Coding for the Many, Transforming Knowledge for All: Annotating Digital Documents

How to cite:


© 2020 The Authors

Version: Accepted Manuscript

Link(s) to article on publisher’s website:

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.

oro.open.ac.uk
ANNOTATION—“A NOTE ADDED TO ANYTHING WRITTEN, BY WAY OF EXPLANATION OR COMMENT”—IS AS OLD AS WRITING ITSELF (“ANNOTATION”). Among the first texts to be written down, Homer’s oral poems survive thanks first to Hellenistic scholars, whose comments and explanations formed the editions that came down to us, and second to later manuscript technology, which enabled the painstaking copying of both the texts and the notes associated with them (see fig. 1). At I Annotate 2019, Gardner Campbell reflected on the meaning of to note and identified as its essence the idea of signing: “A sign that we formulate, a sign that we leave, a sign that points to something, points to a meaning, points to another word, but also points to the pointer. We leave signs; we leave signs; I annotate. The agency in the word note is extraordinary” (00:07:35–58). To note is, as Campbell’s keynote put it, a fundamental act of attention, of sharing, as basic as “water” or “love” (00:06:16–00:07:25).1 To note is an essential human act.

In this essay we reflect on the role and use of annotation in the digital humanities. Just as the manuscript medium provided new opportunities for annotation, the digital revolution has the potential to radically transform what and how (and even why) we annotate.2 In spite of its potential, however, a culture has yet to emerge that broadly supports or makes use of digital annotation in ways that would, like manuscript production in monasteries, establish a mode of behavior—a scholarly practice.3 Rather, communities that practice digital annotation are only just beginning to emerge, and we focus on one here: Pelagios.4 While various methods for, as well as kinds of, annotation can be applied to a digital document, Pelagios has established a process that uses semantic annotation, where concepts (such as places, people, and organizations) are encoded with additional information that is machine-readable.5 Semantic annotation enables humans to identify references to places in individual Web documents and align those references with global place authorities, or gazetteers.6 Through this process, not only can otherwise individually created or curated Web documents be connected, but
heterogeneous materials and resources, from texts to images and databases, can be brought together in dialogue. Through the semantic annotation of place references, Pelagios is building an infrastructure for linking historical or historically related data on the Web.

Annotating place has significance beyond linking data: people are born, live, and die in places; events happen somewhere. Far from being static containers for culture and history, space and place are social entities produced through human agency. The spatial turn in the humanities has accelerated with the advent of digital technology, its new tools and methods giving rise to a new discipline, the spatial humanities. Even though “locating historical and cultural exegesis more explicitly in space and time . . . finds patterns, facilitates comparisons, enhances perspectives, and illustrates data” (Bodenhamer, “Potential” 28), spatial humanities research struggles with the ways digital tools derive from, and tend to reproduce, positivist cartographic methods, making them a poor fit for working with the complexity, uncertainty, and plurality of literary phenomena.

In the rest of this essay we use the process of semantically annotating place to study an early example of spatial representation and to think through issues associated with researching textual constructions of space more generally. Our witness is the second-century CE writer Pausanias, whose Periegesis Hellenidos (A Description of Greece) presents a ten-volume survey of the Greek mainland from Attica to Phocis, following a counterclockwise circuit around the Peloponnese (see fig. 2). His account of the routes through the towns, buildings, monuments and artifacts found in the Greek mainland has been widely used as a guide for interpreting archaeological sites (Hamilakis).

However, using Pausanias as a straightforward guide is problematic, and not only because he shows little interest in describing natural environs or infrastructure. As Pausa-
nias himself points out, in what amounts to a statement of his method: “Such in my opinion are the best-known stories [logoi] and sights [théorēmata] that the Athenians possess, and from the beginning my own narrative [lo-gos] has picked out, from such many things, those that stand out as worth writing-up [sun-graphein]” (1.39.3). Pausanias’s description is predicated on what he finds interesting—past stories from the places through which he passes and the sights he wants us to see. This is not a straightforward guide to topographical realities but a virtual pilgrimage through an imaginary space, a past-present Greece that exists not on the ground but in the text, and in the memories and myths of local landscape(s). The challenge of analyzing the organization of space and place in the Description of Greece is the thickness of its spatial representation, whether Pausanias is taking the reader on a tour of a temple precinct, stopping on a road to take note of a statue, or recalling the mythical stories associated with a rock.

