

Abstract

The way human beings talk about past time creates problems for the digital description of historical information. Computers can parse dates into a common representation so that they can be easily compared, but they can't parse expressions like "the long 18th century" or "the Archaic period." The same is true for geographical spaces – but where people can generally agree that "New York City" is a distinct entity that can be located within coordinate space, they don't necessarily agree on the boundaries of "the long 18th century" either in time or in space. Local metadata systems define historical, art-historical, archaeological, and even literary periods in idiosyncratic ways that can be difficult to align across systems. Centralized defined vocabularies like the Getty Art and Architecture Thesaurus define periods generally, but in the process sacrifice local specificity. Neither approach makes it easy for a user to find information associated with a particular period term or a particular date range across a range of library or museum records. The PeriodO project (<http://perio.do>) seeks to solve this problem through the creation of an open-ended Linked Data gazetteer of period definitions that are authoritative and clearly modeled, while avoiding the establishment of a "canonical" periodization that erases regional differences, scholarly disagreements, alternative voices and disciplinary histories.

During the NEH-funded first phase of the project, the PeriodO team, led by Dr. Adam Rabinowitz of The University of Texas at Austin and Dr. Ryan Shaw of the University of North Carolina, successfully modeled and built a gazetteer of period definitions that now contains more than 3500 unique entries with persistent URIs, all available both through a graphic user interface and as Linked Data (JSON-LD and Turtle) at <http://n2t.net/ark:/99152/p0>. We also established partnerships with two major aggregators of archaeological data, the US-based Open Context and the EU-funded ARIADNE project, who are using PeriodO URIs to describe their records. In this phase, Dr. Lorraine Haricombe, Vice Provost and Director of Libraries at UT Austin, joins the project as co-PI. Our goal in this phase is to expand the usefulness of the PeriodO platform and dataset beyond archaeology to meet the needs of a broader audience of librarians, data managers, and scholars and students across the academic spectrum. The lack of a consistent reference point for the periodization of the past is a shared problem for data access and discoverability across all those communities. The tools we plan to produce will make it possible for a user to search library catalogues not by the date of a publication, but by the dates of the topics covered in that publication; for a student to visually compare changes in the definitions or usage of particular period terms in a discipline across time; for an indigenous group to create and employ in a database a set of non-canonical period definitions of their own, without hindering the interoperability of their periodized information with other more traditional sources; and, eventually, for those interested in text-mining to automatically extract machine-actionable date ranges from texts that only refer to periods by name, just as they now do with named places and geographic coordinates.

This phase of the project will last two years. During the first year, working with our original programmer and a second specializing in visualization and API development, we will complete a set of visualization tools for searching and filtering in the graphic user interface. These will include a map display, enhancements to the existing timeline filter, search tools that allow the user to organize results by date of publication or by source, and visualizations for the representation of temporal uncertainty and the comparison of differing definitions for the same period term, or different terms used in the same places. We will also work with both old and new partners, including DPLA and Europeana, from a wide range of disciplines such as modern history, literature, library science and museum studies, to explore the role of PeriodO might play in the management and discoverability of their data. This will take the concrete form of a focused workshop with representatives of partner projects and other stakeholders, as well as consultation with an advisory board. During the second year, drawing on that workshop and on the user experiences of our partners and advisors, we will turn to the development of an API and a reconciliation service that will allow users to align their own data with period definitions present in PeriodO or identify data with overlapping date ranges in other sources. Since several of our partners will by that time have implemented PeriodO URIs in their own datasets, we also hope to compare the usage of period definitions to the absolute dating of individual records, and leverage those usage statistics to develop a probabilistic approach to the extraction of calendar dates and geographic coverage from named (but undated) periods in other texts and databases. A second workshop in the second year will bring together a different group of stakeholders to discuss both the adoption of PeriodO as a chronological gazetteer and the potential of probabilistic period-parsing and reconciliation tools for researchers and cataloguers.

Narrative

1. Statement of Need

Human beings have been documenting the passage of time since we learned how to tell stories and make symbols. Keeping track of time is important: it helps us to understand our place in the world in relation to our ancestors, and it helps us to plan for the future, especially in terms of the cyclical availability of food resources. Simple ways to describe time are readily available -- for example, the movement of astronomical bodies like the sun, moon, or stars, which keep us attuned to the seasons and years, or contextually meaningful markers like dramatic events or human generations (“in the year of the great snowstorm”, “after the flood”, “in my grandfather’s day”). After the invention of writing, our efforts to describe time expanded in scope. On the one hand, attention to the Earth’s annual cycle led to the development of calendrical systems, with counts that began with a particular event and proceeded on a regular count by lunar or solar year thereafter. On the other hand, people began to organize past time by grouping together certain events or circumstances, eventually associating those phenomena with specific years in calendrical systems. In the Bronze Age Near East and Egypt, this meant dating by dynasties and the regnal years of named kings, which coupled calendrical systems and absolute dates tightly to specific events and individuals. At the same time, however, a parallel, more relative system of organizing the past emerged: the grouping of past time into chunks, often without absolute dates, defined not by specific local kings or events but by larger patterns that reflected some sort of overarching order. The Greek poet Hesiod, for example, writing no later than the 7th century BCE, divides the history of the world into a Golden Age, a Silver Age, a Bronze Age, a Heroic Age, and his own miserable Iron Age.

If this chronological system seems familiar, it should. Similar periodizations recur in a vast series of texts, both ancient and modern, from the Bible to any contemporary art-history textbook. With the Enlightenment and the rediscovery of the prehistoric and Classical past in Europe in the 17th and 18th centuries, and with the establishment of the principles of geological stratigraphy and the attendant rise of more “scientific” approaches to history and archaeology in the 19th century, periodization became one of our primary means of organizing information about the past. Unlike raw calendrical dates, periods have the advantage of expressing a shared set of ideas about a combination of factors ranging from style to culture to technology, from intellectual trends to biodiversity. At the same time, periods -- expressed in purely relative terms of sequence -- side-step discussions of specific chronology and permit a salutary vagueness that has helped them to maintain their currency for decades or centuries (Christian Thomsen established in 1836 the concepts of “Stone Age”, “Bronze Age”, and “Iron Age” still used by prehistorians today). Periodization therefore continues to play a central role in the description and organization of information about the past -- and even the present, as the very current debate about the Anthropocene indicates (Waters et al. 2016). As a result, period terms and concepts are almost universally used as metadata elements in library catalogues (embedded in Library of Congress subject headings, for example) and museum databases.

The very characteristics that make period terms useful for local metadata systems, however, also make them problematic in a broader digital ecosystem. Period terms work effectively when the user and the audience share a common frame of reference: as long as both parties recognize the general time and space where the term applies and agree on a common set of characteristics, minor disagreements about space-time boundaries are largely irrelevant. But when the user moves from the local to the national or international level, shared frames of reference cease to apply, except in the broadest terms, and both local variations and disciplinary disagreements come to the fore. For information available on the global internet, this means that period terms become much less useful for the discoverability, visualization, and reuse of data. Unlike geographic locations or people, both of which may have different names in different systems but refer to unique entities and thus allow alignment through shared gazetteers (e.g. Geonames for places or the Virtual International Authority File for people), periods differ in both label and identity across time, space, and scholarship.

This disjunction is highlighted by the Anthropocene debate: although this is a period concept that should apply globally and, as a contemporary object of scientific analysis, should have precise coordinates in time, scholars disagree about when it should begin (anywhere from widespread deforestation associated with the spread of agriculture in the first millennium BCE to nuclear testing in the 1950s) or even if it exists at all. The use of “Anthropocene” as a metadata label in a library or museum catalogue therefore indicates the adoption of a particular intellectual position -- but without additional information, the user of the catalogue cannot know what chronological or spatial coverage that position entails, or whose opinion it is based on. As a result, internet

users cannot be sure if the results provided by a search for the string “Anthropocene” match their own spatio-temporal definition (or, in a federated search, if they are missing relevant results from databases that use only “Holocene”), while users searching for material within a specific time-range will not necessarily see records that have been associated with that string, but without dates.

Libraries, museums, and data-sharing initiatives have attempted to deal with this with tools that have worked well for other types of metadata: by using internally or externally defined controlled vocabularies, international standards, and -- more recently, and especially with the rise of Linked Data and semantic-web approaches -- alignment to multilingual gazetteers. But periods have proven very resistant to these solutions. Local periodizations offer specific coordinates in time and space, but are often too specific for use outside that local context. Global periodizations can be used by any dataset, but must therefore remain general and avoid the use of specific coordinates in time or space (this is the case for the period vocabulary in the Getty Art and Architecture Thesaurus). And although a single, central, internationally-accepted standard period vocabulary has been proposed several times (see Doerr, Kritsotaki and Stead 2010), it has so far failed to emerge, in part because different schools of thought take irreconcilable positions on the definitions such a vocabulary would include. There is therefore currently no universally recognized gazetteer or authority that identifies, with temporal and geographic precision, a comprehensive, multilingual, multidisciplinary set of periods.

