11 Heritage metadata

A digital *Periegesis*

Anna Foka, Kyriaki Konstantinidou, Nasrin Mostofian, Linda Talatas, John Brady Kiesling, Elton Barker, O. Cenk Demiroglu, Kajsa Palm, David A. McMeekin, and Johan Vekselius

Introduction: *Periegesis Hellados* as heritage data

Thinking of literature as spatial information with Geographic Information System (thereon GIS) is emerging into a science known as Geographic Information Science (Harris, Bergeron, and Rouse 2010). The geospatial information community has been contributing methods, ontologies, use cases and datasets compatible to GIS as means of enabling research in the humanities and social sciences. In praxis, the application of GIS for spatial narratives means essentially unfolding their historical, non-cartesian complexity into layers of meaning-making; it can even facilitate a deeper thinking of place *both* as the locus for exploring human activity particularly as a contested terrain of competing definitions *and* as a linking mechanism for information from disparate sources, e.g., the compatibility of text to the actual archaeological data on the ground.

This chapter provides a novel perspective on GIS as both an epistemic device and a method for information organisation by focusing on the process of creating a digital cartographic edition, essentially a GIS of Pausanias's 2nd century CE ten-volume travellers' guide, the *Description of Greece*. The ten volumes comprise a narrative time machine that binds together *place* and *artefact* with its notional *origin* and *purpose*. Methodical but inconsistent in listing temples, statues, hero shrines, altars and other spaces as "Greek" places, Pausanias constructs an idiosyncratic view of Greek cultural heritage. His method, which he mentions in passing, is overtly personal and selective:

"Such in my opinion are the most famous legends (*logoi*) and sights (*theorêmata*) among the Athenians, and from the beginning my narrative has picked out of much material the things *that deserve to be recorded*". Pausanias, *Description of Greece* HYPERLINK "https://scaife.perseus.org/reader/urn:cts:greekLit:tlg0525.tlg001.perseus-grc2:1.39.3/" 1.39.3

To create a contemporary GIS out of a 2nd century CE non-cartesian, literary description of Greek heritage is a challenging scholarly endeavour

DOI: 10.4324/9781003131816-11

with importance beyond the field of classical studies. To start with, Pausanias's reputation as an actual guide for Greek heritage and archaeological finds has fluctuated over the centuries. Recent work, however, suggests that at least some of his descriptions are compatible with the archaeological record, as demonstrated at Delphi by the École Française d'Athènes and in the Athenian Agora by the American School of Classical Studies (Cundy 2016). Indeed, Pausanias's description of place does not always map easily to the archaeology. However, the gaps and disjunctions can be revealing of biases in his description as well as in contemporary scholarship. Examining the compatibility of Pausanias's ten volumes to the archaeological data on the ground is an important gap that the Digital Periegesis project seeks to fill in relation to the humanistic disciplines of classical studies and archaeology.

Moreover, Pausanias's ten volumes provide an excellent case study for additional research gaps that ought to be addressed in relation to GIS – and information organisation more generally – both from an epistemological and a technical perspective. Identifying and describing heritage, artefacts and objects and their association to cultures and space is never a straightforward task. The description of heritage in Pausanias is nearly two thousand years old and it is a "thick" narrative with a lot of disorganised information. It is a representation of material and immaterial culture and its multiple articulations over time. It constitutes an archive of sorts that, in order to be implemented in the technical environment of a GIS, first needs to be sorted in contemporary information organisation terms. From a technical perspective, GIS, with its ability to enrich and to combine layers of information, provides a possibility of combining disparate data, literary, historical and archaeological information. The project applies GIS as a means to organise heritage information to a deeper understanding of the spatial idiosyncrasies of ancient Greek culture, while responding to the broader epistemological and technical questions arising in the intersection between information organisation and digital humanities (DH).

This chapter's purpose is to highlight how GIS can help gather, organise and present heritage information (Dunn 2019; Foka et al. 2020). However, notions of heritage often concern culture and memory related to a given geographical space. Seeing as space becomes a place through the people and stories associated with it, objective heritage information organisation ideally comes with the responsibility of cultural sensibility. Geographic information, spatial data and their organisation are bound with humanistic inquiry and concepts such as ethnicity, cultural memory, conflict and provenance naturally come to the fore (Dunn et al. 2019). The project imbricates the digital and the humanistic thus opening up to the possibility of a deeper understanding of Greek heritage and archaeology, while posing the additional epistemological and technical challenges concerning the humanistic dimensions of information organisation.

In answering the essentially DH research question – how Pausanias's literary heritage information can be best organised and connected to the archaeological record on the ground – the Digital Periegesis project is charting and analysing the relevant digital tools and methods by which extensive semantic annotation and Linked Open Data (LOD) can facilitate the organisation of heritage information in Pausanias's text and its connection to actual archaeological finds. This chapter also discusses the potential application of GIS for such complex pre-cartesian narrative analysis. Finally, it emphasises the importance of building geo-spatially enriched digital editions collaboratively, involving discipline specialist researchers and information organisation experts, with the aim of holistically interpreting histories of place.

