Making up for Lost Time: Digital Epigraphy, Chronology, and the PeriodO Project

Abstract: Digital epigraphy has made great strides toward interoperability and data integration over the last two decades, and Linked Data approaches are now taking advantage of the spatial information associated with inscriptions for new search and visualization tools. The ability to search across epigraphic collections by time, and especially by relative chronologies, lags behind. The PeriodO project has created a Linked Data gazetteer of structured period definitions that facilitates translation between absolute dates and relative chronologies, creating new possibilities for the interoperability of epigraphic collections and their connection with archaeological databases.

Keywords: periodization, Linked Open Data, gazetteers, reconciliation, interoperability.

1 The promise of digital epigraphy

The field of epigraphy, with its widely-dispersed body of evidence, its longstanding conventions for description and publication, and its bewildering range of publication venues, has been positioned to benefit from digital approaches since the dawn of the digital age. For the Classical world, this was demonstrated by such early projects as the Packard Humanities Institute digital corpus of Greek inscriptions (Iversen, 2007), and has been confirmed by an array of further efforts spurred on by the rise of the internet. On the most basic level, a digital environment makes it possible to assemble and search across collections of inscriptions that are otherwise scattered in both geographic and bibliographic space. In the last two decades, following the development of the EpiDoc extension of the Text Encoding Initiative to permit the encoding of inscriptions in XML (Cayless, Roueché, Elliott, & Bodard, 2009; Bodard, 2010), the publication venues themselves have moved online (Reynolds, Roueché, & Bodard, 2007; Bodard, 2008), and the possibilities for the discovery and integration of epigraphic texts have increased exponentially.

At the same time, new digital tools have enhanced the documentation of the physicality of inscriptions, which had long been neglected in publications in favour of textual content. Some of these, like laser scanning, have presented a fairly high technical and financial bar to entry, but others use cheaper and more widely-available technologies like flatbed scanners or computational photography to create 2.5D or 3D images (Barmoutis, Bozia, & Wagman, 2010; Rabinowitz, Schroer, & Mudge, 2010). These representations capture both the text and the thing-ness of epigraphic objects more fully than photography alone, and with more potential for interactivity. On the other hand, they require more technical investment in online viewing platforms, and their dependence on customized viewers makes them more fragile in the changing online environment. As a result, these techniques have not been incorporated into digital epigraphic practice to the same extent as the EpiDoc text encoding standard.
Developments in these two areas reflect the traditional consideration of an inscription as a combination of text and object. A third area of digital potential, however, reflects a more recent concern not only with the materiality of inscriptions, but also with their archaeological context. In the 19th and early 20th centuries, the epigraphic record was valued for the contributions it could make to our understanding of history, and thus the context in which inscriptions were found was usually considered less important than the texts themselves, especially if the stone bearing the inscription had been moved from its original position or reused. More recently, however, the importance of archaeological context for the understanding of inscriptions has been recognized, both within individual sites and on the level of broader regional landscapes (e.g. Holdenried, Roueché, & Scholz, 2014). Fortunately, this recognition has been accompanied not only by an explosion in the online availability of archaeological data, but also in the emergence of a number of projects focused on the aggregation of such data across datasets, like the ARIADNE infrastructure (Niccolucci & Richards, 2013). It is thus increasingly possible to connect aggregations of epigraphic data like EAGLE with aggregations of related archaeological resources, enriching our understanding of the relationship between text and context.

Epigraphic corpora have used space as a central organizational principle since the 19th century, from the regional division to the single site. It is not surprising, therefore, that space and place have offered the easiest point of entry for data integration. Trying to join databases of objects and inscriptions by place-name strings, however, is a futile endeavour: it is not feasible to connect information by strings across databases in a dozen different languages, especially when place-names are often spelled in different ways within a single language. Many projects that seek to create interoperability focus instead on the use of shared external reference points: “gazetteers” that establish the identity of a spatial entity unambiguously, in a standardized and consistently-structured format attached to a unique and persistent identifier. By describing metadata values within a database in a semantically transparent fashion, and by including in those descriptions links to persistent identifiers that are themselves described in a semantically transparent fashion, a database manager can plug records in to a wider network of related information.