This gap has serious digital consequences for any discipline that periodizes its information. On the most basic level, it makes it difficult or impossible for data aggregators to reconcile on a chronological level the period terms belonging to a group of disparate datasets that use those terms idiosyncratically (and/or multilingually). This makes it harder for those searching aggregated data to employ time as an effective search axis, and leaves them more dependent on absolute dates in the metadata, which may be unavailable, inconsistent, or poorly formatted. It also creates barriers to the visualization of periodized data on a timeline and to the inclusion of spatial elements in both the search and visualization of those data. Finally, it makes the reuse of periodized data for temporal or spatial analysis much more challenging, because it is often difficult or impossible for the re-user to determine what the data provider meant, in chronological and geographic terms, by the use of a specific period label.

As more ambitious efforts to aggregate information and federate databases emerge, the question of periodization is receiving more attention. Initiatives like the US-based Digital Public Library of America and the EU-based Europeana are exploring approaches that expand their current use of date-spans to deal with the presence of period references in the data of their contributors. The need for resources and mappings to connect the usage of period terms with the concrete coordinates expected by metadata fields for spatial and temporal coverage has recently been emphasized in various contexts (Berman 2011; Grossner 2013; Zadorozhny et al. 2013; Niccolucci and Hermon 2015). A number of projects have attempted to address the issue over the last decade, either by using existing thesauri like the Library of Congress subject headings (the Electronic Cultural Atlas Initiative: Petras et al. 2006) or the Getty Art and Architecture Thesaurus (Schreiber et al. 2008), or by creating their own interoperable gazetteers (CommonEras: Isaksen et al. 2009; SENESCHAL 2013: <http://www.heritagedata.org/blog/about-heritage-data/seneschal/>; chronOntology 2014: <http://gepris.dfg.de/gepris/projekt/247900032>). None of these projects, however, has been able to fully address the chaos of period usage across multiple platforms, languages, scholars, and disciplinary boundaries.

The PeriodO project (<http://perio.do>) is an attempt to meet this need, in a non-discipline-specific manner, for all those dealing with periodized digital data. The project is based on two fundamental ideas that will allow it to be of use to a very broad audience of data managers, including library cataloguers and metadata specialists, archivists, aggregators, developers of controlled vocabularies, and administrators of databases containing periodized information, as well as scholars and the lay public. First, we have embraced a Linked Data approach to period information, which we hold is better managed from below by alignment to a shared reference point -- a gazetteer, the entries in which are expressed according to a clear semantic model and given persistent URIs -- than from above by standardization of terms, concepts, or metadata schemata. Second, instead of developing a thesaurus of unitary period *concepts*, like the Getty AAT, we have chosen to develop a gazetteer of unique period *definitions*, each consisting of a period label, a temporal coverage that can be represented in ISO8601 years, a spatial coverage (currently parsed where possible in terms of national boundaries, as expressed in DBpedia), and an authoritative source. By collecting definitions that include these elements, we seek to allow the chronological cross-searching of disparate data sources without eliding scholarly disagreements and

disciplinary evolution. The PeriodO gazetteer offers data managers the opportunity to define their use of period terms clearly, unambiguously, and in a well-modeled, machine-actionable form. It is also intended to be expansible and responsive to the needs of the community: it has been built to allow (and maintain a history of) edits to existing definitions and the addition of new definitions by authoritative contributors. Thus, although the initial dataset has focused on periods related to archaeology and art history, it can easily be expanded to include period definitions related to history, literature, geology, paleontology, music, etc.

The PeriodO project provides a bridge between local efforts to develop controlled vocabularies for particular data-federation initiatives and more global attempts to develop a shared ontological framework for the representation of periods or time-spans, such as that of the CIDOC-CRM or the Extended Date-Time Format. By documenting and providing URIs for an unlimited number of period definitions that include date-range, spatial coverage, and source, PeriodO allows libraries, museums, and archives to describe their periodized data with explicit statements of spatio-temporal coverage, either drawn from existing authorities or minted to match local usage. This will facilitate cross-searching and interoperability, reducing the number of both false positives and false negatives that result from searches for periodized material solely by text string or date range. At the same time, the implementation of PeriodO URIs in large datasets will provide a statistical basis for the examination of period usage, which can in turn be used to parse dates from period references in texts or suggest appropriate period definitions to cataloguers. Furthermore, by embracing, rather than erasing, disagreement and multivocality and through the development of geotemporal visualization tools, PeriodO will enhance the ability of scholars and the public to understand how period definitions have evolved over time, where authorities agree and where they disagree, and how different national or intellectual traditions deal with the same historical phenomena. Finally, the expansion of the gazetteer to include period definitions from a wider range of disciplines will increase the potential for the serendipitous discovery of connections across aggregated data.

2. Impact

The first stage of the PeriodO project focused on the formulation of a functional semantic model for a period definition, on the construction of a web-based platform to store and manage such definitions, and on the creation of an initial period definition dataset. The proposed second phase will focus on making it easier for both data managers and laypeople to find period definitions, use their URIs to describe data, and then find synchronous records across heterogeneous datasets. The most immediate impact on libraries and archives will be a substantial increase in the discoverability of periodized data. When libraries, archives, and databases begin to use PeriodO URIs for collections that are already represented as Linked Data, it will be possible to search by date for records that are periodized but not explicitly dated. It will also be possible to carry out federated searches that will return periodized records that are chronologically and spatially relevant, even if the search term and the database vocabulary are different. Both of these tasks are currently very difficult. Library catalogues usually cannot be searched by date range of subject, but only by subject headings expressed as strings, which only sometimes include period terms and/or dates (and these are often both place-specific and idiosyncratic, as for example <http://id.loc.gov/authorities/subjects/sh85057078>). Conversely, federated platforms like the DPLA make it possible to search by absolute date across a range of datasets with date information in specific formats -- but because of differences in the use of period terms, results for period-term searches can be incomplete, inconsistent, or confusing.

The tools we propose to develop in this phase of the project will make it easier for data managers in libraries, archives, and museums to solve these problems by reconciling their own period vocabularies with PeriodO, harvesting PeriodO URIs, implementing those URIs in their own datasets, and then sharing that periodized information as Linked Data (our model here is the highly successful Pelagios project: <http://pelagios-project.blogspot.com/>). As this ecosystem grows, so will the ability of the user to find periodized information within a single dataset or across many, adding a critical search facet and enhancing data discoverability. Just as the use of spatial gazetteers has made it possible for libraries and archives to create dynamically updated representations of their geographic coverage on a map, and for federated platforms to allow map-based searching and browsing, the PeriodO gazetteer will facilitate visualizations of temporal coverage and timeline-based searches for periodized data across categories and disciplines.

The potential for visualization of period extents in both space and time is another area in which we expect the project to have a significant public impact. We plan to develop more sophisticated visualization interfaces

for the PeriodO client in this phase, using Javascript libraries to allow searching and the display of results by map, by timeline, by stacked timeline (to compare different definitions), and by map-timeline combination. These visualizations will make it easier for students and members of the public to understand scholarly discourse about the past by highlighting areas of disagreement and consensus, local or national traditions, development over time within a single discipline and correlations between developments in separate disciplines. We will also refine the data model to include information about derivative or hierarchical relations between definitions. In addition, as the number of collections implementing PeriodO and the number of records using PeriodO URIs increases, we hope to use data harvested from these records to build statistical models that will allow us to parse undated references to periods in the corpus of digital literature. Automatically finding and annotating period terms in digital texts would make those works more discoverable while also making it possible to represent their contents on a timeline, just as geoparsing a text makes it possible to map the places mentioned in it. By increasing the ability of both scholars and laypeople to find unstructured temporal information by date-range, period term, or a combination of the two, we hope to add a new and powerful layer to the current landscape of digital discovery.

