MAKING SENSE OF THE WAYS WE MAKE SENSE OF THE PAST: THE PERIODO PROJECT

ADAM RABINOWITZ, RYAN SHAW, SARAH BUCHANAN, PATRICK GOLDEN, AND ERIC KANSA

1. Introduction

Digital classicists have been enthusiastic proponents of the 'spatial turn' in the Digital Humanities. Their material is distributed across the broader Mediterranean world, and they have benefited from the precocious digitization of textual corpora carried out by the Perseus Project and the geographic context provided by the *Barrington atlas of the Greek and Roman world* and its online successor, the Pleiades gazetteer.[1] These resources place the field at the forefront of the neo-geographic approaches to literature so important to digital humanists. In particular, texts encoded with Text Encoding Initiative (TEI) markup (and inscriptions with the EpiDoc extension) and the decision to build Pleiades as a Linked Data gazetteer have offered digital classicists the resources not only to make maps from texts, but to connect those maps with a range of external data.[2] Most recently, the Pelagios project has used the Pleiades gazetteer and Linked Data principles to aggregate geolocated information from diverse online sources.[3] The map-based tools and visualizations that have emerged from these efforts are varied and sophisticated.[4] Despite the vast chronological range of the classical past, however, digital classics projects are only beginning to pay attention to time as a factor in data discovery and visualization.

Here digital antiquity lags behind the spatial humanities movement, which has framed time as the next big challenge in mapping and spatial analysis.[5] Scholars of ancient literature

[1] G. Crane, ed., 'Perseus Project', http://perseus.tufts.edu; *Barrington atlas of the Greek and Roman world,* ed. R. Talbert (Princeton 2000); R. Bagnall and R. Talbert, ed., 'Pleiades gazetteer', http://pleiades.stoa.org.

[2] See the online companion to D. Roller's edition of Strabo's *Geography* (Cambridge 2014) or the work on Herodotus carried out by the Hestia Project, http://hestia.open.ac.uk/: E. Barker, L. Isaksen, N. Rabinowitz, S. Bouzarovski, and C. Pelling, 'On using a digital resource for the study of an ancient text: the case of Herodotus' *Histories*', in *The Digital Classicist*, BICS Supplement 122 (2013) 45–62; E. Barker and S. Bouzarovski, 'Between east and west: movements and transformations in Herodotean topology', in *New worlds out of old texts*, ed. E. Barker *et al.* (Oxford 2016) 155–80.

[3] L. Isaksen, R. Simon, E. Barker, and P. de Soto Cañamares, 'Pelagios and the emerging graph of ancient world data', *Proceedings of the ACM conference on web science* (2014) 197–201.

[4] See for example C. Evans and B. Jasnow, 'Mapping Homer's Catalogue of Ships', *Literary and Linguistic Computing*, 29 (2014) 317–25; 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-Perimetron*, 10 (2015) 49–59; and see soon the Peripleo interface at http://pelagios.org/peripleo.

[5] I. Gregory, 'Exploiting time and space. A challenge for GIS in the Digital Humanities', in *The spatial humanities*, ed. D. Bodenhamer, J. Corrigan, and T. Harris (Bloomington 2010) 58–75; E. Ayers, 'Mapping time', in *Geohumanities. Art, history, text at the edge of place*, ed. M. Dear *et al.* (London 2011) 215–25.

and history, who usually deal with specific dates, may not see what the fuss is about. After all, tabular representations of dated events have been a feature of Graeco-Roman scholarship since the *fasti consulares* were inscribed on monuments in the Roman Forum, and there are now digital tools that can be adapted from fields like journalism to visualize events along a temporal axis.[6] This reaction, however, ignores two important theoretical issues with major implications for digital practice.

The first is the recognition that space and time are no more separate in human cognition than they are in theoretical physics. Peta Mitchell, writing in a recent volume on the 'spatial turn', has pointed out that from Bakhtin to Foucault, the spatial and the temporal have been inextricably intertwined in critical theory. Rejecting, with Foucault, Kant's separation between 'space as geography' and 'history as time', Mitchell embraces Bakhtin's idea of the 'chronotope', or a fusion of space and time.[7] Although Mitchell applies this concept to a spatio-temporal understanding of modern literary works, the interconnection is equally applicable to digital antiquity. In fact, several recent digital classics projects have integrated space and time through mash-ups of timeline and map-based visualizations.[8] However, these projects have been more successful in integrating space with narrative 'time' (sections, chapters, pages) within a given text than they have in combining geography and historical time. Where absolute dates are absent or uncertain, as they often are for the ancient world, scholars fall back on verbal descriptions of time that do not lend themselves to digital manipulation.

This raises the second theoretical barrier to the use of time as an axis for digital discovery and visualization: our approach to past time is as much qualitative as quantitative. It is not just a matter of terminology and differences in calendrical systems. This is certainly a complex issue in the ancient world, as exemplified by Thucydides' famous expression of the date of the beginning of the Peloponnesian War: 'In the forty-eighth year of Chrysis being priestess [of Hera] in Argos, with Aenesios being ephor at Sparta, with two months left to go of the archonship of Pythodoros at Athens, in the sixth month after the battle at Plataea and right at the beginning of spring' (2.2). But, as Thucydides shows us, these calendars can be reconciled. After all, they are all based on linear counts of the same cyclical astronomical phenomena: either the waxing and waning of the moon or the revolution of the earth around the sun.

