White Paper

Report ID: 111875

Application Number: HD-51864-14

Project Director: Adam T. Rabinowitz

Institution: University of Texas, Austin

Reporting Period: 3/1/2014-12/31/2015

Report Due: 3/31/2016

Date Submitted: 4/4/2016

NEH Project White Paper
Grant HD-51864-14

Periods, Organized (PeriodO): a gazetteer of period assertions for linking and visualizing periodized data

Project Directors:

Adam Rabinowitz

Ryan Shaw

Grantee Institution:

The University of Texas at Austin

Report Contributors:

Adam Rabinowitz, Ryan Shaw, Patrick Golden, Sarah Buchanan, Eric Kansa

Submitted: April 4, 2016

Origins and Objectives

The PeriodO project arose from the recognition of a gap in the infrastructure for Linked Data. Linked Data relies on the use of persistent unique identifiers for concepts, as well as resources, and the description of those concepts using structured data. Opinions vary regarding how that data should be structured, but one option that has found wide acceptance among creators of scholarly and scientific Linked Data is the RDF (Resource Description Framework) subject-predicate-object "triple" (Berners-Lee 2009; Heath and Bizer 2011; W3C Consortium 2015). When combined in a "triple store", data from and relations between concept identifiers in heterogeneous datasets can be searched together. Persistent unique identifiers in the form of Uniform Resource Identifiers (URIs) provide shared reference points for the individual members of particular groups of entities (such as people, places, biological taxa, etc.) or any groupings of these. Collections of structured data about such entities that provide persistent unique identifiers for each entity are sometimes referred to as "gazetteers" (which typically organize collections of places). Gazetteers act as phone books, allowing both human users and computers to look up and find identifiers for an entity, and to use an identifier to find descriptions of the entity it identifies. Because an identifier refers to an entity, not to a particular name, gazetteers can collect multiple names for the same entity, allowing a user to connect datasets that might refer to the same entity by different names (for example, "Mark Twain" vs. "Samuel Clemens").

Linked Data gazetteers have been very successful in creating an infrastructure that allows both connections between heterogeneous datasets and the disambiguation and digital visualization of named entities in texts. For example, a software program like the Edinburgh Geoparser (https://www.ltg.ed.ac.uk/software/geoparser/) can look in a text for strings of characters that match strings in the names of geographic places in a gazetteer, and then populate a map of the places in the text using spatial coordinates for those places drawn from that gazetteer. These gazetteers work best with concrete entities like people and places, or with abstract concepts about which communities can agree, at least on a broad level. Concepts that communities disagree about, however, present greater challenges – and periods in the past are a prime example.

The way human beings talk about past time creates problems for the digital description of historical information. Computer programs 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." 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 many terms in the period vocabulary of the Getty Art and Architecture Thesaurus). 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. And although an ontological frame for the standardization of heterogeneous thesauri of period concepts has been circulating for some time (see Doerr, Kritsotaki and Stead 2010), an ecosystem centered on this ontological approach has

so far failed to emerge. There is therefore currently no universally recognized gazetteer or authority (or set of authorities) that identifies, with temporal and geographic precision, a comprehensive, multilingual, multidisciplinary set of periods.

The PeriodO project 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 than from above by standardization of terms, concepts, or metadata schemata. These shared reference points are provided by a gazetteer, the entries in which are expressed according to a clear semantic model and given persistent URIs. Second, instead of identifying unitary period *concepts* (as for example the Getty AAT does, and as the CIDOC-CRM model insists) 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 gathering together and providing unique identifiers for 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.

PeriodO is thus meant to provide 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, archives, and data managers to describe their periodized data with explicit statements of spatio-temporal coverage, either drawn from existing authorities or minted to match local usage. We hope that this will facilitate cross-searching and interoperability among periodized datasets, 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.

Partners and Data Sources

The PeriodO team is especially grateful to its partners and data providers, both those who signed on to the project in its conceptual stage and those who contributed period definitions after the project began. Without their contributions – often the fruit of years of careful thought about archaeological periodization, and reflecting a tremendous investment of intellectual energy – PeriodO would still only be an idea. Our original partners were the GeoDia interface at The University of Texas at Austin, directed by PI Rabinowitz (http://geodia.laits.utexas.edu); the archaeological data-publication platform Open Context, directed by Eric Kansa, and through it the Digital Index of North American Archaeology, directed by David Anderson and Josh Wells (http://opencontext.org; http://ux.opencontext.org/archaeology-site-data/); the British Museum and the Portable Antiquities Scheme, with the help of Dan Pett (https://finds.org.uk/); the SENESCHAL project, led by Doug Tudhope and Ceri Binding (http://www.heritagedata.org/blog/about-heritage-data/seneschal/); the Pleiades project, (http://pleiades.stoa.org); the Pelagios project, directed by Elton Barker and Leif Isaksen (http://pelagios-project.blogspot.co.uk/); the ArcheoInf project directed by Johannes Bergemann (http://www.ub.tu-dortmund.de/archeoinf/); the CLAROS project at the University of Oxford (http://www.clarosnet.org/XDB/ASP/claroshome/index.html); Fasti Online, a project of the Associazione Internazionale di Archeologia Classica managed by L – P : Archaeology, under the supervision of Guy Hunt, Stuart Eve, and Jessica Ogden (http://fastionline.org/); Arachne, the object database of the Deutsches Archäologisches Institut, managed by Reinhard Foertsch (http://arachne.uni-koeln.de/drupal/); the Archaeology Data Service (ADS) of the UK, directed by Julian Richards (http://archaeologydataservice.ac.uk/); and the UCLA Encyclopedia of Egyptology, under the direction of Willeke Wendrich (http://uee.ucla.edu/).

Of those original partners, **GeoDia** provided about 700 period definitions with coordinates in both space and time, and formed the initial core of the PeriodO gazetteer. **Open Context** provided 522 period definitions drawn from contributors to the Digital Index of North American Archaeology, and began to implement PeriodO URIs in its own datastructure as envisioned by the original proposal (see, for example, http://opencontext.org/subjects/52FB12D0-09CD-453E-91F3-58A9A8B724CF). **The British Museum** provided an XML document containing all of the period terms in its internal vocabularies; not all of these had all the necessary information, but we were able to include 944 definitions, to which we added 31 period definitions related to the UK from the **Portable Antiquities Scheme**. The PAS has also now implemented PeriodO URIs in its own dataset (see, for example, RDF representation here: https://finds.org.uk/database/artefacts/record/id/741974/format/rdf). The SENESCHAL project provided another 42 definitions from English Heritage, among which were several that were reused in the PAS, and the **Archaeology Data Service** provided both 48 definitions and a dump of periodized data that we plan to use in the next phase of the project to explore period term usage. **Pleiades** offered us the interesting opportunity to determine spatial coverage on the basis of the geographic coordinates of sites where the 116 terms in its period vocabulary were applied, rather than by verbal assertion. This differed from the contribution of **Fasti Online**, which defined the spatial extent of its 212 period terms by geographic polygons mapped onto modern

national boundaries. The **UCLA Encyclopedia of Archaeology** turned out to be in the process of reworking its period definitions, so we had to be content in this case with an older set of 35 (but these are the definitions used by several other platforms, so even if they change, it will be useful to have documented the earlier versions).