PeriodO has been conceived since the beginning as an open, community-focused platform. The potential for community engagement is built into its architecture, and the tools we propose to build in this phase will foster the development of that community by making it easier for data managers to align their own data with PeriodO and add PeriodO URIs to their period vocabularies. Because the client runs in a local browser, users are able to create local period databases in the PeriodO format, either by collecting existing definitions from the canonical dataset or by adding their own new collections and definitions. Users authenticated through ORCID can submit these new collections and definitions as a patch to the canonical dataset; that patch is reviewed by the editors and, if accepted, merged with the canonical dataset, at which point the new definitions receive URIs, using the ARK identifier system deployed by the California Digital Library's EZID infrastructure. Thus specialists and scholars working with periodized data are encouraged to submit their own definitions and correct existing definitions that contain errors of fact (not interpretation -- differences of opinion are expressed by new definitions). We deliberately chose this strategy so that the gazetteer would be expansible and available to any discipline that can assign temporal and spatial coverage to its period terms. Currently the project managers are solely responsible for approving and merging patches, but if, as we hope, the user base increases substantially, we plan to establish a formal editorial workflow and an editorial college along the lines of the Pleiades spatial gazetteer. Moreover, since the definitions themselves are modeled simply and do not insist on any particular ontological framework, they can be easily adapted to a range of metadata approaches, from Dublin Core to the CIDOC-CRM.

Beyond our own maintenance of the dataset, the code for the PeriodO server and client, as well as the complete dataset, are available for download, reuse, and remixing under a Creative Commons 0 1.0 license. This means that anyone with access to the internet can download the dataset, use it in another digital project, build a separate instance on a separate server, etc. The code is currently hosted on GitHub, and an archival copy of the code will be deposited in the institutional repositories of both The University of Texas and the University of North Carolina in case this arrangement becomes impossible. Backup copies of the canonical dataset itself, as JSON-LD, will also be deposited in those institutional repositories on a regular basis, ensuring that copies of the dataset remain available indefinitely. The ARK IDs will persist even if the server goes offline, and the details of any period definition can be recovered from the static backup files if necessary.

The progress of the project in its second phase will be assessed by both quantitative and qualitative measures. The easiest performance metric will be the expansion of the dataset: by the end of the first year of the project in May 2017, we will have added at least 1500 new period definitions. Some of these will be from our partner projects, but we also plan to add a significant number of period definitions from scholarship of the 19th and early 20th centuries. Many of these definitions are now outdated, but their inclusion will allow us to trace the evolution of period concepts across time in a more comprehensive manner. We will also fill in coverage of some geographic areas that are currently underrepresented (especially Russia, Australia, Africa, and Central and South America), and add period definitions from a new set of disciplines, including geology, paleontology, and literature. In the latter areas we will be guided by new members of our Advisory Board: Laura Mandell, a literary scholar and digital humanist who directs the Advanced Research Consortium (ARC): <http://idhmc.tamu.edu/arcgrant/> at Texas A&M, and Denné Reed, a paleoanthropologist who directs

PaleoCore, an ontology and digital infrastructure project for human origins (<http://paleocore.org/>). The beneficiaries of this expansion will be the scholars and lay users who consult the site; we intend to keep track of the number of unique visitors using web analytics, and to use this as a rough measure of the impact of the dataset itself. These visitors are also the intended beneficiaries of the second performance metric: the completion of the suite of search and visualization tools, which will allow users to search and view data using a range of chronological and spatial filters and displays. These tools will also take advantage of refinements to the data model that will allow the representation of hierarchical (period/subperiod) and derivative relationships between definitions. The effectiveness of our visualization tools will also be evaluated by the members of our Advisory Board, many of whom are specialists in data visualization and user-interface design.

More directly relevant to the library and museum world will be the third performance metric: the implementation of PeriodO URIs for all periodized records in at least two of our partner projects, Open Context and Pleiades. Open Context is a publication platform for archaeological data that contains nearly a million records from more than 60 archaeological projects, including the site-files contributed by US states as part of the Digital Index of North American Archaeology. Many of these records can be provided with PeriodO URIs. Pleiades, a gazetteer of ancient places, has a smaller dataset -- roughly 35,000 places -- but most of those also have locations in coordinate space, so this dataset will enable us to experiment with map representations and help us explore ways to extract spatial information directly from periodized historical gazetteers. We have already ingested both DINAA and Pleiades periodizations into PeriodO. By the end of January 2017, during year one, we will have helped Open Context and Pleiades to attach those URIs to all of their periodized records, which we expect to total more than half a million. We will measure success in this area by the number of records assigned URIs and by feedback from the directors of these partner projects.

The implementation work for Open Context and Pleiades will lay the foundations for the API, reconciliation, and federated search tools to be developed in year 2. Our products for the second year will be an API (usable with data-improvement tools such as OpenRefine) that will allow the PeriodO dataset to be used for mostly-automatic alignment and reconciliation of period terms found in datasets, as well as a cataloguer toolkit that will allow a data manager with a period vocabulary to manually reconcile it with definitions in PeriodO on the basis of overlaps in label, time, and space, and then to join vocabulary terms to the corresponding period URIs, date ranges, and geographic coverage. We will also produce a proof of concept or prototype tool for the statistical matching of period definitions with undated period terms used in databases or publications. This tool will be intended to serve as both a suggestion engine for database managers, to enhance reconciliation, and as a text-mining resource for the recognition, disambiguation, and visualization of references to named periods. The API and the reconciliation toolkit will be evaluated by the alignment consultant, who will be a senior specialist with extensive experience in the alignment of vocabularies across federated datasets. A final measure of progress will be the application of these tools to at least one more of our partner datasets before the end of the second year. We will also solicit detailed feedback from other partners and our Advisory Board on these tools.

Our trajectory across the two years of the grant will be strongly shaped by two workshops, one to be held in Austin at the end of the first summer and the other to be held in Chapel Hill at the beginning of the second summer. These small, intensive workshops will bring together representatives of our partner projects, members of our advisory board, and digital strategy staff from major national or international libraries and museums. We will use them not only to introduce PeriodO and explore ways in which it might be used to support participants' datasets, but also to solicit feedback about the needs of the museums and libraries represented, about our planned tools, and about different ways of conceptualizing time in different disciplines. We will let the needs and conceptual frameworks of our partners and advisors shape the project's agenda, while sticking to the two basic principles mentioned above. The first workshop will focus on visualization and reconciliation tools, while the second workshop will explore more technical elements of API design and statistical modeling. The outcomes of these workshops will include a greater interest on the part of our partners in the integration of the PeriodO gazetteer into their work, guidance on tool functionality and statistical approaches, and a clearer sense of the documentation and resources potential adopters will need to use PeriodO in their own data management. We will write up the results of these workshops and post them, along with tutorials and documentation, on the PeriodO website. As an added benefit, the workshops will help us to formulate particular metadata competencies that PeriodO users can acquire through the interface and dataset. We have already been invited to propose PeriodO-related competencies to add to the competency index maintained by the IMLS-funded LD4PE

project (<http://explore.dublincore.net/theory/briefing-papers/ld4peoverview/>). By linking PeriodO tools and tutorials to those competencies, we can raise the project's profile and make it attractive for use in training courses and information-science curricula.

We anticipate that the gazetteer itself, the reconciliation tools, and the statistical approach to period recognition and disambiguation will have a substantial impact on those libraries, museums and archives that are already beginning to embrace Linked Data strategies. The gazetteer adds a missing temporal dimension that will allow much richer visualization and discovery within and across datasets. The API and reconciliation toolkit will make it easy for cataloguers to include these Linked Data URIs in their own datasets, dramatically increasing the interoperability of those datasets with other Linked Data sources. And statistical approaches to unstructured period expressions may reveal unseen temporal patterns in texts and databases, including hidden inconsistencies, disagreements, and errors. If our current partners, all of whom have adopted or are exploring semantic strategies, and all of whom are leaders in their fields, adopt PeriodO as their shared point of reference for periodization, they are likely to form a critical mass that will encourage further adoption well beyond these partners and the term of this grant. We are confident that funding for this phase of PeriodO will allow it to achieve that critical mass of users, and once it does, we feel that it will become a critical piece of infrastructure for the National Digital Platform. More adoption will mean a more active user community, and a more active user community will mean that the gazetteer will continue to grow and that the dataset -- copied many times and stored in many places -- will remain accessible and stable long after the project is complete.

3. Project Design

The PeriodO project was funded from March 2014 to December 2015 by an NEH Level II Digital Humanities Start-up Grant (HD-51864-14). With that grant, co-PIs Rabinowitz and Shaw, together with Lead Developer Golden and PhD student Sarah Buchanan, were able to build the PeriodO server and client, and to create an initial dataset that now stands at just over 3500 entries. With the help of our Advisory Board, we developed a robust and flexible datamodel and a basic interface for viewing, searching, editing, adding, and downloading period data. This phase of the project exceeded some of its goals, especially in the size of the dataset, and met most of the others. The only part of the project that remained incomplete was the visualization suite, since the construction of the editing interface and patch submission workflow took longer than expected. The project has resulted in more than ten presentations and four publications to date (see Publications and Presentations in Supporting Documents), and a fifth, for the Bulletin of the Institute of Classical Studies, is underway. We are now requesting IMLS funding to embark on the next phase of the project, which builds on those existing accomplishments, on user and Advisory Board feedback, and on our experience during the past year and a half.