The problem, instead, is the way we conceptualize time when dealing with broader historical trends and material remains. Within the growing digital dataset related to the ancient world, absolute calendrical dates are a rarity. It is no coincidence that the timeline filter in Pelagios' new Peripleo search interface returns results consisting mainly of coins. Coins are an ideal subject for geotemporal visualization: most are given precise calendrical

[6] For example, the Knight Lab TimelineJS interface: https://timeline.knightlab.com/.

[7] P. Mitchell, '"The stratified record upon which we set our feet": the spatial turn and the multilayering of history, geography, and geology', in *Geohumanities*, ed. M. Dear *et al.* (n. 5, above) 71–83 (72–74 and n. 9).

[8] Many of these projects are based on Nick Rabinowitz's timemap.js Javascript library (https://github.com/datadesk/timemap), which combined MIT's SIMILE timeline and web-mapping services. This served as the base for the GeoDia project (http://geodia.laits.utexas.edu: A. Rabinowitz, 'GeoDia: or, navigating archaeological time and space in an American college classroom', in *Archaeology in the Digital Era*, ed. G. Earl *et al.* (Amsterdam 2013) 263–72) as well as for GAPVis (https://github.com/nrabinowitz/gapvis: E. Barker, K. Byrne, L. Isaksen, E. Kansa, and N. Rabinowitz, 'The Geographic Annotation Platform–a framework for unlocking the places in free-text corpora', in *NeDiMAH workshop at Digital Humanities 2012 Conference (DH2012)* (2012); L. Isaksen, E. Barker, E. Kansa, and K. Byrne, 'GAP: A NeoGeo approach to classical resources', *Leonardo* 45 (2012) 82–83), used by both Hestia and the Hellespont Project, http://gapvis.hellespont.dainst.org/#index.

dates, and even when they lack findspots, they can be associated with mint locations. In contrast, other artefacts from the classical world can be connected with places, but many can be dated only broadly. When absolute chronological coordinates for objects or phenomena are absent or problematic, scholars organize evidence into chronological groups according to shared characteristics, and then attempt to place those groups in temporal relation with each other. These chronological groups are what we call 'periods'. Unlike calendar dates or historical people or places, which existed in the past, periods exist only in the discourse of those who are looking back at the past from a later vantage point. And since different phenomena occurred at different times in different parts of the world, periods, like Bakhtin's chronotopes, combine both time and space.

Because periods are discursive constructs that have no objective existence in the past, their boundaries and characteristics are often implied rather than expressed in scholarship, on the assumption that a period concept will be semantically transparent to a particular community. But the particular fusion of time and space that characterizes a period concept for one scholar is not necessarily the same as that which defines the same concept for another. Moreover, the same period term can mean wholly different things in different geographic contexts ('Archaic' is construed very differently in North America than in the Mediterranean). As a result, it is hard even for human specialists to translate an undefined period term into a set of definite space–time coordinates without additional contextual information, assumptions, or guesswork. It is vastly more difficult for a computer to transform a qualitative concept into the quantitative data necessary for graphic display in a digital interface.

Fuzziness, inconsistency, and spatial variability create problems not just for digital visualization, but — more importantly — for the interoperability of periodized data across different databases. That does not mean that periods have been ignored by scholars working on metadata and ontologies; on the contrary, there have been significant advances in recent years. Sophisticated work has been done on the semantic modelling of the concept of the period:[9] general period thesauri are now published in machine-readable form on the web (*e.g.* the period terms in the Getty Art & Architecture Thesaurus), as are more precisely-defined and well-modelled thesauri for specific regions (*e.g.* the English Heritage period vocabulary presented through the SENESCHAL project).[10] As early as 2003, specifications for period gazetteers were suggested by the Electronic Cultural Atlas Initiative.[11] Yet despite more than a decade of effort, no period resource has emerged that allows us to carry out analysis, connect related data, and create visualizations with time — as has been the case for spatial gazetteers with place — let alone examine the diachronic development of period concepts across the disciplinary traditions of history, classics, and archaeology.

[9] M. Doerr, A. Kritsotaki, and S. Stead, 'Which period is it? A methodology to create thesauri of historical periods', in *Beyond the artifact. Digital interpretation of the past*, ed. F. Niccolucci and S. Hermon (Budapest 2010) 70–75; M. Doerr, A. Kritsotaki, and S. Stead, 'Thesauri of historical periods–a proposal for standardization', in *Proceedings of CIDOC Conference* (2005); C. Binding, 'Implementing archaeological time periods using CIDOC CRM and SKOS', in *The semantic web: research and applications, Part I*, ed. L. Aroyo *et al*., Lecture Notes in Computer Science, 6088 (Berlin 2010) 273–87; F. Niccolucci and S. Hermon, 'Representing gazetteers and period thesauri in four-dimensional space–time', *International Journal on Digital Libraries* (2015) 1–7.

[10] Heritage Data, 'SENESCHAL', http://www.heritagedata.org/blog/about-heritage-data/seneschal/.

[11] M. Feinberg, R. Mostern, S. Stone, & M. Buckland, 'Application of geographical gazetteer standards to named time periods', http://www.ecai.org/imls2002/time_period_directories.pdf (2003); V. Petras, R. Larson, & M. Buckland, 'Time period directories: a metadata infrastructure for placing events in temporal and geographic context', in *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries* (2006) 151–60.