These are the principles that characterize the Linked Data ecosystem (Heath & Bizer, 2011), and several ancient-world initiatives have already made significant advances by adopting them (Depauw & Gheldof, 2013; Elliott, Heath, & Muccigrosso, 2014; Isaksen, Simon, Barker, & de Soto Cañamares, 2014). The Pelagios project demonstrates the potential of this approach: its Recogito tool associates place-names in texts with entries in gazetteers, while its Peripleo browser aggregates data from a variety of datasets that refer to shared historical gazetteers to permit cross-search by ancient place.¹ The datasets aggregated by Peripleo already include the Epigraphische Datenbank Heidelberg, and more coordinated efforts to integrate inscriptions into the larger Linked Data environment are beginning to materialize (Álvarez, Gómez-Pantoja, & García-Barriocanal, 2011; Blanke et al., 2012). These efforts focus on named entities, which are the most susceptible to disambiguation, unique identification, and manual or automated extraction from text. This work again requires shared points of reference for identification, which are currently

provided by spatial gazetteers for place and are in development for past people (Lawrence & Bodard, 2015; Depauw et al., 2017). Temporal periods, however, despite being the other named entity most frequently encountered by scholars of the past, have until recently been conspicuously absent from this emerging ecosystem.

2 The trouble with time

Both epigraphy and archaeology have long traditions of arranging information according to geographical space, so place-based data aggregation comes very naturally to these disciplines. Both are also deeply engaged in questions of time – but here the two diverge in the nature of their evidence. Inscriptions sit at the intersection between the world of absolute dates common to textual sources and the world of relative chronologies based on style, more closely associated with archaeology and art history. On the one hand, calendrical expressions, names and titles of rulers or officials, and particular letter-forms are often very closely dated, to the point where inscriptions, like coins, are used to provide absolute dates for archaeological contexts. On the other hand, inscriptions that lack clearly datable features, or that were produced in periods for which absolute dates are less well-established, are often organized in broader stylistic classes. In some cases, those relative chronologies are the same as those used to classify archaeological material; in others, they were developed specifically for the epigraphic record. In some cases, they are shared widely across multiple geographic regions (for example, the classification “Roman period”); in other cases, they are unique to a single region or language group. And in some cases, these relative chronologies are attached to absolute dates, while in others their dating is either left open or inferred from absolute dates ascribed to the inscriptions themselves.

Archaeological remains, on the other hand, are much more commonly classified by relative chronologies based on a complicated and idiosyncratic combination of historical, stylistic, and material features. The defining characteristic of these chronologies is their division into “periods”, blocks of time that the scholarly community assumes to be characterized by distinct and consistent qualities or phenomena. While these periods can appear at first glance to be fairly consistent across regions and projects – “Roman”, for example, seems like a transparent term on its face – the apparent agreement masks a vast number of chronological inconsistencies and disagreements based on factors like geography (“Roman” in the UK does not have the same temporal range as “Roman” in Italy, for example) or school of thought (where does “Roman” stop and “Late Antiquity” begin?).

As a result, although archaeologists, epigraphers, and historians alike group material by time as often as they do by space, time has resisted the integration strategies applied so effectively to space by the Pelagios project. Variation in the usage and meaning of period terms makes it difficult to integrate archaeological records chronologically across multiple databases, and it makes it even more difficult to integrate those records with the contents of epigraphic databases, which often eschew periodization altogether, or use it only in the absence of tight absolute chronologies. Some epigraphers might not see this as a real problem: after all, absolute dates can be easily searched both within and across databases, as long as some basic standards for date format are observed. But to ignore
the issue is to discard one of the greatest benefits of the emerging digital ecosystem for the study of the past: the combination of different strands of evidence to create new understandings of ancient societies. In some cases, integration might even lead us to reconsider long-standing knowledge categories. What would we find, for example, if we could compare current epigraphic work to redefine the meaning of “Late Antiquity” (Tantillo, 2017) with objects described with terms analogous to “Late Antique” across multiple languages and databases? Furthermore, better strategies for navigating between relative and absolute dating systems might help to expand context for inscriptions currently isolated within idiosyncratic local chronologies.