We were not able to include contributions from all of our partners. **ArcheoInf** contributed its controlled vocabulary for periods, but the form in which we received this vocabulary did not include dates or statements of spatial coverage, so we could not integrate it with the PeriodO dataset. A similar problem arose with the period vocabulary of **Arachne**, the database of the Deutsches Archäologisches Institut, and this was compounded by the fact that Arachne is working on an entirely new period-management system ("ChronOntology"), which is not yet complete. They have expressed a desire to cooperate in a more concrete manner in the future. We did not receive any period definitions from the Oxford-based **CLAROS** project, in part, we think, because that platform expresses periods entirely in terms of date ranges, without period terms or spatial designations. Direct collaboration with **Pelagios** did not materialize during this grant term, although one of its PIs was on our board, but this is largely because we were not able to reach a point where Pelagios could begin to incorporate PeriodO data. We have been in close touch with them about a visual browser platform (Peripleo) that they are developing, however, and we think that we will be able to implement a more direct collaboration in the next phase of the project.

The issues with this group of partners were varied. The biggest challenges had to do with the compatibility of data models for dated/periodized data: some of our partners managed this information in ways that were incompatible with our data model, and others had suitable data but were not able to convert their periodizations into a format that would allow us to incorporate them easily in PeriodO. We might have mitigated some of these challenges if we had been able to start the project with a more clearly-defined and stable data model, but part of our strategy was to adapt the model on the basis of feedback as we went, so this was largely unavoidable (although we could have anticipated some of these incompatibilities better). Also unavoidable – and not always anticipatable – were problems with timing. In some cases, projects had not reached the necessary level of development by the end of our grant term; in other cases, we were not able to develop reconciliation and bulk ingest tools necessary for some partner datasets by the end of that term.

On the other hand, we managed to recruit a number of willing new collaborators, who provided us with more than a thousand additional period definitions: the Spanish **Institute of Heritage Sciences**, the **Levantine Ceramics Project**, the Dutch **Rieksdienst voor het Cultureel Erfgoed**, the Swedish **Historiska Museet**, the **China Historical GIS**, and, most importantly, the EU-based **ARIADNE** project for archaeological data integration, which supplied us with 659 additional curated period definitions from more than a dozen countries and also undertook to use PeriodO URIs to manage period vocabularies for all of its contributors (see http://www.ariadne-infrastructure.eu/Resources/PeriodO). We are also currently in negotiation with the **Archnet** project to include period definitions from the Aga Khan Documentation Center, and we have established a connection with the NEH-funded **Project Andvari**. Finally, the **Agora Excavations** of the American School of Classical Studies in Athens and **the Digital Archaeological Record (tDAR)** have both given us test datasets containing a combination of

period terms and absolute dates that we hope to use in the next phase to compare definitions with absolutely-dated archaeological material.

Project Accomplishments

Our proposal to the NEH envisioned the creation of a Linked Data gazetteer of period definitions provided by authoritative sources, which would allow curators of periodized data to make transparent, machine-readable statements about the temporal and geographic boundaries of the period terms they used. The PeriodO gazetteer was to include a series of period definitions that included start and end dates, statements of spatial coverage, attribution to an authority, and persistent unique identifiers minted through the EZID system of the California Digital Library. These definitions would be input through a browser-based graphic user interface, with ongoing contributions managed through a patch-submission process, and accessed both as human-readable records in the graphic user interface and as a machine-readable serialization in JSON-LD. In our proposal, the graphic user interface would include visualization tools such as a map and a timeline to enable both professional and lay users to search and browse period records. The original proposal also envisioned that the dataset of definitions would be managed in GitHub, which would allow us to maintain a record of revisions and changes. Both data-contributing partners and an Advisory Board were assembled, with the expectation that partners would provide their period thesauri or definitions in a form that could be adapted to PeriodO, and that the Advisory Board would guide and evaluate the development of the platform across a series of teleconferences.

In keeping with that plan, our primary activities over the course of the grant included:

- refining and finalizing the period definition model proposed for our dataset, and representing that model as properly formed JSON-LD, Terse RDF Triple Language (Turtle), and CSV
- programming the user interface, browser client, and server architecture to permit the display, filtering, editing, and contribution of period definitions
- ingesting the period definitions from our initial set of contributors into the dataset on the server, and documenting the process
- using the user interface and patch submission process to add new period collections and definitions from both new contributors and published sources
- in conjunction with the development of the patch submission process, implementing a provenance-documentation model, also expressed as JSON-LD, to be maintained on the server to track and attribute the editing of existing definitions and the addition of new ones
- establishing a method for the minting of persistent, globally unique URIs through the California Digital Library EZID service
- holding four advisory-board teleconference calls via Skype to solicit feedback, advice, and comments on the developing platform and interface
- working with Open Context to begin implementation of PeriodO URIs in a selection of Open Context records

- presenting the project in various venues (conferences, workshops, etc.) and preparing articles on the project for publication in peer-reviewed outlets (see "Online material, presentations, and publications", below)
- in the final months of the grant, preparing proposals for next-phase funding

Our application envisioned the creation of a Linked Data gazetteer of period definitions from authoritative sources that were explicit in their statements of chronological and geographic coverage, "working tools for organizing, publishing, maintaining, and visualizing" those definitions, and a community of practice among individuals and groups concerned with the creation and management of periodized data, especially in Classical studies. Of these goals, only the full extent of the visualization platform was not achieved. We did not set quantitative goals for the PeriodO dataset in our proposal, but if we had, we would certainly have exceeded them: in our internal conversations at the beginning of the project, we envisioned a dataset of perhaps 2000 entries. By the end of the grant term, we had nearly twice as many.

The gazetteer we created is fully modeled and usable; it can be found at http://n2t.net/ark:/99152/p0 (see brief user guide below). The canonical dataset currently contains 3,672 authoritative period definitions with coordinates in space and time, from 80 sources in 17 languages. These definitions go well beyond our original focus on the Classical World: we have hundreds of definitions of North American periods from the Digital Index of North American Archaeology, more than 50 definitions of Chinese periods from the China Historical GIS at Harvard, and dozens of definitions that span the globe from the period vocabulary of the British Museum. There are some geographical areas where our coverage is still thin – for instance, Central America and sub-Saharan Africa – but we are working to fill in those gaps, and this is one of the explicit components of our second-phase grant proposals, along with the intent to expand coverage deeper into the past, both in terms of periods (e.g. geological definitions from the International Commission on Stratigraphy) and in terms of sources (we plan to gather period definitions from 18th and 19th-century published works to permit users to trace the disciplinary use of period terms across time).

