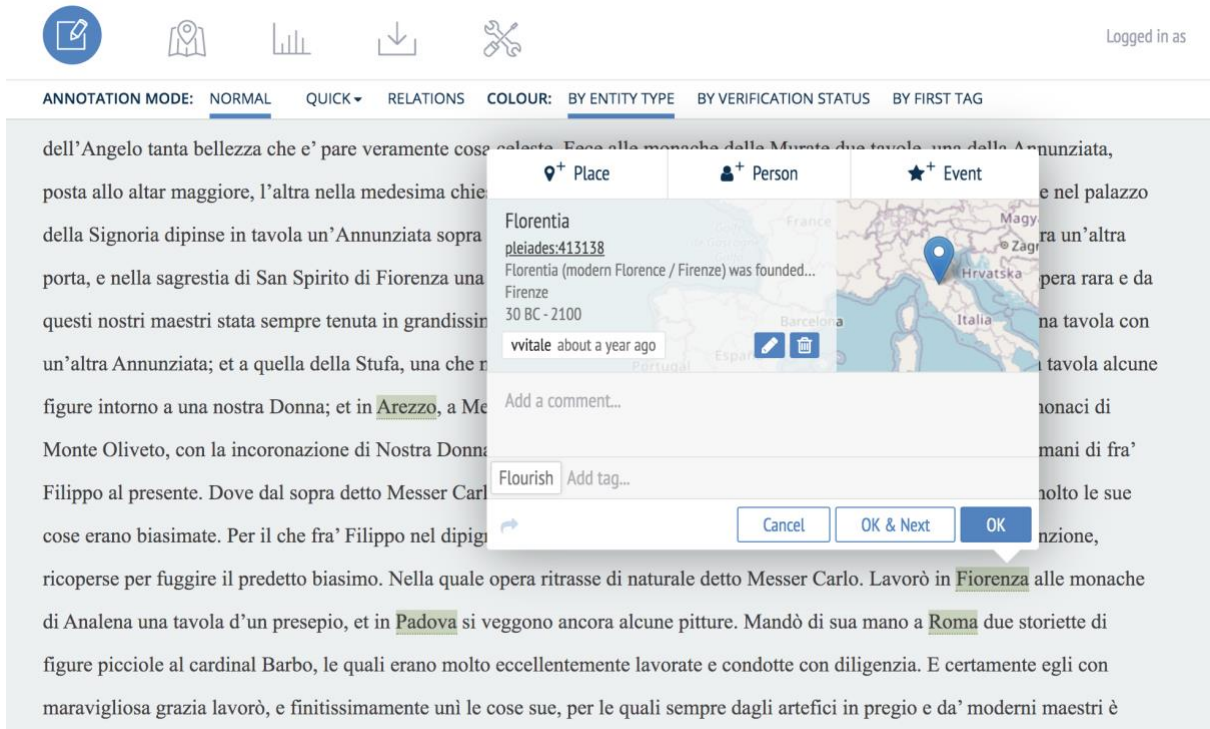


Figure 4 A screenshot from Recogito showing identification of the toponym 'Fiorentina' with a gazetteer URI.