3 The PeriodO temporal gazetteer

The reconciliation of relative and absolute chronologies, and the clarification of scholarly usage of period terms, is the goal of the PeriodO project. PeriodO offers a Linked Data gazetteer, not of spatial entities, but of definitions of periods located in both space and time. It emerged from the recognition that the spatial and temporal coordinates of period terms, as these terms are used in the study of the human past, are deeply entangled, and that the terms themselves are discursive constructs subject to disagreement and diachronic change (Morris, 1997; Rabinowitz, 2014; Rabinowitz, Shaw, Buchanan, Golden, & Kansa, 2016). On a chronological (and, arguably, phenomenological) level, there is no single “Roman period” in modern scholarship or datasets: there are a series of related “Roman” periods with different temporal boundaries in different places, and if we want to be able to aggregate data along a temporal axis, it is critical for scholars or data-managers to be able to make transparent statements about which of those meanings of “Roman period” is in play in a particular context. The PeriodO project considers three pieces of information to be critical for a transparent period definition: coordinates – even vague coordinates – in time (that is, an earliest start and a latest stop); coordinates in space (that is, in what part of the world the term is applied with that chronological meaning); and an authoritative source for the association of those coordinates with that period term (Figure 1). By modelling both sources and definitions as structured data, and by providing both with unique, persistent identifiers, PeriodO makes it possible for a dataset to make an unambiguous statement about its usage of a given period term (“By ‘Archaic’, we mean the period between 700 BC and 480 BC within the bounds of modern Greece and Turkey, as defined by scholar X”). This in turn makes it easier to visualize and search the contents of that dataset by both time and space, and to understand how the chronology used in one dataset relates to the chronology in another, which might assign different dates to “Archaic” or use a different term (e.g. “Orientalizing”) for part or all of the same date range.

2 http://perio.do; the permalink for the client interface is http://n2t.net/ark:/99152/p0. PeriodO has been generously funded by grants from the US National Endowment for the Humanities (grant HD-51864-14) and the US Institute of Museum and Library Services (grant LG-70-16-0009-16).
Figure 1: Diagram of PeriodO data model

In documenting usage through the collection and modelling of period definitions, PeriodO does not intend to create a centralized, authoritative, prescriptive vocabulary for periods. Instead, the set of required attributes are meant to encourage multivocality: as long as a definition has a date range, a spatial extent, and an authority, and as long as it is not identical to an existing definition in the dataset, it can be added on an equal footing with other definitions. Although the initial content of the dataset was gathered by the project team from published work and from the formal vocabularies contributed by a group of generous partners, our goal is to expand that content in the future through user submissions. If a user interested in deploying PeriodO period identifiers in a dataset does not find definitions that match his or her own, new definitions can be submitted to the dataset as a data “patch” that is then merged into the “canonical” dataset (Shaw, Rabinowitz, Golden, & Kansa, 2016). As the user community grows, and as the project team continues to add periodizations from new disciplines and more diverse sources (including works written in the 18th or 19th centuries), we hope that the dataset will serve not just as a source of structured temporal data and transparent identifiers, but also as a representation of the broader scholarly discourse about periodization. In order to make this possible, the dataset includes not only sources and definitions, but also semantically-transparent formal relationships between definitions (specifically, that a period definition in one source is derived from a period definition in another source, or that a definition is the same as that described by a Linked Data identifier in another dataset), and a full provenance history describing who submitted data to the dataset, who approved it for inclusion, and when it was merged (Golden & Shaw, 2016).