The community of participants we proposed is solid and rapidly expanding: we have strengthened our connections with most of our original partners, and two of those partners (Open Context and PAS) have begun to implement PeriodO URIs in their own data; we have found new partners, most importantly the EU-funded ARIADNE initiative for archaeological data harmonization; and we have begun to attract attention and collaboration beyond our original circle, playing important roles in existing and proposed projects with focuses ranging from Mesopotamian prosopography to hominin fossils. We have also begun to receive inquiries from individual data managers, especially on the Linked Data side, who are helping us to test the robustness of our user contribution system.

Finally, we stated in our proposal that one of our project goals was the development of funding proposals that would bring us beyond the start-up stage and allow us to develop the complex reconciliation tools and web-services necessary to make PeriodO a more useful resource for data-managers. We also indicated that we would seek support for the project within our own institutions. We have accomplished these objectives as well: in January and February, we submitted proposals for the next stage of PeriodO to both the IMLS (National Leadership Grants for Libraries) and the NEH (Digital Humanities Implementation Grants), and in the context of

these proposals, we established collaborations with and long-term commitments from the University of Texas Libraries and the iBiblio project at the University of North Carolina.

Modifications and Lessons Learned

Like most digital projects, PeriodO underwent some modifications in the course of its development, and like most digital project managers, we were somewhat too optimistic about the amount of software and feature development we could accomplish within the time and budget we had available. Modifications to the data model, workflow, and platform were minimal, however, and with the help of our advisory board, we were able to establish clear priorities and well-defined work packages that could be deferred to a second project phase without harm to our objectives in this phase.

On the technical side, the development of PeriodO largely followed the path laid out by the original proposal. There were only a few relatively minor technical changes that did not affect the scope or functionality of the project. Probably the most significant of these was the decision to abandon our original plan to use Julian Day notation to express dates and levels of uncertainty. As we prepared our initial data for inclusion, it became clear that the notation we had proposed would make assumptions about the degree of uncertainty in a definition that were not specified by the definition's source (for example, for dates beginning with "circa"). Since a core principle of our data model was not to add information to the assertions we documented, we decided to take a more neutral approach and to use ISO8601 date notation (see Golden and Shaw 2016). This was also less complicated on a computational level, and made it easier and faster for us to finalize the input interface. The variety of date notations used in our sources made it difficult to build a parser that could handle formats beyond the most common (e.g. BC, BCE, "6th century", "beginning of...", etc.), so it was important to make it possible for users to enter dates manually – but we could not expect most users to be able to convert standard formats to Julian Day notation.

On the level of infrastructure, we realized at the beginning of the project that our plan to manage patches to the canonical dataset through the GitHub interface was not going to provide the level of documentation and control that we needed. Therefore, although we continued to use GitHub as the repository for both project code and an updated copy of the dataset, and as a critical part of our workflow for issue tracking and code patching, we decided to manage the active dataset and user contributions through a server we run ourselves. In the next phase of the project, if not before, we will move this piece of infrastructure to more permanent hosting at either the University of Texas Libraries or the iBiblio project at the University of North Carolina (or perhaps both, with the main server at one and a mirror at the other). This will provide a more stable and secure long-term home for the dataset. At the same time, during the current phase of the project, having our own server allowed us to create and implement a comprehensive schema to record the provenance of new submissions to the canonical dataset: who submitted patches, who accepted or rejected patches, what additions a patch contained, and, if it also contained edits, what material had been edited and what edits had been made. We consider this to be a critical improvement to our model of transparency and attribution, which more than makes up for the burden of maintaining a separate server outside GitHub.

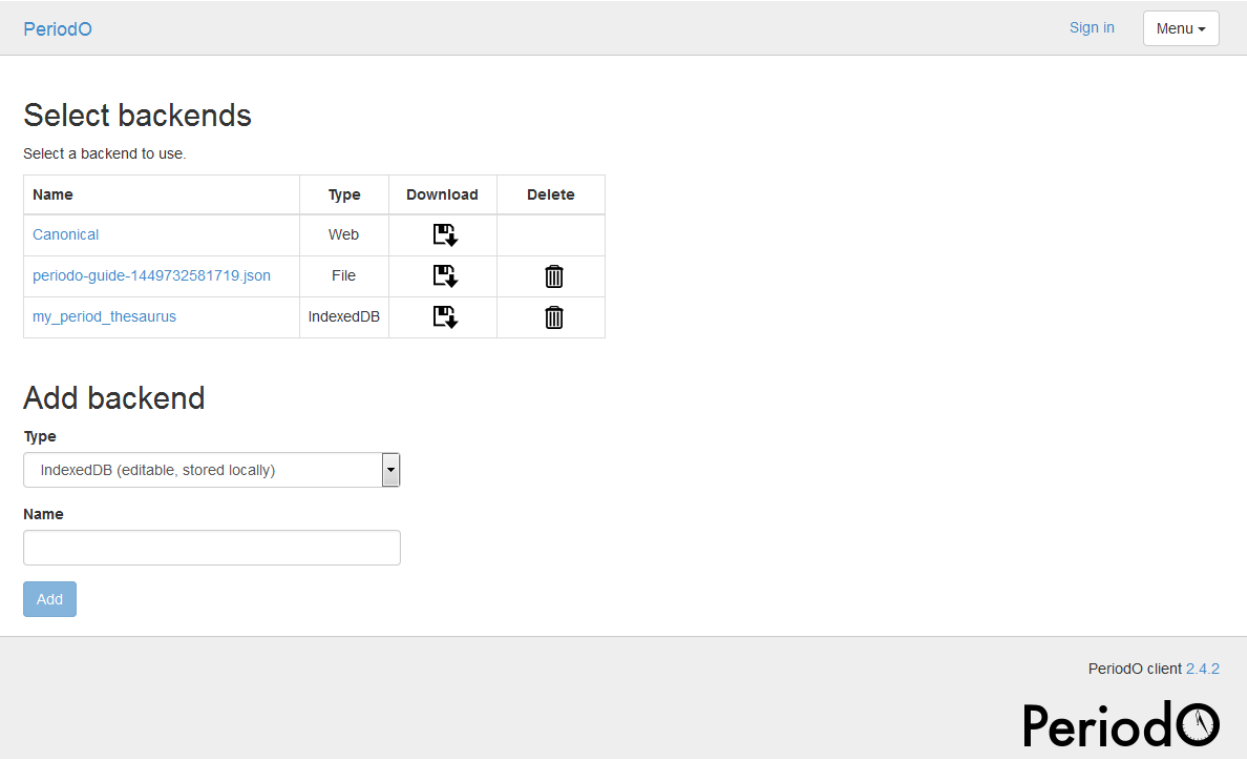
A last minor change was our decision to use ARK IDs with suffix pass-through to mint our unique identifiers for period collections and definitions. We had originally planned to use DOIs, but this would have involved a transaction with the EZID system for each new identifier minted – and since we are envisioning a dataset that may have tens of thousands of definitions, this would have become costly on both financial and computational levels. Instead, we minted a single ARK ID through EZID for the dataset as a whole, and now use the suffix pass-through property of the ARK format to create individual URIs for collections and definitions. Using the California Digital Library's Name to Thing Resolver, one can resolve a PeriodO URI through EZID to a visual representation of the entity identified by that URI.