This chapter aims to review the state-of-field for using digital heritage metadata in the context of GIS mapping and LOD and to identify key challenges from both theoretical and practical perspectives. The chapter illustrates these challenges and how they can be dealt with a case study of a project using cutting-edge methodologies, the Digital Periegesis project. This allows us to answer research questions about how to organise and link textual data in relation to archaeological material culture, generally, and with regard to Pausanias's *Description of Greece* and places mentioned by him, specifically. This endeavour makes it possible to approach an overarching purpose and address larger issues related to information organisation from epistemological and technical perspectives.

In what follows, we assess Pausanias's ten books from the perspectives of DH and information organisation and in relation to the Wallenberg Foundation project: Digital Periegesis (2018–2021). We begin by drawing together previous scholarship on information organisation in relation to heritage, literature and archaeology. We then proceed to address specifically contemporary heritage initiatives that are preoccupied with spatial information organisation; we describe our case study, more precisely the process of applying computational methods to extract, to organise and to enrich heritage information, monuments and artefacts mentioned in the text. Using the open-source semantic annotation platform Recogito (2021), we record the different aspects that make up "Greek heritage" – the built environment, objects, people, events and stories and how their spatial information is organised. With a focus on marking the location of heritage information, we use Recogito to align Pausanias's places and objects in space to records in global authority files (gazetteers), as well as archaeological databases. As an example of the kind of complexity enabled by Recogito's "free tagging" capability, we discuss the use of relational tags to generate formal data statements that can enrich a broader corpus of organised heritage information. We conclude with reflections on the new knowledge gained by interdisciplinary endeavours at the intersection of information organisation and DH.

Background and related research

Heritage is essentially information organisation in praxis and has come to mean the events, materials or processes that have a special meaning for the memory and identity of certain groups of people. While definitions may vary, heritage is understood as present cultural production that has a resource to the past (Kirshenblatt-Gimblett 1998). As such, heritage springs from modernity's ambitions in information organisation: selecting, ordering, classifying and categorising the world, and simultaneously from threats that force humanity to recognise identities and their tangible or intangible representation (Harrison 2013). With the advent of the nation state in the 19th century, heritage became a challenging and a contested subject. The constant transformation of cultural identities globally due to conflict, migration and colonisation has further contributed to a complexity in understanding what heritage is and if it belongs to someone. Concepts such as a "transnational heritage" or even a "difficult heritage" have not always been specified but are present in disciplines like anthropology, archaeology, history, geography, architecture, urbanism and tourism, constituting a framework that drives applied research internationally (Silverman 2011). Heritage is thus not so much a selection of values as it is a contested subject. Who values what, where and why? And how can these values be described, organised and represented as objectively as possible through the lens of the peoples, places and stories associated with them?

In relation to the organisation of geographic information concerning peoples and cultures, heritage institutions and collections have a legacy in representing complex layers of place, before the utilisation of digital technology. Analogue information such as museums and museum catalogues have a long history of organising, curating and representing place. Spatial information is a part of nearly any curatorial practice or exhibition, more recently addressing questions of complex provenance of fragmented and disembodied artefacts as object "biographies" or "itineraries". The negotiation, organisation and representation of spatial information has always been central to the mission of any heritage institution from their early modern period origins to the Internet (Dunn et al. 2019). The increasing use of digital methods and tools in heritage information management has merely reinvigorated these questions. Indeed, the stark transformation in the way cultural heritage information is now described, communicated and experienced, especially in relation to spatial information raises complex issues pertaining to ownership and authenticity.

Over the past decades the extraordinary growth of new technologies has made it possible to aggregate, organise and analyse archaeological spatial information with GIS (Conolly and Lake 2006; Landeschi 2019; Foka et al. 2020; Trepal, Lafreniere, and Gilliland 2020; Rajani 2021). In praxis, and concerning the information organisation work of heritage institutions, this idea of organising geographic information has been utilised by large

archaeology, architecture, art and heritage stakeholders and their associated entities, most notably, the Getty Thesaurus for Geographic Names (TGN; Getty Thesaurus of Geographic Names 2017). The purpose of the TGN as a structured and organised resource for spatial data is to improve access to geographic information about art, architecture and material culture more generally. The Getty Thesaurus is in essence an organised information system aimed at providing rich spatial metadata descriptions for digital art history and related disciplines. TGN is constructed using national and international standards for thesaurus construction; its hierarchy has tree structures corresponding to current and historical worlds; it is validated by use in the scholarly art and architectural history community; and it is compiled and edited in response to the needs of the user community. All releases are available under Open Data Commons Attribution License (ODC-By). The focus is on historical art architecture and archaeological information and organisation including more recently 1 archaeological sites, lost sites, and other historical sites and 2 building concept hierarchies for historical nations and empires, where a concept hierarchy defines a sequence of low-level concepts to higher-level, more general concepts, e.g., ancient Greece (a country concept) – Peloponnese (a regional concept) – Sparta (a town concept). Thus, information organisation for monuments and artefacts is a well-articulated and documented activity in both scholarly terms and implementation in praxis.