The PeriodO project (http://perio.do) is an attempt to bridge this gap with a Linked Data gazetteer — but a gazetteer in a special sense. Unlike gazetteers of places, which provide identifiers for unique entities with multiple names, the PeriodO gazetteer documents the usage of period terms in scholarship. That is, it catalogues not global period *concepts*, but specific period *definitions*: authoritative assertions about the chronological and geographical coverage of period concepts, expressed using machine-readable coordinates (including start and end dates as well as geographic boundaries). By transparently modelling period definitions with quantitative spatio-temporal coverage and dereferenceable unique identifiers, it aims to facilitate the discovery of chronologically related data across heterogeneous digital resources. Existing period vocabularies offer a single, non-overlapping set of definitions, but these cannot cover all use-cases or deal with disagreement, so either precision in time and space or broad applicability must be sacrificed. The PeriodO gazetteer takes a multivocalic approach that cuts this Gordian Knot by embracing differences of opinion and documenting period definitions from different scholars, times, and geographic regions. In the following sections, we present the rationale, structure, and future of the project, which has just finished a phase of development funded by the National Endowment for the Humanities in the US (grant number HD-51864-14) and is entering a second phase funded by the Institute of Museum and Library Services (grant number LG-70-16-0009-16). First, however, we must review the history of the periodization of the past to understand why periods present such a challenge for digital interoperability.

2. A brief history of periodization

The desire to organize the human past into distinct stages appears in the Mediterranean world with the earliest works of classical literature. Although Homer is more concerned with geography, a rough division of time can be glimpsed in the *Iliad*, in the narrator's consciousness of the difference between the heroes of the Trojan War and men 'such as they are now' (*e.g.* 5.303-04). This division is made more explicit in Hesiod's *Works and Days,* which outlines an historical sequence of five ages — or rather γένη, 'generations' or 'races'. The first three of these γένη were associated with metals, arranged in declining order by value: gold, silver, bronze. The fourth included Homer's heroes, and the fifth, Hesiod's own age, turned back to a metal (iron) that represented both contemporary technology and the lowest position on a scale of metallic value. For Hesiod, the succession of generations represented a process of degeneration (*Erga* 174–78). The idea that human existence tends toward decline is hardly unique to Greek myth, but its explanation in terms of successive phases with distinct individual characteristics provides the template for the notion of the 'period'. This template is a relative one, for Hesiod's formulation, like Homer's, only identifies three temporal categories: 'earlier' (in the past), 'later' (in the past), and 'now.'[12] But absolute dates can be attached to it: two hundred years later, Herodotus would use Egyptian historical records to pin the Trojan War, and with it the 'generation of heroes', to a date between 800 and 900 years before his own time (Hdt. 2.145).

The Romans were equally interested in the division of time into distinct ages. By the late Republic, this found voice in the poetry of the Epicurean Lucretius, who revised Hesiod's ages to reflect a technological progression, replacing gold and silver with stone, and then proceeding from bronze tools to iron.[13] By the time of Augustus, a triumphalist perspective

[12] M. Scott, *Space and society in the Greek and Roman worlds* (Cambridge 2012) 145.

[13] Lucr. *De rerum natura* V.1281ff.

supplanted Hesiod's pessimism, as Vergil's fourth *Eclogue* demonstrates: 'The last age of the Sibyl's song is come; / a great new order of ages is born anew: / now the Virgin returns, the kingdom of Saturn returns; / now a new generation is sent down from the lofty sky'.[14] The tension between qualitative and quantitative descriptions of time, however, persisted. Vergil's diverse terms for the new age — *aetas*, *saeculum, ordo*, *regna*, *progenies* — all emphasize the notion that something clearly different from the past has appeared, but their metaphorical connotations (political, dynastic, genetic) resist description in terms of absolute chronology. Even the term *saeculum*, borrowed from Etruscan cosmology and often rendered in English as 'century', is in fact a general reference to periods spanning around a hundred years (but sometimes more), and is better translated by the more indefinite 'age'.[15] At the same time, the Roman state maintained lists of magistrates by year (the *fasti*), and Roman historians were concerned with absolute dating of periods and events, usually in relation to the mythical founding date of the city of Rome (*ab urbe condita*). Augustus himself may have arranged for the reinscription of the *fasti consulares* and *triumphales* on a triple arch in the Roman Forum, firmly placing the new order in absolute time.[16]

The Roman concern with absolute dating intensified in the work of the Late Antique chronographers, who sought to reconcile a thousand years of historical research from different traditions. The best known is Eusebius, whose approach to the past was only marginally 'periodic', to the extent that his *Chronicle* listed events organized by dynasties and punctuated by rubrics noting certain turning points (for example, the fall of Troy) and the time that had passed since the previous turning point. Nevertheless, the *Chronicle* visually expressed the connection between place and time that characterizes periodization: the charts in Eusebius' second book documented historical synchronisms by arranging political/cultural entities (which of course have spatial correlates) in columns, and events related to those entities in rows. Thus events like the fall of Troy permit the calibration of the chronological accounts of the Assyrians, the Hebrews, the Egyptians, and the Sicyonians.[17]

The *Chronicle* was very influential in the development of the study of antiquity in the modern period. It provides us, for example, with many of the canonical dates for the foundations of Greek colonies. Furthermore, it offers a visual account of history that clearly separated the timelines of different ancient cultures and, by extension, their territories, affirming a connection between time, place, and culture that persists in our own study of the past. However, its focus on the synchronization of ancient and biblical history discouraged further attempts to distinguish sets of events or associated phenomena in human history. This came only with the rise of humanism in the Renaissance, the development of antiquarianism and ancient art history in the seventeenth and eighteenth centuries, and the formalization of archaeology in the nineteenth.[18]

[14] Verg. *Ecl*. IV 4–7.