We are working on a project to produce a digital Periegesis, which aims to identify, trace, and explore the spatial form of, and the forms of space in, Pausanias’s narrative—the ways in which place, objects, and peoples in space are described and how the narrative is organized spatially. In its similar investigation of the fifth-century BCE Histories of Herodotus, the 2008–10 project Hestia had demonstrated the value of digitally mapping literary constructions of space in ways that challenged normative views of that space, mapping and studying places according to how Herodotus related them to other places rather than according to topographical location (see fig. 3). Challenges remained, however, not least of all because spatial information had to be encoded by hand in an Excel spreadsheet (Bouzarovski and Barker) and was of such complexity that visualizations resembled “spaghetti monsters” (Malkin 18).
This brings us back to the idea and practice of semantic annotation. In the effort to facilitate linking data, Pelagios has developed an open-source browser-based tool called Recogito, which enables researchers to semantically enrich their material without the need for coding skills. Using this tool, the team working on the digital Periegesis is annotating both the entities (places, people, and events) that Pausanias describes and the relations among them, as he moves through space (and time) linking places, objects, and stories to each other. Annotating the places themselves follows the two-step process outlined above, first identifying a word as a place and then aligning that reference to a gazetteer, for which Recogito helps by providing a map-based pop-up of options. Its tagging feature further enables us to provide additional information on, and construct a schema for thinking about, place in more depth, such as specifying whether the place is physical (a river, mountain, etc.), built (a city, temple, altar, etc.), regional (a wider geographical area), or mythical (fig. 4). By developing this schema, we are in effect producing a searchable database created from, and directly linked to, our annotated text.

While annotating people provides important context for analyzing the text’s spatiality, using the “event” category to identify relations is a good example of how Recogito can be customized. Where Pausanias moves through or over space, describing places (or objects) in the landscape, we mark his description as topographic. Where he moves through time, noting the history of a particular place or object, we use the category chronotopic (Bakhtin’s term). Finally, where Pausanias compares one place with another, we annotate it as analogic. The first two categories can be further defined with tags. Pausanias’s description of place or objects in space (topographic) can be understood as either hodologic—that is, a description of a place in its landscape as the narrator moves through it (e.g., “A little
farther away from the gateway, on the right as you go in, is a bronze Heracles,” [2.3.2])—or synoptic, as when a place is described as if from above (e.g., “Corinth is part of the Argolid” [2.1.1]). His movement through time (chronotopic movement) is potentially even more revealing of the depth of his description, since his (re)use of the past builds the sense of place. From our initial work, movement appears to be a key criterion: when Pausanias introduces a historical and mythical account, it is largely in terms of how a person representing one place moves to another, intervenes in it, or even transforms it, actions that can be categorized using the relational tagging feature (see fig. 5). Annotations like these present a picture of how the narrative moves not only through space and time but through different spaces at different times.

While Recogito offers a suite of different download formats to enable other applications to further analyze the annotations and place references, insights can be gleaned simply from its built-in map visualization. While figure 2 reproduces Pausanias’s tour of the Greek mainland, and shows the overt structure of his narrative, it fails to capture the breadth and relatedness (let alone depth) of his text’s spatial construction. Contrast figure 6, which displays all the places currently annotated in the digital Periegesis. Though by no means comprehensive, it is enough to demonstrate the extent to which Pausanias relates this core territory to places all around the Greek world.

Annotating in Recogito has raised three key interrelated points. First, reading. Much of the new, exciting work of the digital humanities has focused on big data and seemed to emphasize distant reading. Paradoxically, our experience of working on a digital platform—and developing a practice of annotation—has meant thinking about individual places in terms of a broader schema and their relations to each other, bringing us closer to the text and slowing our progress through it. Second, visualization. The ability to “place”
places enables a narrative to be experienced spatially, the value of which extends far beyond simply mapping text. By creating a dialogue between the text and the map (through annotation) we can better trace and analyze how narrative is grounded in cultural ideas of space, and how, in turn, those ideas are deeply embedded in ideological structures. Lastly, collaboration. The work that we have been describing here far exceeds the work of an individual scholar. Not only are we ourselves a team; we are using tools built by others and developing a methodology informed by scholars working on similar issues. It is this vision of interdependent groups working alongside each other that we believe to hold the greatest potential for the digital humanities. It is a legacy that goes back to the many hands that, annotating alone and together, helped preserve the texts that we have in the first place.

Notes

1. According to Campbell, the noun form of note is one of the earliest 2% of entries recorded in the Oxford English Dictionary, along with such words as fall, run, and water, and the verb form of note is one of the earliest 3% of entries, along with call, love, and put (00:06:16–28).

2. On the multitextual reading facilitated by manuscript technology, and the capacity for digital annotation to enable similar linking strategies, see Ebbott. See also Dué and Ebbott, which takes the long view on how various new technologies (the manuscript, the digital text) affect editorial choices.

3. On the interdependence of literary and sociopolitical form, see Levine, especially the discussion of the importance of the monastery in the formation of the university (57–65). On form structuring everyday experience, see Foucault. On communities of practice (habitus), see Bourdieu.

4. Hypothesis, the sponsor of I Annotate, is another such community (web.hypothes.is/about/).

5. For information about Pelagios, see its Web site, pelagios.org/about-us/; Barker and Isaksen; and Vitale et al. On semantic annotation, see “What Is Semantic Annotation?”
6. As well as providing information about places that are useful for their understanding—such as coordinates, type, and bibliography—gazetteers critically provide Uniform Resource Identifiers (URIs), which are essentially unique code numbers used to identify and disambiguate different places. See Simon et al., “The Pleiades Gazetteer”; and the World-Historical Gazetteer (whgazetteer.org/).