Since the 2010s, the discipline of Geographic Information Science has focused on information organisation and visualisation of non-cartesian textual narratives. The need to combine the organisation of information with complex historical humanistic reasoning has been iterated as a necessary approach: thinking broadly in terms of Geographic Information Science and the complex epistemological concepts of space rather focusing on GIS as a system: "it is in the arena of GISc that the more substantive intellectual engagement and reciprocity between geography, GIS and the humanities will emerge" (Harris, Bergeron, and Rouse 2010). Similarly, the geospatial semantics community has contributed information organisation methods such as folksonomies, use cases and datasets targeting Semantic Web principles and LOD (Janowicz et al. 2012; Mai et al. 2019)

Research on geographic information and its organisation, focusing on traveling literature in particular, has been conducted on historical texts. Examples include the Corpus of Lake District Writing project (CLDW), a corpus of digitised and annotated texts (from 1622 and 1900), in which geographic information was aggregated and organised using automated approaches such as Named Entity Recognition (NER). The project has led to a new methodology, called Geographical Text Analysis. This methodology combines GIS applications with corpus linguistics and Natural Language Processing (NLP), targeting aesthetics, literature and physical geography used in writing about the English Lake District (Rayson et al. 2017; cf. Foka et al. 2020). The organisation of information about place names in novels

published between 1800 and 1914, working with street names in Paris, is another similar project, albeit focused at an urban context. The project combined NLP and NER with textometric tools thus facilitating automated geoparsing of street names (Moncla et al. 2017). Related work focusing on traveling itineraries is the Ben Johnson Walk project focusing on narratives concerning travels in the summer of 1618 (Ben Jonsons Walk 2020) and the City of Edinburgh project – an intra-city geographic information project collecting and organising narratives about the city of Edinburgh (Alex et al. 2019). Finally, according to Barker's Hestia Project (Barker, Isaksen, and Ogden 2016, 181–224), network graphs, by which places were organised and visualised relationally in terms of their action and influence, were a better means of identifying the links and the underlying spatial structure of the narrative, than topographic representation.

Thus, literary narratives seen through the prism of GIS, highlights human complexity, pluralism and the ambiguity of historical concepts of space and time (Foka et al. 2020). In what follows we address how the Digital Periegesis project tackles archaeology on the ground, contemporary technological frameworks for geographic information organisation and exceptionally complex ancient narratives about place and culture. We also reiterate the purpose and aims of the project focusing on methodology, results and discussion. We show how our information organisation schema is rather similar to that used by cultural heritage curators, however, categories and hierarchies are based on the concepts and terminology found in Pausanias.

Case study: Purpose and aims, methodology, results, discussion

Purpose and aims

The purpose of laying out the case study is to demonstrate the application of GIS for documents in praxis, while its central aim is to show how Pausanias's literary heritage information can be best organised and connected to the archaeological record on the ground. In doing so, the team performed an extensive semantic annotation of the volumes and applied LOD principles to facilitate the organisation of heritage information in Pausanias's text and its connection to actual archaeological finds.

Text, methodology and the technical environment

In creating a heritage-data rich version of Pausanias's *Periegesis*, we have focused on (re)using materials and resources already established in the DH community. The text we have used is available in open-license (CC-BY) (in both Greek and English) from the Scaife Digital Library (2021), a reading environment for premodern text collections in both original languages and in translation. The text itself is prepared for organising and linking

information. Documents in the Scaife Library follow the Text Encoding Initiative (TEI), the industry standard for digital texts, which uses a robust interoperable XML-schema to provide enriched and organised information, metadata, such as provenance, edition, book structure and named entities; places and peoples e.g., (for TEI and its evolution, see Burnard 2013). To be able to organise spatial information throughout the ten volumes and in collaboration, Recogito, an open-source web browser platform for semantic annotation was selected, which enables users, without coding expertise, to semantically annotate place information with Uniform Resource Identifier (URI)-based gazetteers to produce annotations as LOD. Recogito is particularly effective in collaborative work, since it keeps track of version history and edit provenance, as well as supporting the downloading of annotations in a range of different data formats.

The method of semantic geo-annotation in Recogito is twofold: (a) reading the document and manually locating and annotating the words that denote heritage in the online document and (b) then resolving and connecting annotations to a digital authority file with organised information about space (a gazetteer) that provides the means to identify and disambiguate between different places. This process is carried out entirely by the annotator who has the opportunity to review the alignment of a word denoting heritage in a document to a global gazetteer URI. The annotators can choose what they consider the appropriate URI and disambiguate that place and map, according to the Web Annotation Data Model (Web Annotation Data Model 2017). Thanks to global gazetteer initiatives, the procedure for identifying and disambiguating ancient place information from documents in Recogito is relatively robust, and can greatly assist comparison and further analysis.