[15] N. T. de Grummond, *Etruscan myth, sacred history, and legend* (Philadelphia 2006) 42–43.

[16] C. B. Rose, 'The Parthians in Augustan Rome', *American Journal of Archaeology* 109 (2005) 21–75.

[17] D. Rosenberg and A. Grafton, *Cartographies of time* (New York 2010) 15–16.

[18] With the exception of a division of time into periods based on Christian theology and biblical exegesis: for instance, the division of the folios of both Werner Rolevinck's *Fasciculus temporum* (1474) and Hartmann Schedel's *Nuremburg Chronicle* (1493) into seven *aetates mundi*, 'ages of the world', a medieval paradigm derived from the writings of Augustine. Rosenberg and Grafton, *Cartographies* (n. 17, above) reproduce a number of pages from these works but do not comment on this feature (Figs 11–13, 30, 36–38).

The notion that works or events in the past can be grouped into 'periods' determined by style, culture, or technological change, and that those periods can be placed in absolute time, re-emerged during the sixteenth century. The word 'period' itself, in the sense of a marked span of historical time, came into currency in English around then.[19] In 1561, Julius Caesar Scaliger published the *Poetice*, a commentary on ancient poetry in which he divided Greek and Latin poetry into periods or ages: for the Greek poets, he employed a tripartite system of youth, maturity, and senescence, while he divided their Latin counterparts into five periods that followed the same biological metaphor.[20] Later his son, the humanist Joseph Scaliger, having grappled with Eusebius, turned to absolute dates, producing monumental works of chronography such as his *De emendatione temporum* in 1583 and his *Thesaurus temporum* in 1609.[21] Meanwhile, in Italy, Lucretius' three technological ages were put into action by naturalist and antiquarian Michael Mercati for the classification of material remains. Over the seventeenth century, French antiquarians like Bernard de Montfaucon and Nicolas Mahudel applied Mercati's schematic division on an increasingly systematic basis.[22] Only in the eighteenth century, however, did the concept of periodization come into general use in connection with absolute historical chronologies.

Perhaps the most important development for the periodization of antiquity during the eighteenth century was Johann Joachim Winckelmann's division of Greek art into four periods in his *Geschichte der Kunst des Ältertums* (1764), again following the biological metaphor employed by the elder Scaliger. Winckelmann's close examination of Roman copies of Greek sculpture, together with his consideration of literary sources, led him to group the assumed Greek originals into 'four times and four styles' (*vier Zeiten und vier Stile*).[23] These were, in order, an 'older' style (*Ältere Stil*), a 'great and high' style (*Grossen und Hohen Stil*), a 'beautiful' style (*Schöne Stil*), and a 'style of the imitators' (*Stil der Nachahmer*). Following the elder Scaliger and ancient tradition, Winckelmann assigned genealogical dates to these stylistic periods: the older style lasted to Pheidias; the high style ran from Pheidias to Praxiteles, who initiated the beautiful style; and the beautiful style lasted through to Lysippus and Apelles, after which it degenerated into the decadent period of the imitators.[24] This periodization corresponds very closely to the one still in use in Greek art and archaeology, which distinguishes the Archaic, Classical, Late Classical, and Hellenistic periods.

From this point on, periods with both relative and absolute chronological characteristics became a fixture of scholarly approaches to the past. In the late eighteenth and early nineteenth centuries, Christian Thomsen, a Danish prehistorian, applied Mercati's three-age system more rigorously, physically reorganizing the collection of the state museum of Danish antiquities along these lines in 1819.[25] In England, the publication of Charles Lyell's *Principles of Geology* in 1830–1833 explicitly periodized the deep past, strengthening the connection between the concept of periodization and the stratification of deposits in the

[19] *Oxford English Dictionary* (2016) *s.v.* 'period[3a]'.

[20] M. Bizer, 'The genealogy of poetry according to Ronsard and Julius Caesar Scaliger', in *Corona Martiniana*, ed. G. Tournoy, *Humanistica Lovaniensia. Journal of Neo-Latin Studies* 42 (Leuven 1994) 304–18 (308, 312–14).

[21] A. Grafton, *Defenders of the text: the traditions of scholarship in an age of science, 1450–1800* (Boston 1994) 106; R. Burgess, *Studies in Eusebian and post-Eusebian chronography* (Stuttgart 1999) 22.

[22] D. Clarke, *Analytical archaeology*, 2nd edn (New York 1978) 4–6.

[23] J. J. Winckelmann, *Geschichte der Kunst des Ältertums* (Dresden 1764) 213.

[24] Winckelmann, *Geschichte* (n. 23, above) 214.

[25] Clarke, *Analytical archaeology* (n. 22, above) 9.

ground.[26] By 1871, when Schliemann put spade to earth at Hissarlik, the division of material culture and works of art into successive periods, identified by style or culture and assigned absolute chronological boundaries wherever possible, was standard practice.