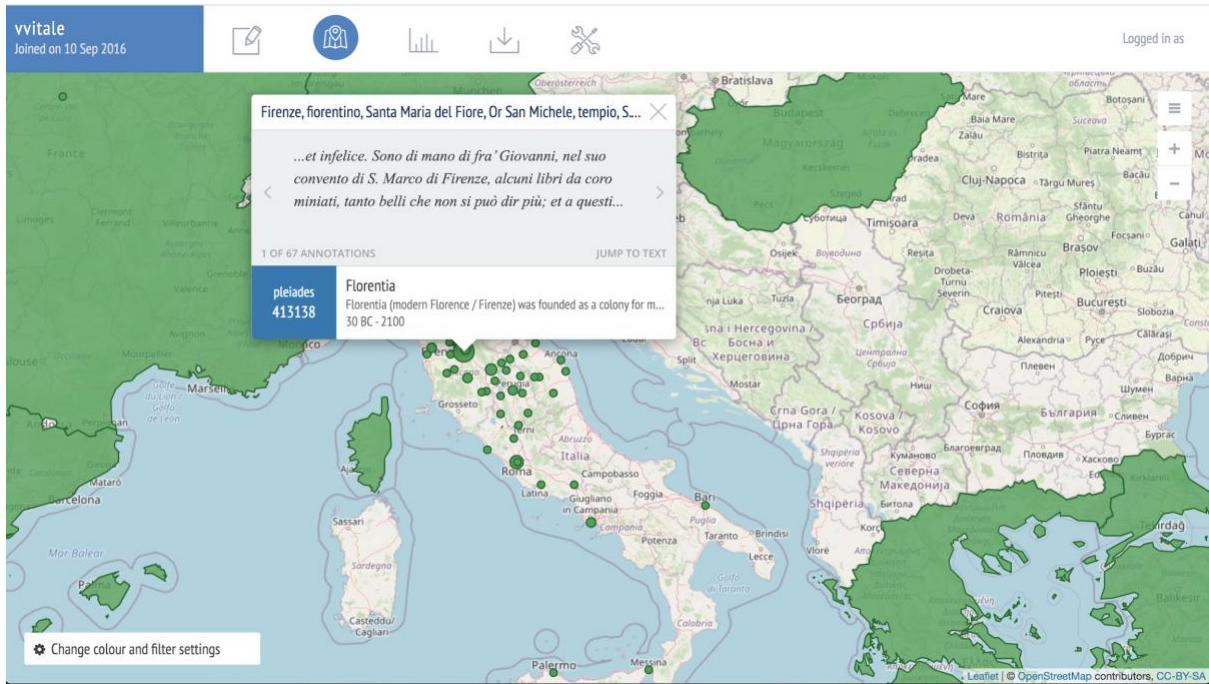


Figure 5 A screenshot from Recogito showing the second book of Vasari's *Lives of the Most Excellent Painters, Sculptors and Architects* annotated in Document View.