One obstacle that we needed to overcome was where Recogito draws on a suite of established global gazetteers, including Pleiades, the gazetteer for the ancient world. Usually, Pleiades would be sufficient when working on a text from the ancient world, since its coverage spans the Roman Empire (and beyond into Persia). Pausanias's *Description of Greece*, however, presents a challenge, because so much of the narrative takes place within settlements – the place (city, town, village) being the customary baseline for Pleiades (2021). Pausanias's deep dive into places includes descriptions of areas within a city (e.g., the Athenian agora, the Acropolis), and, above all, its heritage monuments – buildings (e.g., temples) and objects (e.g., statues). Very few, if any, of these places or *objects in space* have a record in Pleiades. To address this obvious omission that for reasons unrelated to this study where not an option in the global instance, we hosted a local instance of Recogito, to which we could then upload custom gazetteers in addition to Pleiades and the Digital Atlas of the Roman Empire (DARE 2019). To have more granular topographic and heritage data identifiers, we generated and imported three additional gazetteers. From ToposText.org, an indexed collection of ancient texts and mapped places relevant to the history and

mythology of Greece from the Neolithic period to the 2nd century CE, we collected identifiers for ancient Greek sanctuaries and buildings not yet in Pleiades. For art historical artefacts and monuments in Athens, we derived identifiers and extrapolated coordinates from the late J. Binder's *The Monuments and Sites of Athens: A Sourcebook*, as digitised by J. B. Kiesling for the project, Dipylon (2020). Finally, we utilised a detailed database of ancient art objects mentioned by Pausanias, compiled by T. Hölscher et al., *Bildwerke bei Pausanias*, and included in the database of the Deutsches Archäologisches Institut (DAI). Once these additional resources had been added to our instance of Recogito, we then uploaded the Scaife TEI Greek text of Pausanias, dividing it into the ten books that correspond to the ten volumes of the work which were then assigned to different members of the team, reflecting their disciplinary expertise.

The manual process of digital semantic annotation that is used for the project's case study is extensively described elsewhere (Barker, Foka, and Konstantinidou 2020, 195–202) and hence, is only briefly presented here. The general practice is to manually identify and mark up a word that denotes "heritage" in the broader sense, as a tangible or an intangible manifestation of Greek throughout Pausanias's ten volumes. For example, it could be a word for an architectural monument or an artefact, or even a word that denotes a group of people who carry a specific story of origin or culture, e.g., the Spartans, as proxies to a geographic location, e.g., Sparta.

In addition to manual annotations that require specialised knowledge, Recogito offers a NER option, an automated mechanism for the identification and annotation of named entities, as part of a first, automated sweep of the document, before each annotation is checked and verified by the annotator. NER is currently restricted to European languages, with the default (Stanford CoreNLP) trained models for NER for English language texts.

Since the team is working with the Greek text, NER cannot be applied, therefore we focus on manual annotation only. The annotator then remains in full control – a critical feature in a text where a place may be referred to in terms that are clear only in context, e.g., "the temple" ('ναός'), where the annotator must perform the disambiguation by reading above and below to identify the Temple of Hera at Olympia. In Recogito's annotation screen, the user identifies a character string as a monument or an object in space and then aligns that reference to a suitable gazetteer entry. By virtue of this two-step process, the user not only disambiguates their individual place information and links it to an authority record; by using a gazetteer URI, they also produce LOD annotations by which the place referenced in Pausanias can be linked to other resources mentioning the same place. Selecting a gazetteer entry also has the added benefit of automatically providing coordinates (where available) to map the place. An annotator can also provide additional information in a "comment" field or as "tags". Figure 11.1 shows the working interface for semantic annotation in Recogito. In the figure the word eikōn, is marked up and tagged in the Greek version of *Description*

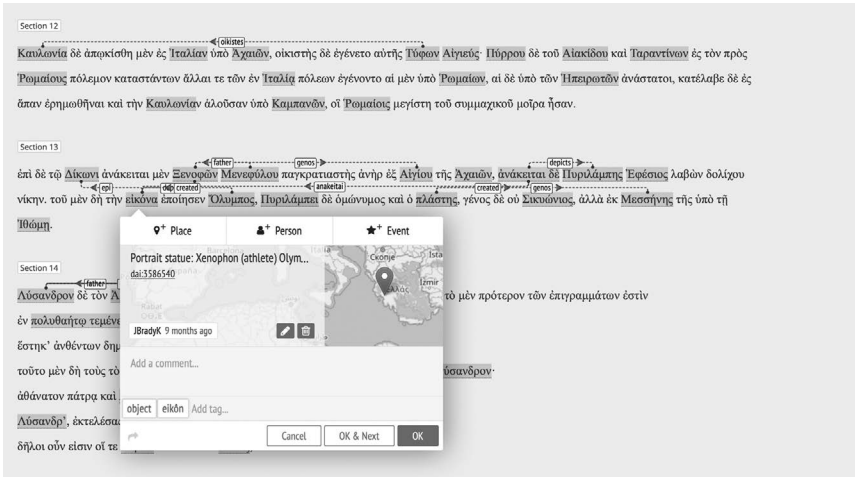

Figure 11.1 Semantic annotation in Recogito of the word eikōn (meaning interchangeably icon, painting, likeness or image) in Pausanias's *description of Greece* 6.13.11, including gazetteer entry and organised free-text tags.

of Greece 6.13.11. As it appears by the interface the user disambiguates and aligns the word with a specific entry. As the bottom line shows the annotator also adds hierarchical free-text tags, object and eikōn, in this case.