3.1 PeriodO and digital epigraphy

These features – embrace of multiple definitions of periods, extensibility by a user community, attention to scholarly provenance and intellectual genealogy – make PeriodO
particularly useful for the integration of digital epigraphic collections into a Linked Data ecosystem. By avoiding a centralized vocabulary, it allows the discipline to document the different period definitions used by epigraphers across dozens of countries over the last several hundred years, while facilitating the reconciliation of a wide range of locally- or regionally-specific periodizations used in current databases. The ability to match periodized material in one database with periodized material in another, by period term or date-range or both, offers significant advantages for scholarship both within the field of epigraphy and outside it.

Within the field, for both experienced and novice epigraphers, there are times when it is useful to assemble a set of inscriptions that are contemporary in date across several corpora. This has always been difficult with the printed record. Although inscriptions in a specific publication or fascicule are usually arranged in chronological order, they are also published as they come to light, which means that in the best case inscriptions of the same general period can be spread across several different volumes in a single series (more frequently, they are spread across multiple series and specialised venues). The situation is somewhat eased in digital collections, records in which can be reorganized according to any criteria included in metadata and considered by the database designer. But the way in which dating criteria are considered differs widely from collection to collection. The Epigraphische Datenbank Heidelberg, for example, currently allows search by periods derived from Roman political history, but simply uses those periods as a proxy for absolute date ranges, and does not include period attribution within the metadata associated with those inscriptions.\(^3\) By contrast, the Europeana EAGLE database, which aggregates inscriptions from several different epigraphic collections, does include “period” as a metadata attribute, but does not have a period search facet.\(^4\) EAGLE and the online publication of the Aphrodisias inscriptions\(^5\) both allow searching by absolute dates, either as a date range alone (for the former) or by either date range or century (for the latter). The PHI database of Greek epigraphy includes absolute dating information drawn from the published corpora, but does not allow any searching or browsing by date or date range.\(^6\) Other online collections include periods as metadata attributes, but because of uncertainty about the relation between relative and absolute chronologies do not include any date information.

A metadata attribute that points to an identifier in an external gazetteer for a structured spatiotemporal representation of a period term has the potential to bring some


order to this chaos. This is especially true when that identifier also offers a transparent record of not only authority, but uncertainty. The date ranges associated with PeriodO definitions are parsed from date expressions in the original source, but we also retain the original date labels, which can be as vague as “around the middle of the second century BC”. Furthermore, the proleptic Gregorian calendar dates, expressed according to the ISO8601 standard and the OWL-Time ontology\(^7\), can be structured as a four-part date range, with earliest/latest start and earliest/latest stop, in order to preserve fuzzy chronological boundaries while still allowing date-based search. PeriodO identifiers thus make it easier to search within a single dataset by both date range and period term, while facilitating cross-searching and aggregation across different datasets that share the gazetteer as a common reference point.

Perhaps even more importantly, reference to a shared temporal gazetteer provides a bridge between inscriptions with absolute dates and archaeological material classified by period, enabling union searches that return both kinds of records. The Pelagios Project’s PeriIeop browser already provides a model for such searches, but since it is only beginning to incorporate periods as a search facet, the current timeline filter is useful primarily for objects with absolute dates, like coins. With the addition of shared external reference points for structured-data representations of periods, we will move closer to a fully integrated spatiotemporal search, within which a single bounded query could return Palmyrene sculptures contemporary with Palmyrene epigraphy, or Pompeiiian graffiti together with Flavian-period wall-painting. Such combinations of the material context and the epigraphic record have the potential to shed much new light on both sides.

### 3.2 Using the PeriodO gazetteer in epigraphic corpora

Before we can reach this point, however, there are more mundane considerations. The most pressing involves the sea of data a user must navigate in PeriodO, which now contains more than 5000 definitions, many of them referring to the same or similar concepts. The PeriodO project provides user documentation both on its current homepage and in a Github repository.\(^8\) While the project’s online documentation should be seen as the definitive guide, it is nevertheless useful to discuss the structure of the dataset and how PeriodO URIs can be added to epigraphic collections.