This short history of periods and periodization demonstrates the uncertainty and inconsistency that characterize efforts to divide human history into comprehensible blocks of time set in logical relation to each other. The divisions are essentially arbitrary, and change according to time, individual, nation, academic questions, *etc*.[27] And their borders are often blurry. Scholars feel compelled to provide absolute dates for relative periodizations, but when they do, their acknowledgment of the nuances of the evidence leads them to use impressionistic expressions like 'around the end of' or 'somewhat before'. This has implications for the visualization and alignment of temporal data. Like the organization of space represented in a map, the organization of time makes more sense to us when it is laid out in schematic form. Attempts to schematize time have moved in lockstep with revolutions in the conceptualization of space since the Renaissance: Julius Caesar Scaliger was a contemporary of both Copernicus and the cartographer Martin Waldseemüller. While space could be represented as independent of time, however, historical time was harder to separate from space. Many nineteenth-century 'cartographies of time' arranged historical phases within a geographic frame, reinforcing the identification of periods as chronotopes.[28] But how do you turn a chronotope into a static image or match it to another if its boundaries in both space and time are not only vague but also contested?

This brings us back to the current revolution in digital representations of space. As in previous scientific revolutions, time is deeply connected with the 'spatial turn' of the digital age. Digital tools and approaches have opened vast new worlds of representation and information organization, and the potential for new avenues of inquiry is tremendous. Because the human classification of time resists quantification, however, periodized time has proven much more resistant than space to the application of digital methods.

3. Digital management of the past

The last two decades have seen an explosion of digital information about the Graeco-Roman past. The digital classical world has grown to include not only most of ancient literature, but also archaeological databases, scans of manuscripts, vast image collections, and the digitized collections of hundreds of museums: a huge collection of information, much of it openly accessible. At the same time, digital tools have created new possibilities for the visual representation of structured data. Approaches echoing Franco Moretti's foundational work on digital literary analysis ('graphs, maps, trees') have been applied to classical literary corpora.[29] Map-based approaches have been especially popular, in large part because the Pleiades gazetteer, by assigning latitude–longitude coordinates to

[26] C. Lyell, *Principles of geology, being an attempt to explain the former changes of the Earth's surface, by reference to causes now in operation* (London 1830–33). Coming full circle, Lyell in his introduction credits the Greeks with the 'theory of great periodical revolutions in the inorganic world' (16).

[27] F. E. Sparshott, 'Notes on the articulation of time', *New Literary History* 1 (1970) 311–34.

[28] Rosenberg and Grafton, *Cartographies* (n. 17, above) 116–30.

[29] F. Moretti, *Graphs, maps, trees: abstract models for a literary history* (London 2005); D. Cline, 'Six degrees of Alexander: social network analysis as a tool for ancient history', *Ancient History Bulletin* 26 (2012) 59–69; F. Mambrini, 'The Ancient Greek Dependency Treebank: linguistic annotation in a teaching environment', in *Digital classics outside the echo chamber*, ed. G. Bodard and M. Romanello (London 2016) 83–99.

ancient places and exposing data in a form that can be read by computers, has facilitated the extraction of maps from ancient texts and museum collections alike. Furthermore, by providing an unambiguous way of describing a reference to an ancient place, Pleiades has made it possible to connect spatial data across disparate, decentralized databases, making it easier for a user to find specific facts in the boundless space of online digital information.

The success of the spatial component of digital classics projects highlights a set of issues that are important not only for visualization, but also for the discoverability of digital information. These issues have to do with metadata, semantics, and ontologies: that is, the way that data are described, the way those descriptions are constructed, and the way that both data and description correspond to the organization of knowledge within a discipline. When we gather information from printed sources, we use knowledge of our discipline and our language(s) to understand that 'Athens' in one case refers to the Greek *polis* and in another case to the American city. When we use computers to aggregate digital information, however, we rely on the consistent formatting of that information and its correspondence to known organizational systems or points of reference. Unfortunately, the long, multilingual history of classical studies makes this very difficult: even in areas where we basically agree, as for instance with the naming of Greek or Latin textual sources, practices vary in terms of abbreviation, numbering, *etc*. The same is true of the study of classical art and archaeology, which deploys a bewildering array of names and labels for objects, types, places — and periods. The traditional way to make order out of this chaos involves the use of centralized controlled vocabularies, so that strings of text in metadata fields are always identical. For example, a cataloguer might always use the name of an artist as it appears in the Library of Congress Name Authority File or the name of a place as it appears in the Getty Thesaurus of Geographic Names.[30] This approach has long been standard among libraries, allowing catalogues to follow patterns that are predictable and easy to integrate across institutions. But it has worked less well for the sea of digital data generated over the last three decades by institutions, projects, and scholars interested in the ancient world.

Two more powerful methods have recently been used to address this digital *apeiron*. The first involves the creation of a centralized, detailed model of the properties, entities, and relationships within a field of knowledge (that is, an ontology), and the establishment of standardized controlled vocabularies that can be used to describe items within it. This approach is best represented for the ancient world by the CIDOC-CRM (the Conceptual Reference Model of the International Committee on Documentation).[31] Two datasets that are both mapped to the CIDOC-CRM can be cross-searched by concept, even if their database structures or terminologies differ. The second approach involves a looser coupling between data sources: in this case, entities in heterogeneous datasets are described using a fairly general semantic system but connect their vocabularies to shared external reference points (gazetteers or thesauri), which provide concept identifiers that are independent of the terms used in a given dataset. Rather than insisting that an entity always be described by the same string of characters, this approach allows a cataloguer to associate it with a unique, non-language-specific identifier (a 'Uniform Resource Identifier', or URI) that points to a semantic description. Thus 'Athens', as a place, can be described in a database by an identifier like 'http://pleiades.stoa.org/places/579885' drawn from an external gazetteer.

[30] *Library of Congress Name Authority File*, http://id.loc.gov/authorities/names.html; *Getty Thesaurus of Geographic Names*, ed. P. Harpring, http://www.getty.edu/research/tools/vocabularies/tgn/.