Tags in particular have the potential to be an extremely powerful means of organising the data. The pros and cons of collaborative, social and coherent tagging (cf. Golub, Lykke, and Tudhope 2014) were considered. The research expertise of the annotating team and initial research questions of the project, however, guided the choices. After some trial and error and multiple presentations to external reference groups, the research team developed a tagging schema that, while based on Pausanias's own description, helped organise and structure place information as a heuristic tool in a way that could be consistently applied. The scheme is as follows: The first two tags are loosely inspired by FISH, the Forum of Information Standards in Heritage Vocabularies (http://www.heritage-standards.org.uk/fish-vocabularies/), and more precisely their three thesauri: the Object Material Thesaurus, the Monument Material Thesaurus and the Archaeological Objects Thesaurus. The point is to identify different types of heritage objects and monuments (an "ontology" or "typology"). That is to say, while being aware of contemporary modes of organising, e.g., heritage or art historical knowledge, the project group chose the original vocabulary that Pausanias uses in the Greek language (and in expert translation) as far as possible, and generated a schema based on his description, rather than impose one from our own culture which would be culture-insensitive and anachronistic.

Therefore, the annotators decided to adhere to the following tagging guidelines. The first tag establishes broad analytical categories to make

entities easier to group and filter: for large structures such as a city, temple, theatre etc., we use the tag "built" (for the "built environment"); for natural features of the landscape, the tag "physical"; for smaller items (in space), like a statue, artwork, dedication, column etc., "object"; and, if the place represents an inherently unmappable space like Hades, we use the tag "mythical". A second tag is used to capture a key element of the description, using vocabulary driven by Pausanias's own word choice: e.g., a "*naos*" (temple), "*hieron*" (sanctuary), "*bomos*" (altar') or an "*agalma*" (statue as divine offering), "*xoanon*" (roughly carved old wooden image), "*anathēma*" (offering) etc. The third set of tags corresponds directly to the research question of the project. We use the tag "Paus" to signify that Pausanias is writing as if he is physically present at the place at this moment of his narrative. We use the tag "opsis" (sight/sighting) when Pausanias writes about a place he knows from direct experience but is outside the geography of his current narrative – when he does not appear present at the time. This data enrichment allows Pausanias's nominal itinerary to be visualised programmatically and defines a set of actual places of which Pausanias gives a more complex historical and geographical account than the mere record of visiting one ancient temple after another.

Tagging persons, tagging time

It is worth mentioning two other features beyond place information that we have also annotated. Recogito's flexibility allows us to markup prosopographical (referring to persons) and temporal information in addition to spatial data, the difference being the lack of an authority file for the latter two. That is to say, where marking a place is a two-step process – identify the reference in the text; align to the gazetteer record – marking people or time only involves the initial step. This is because, at the time of writing, there is no global authority standard for ancient people or for temporal information in the same way as there are with places. The original vocabulary that Pausanias uses in the Greek language (and in expert translation) comes with variations and discrepancies as well as added complexity, e.g., it could be "Dionysos" in one gazetteer and "Dionysus" in another and they may not even be the same person; also, the temporal metric systems from our own culture do not necessarily map onto contemporary dating classifications (e.g., "323 BCE" or "the Hellenistic period").

Still, it seemed to us that it was also important to mark both entities in Pausanias, not least because of their associations with place and their impact on how those places are viewed. In addition, in a similar way to how have approached the challenge of meeting Pausanias's thick place description by incorporating more granular place-based resources in Recogito to align our references, we developed lightweight, practical measures to disambiguate and authorise our prosopographical and temporal data so far as possible. For the former, this has meant manually aligning named persons in

Pausanias to their Wikidata identifier, by which we will be able to track the gods, heroes, artists, athletes and politicians whose names recur throughout the narrative. For efficient workflow, we annotate personal names in Recogito simply as "person" rather than align them individually. We export these annotated names in Greek as batches, match to their English/Latin forms and align to Wikidata using Excel. We then import again to the final annotation file which is enriched with structured data extracted from Wikidata using OpenRefine, a free and open access data cleaning and information organisation tool.

As for time Pausanias's narrative moves rapidly back and forth in time, from the Golden Age of Greek myth, to the wars between Hellenistic monarchs, to his own period. Capturing these varied chronological elements as one moves through the narrative is challenging. Even more difficult is rendering Pausanias's time descriptions as year dates. Again, there is a need to be sensitive and alert to the nuances of Pausanias's description: how he talks about time – as, say, an event like "the Trojan War", or else through the figure of a mythical/historical person, like "Ptolemy Soter" – is an important aspect to investigate for the reader and there needs to be an informed annotation in place that signifies the time and/or the temporal information of that event.