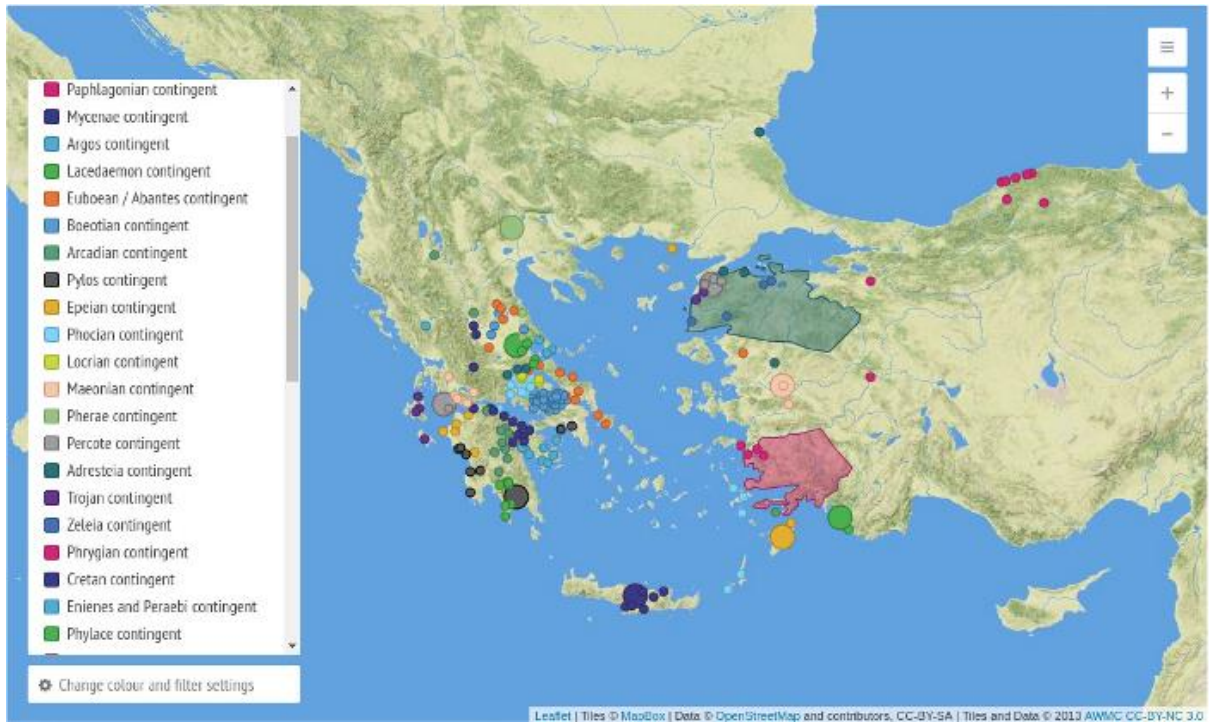


Figure 5 A screenshot from Recogito showing the Homeric catalogue of ships tagged by tribal contingent.



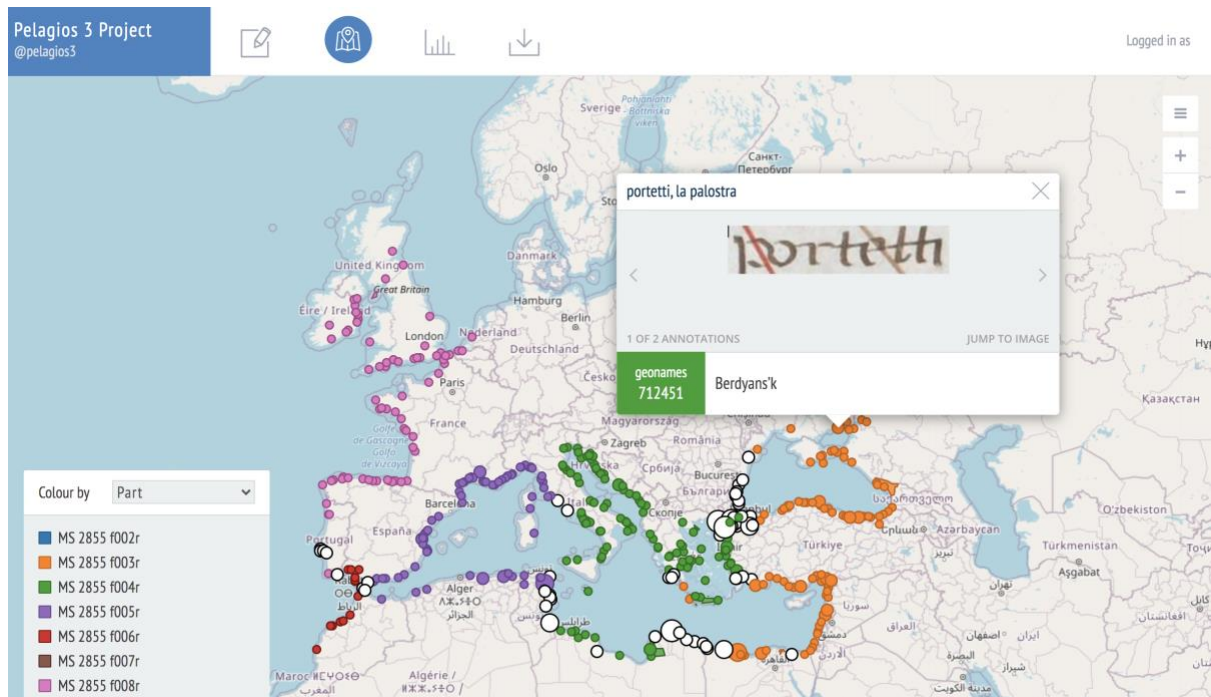


Figure 8 Annotations of the Egerton Atlas displayed in map view. A different colour is assigned to each folio, to visualise the provenance of the annotations.