7. On space and place as active settings for historical action, see Certeau. On the notion that they are social products, see Lefebvre. On the idea that place is space that is valued, see Tuan (1977). See also Cresswell; Massey.

8. On the potential to explore “subjective geographies through the spatial representation of qualitative, or fuzzy, data,” see Gregory and Cooper 89.

9. Geographic information systems (GIS) find it difficult to handle uncertain, incomplete, and ambiguous data (Gregory and Healey). Text-based research raises “fundamental epistemological and ontological issues for GIS applications” (Harris et al. 228). Conventional cartography is “typically totalizing, usually two-dimensional, Cartesian, and very undialectical” (Harvey 4).

10. A greater literary appreciation for Pausanias’s descriptions has emerged only recently, since Habicht. See Alcock and Osborne; Arafat; Pirenne-Delforge; Alcock et al.; Akujärvi; Hutton; Pretzler.

11. Whitmarsh observes how Pausanias presents his target of describing πάντα . . . τὰ Ἑλληνικά (“all that is Greek”) in terms of what is ἀξίων μνήμης (“worthy of memory”; “Mnemology” 49). On Pausanias’s constructed itineraries see Hutton. On the idea of virtual pilgrimage I have benefited from conversation with Jody Cundy; on this topic see also Elsner; Rutherford. On how Pausanias traces paths of knowledge, see Hawes, “Pausanias’ Messenian Itinerary.” On Pausanias and myth see Hawes, “Of Myths.” On so-called glocalization see Whitmarsh, “Thinking Local.”

12. Pausanias presents a thick account by offering a series of local snapshots in a broadly Hellenic framework (Hawes, “Pausanias”).

13. See the project’s Web site: periegesis.org/. Deriving from the verb periēgeisthai (“to lead or show around”), Periegesis has the double sense of description (of place) and movement (through space and time). The Digital Periegesis is funded by the Marcus and Amalia research foundation for three years (2018–21).

14. See the project’s Web site: hestia.open.ac.uk/. For a discussion of the technology involved in the project, see Barker et al., “Telling Stories.” For the argument that Herodotean space is fluid, transformative, and relational, see Barker and Pelling. On literary maps in general, see Moretti.

15. For details about Recogito (recogito.pelagios.org), see Simon et al., “Linked Data Annotation.”

16. People are often proxies for place in that their actions occur in a particular time and place, or else they...
can act as a proxy for place: e.g., when Herodotus mentions Darius, it is most frequently in his capacity as a representative of Persia (as the king) rather than as an individual. See Bouzarovski and Barker. Annotating people in Recogito is currently a one-step process, though attempts are ongoing to create global prosopographies on the model of gazetteers; see, e.g., SNAP:DRGN, a global authority list for ancient people (snapdrgn.net/).

17. The first two aspects have been broadly recognized. For movement through space, see Hutton. For movement through myth, see Hawes, “Pausanias” and “Pausanias’ Messenian Itinerary.” In her PhD dissertation, Cundy identifies an equivalent to our third aspect, when the text, which is for the most part “pedestrian” and “hodological,” undertakes great leaps in space (141). The “hyperbatic” moment disrupts “the topographic flow of the text and takes the reader to a far off-place” (145).

18. On both aspects in Herodotus, see Bakker.

19. For movement and transformation as key markers of spatial relations, see Bouzarovski and Barker.

20. Other kinds of annotation can be made; e.g., tags can mark focalization, locatives, and time phrases.

21. Recogito enables many download options (e.g., CSV, KML, and JSON-LD) for use in a variety of applications, including simple spreadsheets (like Excel), GIS, and Gephi, as well as for consumption as linked data.

22. On maps as part of the investigative process and as a new critical practice, see Bodenhamer, “Narrating Space.”

23. For instance, we are using Brady Kiesling’s Web site topostext.org to build new gazetteers for ancient Athens and Corinth, contributing to the work of Greta Hawes and Scott Smith in compiling a list of mythical names from primary sources, and working with Chelsea Gardner and Rebecca Seifried to establish the ground truth of travel accounts of Mani. See Gardner and Seifried.

WORKS CITED


A portion of the manuscript known as Marcianus Graecus Z.454 (= 822), or Venetus A, which is the earliest extant manuscript of Homer’s Iliad. Shown here is folio 12 recto, which contains book 1, lines 1–25. The image can be found at www.homermultitext.org/hmt-image-archive/venetus-a/VA012RN-0013.tif.
The world of book 5 of Herodotus’s *Histories*, organized by action and influence rather than by topography, visualized using Science of Science (Sci2), a tool provided by Indiana University and SciTech Strategies (http://sci2.cns.iu.edu) with the support of Scott Weingart. The size of the font shows the strength of the relations of each place; the bigger the font, the more relations a place enjoys.

Recogito allows researchers to annotate a place in Pausanias by identifying the character string as an entity, aligning it with a global gazetteer, and defining it by tags.
Annotating the “event” entity to mark all the relations in a given section according to whether they are topographic, chronotopic, or analogic.

All the places to which Pausanias refers in the *Periegesis* annotated using Recogito’s built-in map visualization.