[31] CIDOC-CRM, http://www.cidoc-crm.org/.

That identifier carries with it in turn a set of semantic information including, for example, the coordinates of the centre point of the geographical concept and the alternative names by which the place has been known. This approach, in which digital information is described and connected by unique identifiers and semantic structures, is associated with the Semantic Web and the Linked Data ecosystem.[32]

These three descriptive approaches — controlled vocabularies, ontologies, and Linked Data gazetteers — are increasingly integrated (the Library of Congress authority files and the Getty Thesauri are available as Linked Data; CRM-modelled datasets can refer to Linked Data gazetteers; and datasets that are currently expressed as Linked Data are being mapped to the CRM). But all of them present significant barriers to those who want to use them to describe their own data or to reuse data described with them. Controlled vocabularies require both a system to enforce their use and a way to guard against input errors; an ontological approach requires extensive knowledge of both information architecture and a scholarly domain; a Linked Data approach requires procedures to map entities in a dataset to corresponding entries in a gazetteer and a knowledge of esoteric query protocols. At the moment, the Linked Data framework seems to present the lowest bar to entry for digital classicists, and the shift from terminological standardization to the Semantic Web has been transformative for digital projects focused on the ancient world.[33] Linked Data approaches have greatly facilitated data integration projects based on spatial features: Pleiades, for example, made possible the Pelagios project, which began by aggregating records with spatial associations from a broad variety of databases through the shared use of Pleiades URIs.

Pleiades also made it possible to apply named entity recognition to place-names in English translations of ancient texts, which could then be 'geoparsed' to generate maps.[34] By providing a common point of reference for heterogeneous collections, it has also enabled the automatic combination of different types of data: for example, a page in the HestiaVis Herodotus text can combine place information from Pleiades with a set of spatially-tagged images from Flickr. So Pleiades, as a spatial gazetteer, has made it possible to get data with a spatial component both out of a dataset (named entities from unstructured text, for example) and into it (through the integration of external resources that share a common reference).

Although scholars have repeated for a decade that temporal data are critical for the 'spatial turn' in history, and although archaeologists and art historians are just as interested in discovering data by period as by place, there is no temporal equivalent to Pleiades. Pleiades only works because we agree that whatever name we give it, Athens is a unique space on the surface of the globe that can be described with some degree of accuracy through one or more latitude–longitude coordinate pairs. It is possible to search across information described with calendar dates, as long as some very basic formatting rules are respected: if we know the calendrical system and format used to express a date, we can translate it mathematically into a date in another calendrical system. But we are still unable to agree

[32] T. Heath and C. Bizer, *Linked Data: Evolving the web into a global data space*, 1st edn (San Rafael 2011); 'Linked Data – W3C', W3C, https://www.w3.org/standards/semanticweb/data.

[33] T. Elliott, S. Heath, and J. Muccigrosso, *Current practice in Linked Open Data for the ancient world*, ISAW Papers 7 (New York 2014).

[34] C. Grover *et al.*, 'Use of the Edinburgh Geoparser for georeferencing digitized historical collections', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368 (2010) 3875–3889; B. Alex, K. Byrne, C. Grover, and R. Tobin, 'Adapting the Edinburgh Geoparser for historical georeferencing', *International Journal of Humanities and Arts Computing* 9 (2015) 15–35.

on the boundaries of the Iron Age, although (or perhaps because) we have been using it as a concept and a term for almost three thousand years. Furthermore, natural scientists and geologists are currently arguing not only about the boundaries, but about the very existence of a period called the 'Anthropocene.'[35]

The PeriodO gazetteer was conceived as a potential solution to this impasse. If we disagree about the boundaries of periods, if we use them inconsistently across space and time, and if we nevertheless have felt compelled to assert absolute dates for them from Herodotus' time to our own, then perhaps the answer is to document stated definitions, rather than abstract concepts. Our gazetteer therefore includes period definitions with four core characteristics: at least one specific name or label; explicit temporal coverage, however vague; explicit spatial coverage, at any level of specificity; and an authoritative source. The result is a gazetteer of *claims* that authorities have made about periods, not a gazetteer of period *entities*. It accommodates any authoritative statement that includes those four characteristics, even if it conflicts with other statements, is out of date, or lies outside mainstream usage. This strategy sidesteps the resistance of disciplines like classical archaeology to centralized authorities, makes room for variation across space and time, and fosters transparent statements about the provenance and intellectual genealogy of the period definitions used by a particular project or cataloguer. The insistence on provenance allows us both to collect extensive structured information about the history of human attempts to categorize time and to welcome new period definitions from a community of users. It will, we hope, help us to see where our definitions overlap, where they differ, and how they change over time — and, by including dates in the definitions, eventually facilitate search by date range across multiple datasets with idiosyncratically periodized data.

4. The PeriodO data model

We model period information in PeriodO through two kinds of entities: individual period *definitions* and *collections* of period definitions that share the same source. This allows us to manage the details of individual period definitions separately from the bibliographic description of sources. For example, in her study of the Salammbô tophet at Carthage, Hélène Bénichou defined four new periods for the site (*première, deuxième, troisième,* and *quatrième époque*).[36] We model this fact in PeriodO as four period definitions belonging to a single collection, for which the source is that publication.[37] At a minimum, we record the title, authors or editors, and year of publication of each source. Where possible, we also link sources to their corresponding records in external bibliographic databases: WorldCat for books, and CrossRef for journal articles with Digital Object Identifiers (DOIs).