On the other hand, it is useful if we can also translate those periods into date stamps for visualisation purposes, with which one will be able to explore how the chronological structures of the events described relates to, intersects with, and works against the chronotope of the narrative (e.g., Book 1, chapter 2, paragraph 3). Rich libraries of chronological expressions have been compiled, most noteworthy being the structured authority files for time periods of PeriodO (period.o 2020), a public domain gazetteer of historical, art-historical and archaeological periods. While linking among datasets that define periods differently may be an interesting exercise, the resource is at the time of writing by no means complete, although it helps scholars and students see where period definitions overlap or diverge. However, such terms and their associated date ranges seldom map neatly to Pausanias's narrative which tends to establish a working chronology by using known events such as battles or Olympiads. Fortunately, Wikidata is rich in such items. We can thus annotate the 102nd Olympia mentioned by Pausanias with its Wikidata ID, Q57337793, and extract the year date as a temporal expression, "tx:372 BCE". We can then use relation annotations to link persons, places and events in Pausanias's narrative to a year we can place on a visualisation timeline.

Tagging relations

The aim of the Periegesis project is to not simply catalogue and organise Hellenic heritage according to Pausanias but rather to delve deeper into the meaningful semantic relationships between objects, monuments, people

and events. As noted above, Recogito allows annotations to be linked to one another by any relationship term (e.g., "origin") the project members are interested in defining. The end product of the annotation process is a downloadable nodes and edges CSV format file download that is compatible with social network visualisation platforms such as e.g., Gephi that can be shared and reused on many platforms.

A particularly relation-information rich section is Pausanias's description of the monuments in the sanctuary of Olympian Zeus at Olympia in Books five and six. The Olympic Games brought together elite audiences and performers from the entire Greek-influenced world. Preeminent Greek artists memorialised preeminent personalities there. The relative placement of portrait statues and other dedications within the Altis, the sacred enclosure, in Pausanias's ten volumes is a testimony to dynamic semantic relations of heritage and memory as connected to political power and patronage over the centuries. Pausanias draws distinctions between divine images offered to the gods (*agalma*, *xoanon*) and statues of men (*andrias*, *eikôn*), but the true significance of such terms is not made explicit. Tagging relations using Pausanias's precise nomenclature is thus vital to understanding his description, since it allows us to derive important semantic data from systematic analysis of who is depicted under what circumstance. It is particularly interesting to contrast the role of human portrait statues and divine statues at Olympia, where objects were given a particularly high exposure and had strong social and political implications.

The number of historically charged art objects Pausanias describes, well over three hundred at Olympia alone, and the number of artists, teachers and patrons he mentions, is too large and complex for rigorous organisation without computer assistance. Often, whether through his historical knowledge or the inscriptions he reads on the statue bases, Pausanias provides us with a complete genealogy of the person portrayed. Our relatively basic annotations of the Altis section of Book 6 harvested almost 2,500 instances of 1,110 unique named entities.

Our annotation efforts were designed to distil Pausanias's description into a series of consequent machine-readable statements in Subject -> Verb -> Object form. For the 160 or more portrait statues and statue groups Pausanias lists, the annotations of relations are complex and long and tend to follow the following model: Object A, offered to Zeus by Person B, depicts Person C son of Person D from Place E, in honour of Event F, created by Person G, the student of Person H, at Temporal/Event I, using Material J from Place K, is contained in Place L and located in spatial relation M to Object N and O (which have their own set of similar properties).

Each letter above represents a recognisable Named Entity as the subject or object of our annotation statements: artwork from the Arachne database (DAI 2017); places from ToposText/Pleiades/DARE; persons, events (e.g., Olympiads, battles) and materials from Wikidata. Relationship labels need to be short, to save typing effort, but also unambiguous in their

directionality, since in Recogito they are drawn as arrows from subject to object/target. A tag like "father" is inherently ambiguous, because the relationship could easily run in either direction: is father of, or has as father. To reduce that uncertainty, we regard relationship labels as active verbs, e.g., "father" as "he fathered", "dedicates" (person that dedicates an artwork) and "depicts" (artwork that depicts a person). In most cases, the extracted relations translate directly into Wikidata properties. Our "depicts" maps to Wikidata P180 (https://www.wikidata.org/wiki/Property:P180) "depicts", while "creates" is the inverse of Wikidata P170, "has creator", but in practice can map directly to P800 "has notable work".

When it comes to spatial relationships, to ensure maximum precision and granularity, we elected to retain Pausanias's own terms, transliterated but not translated. Thus, portrait statue A is "*pros*" statue B, that is, close up against it, while statue C is "*ephexes*" statue B, that is, comes next in a series. These relationship tags give us a better understanding of the multiple layers of information that lay within the text: they draw spatial links between monuments on a map, while by referring to space, they illustrate links between people, events and places, thus drawing a lively picture of movements and exchange, and improving our understanding of social, economic and geopolitical relations in Greek antiquity.