The most substantial modification to our plans concerned the visualization tools we proposed to provide. In our original proposal, we indicated that we would build a graphic user interface for searching and browsing period definitions that would include a combination of maps, timelines, and faceted search. In the execution of the project, however, we focused on the construction of an intuitive and functional user interface for the input and management of period definitions, which turned out to be a somewhat more time-consuming task than we had originally assumed. The development of intuitive, browser-based visual data displays is complex, involved, and dependent on the underlying data architecture, and it made more sense for us to put advanced visualization work aside until we had created a solid framework for the management of the dataset itself. While we were successful in the creation of an interface that permits the faceted browsing of period definitions and simple text- and timeline-based filtering, we decided, with the agreement of our Advisory Board, to defer the creation of more elaborate JavaScript data visualizations to a later stage of the project. These visualizations and filtering options will be increasingly important as both the dataset and the user-base grow, but we felt that the creation of a robust and solid infrastructure, including a smooth patch-submission and provenance-documentation process, was our highest priority. The development of user-friendly visualization tools is one of the central components of our plan for the platform's next phase.

For similar reasons, we have deferred the issue of reconciliation with external datasets. We originally proposed to hold a meeting at the end of this phase of the project to begin a discussion of the development of reconciliation services. Because it seemed to us by the end of the project that this would be a discussion best carried out with a wider range of new partners, however, we decided to defer it to the next phase of the project, which includes an early-stage workshop to discuss the needs of PeriodO consumers. This has also effectively deferred our plans for the quantitative evaluation of the platform by users, although we can already measure success by the expansion of the user community and the number and diversity of definitions in the dataset. Reconciliation tools and evaluation are interdependent, since, with the exception of a few of our partners (Open Context, Portable Antiquities Scheme, ARIADNE), most data managers who will be interested in using PeriodO are likely to begin to do so only when there are tools that make it easier to apply PeriodO URIs across an external dataset. Thus, although PeriodO URIs have been partially implemented in Open Context, user evaluation on the level of implementation seemed premature. We have therefore put the development of reconciliation tools at the center of our proposals for the project's next phase, and we will employ a comprehensive evaluation strategy that includes information about the number of visitors to the site, the adoption of PeriodO URIs in external databases, and a selective user survey, beginning with our partners. In the meantime, visitors can judge the achievements of the project for themselves by engaging with the dataset. The following section provides a brief guide to the PeriodO client interface.

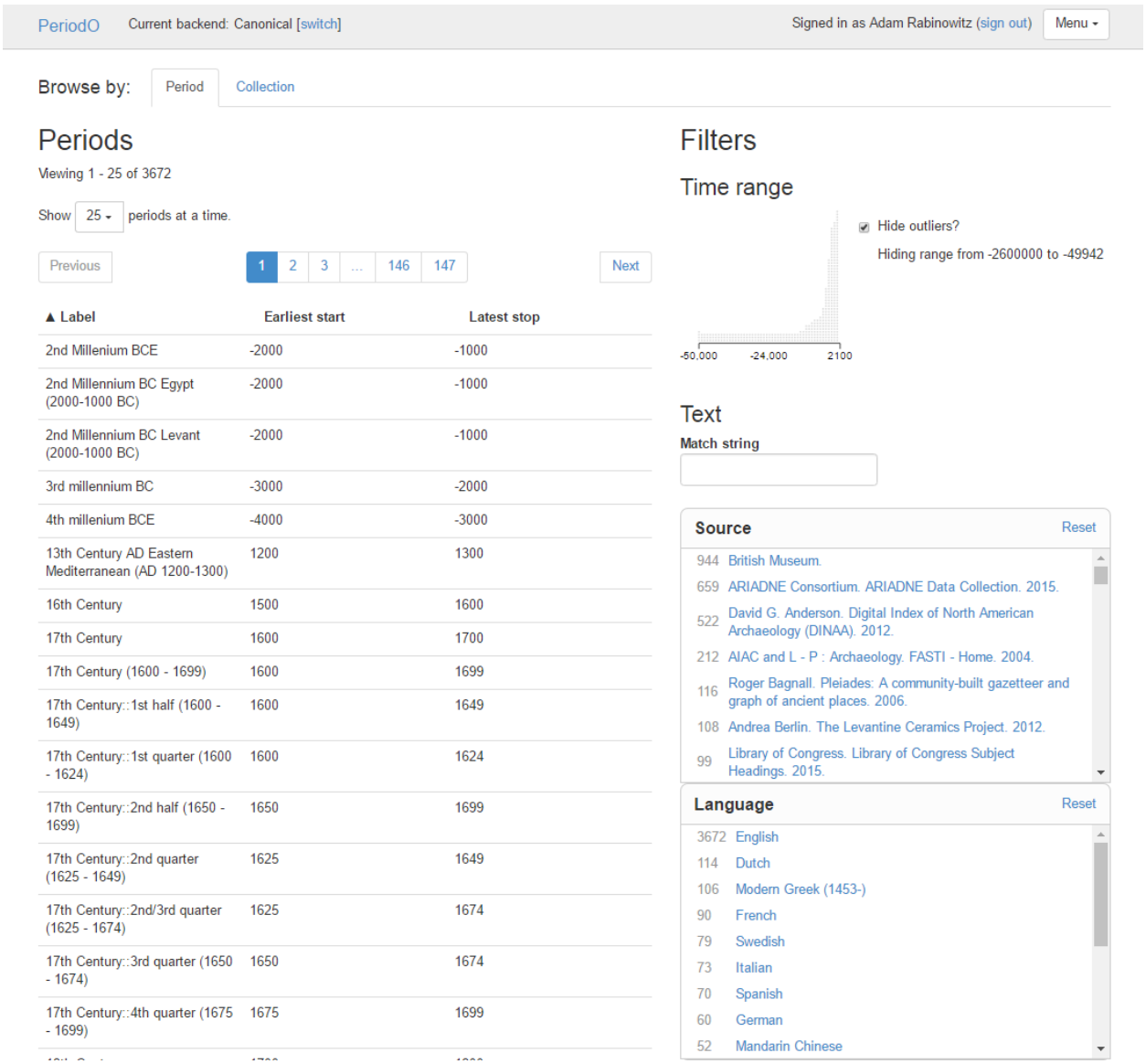
Using the PeriodO Client v. 2.4.2

The PeriodO client (available at http://n2t.net/ark:/99152/p0) provides a user interface that allows visitors to interact with the underlying datastore of period information. It is organized around two classes of data: period *definitions*, which are the expressions of the combination of temporal coverage, spatial coverage, and authoritative source specified above; and period *collections*, which represent a group of period definitions provided by the same authoritative source. A more complete and illustrated user guide is currently under development at the PeriodO website (http://perio.do/guide/), and technical information can be found at the project's GitHub repository (https://github.com/periodo), but this section provides a brief orientation for the casual visitor.