For each period definition, we record the original name given to the period in the source text (*e.g.* '*première époque*'), as well as its language and script. If the original language was not English, we add an English translation (*e.g.* 'First era'). For the temporal extent of the period we record the start ('~800') and end ('675/650') exactly as specified in the source (including the calendrical system used). To normalize that extent and to make it amenable to filtering and visualization, we add a conservative parsing of the temporal expression in

[35] C. Waters *et al.*, 'The Anthropocene is functionally and stratigraphically distinct from the Holocene', *Science* 351 (2016); DOI: https://doi.org.10.1126/science.aad2622.

[36] H. Bénichou, *Le tophet de Salammbô à Carthage: essai de reconstitution* (Rome 2004) 121.

[37] http://n2t.net/ark:/99152/p0rqpwq.

machine-readable ISO8601 years (proleptic Gregorian calendar). Since we cannot know in quantitative terms what the author meant by '~800' or expressions like 'circa', we parse these statements as the numerical date provided. The ISO8601 standard includes a year 0, so '~800' is parsed as '-799'. Where dates are clearly expressed as ranges in the source, we use four-part expressions, so that '675/650' is expressed as the range -674 (earliest stop) to -649 (latest stop).

Similarly, if the spatial extent of the period is indicated in the source, we record that text verbatim, but where possible we also add a link to a record in an external gazetteer approximating the spatial coverage as indicated in the source (in this case, we linked to the DBpedia record for 'Tunisia').[38] Here again our goal is to report the author's claim faithfully and make as few assumptions as possible, while still providing machine-actionable information. Since we are dealing with modern sources, and since periods are used as modern constructs, we currently parse spatial coverage in terms of modern political entities. We seek to document where the period term is currently used, not the ancient political or cultural entities it refers to (so the coverage here is not the Carthaginian Empire, but the modern state where research was carried out). As we increase the specificity and time depth of definitions in the dataset, however, we intend to increase spatial resolution to the level of individual places, to integrate historical gazetteer information, and to improve the representation of modern assertions about the spatial extent of historical political entities like 'the Roman Empire.'

Finally, we record notes about the definition taken from the original source and editorial notes added by the PeriodO contributor. For example, Bénichou's definition of '*première époque*' includes a source note indicating that it is 'derived from Carthage tophet chronology, level Tanit IIa' and an editorial note explaining why we assume that '~800' referred to a date before the turn of the era. Where period definitions are drawn from sources that are themselves Linked Data (for instance, external period gazetteers), we include the original URI in a 'sameAs' relationship.

In addition to period definitions and collections, we also maintain the entire history of changes to the PeriodO, including all changes proposed, who proposed them, when they were accepted, and who accepted them. Contributors are unambiguously identified using ORCIDs.[39] Although this ID system may be new to humanities scholars, we feel that it is more suited to the management of scholarly identities and more durable than a login through commercial services. We believe that making provenance data easily accessible increases trust in the PeriodO curation process, as all users can audit the history of the project and learn who was responsible for creating the definitions they are considering.[40] This will be especially important as the contributing community grows larger and more diverse.

[38] Moving forward, we expect to make more use of Wikidata (https://www.wikidata.org), which now includes both uniquely identified places and periods.

[39] L. Haak, M. Fenner, L. Paglione, E. Pentz, and H. Ratner, 'ORCID: a system to uniquely identify researchers', *Learned Publishing* 25.4 (2012) 259–264.

[40] P. Golden and R. Shaw, 'Nanopublication beyond the sciences: the PeriodO period gazetteer', *PeerJ Computer Science* 2:e44 (2016); DOI: https://doi.org/10.7717/peerj-cs.44.

5. The PeriodO client and server architecture

The software that runs PeriodO consists of two components: a server for managing the canonical dataset, and a browser-based client for viewing and editing PeriodO data.[41] The client can load data that conform to the PeriodO data model from any source on the web or from a local JSON file. Since snapshots of the dataset and its history can be deposited in an institutional repository alongside a copy of the client, future scholars will be able to use the client to view the data whether or not the PeriodO server is still being maintained at that time.

While the PeriodO client supports the informal sharing of period definitions through the export and import of files, the server is needed to establish a curated, canonical version of the PeriodO data.[42] A canonical version of the dataset is necessary to guarantee stable, static records produced through a trustworthy curation process, so that scholars can cite them with confidence. The PeriodO server plays the same role in this curation process that a master source code repository does in a software development process: the dataset itself is freely available, but the server provides a single point of control for managing changes to it. Proposed changes to the PeriodO data can be submitted to the server in the form of *patches* that describe precisely what is being added or removed. Only users with ORCIDs can submit patches, to maintain provenance data integrity. Submitted patches are stored on the server until they can be reviewed by a PeriodO curator. If a patch is rejected by a curator, it will be archived but not added to the canonical PeriodO dataset. If it is accepted, the proposed changes are merged into the canonical dataset and the submitter is added to the list of PeriodO contributors.