PeriodO 2 has **five objectives**. **The first objective is to complete the suite of search and visualization tools for the client interface that we planned for the first phase.** We have already discussed a series of ideas and wireframes for these tools with each other and with the first PeriodO Advisory Board. If we receive funding, the team's first priority is to build these tools, using Javascript visualization libraries such as d3.js. The team has agreed on the core functionality to be added: search by geographic boundaries or interactive timeline, as well as by string and field value; search and order by date of source rather than date of period; compare the extent of several periods that overlap in space and/or time; view period distribution by country (choropleth) or timeline or combination of the two. We will also incorporate refinements to the datamodel, specifically hierarchical and derivative relationships, in the visualization functionality. Work toward this goal will begin at the beginning of the project, informed by both the first Advisory Board conference call and the first workshop, which will take place early in year 1, after prototypes have been created but while there is still time to adjust.

The first workshop will also allow us to collect preliminary input on **the project's second objective: the creation of tools and workflows that will make it easier to align an arbitrary period vocabulary with PeriodO definitions, and facilitate the addition of PeriodO URIs to existing datasets.** We will survey our Advisory Board members, project partners and workshop participants on their needs, and use this information to plan for a gradual approach to the issue of implementation. The first step will be the development of protocols to assist with the integration of PeriodO URIs into the datasets of two partner projects that have already provided periodizations to the PeriodO dataset: Open Context and Pleiades. Once this relatively low bar has been cleared, Shaw and Golden will begin to work on an API and reconciliation service that will make it

possible to align external periodizations with PeriodO definitions. The completion of these tools, which will require some technical expertise, will be followed by their use to reconcile the period vocabulary of another of our partner projects with PeriodO, and to implement PeriodO URIs in that dataset. Here we will draw on the knowledge of a technical consultant with extensive expertise in the alignment of vocabularies and data federation. Finally, this implementation stage will overlap with the development of user-friendly tools and “data cookbooks” to allow data managers, cataloguers and metadata specialists without Linked Data expertise to connect their existing vocabularies with period definitions and URIs from PeriodO.

The **third objective** springs from the second: once PeriodO URIs are implemented in partner datasets that can themselves be expressed as Linked Data, the project will seek to build **a query platform that aggregates those partner datasets and allows a user to carry out a federated search for periodized records using a combination of period terms, spatial entities, and date ranges**. Our model here is the Pelagios aggregation project, which created an API to search place-names across a range of datasets with place identifiers aligned to the Pleiades gazetteer (<http://pelagios.dme.ait.ac.at/api>), and is now developing a more complex interface called Peripleo (<https://github.com/pelagios/peripleo>). Elton Barker, the Community Director of the Pelagios project, is a member of our Advisory Board and will be providing guidance on the development of this functionality.

The **fourth objective** is the simplest, and will be addressed during the first year of the project. This is **the expansion of the canonical dataset to include period definitions from earlier works (18th century if available, otherwise 19th and early 20th), to provide greater time depth among sources; from geographic areas that are not currently well represented (e.g. Africa and Latin America), to provide greater spatial coverage; and from disciplines that we have not yet addressed, especially geology, paleontology, and literary studies, to provide more diverse disciplinary coverage**. These new definitions will be culled from print publications or provided by partner projects, and entered by a graduate student technician trained and employed from the beginning of the project in summer 2016 to the end of the spring semester in 2017. We have set a target of 1500 new definitions, but given our experience in the first phase of the project, this is a conservative estimate, and we will probably be able to include more.

The **fifth objective** is the most complicated and ambitious. The second workshop, although primarily intended to assess and guide progress toward the second and third objectives, will also consider this objective, which envisions **the development of statistical models that use the PeriodO dataset, together with records in partner datasets that have both absolute dates and PeriodO URIs, to recognize and disambiguate unstructured references to periods in external databases or text corpora**. There are a number of precedents for this sort of statistical parsing of unstructured information, both in general and with specific reference to periods (de Boer, van Someren and Wielinga 2010). But the only efforts so far that have had access to such an extensive set of training data are spatial, and the use of PeriodO for this purpose would be a novel and powerful application.

The activities required to achieve these objectives fall into three categories: **programming and development; data entry**; and the **organization and facilitation of conversations with expert advisers**, including the Advisory Board, the alignment consultant, and the participants in the two workshops.

In the **first category**, necessary project activities include the migration of the server architecture and dataset to an AWS server administered by the UT Libraries. This will take place at the beginning of the project, in early summer 2016, under the supervision of Choate and with the participation of Shaw and Golden. Shaw and Golden will also work with Rabinowitz during the first summer to establish a set of design goals for the search and visualization tools, which Golden and the supporting developer will then go on to build over the course of the fall of 2016 and early spring of 2017. Also in the summer of 2016, Shaw will consult with partners at Open Context and Pleiades on the implementation of PeriodO URIs in their datasets, and Golden will offer technical support over the following fall. Development work on the API and reconciliation toolkit will be carried out by Golden and the supporting developer throughout 2017. Final development activities will be carried out by Golden in the spring of 2018, under the supervision of Shaw, and will focus on the cataloguer toolkit and the creation of a probabilistic matching prototype.

In the **second category**, project activities include the training of a graduate student data entry technician, and the collection and addition of period definitions to PeriodO, either from contributed datasets or from print publications. The training of the graduate student will be undertaken by Rabinowitz in the summer of 2016, and

the graduate student will continue to work independently under Rabinowitz's supervision across the 2016-2017 academic year.

In the **third category**, project activities include the organization of five Advisory Board conference calls, to be undertaken by Rabinowitz across the duration of the project; the organization of two workshops, with the first to be carried out in Austin in the late summer of 2016 under the supervision of Rabinowitz, and the second to be carried out in Chapel Hill in the early summer of 2017 under the supervision of Shaw; and consultation with the alignment expert on the application of the reconciliation tool to partner vocabularies, which will take place through a series of remote conversations (Skype, email) in October and November of 2017.

We have formal commitments from external advisors and partners on several levels, all of which are documented in writing in letters included in the Supporting Documents. A number of organizations or projects have agreed to partner with us to explore the application of PeriodO to their data; these include DPLA, Europeana, Open Context, Pleiades, ARIADNE, the Advanced Research Consortium, and Project Andvari. Of these, Open Context, Pleiades, ARIADNE, ARC and Andvari have committed to the most extensive participation, providing period vocabularies for inclusion in PeriodO, attaching PeriodO URIs to their records, and exposing or sharing their datasets for the development of reconciliation and querying tools, as well as offering user-testing and feedback. Representatives of these partners will also serve on our Advisory Board and participate in one or both of the workshops. DPLA and Europeana have committed to exploring the use of PeriodO URIs in their platforms, to user-testing and offering feedback on the tools and interfaces we develop, and to providing representatives to both the Advisory Board and the workshops. We also have ongoing partnerships with archaeological partners from our previous phase (ADS, tDAR), which will continue on a more informal basis (they have provided periodized datasets that we can use for testing).

Additional experts have committed to serve on our Advisory Board, and some of these will also be invited to attend one or the other of the workshops. These experts include Tom Baker of DCMI; Elton Barker of the Pelagios project; Lex Berman of Harvard's China Historical GIS; Karl Grossner of Stanford Libraries; Dan Pett of the British Museum; Denné Reed of the PaleoCore project; and Mia Ridge of the British Library. All of these collaborators have agreed to participate in four to six conference calls and to attend workshops if invited.

Finally, we are planning to invite additional representatives of the library, museum and archives sphere to both of the workshops. We will invite additional participants to the first workshop who are responsible for the management of data with temporal attributes, and who might be interested in tools for visualization, reconciliation, and implementation; these will include representatives of bodies such as the Library of Congress, the Smithsonian, the CIDOC-CRM working group, and the University of Michigan Libraries, as well as participants from local units like the Harry Ransom Humanities Research Center. For the second workshop, we will focus on participants who are involved in the large-scale reconciliation and federation of massive datasets, and who have experience with probabilistic data-mining; these will include representatives of bodies such as the Semantic Computing Research Group at Aalto University, the VU Network Institute at VU University Amsterdam, and the Database Systems Research Group at Heidelberg University, as well as participants from local units like the Carolina Digital Repository and iBiblio.