Conclusion: Extending disciplines, extending data ecosystems

The Digital Periegesis project set out to create a contemporary GIS out of an ancient non-cartesian, literary description of Greek heritage. While the project at the time of writing is in its final, there are important observations and conclusions to be made from a DH research perspective. First, the validation of the description of Greek heritage by connection to the archaeological information record. To this date, numbers are approximate, subject to change as repeat mentions are integrated and slips are continuously corrected. Of 20,081 identified and marked up place mentions, real place information and coordinates can be assigned to 15,670 of them. A key part of the annotation process is to provide an exportable database of all the 4,113 mentions of places or large objects that are not yet catalogued/mapped in any gazetteer. These can be verified in geographical terms by proximity to another verified spatial entity. The latter include many of Pausanias's 366 or so temple mentions, 174 altars, 304 tombs/memorials and 1,058 sanctuaries. Second, the ultimate result will be a densely annotated digital edition, available in several formats, that can be downloaded, reused and re-explored on its own or integrated into a much broader universe of cultural heritage information describing the ancient world through the eyes of the 2nd century traveller Pausanias.

Heritage items, digitised or not, are often built as manifestations of dominant historical and linguistic approaches; as a consequence, thesaurus and vocabulary standards are following anglophone models and thus may fail to encapsulate the meaning of heritage artefacts, including their original uses

and contexts. Another issue is that information organisation standards and classifications are more generally used (see e.g., TGN) but in praxis, and with case studies as specific as Pausanias's description of Greek cultural heritage, there are additional research specific inquiries to take into consideration. The description of heritage in Pausanias is not only a "thick" narrative with a lot of disorganised information, but much of the original language had to be part of the descriptive parameters of each word denoting heritage. With heritage monuments and cultural data in particular, it is important that information organisation is thus embracing the original context with humanistic sensibility. Using GIS to organise a century long narrative may have a similar issue – places and monuments change names and territories change hands over time. Again, geographic information vocabularies need to be often case specific looking at information concerning space as a conjunction of spatial and temporal information. In other words, while a heritage monument or an artefact may seem as a static point on a map, in reality it comes with a flurry of often disparate information, actual, temporal, cultural that needs to be thought through, compartmentalised and organised in a holistic, culturally-sensible and inclusive way.

Tackling technological concerns is equally important: as disciplines and ideas evolve, so does technology, especially pertaining to information organisation. For example, the Pleiades structure continues to evolve as a robust foundation for place data, with ToposText attempting to follow in its wake. Recogito currently supports a range of export formats, which is likely to be expanded. Assigning coordinates taken from authority structured files such as gazetteers is relatively easy to do manually, but the lesson is that one needs to hold the relevant humanistic expertise to implement.

Finally, perhaps the most important lesson to be learnt concerns interdisciplinarity. The Digital Periegesis project is based on interdisciplinary collaboration, involving discipline specialist researchers such as archaeologists and classical philologists, but also geographers and computational linguists alongside information technology and information organisation experts. The feasibility of the open access platform for easy collaborative annotation facilitated interdisciplinary thinking and implementation. In that sense, one of the important lessons to take with is that while subject and case study specific, the Digital Periegesis project aims to be generative and to be used by the wider communities of classicists, archaeologists and heritage experts and institutions. As such, it corresponds to a more general issue rather than being confined to the study of Pausanias – and in doing so, it makes an ancient traveling narrative thought through technology, relevant to this digital day and age.

Bibliography

Alex, B., C. Grover, R. Tobin, and J. Oberlander. 2019. "Geoparsing Historical and Contemporary Literary Text Set in the City of Edinburgh." *Language Resources and Evaluation* 53 (4): 651–675. https://doi.org/10.1007/s10579-019-09443-x

Arachne database (DAI). 2017. https://arachne.dainst.org/

Barker, Elton, Anna Foka, and Kyriaki Konstantinidou. 2020. "Coding for the Many, Transforming Knowledge for All: Annotating Digital Documents." *PMLA/Publications of the Modern Language Association of America* 135 (1). Cambridge University Press: 195–202. doi:10.1632/pmla.2020.135.1.195

Barker, Elton, Leif Isaksen, and Jessica Ogden. 2016. "Telling Stories with Maps: Digital Experiments with Herodotean Geography." In *New Worlds Out of Old Texts: Revisiting Ancient Space and Place*, edited by Elton Barker, Stefan Bouzarovski, Leif Isaksen, and Chris Pelling, 181–224. Oxford: Oxford University Press. https://doi.org/0.1093/acprof:oso/9780199664139.001.0001

Ben Jonsons Walk. 2020. https://www.blogs.hss.ed.ac.uk/ben-jonsons-walk/map/

Burnard, Lou. 2013. "The Evolution of the Text Encoding Initiative: From Research Project to Research Infrastructure." *Journal of the Text Encoding Initiative* 5, 1–15.

Conolly, James and Mark Lake. 2006. *Geographical Information Systems in Archaeology*. Cambridge: Cambridge University Press.

Cundy, J. E. 2016. "Axion Theas: Wonder, Space, and Place in Pausanias' Periegesis Hellados." PhD diss., Toronto: University of Toronto Press.