#### 3.2.1 Technical specifications

The PeriodO dataset is, at the core, a single serialized plain-text file expressed in the Javascript Object Notation (JSON) format, and available as JSON-LD (JSON specifically formatted as Linked Data), Terse RDF Triple Language or Turtle (TTL, another Linked Data format), or comma-separated values (CSV). The data model uses elements from the Simple Knowledge Organization System (SKOS), the Web Ontology Language (OWL) time standard, and the Dublin Core metadata standard. The dataset as a whole is

\(^7\) [https://www.w3.org/TR/owl-time/](https://www.w3.org/TR/owl-time/).

\(^8\) [https://github.com/periodo](https://github.com/periodo).
identified by an Archival Resource Key (ARK) identifier from the California Digital Library EZID system, and persistent unique resource identifiers (URIs) for each period collection (the authoritative source for one or more period definitions) and each period definition are generated by the suffix-passthrough protocol for the ARK ID system (Kunze & Rodgers, 2013).

Acronyms and jargon aside, this means that the dataset is lightweight, flat in structure, standard in format, human- and machine-readable, and provided with persistent, globally unique identifiers for its contents. Snapshots of the dataset will be preserved in a long-term institutional repository under open-access terms, so that if the web front-end ever ceases to work, the ARK ID will always point to a final version of the dataset, and the identifiers will always remain globally unique and persistent, even if the URI cannot be resolved as a URL. The structure of the PeriodO dataset also means that it is easy to download and reuse, adapt, and repurpose it, or to run it from a local server. Long-term preservation will be handled by the University of Texas Libraries, so there is very little risk that access to PeriodO data will be compromised in the foreseeable future.

3.2.2 Reconciliation

While it is possible to find period definitions by browsing the dataset through the PeriodO client, and to add their URIs to an epigraphic dataset manually by copy-pasting, it is not the most efficient process when a large number of period terms are involved. A reconciliation service is a digital tool that uses an algorithm to automatically match values in one dataset (for example, a column containing place-names in a spreadsheet) to similar values in another (for example, a gazetteer of historical places). Such services can be web-based, like the Geocollider tool recently developed to facilitate the matching of place-names in user-submitted structured data with Pleiades identifiers, or they can be integrated into another data-cleaning tool like OpenRefine. This makes it easier for a data manager to match a large number of values at once to an external reference point, rather than copying and pasting one URI at a time. PeriodO has developed a reconciliation service for OpenRefine, instructions for which are available on Github. Using the PeriodO reconciler, a user can match period terms in a structured-data document (in formats such as CSV, XML, JSON, etc.) to period definitions in PeriodO, using not only the term itself but also values in other columns like start or stop date and spatial coverage to refine the matching process (Figure 2). If, then, the manager of an epigraphic database wishes to add PeriodO identifiers to periodized records or a list of period terms, the reconciler makes the process simpler and faster.

---


11 [http://openrefine.org/](http://openrefine.org/). The Geocollider tool is also offered as a service through OpenRefine.

3.2.3 Adding data to the gazetteer

Inevitably, however, some of the periods used in any given dataset will not match any existing values in PeriodO. A near-match might be sufficient for a data manager in some cases, but in others there may be a local period definition that has to be expressed as-is. The PeriodO gazetteer has been designed with the expectation that new users will find new gaps, and therefore it has a process to allow users to fill in missing pieces. Any user with an ORCID\(^\text{13}\) may log in to the PeriodO client and submit a local indexed database with new or revised period entries as a patch to the server. If the new definitions meet the basic requirements of the dataset (authority and spatiotemporal coordinates), and if they are formatted correctly (specifically, if they include the original wording and values used by the source cited for spatial coverage and dates, rather than an interpretation of either by the user), the patch is merged with the “canonical” dataset on the PeriodO server, and persistent URIs are minted for the new definitions. The patch process not only guarantees that new data will meet the criteria and formatting expectations of the platform, but provides a clear documentation trail for the process of submission and approval. This trail itself, including the actors involved, is modelled using the Provenance Ontology and added to the dataset itself, so that any definition can be associated with the individuals who proposed or approved it.