4. Diversity Plan

We have no formal diversity plan, but the structure of the PeriodO gazetteer is by design inclusive of alternate or non-traditional definitions of periods, and the directors of the DINAA project have noted that they are interested in soliciting indigenous period definitions for North American data that might be included in PeriodO.

5. Project Resources: Personnel, Time, Budget

Dr. Adam Rabinowitz will serve as project director. He will be responsible for the overall organization of work; the oversight of the budget; the recruitment, training, and supervision of the graduate student technician; the recruitment (with Shaw and Golden) of a supporting programmer; the organization of the Advisory Board calls and the first workshop; and design, user-testing and evaluation of the tools and interfaces developed during this stage of the project. Dr. Rabinowitz was the PI of the first PeriodO grant from NEH, for which he performed the same duties, including overall project management and financial oversight. As Assistant Director of the Institute of Classical Archaeology, he has ten years of experience in project management, and from 2004-

2008 was responsible for a portfolio of excavation and digitization projects in Crimea with a total annual budget of up to a million dollars. As a field archaeologist responsible for a large digital dataset and the director of the GeoDia project (<http://geodia.laits.utexas.edu>), one of the inspirations for PeriodO, he has extensive experience with the application of periods to digital data and with geotemporal visualization.

Dr. Lorraine Haricombe will serve as project co-director. She will organize and oversee the project's integration with UT Libraries, including the migration of the PeriodO server and data to a UT Libraries server, the archiving of the data in UT's Institutional Repository, the logistical arrangements for the first workshop, and the management of project finances by the UT Libraries program office. As the head of the research library of one of the largest public universities in the US, Dr. Haricombe will also act as a liaison between the project and the library community, expanding the project's exposure and offering the perspective of a library administrator in the conversations of the Advisory Board and at the workshops.

Dr. Ryan Shaw will serve as project co-director. He will be responsible for the oversight of the subaward to the University of North Carolina; the establishment of goals and strategies for the completion of the development tasks envisioned by the project, and the supervision of Golden's work toward those goals; the recruitment (with Rabinowitz and Golden) and supervision of a second supporting programmer; the organization of the second workshop at Chapel Hill; the recruitment of, and collaboration with, an expert consultant on vocabulary alignment; and the theoretical and statistical groundwork toward the prototype probabilistic matching tool. Dr. Shaw was the co-PI of the first PeriodO grant from NEH, for which he performed many of the same duties. He brings extensive experience with Linked Data, semantic approaches, and the digital representation of historical events and periods.

Patrick Golden, PhD candidate in the School of Information at the University of North Carolina, will serve as Lead Developer. He will be responsible for building the visualization interfaces, expanding the functionality of the client, and constructing, with Shaw, the API and reconciliation and federated query tools. He will also help to determine the work carried out by the supporting developer. Mr. Golden has extensive experience with software and web development, and programmed the current PeriodO server and client during the first phase of the grant.

Aaron Choate, Interim Assistant Director for Technology Innovation & Strategy at the University of Texas Libraries, will provide server administration and support. He will carry out the installation of an AWS server to run the PeriodO back-end on a long-term basis through The University of Texas, and the migration of the PeriodO canonical dataset to that server. Choate has extensive experience with information architecture at UT Libraries and is responsible for server management.

A **Software Developer**, to be recruited when the project begins, will be responsible for support Golden's development activities, contributing to the design and programming of the visualization and search interfaces, the client architecture, and the construction of the API and reconciliation tools as necessary. This individual will have experience in web development, Javascript, RESTful APIs, data reconciliation and alignment, and Linked Data. S/he will be recruited in Austin or North Carolina through professional networks and/or the advertisement of the position on university job sites.

An **Alignment Consultant**, to be recruited in the second year of the project, will offer expert advice on the alignment of partner vocabularies to the PeriodO gazetteer. We will seek to recruit an individual with extensive experience in the Linked Data ecosystem and in vocabulary alignment and data harmonization, with knowledge of the issues surrounding the representation of time in digital data. We will rely on our partners for guidance in the identification of appropriate candidates. The candidate selected will be responsible for assisting the project team in the development of a workflow to align partner data with the PeriodO gazetteer.

The project timeline is presented in the Schedule of Completion. The project will run from June 2016 to May 2018. Development of the visualization and search interface will begin immediately, and will be complete by January 2017. Over the summer of 2016, the server and dataset will be migrated to an AWS server at UT Austin. From June 2016 to May 2017, a graduate student employee will add 1500 or more new period definitions to the canonical dataset. The first stakeholder workshop will be held in Austin in late summer of 2016, at a point when the participants can discuss both visualization and reconciliation tools. During the fall of 2016, Shaw and Golden will consult on the implementation of PeriodO URIs in Open Context and Pleiades; this work will give them a better sense of the requirements of the API and reconciliation tools, which they will begin

to develop in the spring of 2017. The second stakeholder workshop will be held in Chapel Hill at the end of that spring, preparing the team to move into the summer with a clear vision of the tools to develop and the needs of potential adopters. The API and reconciliation tools will be complete by December; work on alignment and cataloguer tools will begin in mid-fall of 2017, as the API work draws to a close. Shaw and Golden will experiment with the reconciliation of partner data in the early spring of 2018, and as more partner data is aligned with PeriodO, we will begin to use this information to scope and design the probabilistic matching methods and tools. The project will conclude at the end of spring 2018 with the completion of a probabilistic matching prototype or proof of concept.

Equipment and facilities are limited to the setup, hosting, and administration of the AWS server, which will be provided by UT Libraries. The University of Texas at Austin and the University of North Carolina will provide space and internet access for the participants in the two workshops. Cost-sharing for the Austin workshop and Rabinowitz's travel to the Chapel Hill workshop will be drawn from the UT Libraries budget.

6. Communications Plan

It is the project's intention to reach the widest possible audience among libraries, museums, and archives, as well as individual scholars who may be working with periodized datasets. We have therefore included in the grant request funds for PI attendance at two domestic and two international conferences. We will choose these as we go, according to audience, location, and theme, in order to maximize our coverage, but some candidates are the Digital Heritage conference, the Digital Humanities conference, the annual meeting of the American Library Association, the ALICE conference, the Dublin Core conference, etc. We will also maintain an active online presence through the PeriodO website (<http://perio.do>), the project's Twitter feed (@perio_do), and guest posts on other websites, such as the blog of the American Historical Association. These online activities will not only be promotional; they will also direct attention to the tutorial materials on our website and to technical documentation there and on GitHub that we hope to integrate into instructional resources such as LD4PE. The project team will continue its publication activities in relevant disciplinary journals, and we will continue to welcome and actively solicit new contributors of period information. Finally, we will rely on our partners and Advisory Board members to make their own communities aware of our work and the resources we are developing, as many of them have already begun to do.

7. Sustainability

We have discussed the sustainability of our software and dataset above, and we provide additional technical details in the Digital Stewardship Supplementary Information Form. To summarize, we are taking the following steps to ensure that the resources we have created already and those that we create during the next stage of the grant continue to be useful and accessible on an indefinite basis:

- We are working to create a multidisciplinary community of practice around PeriodO that will maintain the model and data (under a CC0 1.0 license) through copying, reuse and reference;
- The EZID system will permanently maintain Period URIs in the form of persistent ARK IDs with suffix pass-through, which can be resolved using the Name-to-Thing resolver and redirected to the dataset at whatever address is current;
- UT Libraries have taken on a long-term hosting commitment for the server architecture and the dataset, and copies of both the code and the dataset will be archived on a regular schedule in Texas ScholarWorks, the Merritt Repository at CDL, and the Carolina Digital Library;
- The code and data will also be hosted in an open-access repository on GitHub for as long as this is possible, under a CC0 1.0 license;
- The software products will be thoroughly documented, and detailed tutorials will be produced; this technical documentation will be preserved along with the code in both Texas ScholarWorks and GitHub;
- We will formulate Linked Data competencies that can be acquired through engagement with PeriodO, propose these to the evaluation committee at LD4PE, and, if they are accepted, add our tutorial material to the LD4PE index so that it can be incorporated into Linked Data training and curricula.