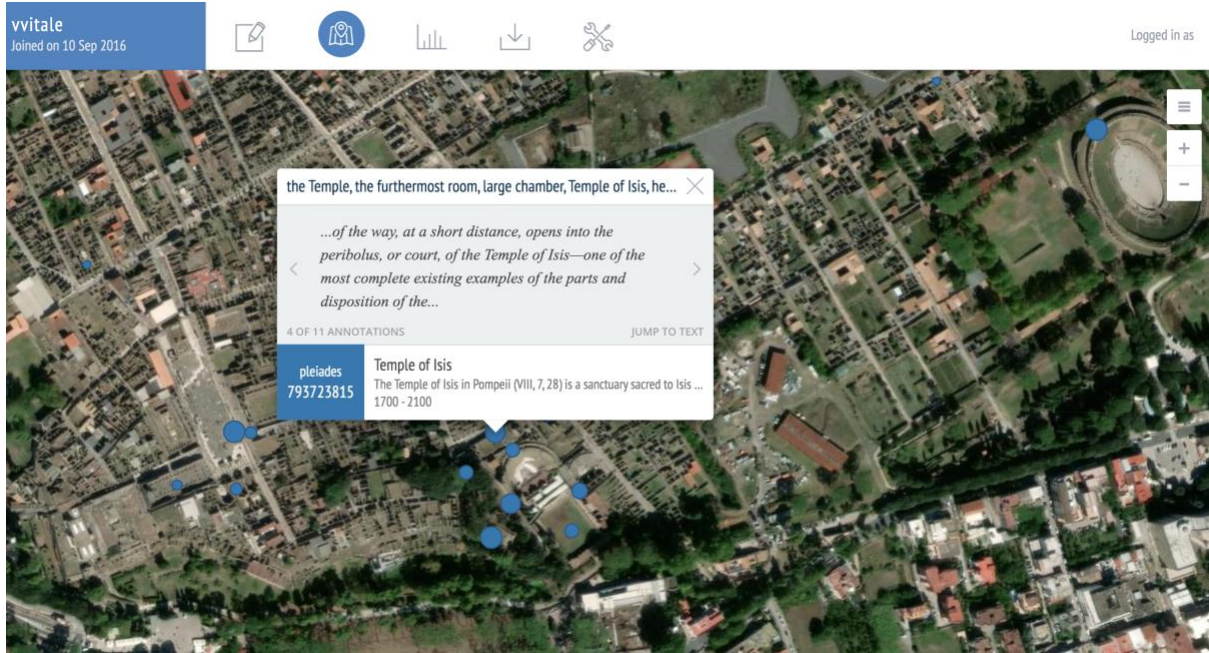


Figure 9 A screenshot from Recogito, showing the visualisation of city-level annotations of the 1827 guidebook *Pompeii Illustrated with Picturesque Views*, by T.L. Donaldson, using the satellite map option.





ANNOTATION MODE: NORMAL QUICK ▾ RELATIONS COLOUR: BY ENTITY TYPE BY VERIFICATION STATUS BY FIRST TAG

if possible, to conquer **the island**; most of them being ignorant of its size and of the number of its inhabitants, Hellenic and barbarian, and of the fact that they were undertaking a war not much inferior to that against the Peloponnesians. For the voyage round **Sicily** in a merchantman is not far short of eight days; and yet, large as the island is, there are only two miles of sea to prevent its being mainland.

It was settled originally as follows, and the peoples that occupied it are these. The earliest inhabitants spoken of in any part of the country are the **Cyclopes** and **Laestrygones**; but I cannot tell of what race they were, or whence they came or whither they went, and must leave my readers to what the poets have said of them and to what may be generally known concerning them. The **Sicanians** appear to have been the next settlers, although they pretend to have been the first of all and aborigines; but the facts show that they were **Iberians**, driven by the **Ligurians** from the river Sicanus in **Iberia**. It was from them that **the island**, before called **Troia**, took its name of **Sicania**, and to the present day they inhabit the west of **Sicily**. On the fall of Ilium, some of the **Trojans** escaped from the **Achaean**, came in ships to **Sicily**, and settled next to the **Sicanians** under the general name of **Elymi**; their towns being called **Eryx** and **Egesta**. With them settled some of the **Phocians** carried on their way from Troy by a storm, first to **Libya**, and afterwards from thence to **Sicily**. The **Sicels** crossed over to **Sicily** from their first home **Italy**, flying from the **Opicans**, as tradition says and as seems not unlikely, upon rafts, having watched till the wind set down the strait to effect the passage; although perhaps

Figure 10 An example of the use of the relations annotations in Recogito: the relationships of between Sicilian tribes as described by Thucydides in the History of the Peloponnesian War