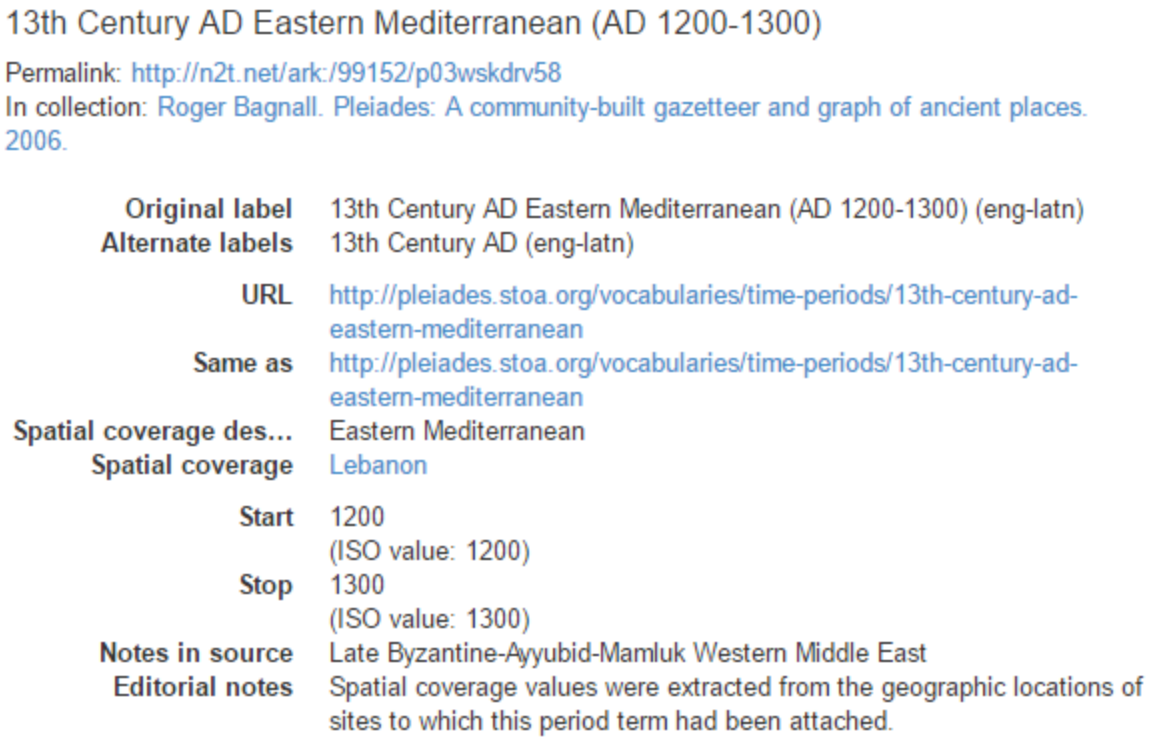
1. Backend selection page for the PeriodO client.

A user begins by choosing the "backend", or dataset, that s/he would like to browse or search. The landing page for the client offers, by default, the *canonical* dataset: that is, the core set of definitions that have been carefully evaluated by the PeriodO team, ingested into the PeriodO server, and published with stable, persistent URIs. We have concrete plans to maintain the accessibility of this dataset in the long term, and the URIs can be used with confidence. Edits can be made to the canonical dataset, but as a matter of policy, we limit those edits to the correction of typographical errors or mistakes in the documentation of the definition. The original assertions of the source about dates and spatial coverage are permanently maintained as they appeared – even if the same authority later changes a period definition, the modified definition will be added as a separate entity (though we will note where appropriate that it is derived from the earlier definition).



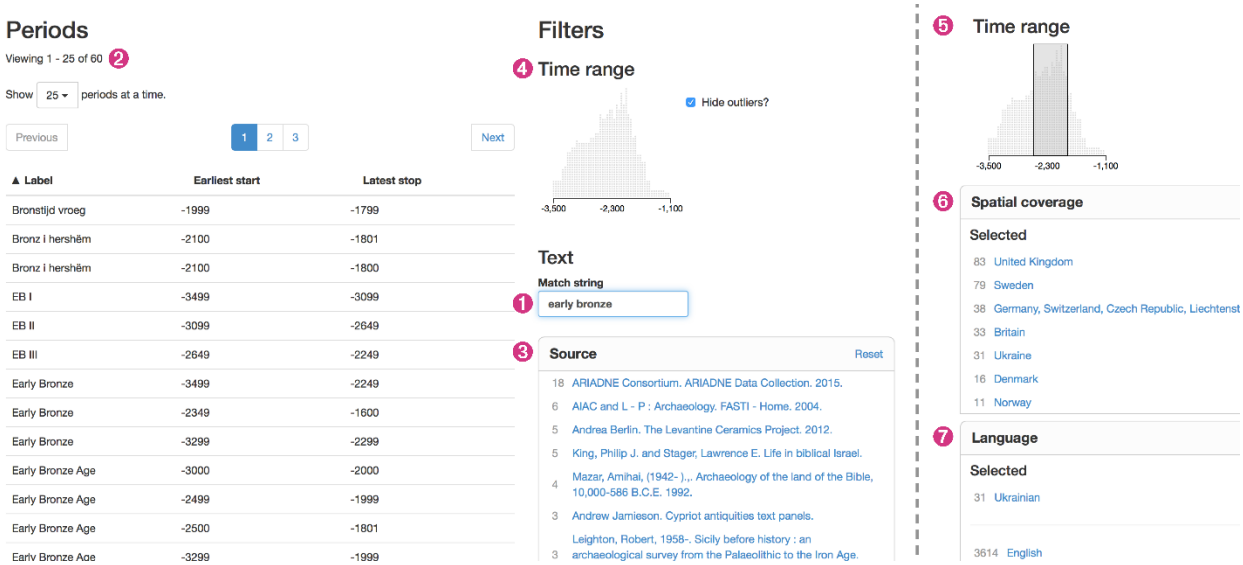
2. Initial view of canonical dataset.