University of Texas Libraries - Periods, Organized (PeriodO) 2: Linking, Discovering, and Reconciling Information about the Past

Activity (Responsible)	June	July	August	September	October	November	December	January	February	March	April	May
First Advisory Board call: discussion of search and filter functionality, visualization tools (Rabinowitz)												
Programming of new client functionality and visualization tools (Shaw, Golden, developer)									milestone: completion of map, timeline, and comparison visualization tools			
Data entry of definitions contributed by partners, gathered from 19th-early 20th c texts (Rabinowitz, graduate student)												milestone: at least 1500 new definitions entered
First stakeholder workshop in Austin (Rabinowitz)												
Implementation of PeriodO URIs in Open Context and other partner projects (Shaw, Golden)									milestone: URIs fully implemented in Open Context and Pleiades			
Second Advisory Board call: discussion of reconciliation service, API (Rabinowitz)												
Development of API/ reconciliation and aggregation tools (Shaw, Golden, developer)												
Conference presentations												
Second stakeholder workshop in Raleigh (Shaw, Golden)												

University of Texas Libraries - Periods, Organized (PeriodO) 2: Linking, Discovering, and Reconciling Information about the Past

Activity	June	July	August	September	October	November	December	January	February	March	April	May	
First Advisory Board call: discussion of outcomes of workshop, planning for API and reconciliation tools (Rabinowitz)													
Development of API/ reconciliation and aggregation tools (Shaw, Golden, developer)								milestone: completion of API and reconciliation tools					
Conference presentations													
Second Advisory Board call: discussion of reconciliation service, API, alignment, probabilistic tools (Rabinowitz)													
Work with alignment consultant to develop strategy for the reconciliation/alignment of partner vocabularies (Shaw, Golden, consultant)													
Application of API and reconciliation/aggregation tools to partner datasets (Shaw, Golden)								milestone: reconciliation of at least one partner periodization with PeriodO tools					
Development of user-friendly tools or workflow to facilitate use of PeriodO gazetteer by cataloguers (Shaw, Golden, Rabinowitz)											milestone: cataloguer tool available for beta-testing		
Exploration of partner data to assess potential for use in machine-learning/probabilistic parsing tools (Shaw, Golden, Rabinowitz)													
Third Advisory Board call: assessment of API and tools, discussion of probabilistic approach, planning for next phase (Rabinowitz)													
Development of proof-of-concept or prototype for probabilistic matching/parsing of period terms in texts or databases (Shaw, Golden)												milestone: proof of concept or prototype of probabilistic matching tool	

Supporting Document:
List of PeriodO publications, presentations and public talks, 2014-2015

Publications

1. Shaw, Ryan, Adam Rabinowitz, Patrick Golden, and Eric Kansa. 2015. "A Sharing-Oriented Design Strategy for Networked Knowledge Organization Systems." *International Journal on Digital Libraries*. doi:10.1007/s00799-015-0164-0 Preprint at <https://www.researchgate.net/publication/280529967>.
2. Golden, Patrick, and Ryan Shaw. 2015a. "Period assertion as nanopublication." In *Semantics, Analytics, Visualisation: Enhancing Scholarly Data Workshop Co-Located with the 24th International World Wide Web Conference*. Florence, Italy. <http://cs.unibo.it/save-sd/2015/papers/html/golden-savesd2015.html>.
3. ———. 2015b. "Nanopublication beyond the Sciences." PeerJ PrePrints. doi:10.7287/peerj.preprints.1284v1. (accepted, appearing shortly)
4. Rabinowitz, Adam. 2014. "It's about time: historical periodization and Linked Ancient World Data." *ISAW Papers* 7 (22). <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/rabinowitz/>.

Presentations and Public Talks

1. Golden and Shaw, "Period assertion as nanopublication." "Semantics, Analytics, Visualisation: Enhancing Scholarly Data Workshop" at the 24th International World Wide Web Conference, Florence, Italy, May 2015.
2. E. Kansa, "Challenges in Archaeology, Linked Data, and Publishing Geospatial Data on the Web." (PeriodO featured prominently in talk) Center for Geospatial Analysis, Harvard University, Boston, Massachusetts, September 2015.
3. E. Kansa, demonstration of PeriodO in lecture and workshop "Methods in Archaeological Data Publishing." NEH-funded Digital Archaeology Institute, Michigan State University, Lansing, Michigan, August 2015.
4. E. Kansa, "Open Context and PeriodO." Lightning talk at Society for American Archaeology annual conference, San Francisco, California, April 2015.
5. E. Kansa, "Contextualizing Digital Data as Scholarship in Eastern Mediterranean Archaeology." (PeriodO featured prominently in talk) Center for Hellenic Studies, Harvard University, Boston, Massachusetts, April 2015.
6. Rabinowitz, "Periods, Organized (PeriodO)." Digital Index of North American Archaeology "Radiocarbon and Temporality" workshop, University of Wyoming, Laramie, Wyoming, June 5, 2015.
7. Rabinowitz, "PeriodO: a gazetteer of period assertions for linking and visualizing data. Why is it important to include periods in a Linked Data infrastructure, and how do we do it?" Mellon-funded "Linking the Middle Ages" workshop, The University of Texas at Austin, Austin, Texas, May 11, 2015. White paper from workshop at <http://dx.doi.org/10.15781/T2MW2C>.

8. Rabinowitz, "PeriodO: a gazetteer of period assertions for linking and visualizing data." Lightning talk at "Mapping the Digital Landscape at the University of Texas at Austin: A Campus Digital Preservation Symposium", The University of Texas at Austin, Austin, Texas, May 1, 2015.
9. Rabinowitz, "Managing Time: PeriodO, a Linked Data approach to the interoperability of periodized data." "Fostering Transatlantic Dialogue on Digital Heritage and EU Research Infrastructures: Initiatives and Solutions in the USA and in Italy" workshop, Library of Congress, Washington, D.C., December 2, 2014.
10. Rabinowitz, Shaw, and Kansa, "Periods, Organized (PeriodO): a Linked Data gazetteer to bridge the gap between concept and usage in archaeological periodization." 2014 Computer Applications and Quantitative Methods in Archaeology Conference, University of the Sorbonne, Paris, France, April 24, 2014.
11. Shaw, "Hybrid online/offline scholarly information resources." Coalition for Networked Information Fall 2015 Meeting, Washington, DC, December 15, 2015. <https://goo.gl/ntvSY4>
12. Shaw, "Report on & demonstration of the PeriodO period gazetteer." 14th European Networked Knowledge Organization Systems Workshop, Poznań, Sept. 18, 2015. <https://goo.gl/a7Mgvv>
13. Shaw, "An ecosystem of time periods: PeriodO." Linked Pasts colloquium, Pelagios project, King's College London, July 20, 2015. <http://pelagios-project.blogspot.com/2015/03/linked-pasts.html>

**Supporting Document:
Proposal Bibliography**

- Berman, Merrick. 2011. "Time in Gazetteer Data Models." In *Symposium on Space-Time Integration in Geography and GIScience*. Seattle. https://cga-download.hmdc.harvard.edu/publish_web/2011_AAG_Gazetteer/Berman.pdf.
- de Boer, V., M. van Someren, and B.J. Wielinga. 2010. "Extracting historical time periods from the Web." *Journal of the American Society for Information Science and Technology* 61 9 : 1888–1908. doi:10.1002/asi.21378.
- Deutsche Forschungsgemeinschaft. 2014. "chronOntology: ein Zeit-Gazetteer für die historisch-kulturwissenschaftlichen Fächer." <http://gepris.dfg.de/gepris/projekt/247900032>
- Doerr, Martin, Athina Kritsotaki, and Steven Stead. 2010. "Which Period Is It? A Methodology to Create Thesauri of Historical Periods." In *Beyond the Artefact: Digital Interpretation of the Past*, ed. Franco Niccolucci and Sorin Hermon. Budapest: Archaeolingua.
- Grossner, K. 2013. "A space-time datatype for historical place?" Paper delivered at the Digital Humanities 2013 conference in Lincoln, Nebraska. https://cga-download.hmdc.harvard.edu/publish_web/SpaceTime/DH2013_Grossner_SpaceTimeDatatype.pdf.
- Heritage Data. 2013. "SENESCHAL." <http://www.heritagedata.org/blog/about-heritage-data/seneschal/>.
- Isaksen, Leif, Chris Gutteridge, Gianluca Correndo, Zurina Muda, Lin Xu, and Thanassis Tiropanis. 2009. *CommonEras*. University of Southampton. <http://commoneras.ecs.soton.ac.uk>.
- Niccolucci, F., and S. Hermon. 2015. July 21. "Representing gazetteers and period thesauri in four-dimensional space–time." *International Journal on Digital Libraries* (July 21): 1–7. doi:10.1007/s00799-015-0159-x.
- Petras, Vivien, Ray R. Larson, and Michael Buckland. 2006. "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 (JCDL)*. Chapel Hill: ACM Press. <http://dx.doi.org/10.1145/1141753.1141782>.