\(^{13}\) https://orcid.org/.
3.2.4 EpiDoc guidelines

The previous paragraphs have described how the manager of a digital epigraphic collection can associate PeriodO URIs with local period terms contained in a spreadsheet or XML document. For collections that are already being expressed in the EpiDoc extension to TEI-XML, it is also important to understand how PeriodO URIs should be represented in that convention. Fortunately, the EpiDoc extension has a property class for named historical periods, which is described in the current version of the EpiDoc guidelines.\(^\text{14}\) Such periods can be encoded in an EpiDoc representation within the “origDate” element using the “period” attribute, according to the example given:

\[
<\text{origDate notBefore="-0332" notAfter="-0200" precision="medium" period="http://n2t.net/ark:/99152/p0m63njc4hd" evidence="lettering"> Early Hellenistic (lettering)</origDate>
\]

PeriodO is accepted in the convention as an authoritative source of URIs for period terms in this context.\(^\text{15}\)

4 Conclusions

Just as the shift from print to digital epigraphic corpora opened a world of new possibilities for searching and aggregation in the 1980s and early 1990s, and just as the shift from CD-ROMs to online databases did this again for the discipline in the early 2000s, the maturation of semantic-web approaches in the 20-teens has begun to reveal the potential of Linked Data for discovery and data integration. This is an exciting development, since it promises to allow us to find unexpected conjunctions between inscriptions in different collections, and between inscriptions and archaeological material, in ways that were barely imaginable a few decades ago. With the Pelagios project, the spatial component of this process of linking and aggregation has taken off. The temporal component still lags behind, however, simply because – unlike places, which exist in physical space – periods are discursive constructs that emerge from the needs of scholarly studies of the past to create order. As discursive constructs, they change over time and inspire revision, disagreement, and critique. This makes them difficult to manage in a structured-data environment: capturing the diversity of usage can create an impression of chaos, while smoothing out disagreement both excludes critique and erases some of the history of historical disciplines. One can see why absolute dates or generic period expressions might be more attractive for managers of digital epigraphic collections.

We hope, however, that we have shown some of the benefits that come with entry into the fray, and the goal of the PeriodO project is to continue to make it easier to do so. If the digital epigraphic community begins to include periods systematically in its data


structures, it will be rewarded with better interoperability across datasets, better ways to find information about inscriptions, and – perhaps most importantly of all – better opportunities to reunite inscribed texts with archaeological context at various scales. The flexibility of the PeriodO gazetteer should be able to meet the needs of a wide range of period uses in epigraphic corpora, from the relatively straightforward chronology of the Inscriptions of Israel/Palestine, which is largely satisfied with the period definitions used by the Levantine Ceramics Project, to the highly specific linguistic/stylistic periods that appear in some of the corpora of the Digital Archive for the Study of pre-Islamic Arabian Inscriptions.\(^{16}\) While the usefulness of period metadata may not appear immediately to the early adopters, it will become increasingly evident as more collections incorporate it and as temporal search and visualization tools become more robust. Today we cannot imagine how we managed without the PHI database of Greek epigraphy or the EpiDoc standard; tomorrow, we will not remember what it was like to be able to search easily across dozens of epigraphic collections for Archaic inscriptions alone, or visualize on a map and timeline how different corpora differ in their definitions of “Late Antiquity”. The transparent association of period definitions with material with absolute dates, like inscriptions, may even lead us to a fundamental reconsideration of the way we periodize the past.

**Bibliography**