When the canonical dataset is selected, the browse/search page will appear. By default, the "browse periods" tab will be selected, and the user will see a list of all the period definitions in the dataset, ordered alphabetically A-Z by label. The dataset can be reordered Z-A by clicking on the column header for "label", or it can be put in ascending or descending numerical order according to earliest start or latest end date by clicking on the appropriate column header. Clicking on one of the definitions in this view will expand it to show a human-readable representation of all the information in the record, including a URI that can be pasted into a user's spreadsheet or database.



3. Expanded view of an individual period definition. The permalink is the URI.

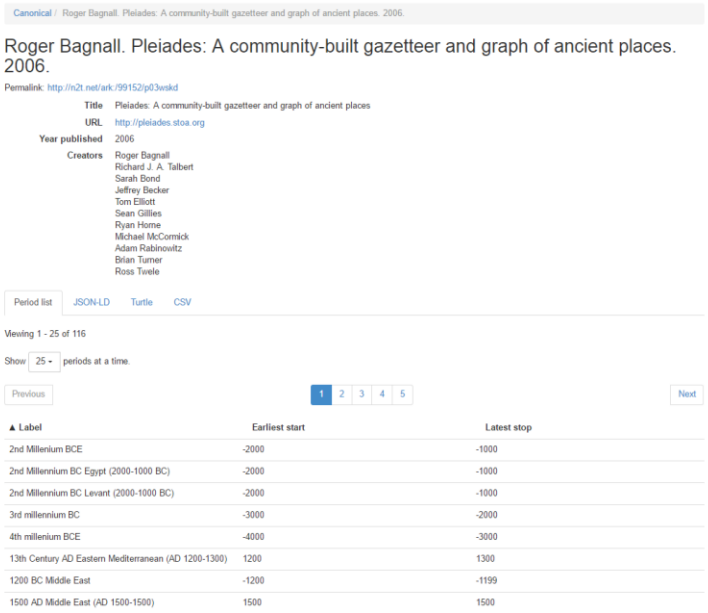
On the right-hand side of the page are a series of simple filter and search tools that allow the user to narrow down the set of period definitions on display. A basic timeline filter allows the user to set the upper and lower temporal limits of the period definitions to be shown, and a checkbox ("Hide outliers?"), checked by default, allows the user to show or hide the long tail of periods with very early start dates (as we begin to add geological periods, this tool will be replaced with a more sensitive and scalable timeline tool). A free text search box searches for text string matches in period labels (more advanced text search tools will be developed in the next phase of the project). Below the search box are a series of faceted lists: collections, languages, and spatial coverage (these values are the expressions provided by the original authorities; the mappings to national boundaries that appear in the individual definitions will be used in the future for map-based searches). Selecting one or more values from any of these lists will limit the list of definitions to those that match those criteria. The user can limit the search according to as many facets as desired, until the results set is a single definition (so, for example, one could choose a particular collection, then a particular language, then a particular spatial coverage, and at each step the results set would narrow further). Each of these filters can be reset to include other values potentially available in the collection(s) displayed.



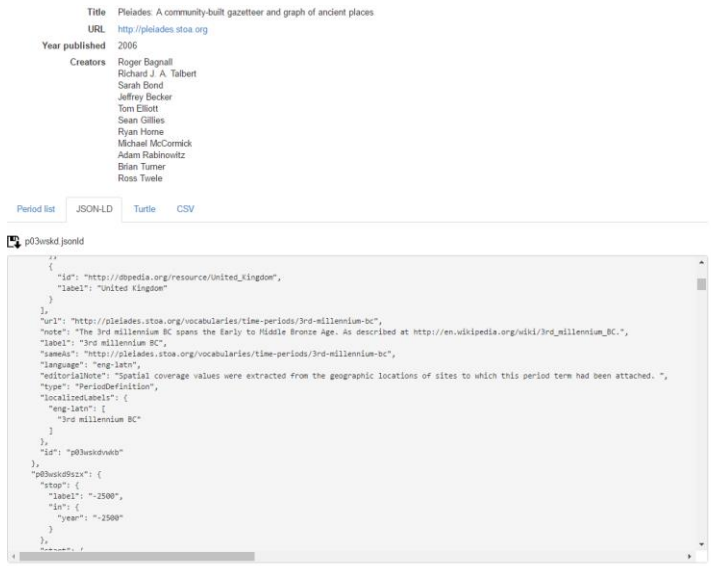
4. Finding and comparing period definitions in PeriodO. Searching for "early bronze" (1) results in sixty period definitions with matching labels (2), from a variety of sources (3). The time range facet (4) updates to show the distribution of temporal extents defined by these various sources. Users can query for period definitions with temporal extents within a specific range of years using the time range facet (5), period definitions with spatial extents within a named geographic area using the spatial coverage facet (6), or period definitions in specific languages using the language facet (7). Queries may combine values from any of these facets. (Illustration from Shaw and Golden 2016)

Users can also explore period information by collection, by clicking on the "Collection" tab at the top of the browse page. Collections cannot be sorted or searched at the moment, although this functionality is also planned for the second phase. Clicking on a collection will bring up information about the source and a complete list of periods from that source (one can also reach this page by clicking on a collection title in a period definition in the "Period" browse view). This list of definitions can be sorted by label and start and end dates. In addition, the collection page allows the user to view and download different expressions of the data in that collection: clicking on the tabs for JSON-LD, Turtle (TTL), and CSV will bring up visual displays of the information in those formats, as well as download buttons (the disk icon) that will download the information as a file in that format.

The entire dataset can also be downloaded as a single JSON file from the backend selection page. The option to download and work with PeriodO data locally is a critical component of the system: not only does it allow a user to use the canonical PeriodO dataset for local data-management purposes, such as the maintenance of a controlled vocabulary, but it also makes it possible for users to edit existing definitions for typographical mistakes and errors of documentation, and to add new period collections and definitions that they have created themselves. These edits and additions take place through a patch submission process in which user identities are managed by the ORCID system (so contributors must be registered with ORCID), patch provenance is tracked and documented by the PeriodO server architecture, and patches are accepted or rejected by a group of PeriodO editors (currently only the project PIs, but we hope this group will grow as the project picks up momentum).