Digital Atlas of the Roman Empire (DARE). 2019. https://dh.gu.se/dare/

Dipylon organization. 2020. https://dipylon.org/en/

Dunn, S., G. Earl, A. Foka, and W. Wooton, 2019. "The Birth of the Digital Object Itinerary." In *Museums and Digital Culture: New Perspectives and Research*, edited by Tula Giannini and Jonathan P. Bowen, 253–271. Berlin, Heidelberg: Springer.

Dunn, Stuart. 2019. *A History of Place in the Digital Age*. Routledge: London https://doi.org/10.4324/9781315404462

Foka, Anna, Elton Barker, Kyriaki Konstantinidou, Nasrin Mostofian, Osman Cenk Demiroglu, Brady Kiesling, and Linda Talatas. 2020. "Semantically geo-annotating an ancient Greek" travel guide" Itineraries, Chronotopes, Networks, and Linked Data." In *Proceedings of the 4th ACM SIGSPATIAL Workshop on Geospatial Humanities*, edited by Ludovic Moncla, Patricia Murrieta-Flores, Carmen Brando, 1–9. New York, NY: Association for Computing Machinery.

Foka, Anna, Coppélie Cocq, Phillip I. Buckland, and Stefan Gelfgren. 2020. "Mapping Socio-Ecological Landscapes: Geovisualization as Method." In *Routledge International Handbook of Research Methods in Digital Humanities*, edited by Kristen Schuster and Stuart Dunn, 203–217. London: Routledge.

Getty Thesaurus of Geographic Names. 2017. https://www.getty.edu/research/tools/vocabularies/tgn/index.html

Golub, Koraljka, Marianne Lykke, and Duglas Tudhope. 2014. "Enhancing Social Tagging with Automated Keywords from the Dewey Decimal Classification." *Journal of Documentation* 70 (5): 801–828. http://dx.doi.org/10.1108/JD-05-2013-0056

Harris, Trevor M., L. Jesse Rouse, and Susan Bergeron 2010. "The Geospatial Semantic Web, Pareto GIS, and the Humanities." In *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, edited by David J. Bodenhamer, John Corrigan, and Trevor M. Harris, 124–142. Indiana: Bloomington University Press.

Harrison, R. 2013. *Heritage: Critical Approaches*. London: Routledge.

Janowicz, K., S. Scheider, T. Pehle, and G. Hart. 2012. "Geospatial Semantics and Linked Spatiotemporal Data. Past, Present, and Future." *Semantic Web* 3 (4): 321–332. https://doi.org/10.3233/SW-2012-0077 (accessed May 8, 2021).

Kirshenblatt-Gimblett, B. 1998. *Destination Culture*, Berkeley, CA: University of California Press.

Landeschi, Giacomo. 2019. "Rethinking GIS, Three-Dimensionality and Space Perception in Archaeology." *World Archaeology* 51(1), 17–32.

Mai, Gengchen Krzysztof Janowicz, Bo Yan, Simon Scheider. 2019. "Deeply Integrating Linked Data with Geographic Information Systems." *Transactions in GIS* 23 3: 579–600. https://doi.org/10.1111/tgis.12538

Moncla, Lodovic, Mauro Gaio, Thierry Joliveau and Yves-François Le Lay. 2017. "Automated Geoparsing of Paris Street Names in 19th Century Novels." In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*. 1–8. https://doi.org/10.1145/3149858.3149859

Pleiades. 2021. https://pleiades.stoa.org/.

Pretzler, Maria. 2007. *Pausanias: Travel Writing in Ancient Greece*. London: Duckworth.

Rajani, M. B. 2021. "GIS: An Array of Tools for Archaeology." In *Patterns in Past Settlements: Geospatial Analysis of Imprints of Cultural Heritage on Landscapes*. 83–110. Singapore: Springer.

Rayson, Paul, Alex Reinhold, James Butler, Chris Donaldson, Ian Gregory, and Joanna Taylor. 2017. "A Deeply Annotated Testbed for Geographical Text Analysis: The Corpus of Lake District Writing." In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*. 9–15. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3149858.3149865

Recogito. 2021. https://recogito.pelagios.org/

Scaife Digital Library. 2021. https://scaife.perseus.org/

Silverman, H. 2011. "Contested Cultural Heritage: A Selective Historiography." In *Contested Cultural Heritage*, edited by Helaine Silverman, 1–49. New York, NY: Springer.

Stanford CoreNLP. 2020. https://stanfordnlp.github.io/CoreNLP/

ToposText. 2019. https://topostext.org/

Trepal, Dan, Don Lafreniere, and Jason Gilliland. 2020. "Historical Spatial-Data Infrastructures for Archaeology: Towards a Spatiotemporal Big-Data Approach to Studying the Postindustrial City." *Historical Archaeology* 54: 424–452. https://doi.org/10.1007/s41636-020-00245-5

Web Annotation Data Model. 2017. https://www.w3.org/TR/annotation-model/