Schreiber, G., A. Amin, L. Aroyo, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, et al. 2008. November. "Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator." *Web Semantics: Science, Services and Agents on the World Wide Web* 6 4. Semantic Web Challenge 2006/2007 (November): 243–249. doi:10.1016/j.websem.2008.08.001.

Waters, C.N., J. Zalasiewicz, C. Summerhayes, A.D. Barnosky, C. Poirier, A. Gałuszka, A. Cearreta, et al. 2016. "The Anthropocene is functionally and stratigraphically distinct from the Holocene." *Science* 351 6269: aad2622. doi:10.1126/science.aad2622.

Zadorozhny, V., P. Manning, D.J. Bain, and R. Mostern. 2013. "Collaborative for Historical Information and Analysis: Vision and Work Plan." *Journal of World-Historical Information* 1 1: 1-14. doi:10.5195/jwhi.2013.2.

DIGITAL STEWARDSHIP SUPPLEMENTARY INFORMATION FORM

Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded research, data, software, and other digital products. The assets you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products is not always straightforward. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and best practices that could become quickly outdated. Instead, we ask that you answer a series of questions that address specific aspects of creating and managing digital assets. Your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions

If you propose to create any type of digital product as part of your project, complete this form. We define digital products very broadly. If you are developing anything through the use of information technology (e.g., digital collections, web resources, metadata, software, or data), you should complete this form.

Please indicate which of the following digital products you will create or collect during your project
(Check all that apply):

	Every proposal creating a digital product should complete	Part I
	If your project will create or collect	Then you should complete
<input type="checkbox"/>	Digital content	Part II
<input checked="" type="checkbox"/>	Software (systems, tools, apps, etc.)	Part III
<input checked="" type="checkbox"/>	Dataset	Part IV

PART I.

A. Intellectual Property Rights and Permissions

We expect applicants to make federally funded work products widely available and usable through strategies such as publishing in open-access journals, depositing works in institutional or discipline-based repositories, and using non-restrictive licenses such as a Creative Commons license.

A.1 What will be the intellectual property status of the content, software, or datasets you intend to create? Who will hold the copyright? Will you assign a Creative Commons license (<http://us.creativecommons.org>) to the content? If so, which license will it be? If it is software, what open source license will you use (e.g., BSD, GNU, MIT)? Explain and justify your licensing selections.

We have opted out of copyright protection, dedicating all PeriodO software and data to the public domain, and waiving all of our rights to the work under copyright law, to the extent allowed by law. Anyone can copy, modify, and distribute PeriodO software and data, even for commercial purposes, without asking our permission. We clearly assert the dedication of our work to the public domain by including and linking to the the CC0 1.0 Universal (CC0 1.0) Public Domain Dedication from our datasets and software repositories. This makes it maximally easy for others to use and build upon our work without being hampered by the complexities of intellectual property law.

A.2 What ownership rights will your organization assert over the new digital content, software, or datasets and what conditions will you impose on access and use? Explain any terms of access and conditions of use, why they are justifiable, and how you will notify potential users about relevant terms or conditions.

We do not assert any ownership rights over the PeriodO data or software, nor do we impose any conditions on access and use.

A.3 Will you create any content or products which may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities? If so, please describe the issues and how you plan to address them.

We do not plan to create any content or products that involve privacy concerns. When our partners and users contribute data, we make it clear to them that they are placing their contributions in the public domain. We do not foresee any issues regarding cultural sensitivities.

Part II: Projects Creating or Collecting Digital Content

A. Creating New Digital Content

A.1 Describe the digital content you will create and/or collect, the quantities of each type, and format you will use.

A.2 List the equipment, software, and supplies that you will use to create the content or the name of the service provider who will perform the work.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to create, along with the relevant information on the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

B. Digital Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance (e.g., storage systems, shared repositories, technical documentation, migration planning, commitment of organizational funding for these purposes). Please note: You may charge the Federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the Federal award. (See 2 CFR 200.461).

C. Metadata

C.1 Describe how you will produce metadata (e.g., technical, descriptive, administrative, or preservation). Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, or PREMIS) and metadata content (e.g., thesauri).

C.2 Explain your strategy for preserving and maintaining metadata created and/or collected during and after the award period of performance.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of digital content created during your project (e.g., an API (Application Programming Interface), contributions to the Digital Public Library of America (DPLA) or other digital platform, or other support to allow batch queries and retrieval of metadata).

D. Access and Use

D.1 Describe how you will make the digital content available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

D.2 Provide the name and URL(s) (Uniform Resource Locator) for any examples of previous digital collections or content your organization has created.

Part III. Projects Creating Software (systems, tools, apps, etc.)

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) this software will serve.

Additional visualization and search capabilities in the PeriodO client, for end-users trying to find and compare period definitions. An API (usable with data-improvement tools such as OpenRefine) that will allow the PeriodO dataset to be used for mostly-automatic alignment and reconciliation of period terms found in datasets. Cataloguer toolkit that will allow a data manager to reconcile a period vocabulary on the basis of overlaps in label, time, and space. Prototype tool for the statistical matching of period definitions with undated period terms used in databases or publications.

A.2 List other existing software that wholly or partially perform the same functions, and explain how the tool or system you will create is different.

We are unaware of any existing software that can perform these functions.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software (systems, tools, apps, etc.) and explain why you chose them.

The PeriodO client is implemented in JavaScript using the React framework. The PeriodO server is implemented using Python, RDFLib, and the Flask framework. We use SQLite to store data on the server, and IndexedDB to store data in the browser. The planned visualizations will be implemented using the D3 framework. We use these tools because they are widely used, well-documented, actively developed, and we are familiar with them.

B.2 Describe how the intended software will extend or interoperate with other existing software.

The PeriodO client currently interoperates with WorldCat and Crossref to help users manage bibliographic metadata, and with DBpedia to help users specify geographic regions. We plan to extend this to support additional Linked Data gazetteers. Finally, as a Linked Data application using standard non-proprietary vocabularies, any software that can process RDF is interoperable with PeriodO.

B.3 Describe any underlying additional software or system dependencies necessary to run the new software you will create.

The PeriodO server can run on any platform supporting Python 3 and SQLite. The PeriodO client runs in any modern web browser, although there are some bugs in Safari that prevent it from being fully functional.

B.4 Describe the processes you will use for development documentation and for maintaining and updating technical documentation for users of the software.

Development documentation is checked into our GitHub repositories alongside our software. Technical documentation for users is kept up-to-date on our website (<http://perio.do>).

B.5 Provide the name and URL(s) for examples of any previous software tools or systems your organization has created.

Source code: <https://github.com/periodo>

Use the PeriodO client: <http://n2t.net/ark:/99152/p0>

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under an open-source license to maximize access and promote reuse. What ownership rights will your organization assert over the software created, and what conditions will you impose on the access and use of this product? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain any prohibitive terms or conditions of use or access, explain why these terms or conditions are justifiable, and explain how you will notify potential users of the software or system.

We use the CC0 1.0 Universal (CC0 1.0) Public Domain Dedication to waive all of our rights to the work under copyright law, to the extent allowed by law. Anyone can copy, modify, and distribute PeriodO software, even for commercial purposes, without asking our permission. This makes it maximally easy for others to use and build upon our work without being hampered by the complexities of intellectual property law.

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

The PeriodO client will continue to be available on the web at <http://n2t.net/ark:/99152/p0>, and all the additional tools we plan to build will also be openly accessible on the web. We will continue to make all of our source code available through GitHub.

C.3 Identify where you will be publicly depositing source code for the software developed:

Name of publicly accessible source code repository: PeriodO
URL: <https://github.com/periodo>

Part IV. Projects Creating a Dataset

1. Summarize the intended purpose of this data, the type of data to be collected or generated, the method for collection or generation, the approximate dates or frequency when the data will be generated or collected, and the intended use of the data collected.

Our data is a gazetteer of unique period definitions, each consisting of a temporal coverage, a spatial coverage, and an authoritative source. By collecting definitions that include these elements, we seek to allow the chronological cross-searching of disparate data sources that use period terms.

2. Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

We do not collect or plan to collect any personal data or do anything else that requires IRB approval.

3. Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

We do not collect or plan to collect any personally identifiable information, confidential information, or proprietary information.

4. If you will collect additional documentation such as consent agreements along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

N/A

5. What will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

We use the PeriodO client to collect data entered by end-users. Occasionally we will bypass manual entry and write a script to import a large amount of structured data from a partner, but this is relatively rare. Only a modern web browser is required to retrieve, display, and edit the PeriodO data.