If accepted changes include the addition of new period definitions or collections of definitions, the PeriodO server is also responsible for minting new persistent, globally unique identifiers (URIs) for these entities. The identifiers are intentionally opaque so that the PeriodO data model or server location can change independently of the identification scheme. They are also designed to be robust to typographical errors: just as for credit card numbers, a checksum can be generated to check whether a name is valid. Persistence and global uniqueness are assured by the combination of a PeriodO ARK ID assigned and maintained by the California Digital Library EZID service and the generation of unique suffixes for each collection and definition by the server; the resulting strings can be resolved using a suffix pass-through protocol.[43]

6. Building community

Although we refer to the master dataset as ‘canonical’, PeriodO is not meant to establish a canon in the sense of a comprehensive, univocal authority. It was initially seeded with definitions from sources and partners with data that happened to match our data model, and it makes no claims to completeness. Instead, it opens itself to new contributions from the scholarly community: if a user cannot find a suitable period definition, he or she can make one. As the community grows, so will the dataset. The first major step toward this goal

[41] The source code for both the client and the server are in the public domain (CC0 Public Domain Dedication) and can be downloaded from https://github.com/periodo.

[42] R. Shaw, A. Rabinowitz, P. Golden, and E. Kansa, ‘A sharing-oriented design strategy for networked knowledge organization systems’, *International Journal on Digital Libraries* 17.1 (2016) 49–61.

[43] J. Kunze and R. Rodgers, ‘The ARK identifier scheme’ (2013), https://tools.ietf.org/html/draft-kunze-ark-18.

after the completion of the client came with the establishment of our collaboration with the ARIADNE data portal.[44] Not only did ARIADNE use PeriodO to manage heterogeneous definitions from multiple partners, but those definitions helped us to refine our data model, which will expand in the current phase of the project to include derivative and hierarchical relationships, to facilitate searching, and to generate a robust picture of intellectual influence.

The expansion of the data model will, we hope, make it more attractive for data managers both to use PeriodO URIs and to contribute their own period definitions. The more users attach PeriodO URIs to their periodized data, the more useful those URIs will become as a way of harmonizing temporal information across platforms. And the more users from different disciplines contribute period definitions, the more comprehensive the dataset will become. Since PeriodO is meant to preserve definitions as they were originally made by an authority, we do not permit edits related to disagreements about the coverage or label of a period. Disagreements should be represented by separate period definitions, which will highlight scholarly debate. To foster this community, we are in the process of developing tutorials that will guide users through the interface and the creation of new period records.[45]

7. The future of the past

We hope that enough community momentum will grow around PeriodO to require the constitution of an editorial board to oversee the review process, as with Pleiades. The development of that broader community is one of the central goals for the current IMLS-supported phase of the project. We are actively recruiting partners beyond classics and archaeology, including paleoanthropologists, modern historians, and literary scholars. We also intend to expand our coverage beyond the ancient Mediterranean and North America; we already have collections that deal broadly with other parts of the world, and although we have specialized coverage for China, we would like to have more information for Africa, India, Australia, and Central and South America. Finally, in this phase we will begin to add period definitions from historical sources, reaching back as far as we can find authorities who identify discrete spans of time according to calendar dates — perhaps even to ancient chronographers like Manetho. This expansion of the dataset in time and space, together with the explicit modelling of intellectual genealogy, will make it a rich resource for the study of time within and across different disciplines.

To enhance the relevance of the PeriodO gazetteer to the broader information-management field, we are developing tools that permit users to match their own period vocabularies to PeriodO URIs, and eventually to carry out date-range searches across multiple heterogeneous datasets that use those URIs. This will improve the discoverability of periodized material across national and disciplinary boundaries, and will open new doors for the integration of different types of evidence in research and teaching. Through alignment and integration, we hope to make it easier to get new data *in* to the research or visualization platforms of the future. But we would also like to be able to make it easier to get data *out*, and here community engagement will be critical. As the dataset of period definitions grows, it will provide useful training data for machine-learning algorithms

[44] 'PeriodO', ARIADNE, http://www.ariadne-infrastructure.eu/Resources/PeriodO.

[45] See the PeriodO website (http://perio.do) and Github repository (https://github.com/periodo) for tutorials, detailed explanations of technical features, and a list of partners and advisors, to whom we are deeply grateful.

focused on the extraction of dates from natural-language sources.[46] A large enough dataset of period definitions could allow a probabilistic temporal parser to conclude that reference to the 'Jungsteinzeit' in a book written in Germany in 1950 is most likely to refer to a particular date range — which could then be represented on a timeline, or return that book when used as a search criterion.

The construction of a prototype of such a probabilistic temporal parser is the final goal of this phase of PeriodO. The more periodized information we can aggregate — especially from datasets containing material with both absolute dates and period associations — the more statistical methods we can use to untangle the diverse and confusing ways scholars have tried to make sense of the past. Through the development of better search and visualization tools, which will allow not only search by date range and map coordinates, but also the comparison of different period definitions by date or place of publication, we hope to illuminate the evolution of historical disciplines over time. We welcome new partners in this endeavour, and we invite the reader to join us on this Janus-headed journey into the past and the future.

Adam Rabinowitz, Department of Classics, The University of Texas at Austin

Ryan Shaw, School of Information and Library Science, The University of North Carolina, Chapel Hill

Sarah Buchanan, School of Information, The University of Texas at Austin

Patrick Golden, School of Information and Library Science, The University of North Carolina, Chapel Hill

Eric Kansa, Open Context and the University of California at Berkeley

[46] Some work has already been done with collocation of period terms and dates in text corpora: V. de Boer, M. van Someren, and B. Wielinga, 'Extracting historical time periods from the Web', *Journal of the American Society for Information Science and Technology* 61 (2010) 1888–1908; M. Mouroutsou, S. Markantonatou, and V. Papavasiliou, 'The development of vocabularies of historical period names from web acquired corpora', *Mediterranean Archaeology & Archaeometry* 14 (2014) 165–74.