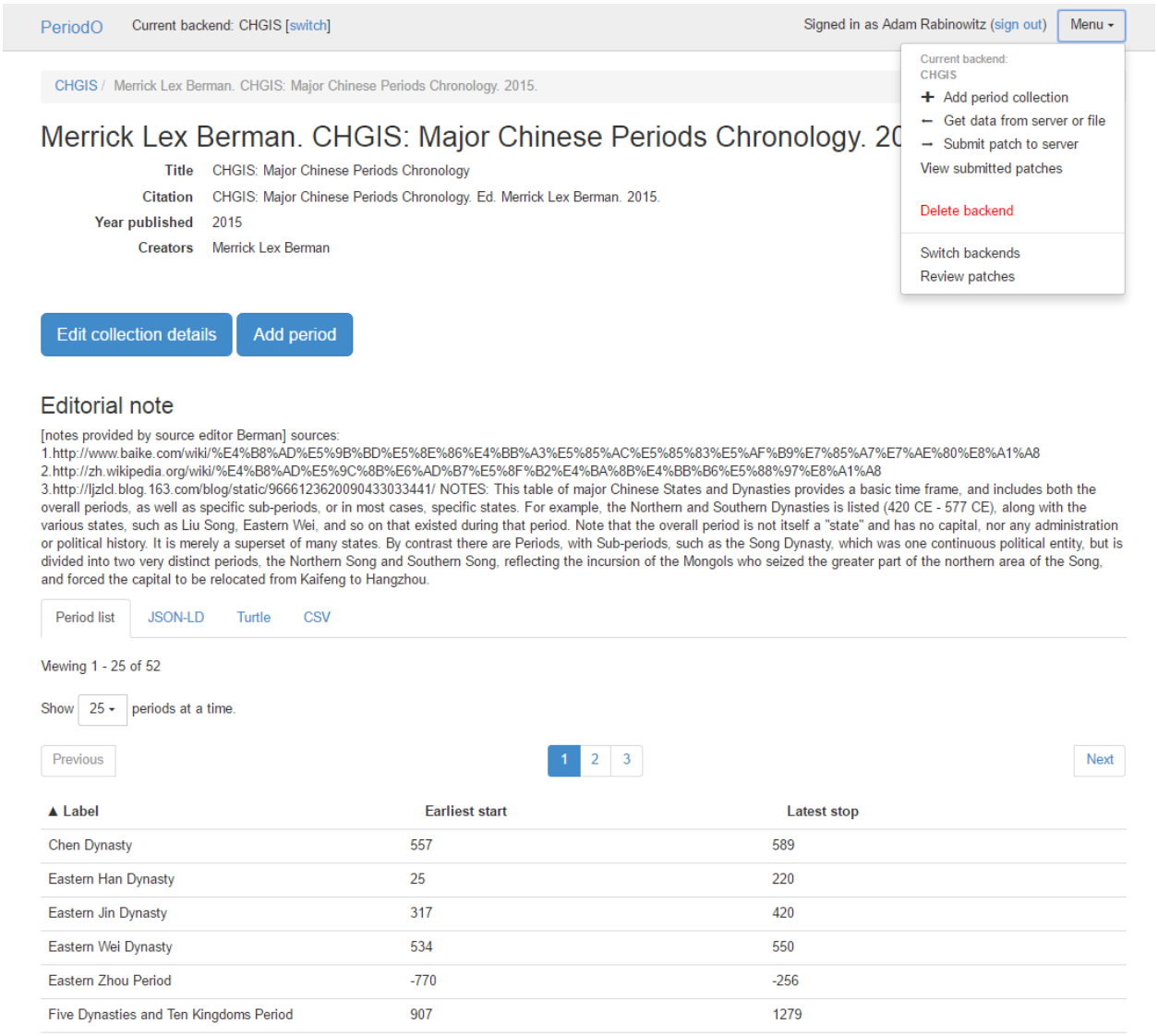
5. Collection page view with graphic interface.



6. Collection page view with JSON-LD representation.

Detailed instructions for the creation of new period collections and definitions and the editing of existing ones will soon be available on the project website, to which we refer those readers who may have contributions to make. For general audiences, however, it useful to point out that the backend selection page allows two additional options beyond viewing the canonical dataset. The user may load a read-only JSON file that follows the PeriodO schema; this will usually be a file exported from the client, which makes it easy to share user-generated periodizations (one can simply send a colleague the export file, and the colleague can load it into the client to look at it).

More importantly, the user can create a locally-persisted database using the Indexed Database (IndexedDB) feature available in standards-compliant Web browsers. IndexedDB databases are stored offline by Web browsers and thus can be used with or without an Internet connection. In a local database, it is possible to add and edit period definitions and collections. Local databases can be populated by downloading a configurable subset of collections and definitions from the PeriodO Server (available from the menu of actions at the top right of the interface), by loading a JSON file containing PeriodO-compatible data, or by adding new user-defined content through a form interface. This is also the workflow by which patches are submitted to the canonical dataset: when a user has carried out edits and additions in a local database, those changes can be submitted as a patch for review (again, through the menu options on the upper right corner of the page).



7. Local IndexedDB: edit view with menu options for data load and patch submission.

Use Cases

We recognize that a lay reader may wonder why all this is necessary. Don't scholars basically agree on periodizations, and don't they have a clear idea of, and general consensus on, the absolute dates and spatial extents of particular periods? Unfortunately, the answer is no, and there are a number of ways in which a gazetteer of period definitions can be helpful not only for abstruse issues of data management, but also for education, research, and the understanding of knowledge production in a range of disciplines that deal with the past. In this section, we lay out an overview of some of the use-cases we envision for the PeriodO gazetteer.

Use case 1: the student
In this use case, a student would like to find more information about the different chronological boundaries assigned to the same period by different scholars, or find out what periods are called in another language or geographic area. This user could acquire clear, authoritative information about period terms and extents simply by browsing and/or searching the database. This use case came up last week in the context of a Classics class one of us is currently teaching: a student was confused by the inconsistency of references to the Late Helladic IIIC period, and wished out loud in a conversation that there were a tool or resource to help explain what the actual chronological boundaries are, and point to the authoritative sources of different definitions.

Use case 2: the data manager
In this use case, the person responsible for a database of periodized information is seeking authoritative period definitions to use in his or her dataset. The data manager would search and browse like the student in use case 1, but having found appropriate matches in PeriodO for the period usage in the dataset, could include the URIs of those matches in the database itself, making the geographic and temporal extent of this local period usage explicit and the source of its authority transparent. A librarian might be a similar user – a PeriodO URI could be attached to a term in a defined vocabulary, which would then permit searching for bibliographic subjects by date range as well as string. Certain users in this case might not be satisfied with the period definitions already in PeriodO, but might still wish to connect their data with a Linked Data URI in a gazetteer – in which case they could add their own periods to the PeriodO dataset.

Use case 3: the disciplinary historian
In this use case, a student of the history of a particular discipline concerned with the past would like to explore changes in the understanding of the chronological boundaries of a particular period term over time, or the differences in period usage in diverse national traditions of scholarship. This user would be able to search for the same period term in sources that ranged across time and compare the results, or search for and compare period terms used for the same chronological horizon in several different countries. The more users in use case 2 add new period collections and definitions, the richer the dataset available for this user's research becomes.