6. What documentation (e.g., data documentation, codebooks, etc.) will you capture or create along with the dataset(s)? Where will the documentation be stored, and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

Human-readable documentation of the data is available on the PeriodO website (<http://perio.do>), and in the PeriodO client. The data is also documented in a machine-readable way using a Vocabulary of Interlinked Datasets (VOID) description file, available at <http://n2t.net/ark:/99152/p0/.well-known/void>.

7. What is the plan for archiving, managing, and disseminating data after the completion of the award-funded project?

PeriodO data will be periodically snapshotted and archived in the California Digital Library's Merritt repository, the UT Libraries, and the Carolina Digital Repository. The UT Libraries have agreed to host the PeriodO server for the long term.

8. Identify where you will be publicly depositing dataset(s):

Name of repository: Texas ScholarWorks
URL: <https://repositories.lib.utexas.edu/>

9. When and how frequently will you review this data management plan? How will the implementation be monitored?

We will review this data management plan annually. UT Libraries will monitor the PeriodO server to ensure that it is accessible and secure.

Original Preliminary Proposal

Periods, Organized (PeriodO) 2: Linking, Discovering, and Reconciling Information about the Past

The way human beings talk about past time creates problems for the digital description of historical information. Computers can parse dates into a common representation so that they can be easily compared, but they can't parse expressions like "the long 18th century" or "the Archaic period." The same is true for geographical spaces – but where people can generally agree that "New York City" is a distinct entity that can be located within coordinate space, they don't necessarily agree on the boundaries of "the long 18th century" either in time or in space. Local metadata systems define historical, art-historical, archaeological, and other periods in idiosyncratic ways that can be difficult to align across systems. Shared metadata systems like the Getty Art and Architecture Thesaurus define periods generally, but in the process sacrifice local specificity. Neither approach makes it easy for a user to find information associated with a particular period term or a particular date range across a range of library or museum records. The PeriodO project (<http://perio.do>) seeks to solve this problem through the creation of an open-ended Linked Data gazetteer of period definitions that are authoritative and clearly modeled, while avoiding the establishment of a "canonical" periodization that erases local differences, scholarly disagreements, alternative voices and disciplinary histories.

We seek a **Project Grant** in the amount of **\$240,120** from the IMLS to help us expand the project to a more ambitious scale, increasing its usefulness for a broader range of libraries, museums, and cultural institutions and contributing to the National Digital Platform. In this phase of the project, we will focus on specific tools and services to improve linking, discoverability, and the alignment and reconciliation of periodized information, while continuing to build partnerships and expand the dataset in time, space, and areas of research. The project already has strong partnerships within the cultural heritage sphere, and has been adopted as a key piece of infrastructure for the EU-based ARIADNE project, which seeks to integrate and harmonize archaeological datasets from a large number of contributors in different EU nations. We are currently in discussions with both the DPLA and Europeana about how Linked Data period definitions might improve discoverability for the diverse information both are collecting.

The first stage of this work has been carried out with support from the NEH Office of Digital Humanities. Over the last 15 months, the PeriodO team has built a platform for version-controlled publishing of period definitions as Linked Data, constructed a user interface for collaborative editing and management of these definitions, and used partner-contributed data from a range of projects and institutions to assemble almost 3000 period definitions covering much of the world from the Paleolithic to the present (data and interface at <http://n2t.net/ark:/99152/p0>). PeriodO already features a stable and permanent system for the generation of unique, persistent identifiers for period definitions using ARK IDs issued by the California Digital Library, we have an institutional home for the client server with iBiblio at the University of North Carolina, and we are making arrangements to house copies of the dataset in several different institutional repositories, including those at the University of Texas and the University of North Carolina. Both the client interface and the dataset are open source and freely available under a CC0 license (<https://github.com/periodo>).

Project Directors. The directors of the project will be Adam Rabinowitz, Associate Professor of Classics at The University of Texas at Austin; Ryan Shaw, Assistant Professor in the School of Library and Information Science at the University of North Carolina, Chapel Hill; and Lorraine Haricombe, Vice Provost and Director of Libraries at The University of Texas at Austin. Rabinowitz and Shaw, the co-PIs of the NEH-funded project, bring complementary expertise to the project. Rabinowitz is a Classical Archaeologist with extensive experience in the past cultures of Europe and the Mediterranean, as well as in the disciplinary history of archaeology. He has also been involved in several projects focused on the pedagogical use of geotemporal visualization. Shaw is an information scientist who works with ontologies, Knowledge Organization Systems, and the modeling of historical events and periods, and has participated in a number of Digital Humanities initiatives. Haricombe is a distinguished director of research library systems, whose work has centered on issues of open access to information; she is a co-founder of the Coalition of Open Access Policy Institutions. Other members of the team will include Patrick Golden, a PhD student at the University of North Carolina who built the current version of the PeriodO client, and Eric Kansa, the director of Open Context. We will also to continue

to draw on the expertise of members of our current advisory board, and we plan to expand that board for this phase to include representatives of libraries and major data-integration initiatives.

Proposed Workplan. We seek support for a two-year project to begin on June 1, 2016 and to end on May 31, 2018. During this time, we will work to expand and improve the digital platform. We will also hold two workshops to introduce the gazetteer to potential users and collect feedback. On the platform side, we will build a powerful, intuitive interface that will allow search and visualization of period definitions from a number of different angles, including maps, timelines, and graphic representations of scholarly consensus, disagreement, or change over time. We will also create tools and/or an API to facilitate the use of PeriodO period definitions by projects aggregating data from diverse sources; this will include a reconciliation service that will make it easy to identify period definitions according to user-specified parameters (string, date range, spatial coverage, authority, etc.) and align those with external defined vocabularies. Finally, we will explore probabilistic approaches that may make it possible, using the PeriodO dataset, to suggest spatial and/or temporal coordinates for period terms that appear as natural language in sources that do not offer definitions. The visualization interface will be completed during the first year; the reconciliation service will be developed over both years, with extensive feedback from partners; and probabilistic approaches will be investigated in the second year. We will hold one workshop in each year, with one in Austin and one in Chapel Hill. These workshops will bring together senior staff from libraries and data integration projects both in the US and abroad to explore the ways in which PeriodO can be useful to them, to refine our tools and services through user testing and feedback, and to identify needs or challenges we have not already considered.

Relevance to IMLS Priorities. We propose this project as a contribution to the National Digital Platform. PeriodO aligns itself naturally with a number of NDP priorities: with ARIADNE, it has already demonstrated its ability to deploy Linked Data to facilitate collaboration and connect content across institutions and datasets; it is built to encourage crowdsourcing and public participation (new definitions can be contributed by registered users through a patch submission process, and integrated into the dataset after editorial review); and it is expressly designed both to increase the discoverability of periodized data for non-specialist audiences and to help those audiences understand the complex, constructed nature of historical periodization itself. The project's international partners will also allow it to contribute to NDP goals of international collaboration and integration.

Potential Impact. PeriodO 2 is positioned to have a very significant impact on public access to information that is currently periodized but not explicitly dated, especially in aggregated collections like that of the DPLA. It will make it easier for these collections to model and align the records they ingest without insisting on the use of specific vocabularies, and it will make it much easier for their users to find items from particular places and/or date ranges. As the dataset and its implementation by partners expands, it will also provide incomparable access to data on period *usage*, which will eventually enable machine-learning and natural language processing to extract date information from any digitized text that uses a period term – and show that coverage on a timeline.

Project Goals and Outcomes. The project's two goals are the creation of additional tools and services to facilitate the exploration and use of PeriodO data, and the expansion of partnerships that would put PeriodO definitions to work to improve the description, alignment, and linking of periodized information across data sources. Our expected outcomes are a new, powerful and intuitive visualization interface for the data, new services and tools for alignment and reconciliation, a larger and more diverse dataset that encompasses a wider range of disciplines and organizations, and the broad adoption of PeriodO as a gazetteer for periods in Linked Data projects along the same lines as the Getty Thesauri or the Library of Congress Subject Headings.

Budget. We request **\$240,120** from IMLS. This sum includes ██████ in stipends for two programmers (Golden at ██████/year; TBD at ██████/year); ██████ for summer salary and fringe for Rabinowitz (3 weeks in each of two years); ██████ for consultant stipends (Shaw, Kansa, and TBD for alignment); ██████ for graduate student salary and fringe (data entry); \$4,000 for participant travel to present project at conferences/workshops (domestic and international); \$37,500 to hold two workshops in Austin and Chapel Hill (travel and lodging for 5 domestic and 5 international participants at each); and \$66,120 indirect cost at a 38% federally negotiated rate.