Use case 4: the database aligner or Linked Data aggregator
As the dataset grows in size and scope, we hope it will attract the users in this use case, who are trying to align idiosyncratic period definitions across a group of heterogeneous datasets by identifying chronological relationships of overlap, bounding, or other topological interactions, or who are seeking to use a shared gazetteer to bring together information expressed as Linked Data

from a variety of different sources. The best parallel for the user in this use-case is the highly successful Pelagios project, which has been doing this work to bring together information from a broad set of databases that have connected their records to a group of aligned historical spatial gazetteers and then expressed them as Linked Data. We know that the community of data providers and managers who deal with the past are interested in these approaches, and we know that they lack a gazetteer that will act like Pleiades or another spatial gazetteer for the reconciliation and alignment of temporal attributes expressed as words rather than ISO dates. We intend to make this use case the focus of PeriodO's next phase.

Use case 5: the natural-language processor
The last use case is the most complex, and will also be developed during the project's next phase. Once the PeriodO dataset has grown through the contributions of users from use case 2 and has taken on historical depth through the research (and, we hope, contributions) of users from use case 3, and once PeriodO URIs have been attached to a broad range of data sources, many of which will combine period terms with the absolute dates of objects, monuments, and events, we hope to use this web of data as a probabilistic training tool for natural-language-processing approaches to the use of period terms in texts. Just as a geoparser refers to a gazetteer to extract coordinate values from place names in a book, we hope to use the PeriodO dataset and the records to which it is linked to mine texts for references to time in order to discover and visualize latent chronological information in the written record. Several groups have already set out to do this, using combinations of words and dates in written sources (de Boer et al. 2010; Mouroutsou et al. 2014), but they have not had the advantage of both a large, chronologically specific dataset reflecting period usage across time and a group of external datasets providing a connection between absolute and relative dates.

Next Steps

The success of the first phase of the project, and the engagement of a growing community of potential users, has encouraged us to plan for its continuation. We hope that PeriodO will become a long-term resource, along the lines of GeoNames or VIAF, for the management of periodized data. We are already having an impact in the area of European archaeology, and we mentioned above a series of emerging projects in diverse fields that are treating PeriodO as part of their ecosystem. In our proposals for next-phase funding to the IMLS and NEH, we have expanded the project to include a range of disciplines beyond archaeology: new partners we have enlisted for the next stage include historians, literary scholars and projects (e.g. the Advanced Research Consortium, led by literary scholar Laura Mandell), and large-scale digital libraries, including both the Digital Public Library of America and the Europeana project. In the meantime, we are pursuing various additional audiences through an ongoing program of presentation and outreach, both to individual data managers and to larger communities at national and international conferences.

Online Material, Publications, and Presentations, 2014-2016

Website, data, and code

- The PeriodO home website (information): http://perio.do
- The PeriodO dataset and client interface (data): http://n2t.net/ark:/99152/p0
- The PeriodO Github repository (code): https://github.com/periodo
- The PeridoO Twitter feed (updates and announcements): @perio_do

Publications

- Rabinowitz, Adam. 2014. "It's about time: historical periodization and Linked Ancient World Data". In T. Elliott, S. Heath, and J. Muccigrosso, *Current Practice in Linked Open Data for the Ancient World* (ISAW Papers 7). http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/rabinowitz/.
- 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.
- Golden, Patrick, and Ryan Shaw. 2015. "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.
- Golden, Patrick, and Ryan Shaw. 2016. "Nanopublication beyond the sciences: the PeriodO period gazetteer." *PeerJ Computer Science* 2:e44 https://doi.org/10.7717/peerj-cs.44.
- Rabinowitz, Adam, Ryan Shaw, Sarah Buchanan, Patrick Golden and Eric Kansa. "Making sense of the ways we make sense of the past". Under review at *Bulletin of the Institute of Classical Studies.*

Presentations

- Rabinowitz presented "Periods, Organized (PeriodO): a Linked Data gazetteer to bridge the gap between concept and usage in archaeological periodization" at the *Computer Applications and Quantitative Methods in Archaeology* conference at the Sorbonne in Paris in April 2014
- Eric Kansa presented PeriodO at the *DH2014* conference in Lausanne in July 2014
- Rabinowitz presented "Managing Time: PeriodO, a Linked Data approach to the interoperability of periodized data" at the workshop "Fostering Transatlantic Dialogue on Digital Heritage and EU Research Infrastructures: Initiatives and Solutions in the USA and in Italy" held at the Library of Congress in December 2014
- Kansa presented "Open Context and PeriodO" as a lightning talk at the Society for American Archaeology annual conference in San Francisco in April 2015
- Kansa included PeriodO in his presentation "Contextualizing Digital Data as Scholarship in Eastern Mediterranean Archaeology" at the Center for Hellenic Studies at Harvard University in April 2015

- Golden presented "Period assertion as nanopublication" (coauthored with Shaw) at the *Semantics, Analytics, Visualisation: Enhancing Scholarly Data Workshop* at the 24th International World Wide Web Conference in Florence, Italy in May 2015
- Rabinowitz presented "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?" at the Mellon-funded *Linking the Middle Ages* workshop at the University of Texas, Austin in May 2015
- Rabinowitz presented PeriodO at the DINAA *Radiocarbon and Temporality Workshop* at the University of Wyoming, Laramie in June 2015
- Shaw presented "An Ecosystem of Time Periods: PeriodO" at the *Linked Pasts* workshop at King's College London in July 2015
- Kansa included a demonstration of PeriodO in his lecture and workshop "Methods in Archaeological Data Publishing" at the NEH-funded Digital Archaeology Institute at Michigan State University, Lansing in August 2015
- Shaw presented a demonstration of PeriodO at the *14th European Networked Knowledge Organization Systems (NKOS) Workshop* at the TPDL conference in Poznan, Poland in September 2015
- Kansa included PeriodO in his talk "Challenges in Archaeology, Linked Data, and Publishing Geospatial Data on the Web" at the Center for Geospatial Analysis at Harvard University in September 2015
- Shaw and Golden presented PeriodO at the *Coalition for Networked Information* conference in Washington DC in December 2015
- Rabinowitz presented PeriodO at the University of Texas School of Information Research Colloquium in February 2016
- Shaw represented PeriodO at the PHOIBOS2 identifier workshop at the Biosphere in Arizona in February 2016
- Rabinowitz will present PeriodO at the annual meeting of the Society for American Archaeology in Orlando in April 2016
- Rabinowitz will represent PeriodO at the inaugural meeting of the Big Ancient Mediterranean project at the University of Iowa in June 2016

Works Cited

Berners-Lee, Tim, 'Linked Data - Design Issues', *www.w3c.org* (2009). Retrieved from https://www.w3.org/DesignIssues/LinkedData.html

de Boer, V., M. van Someren, & B.J. Wielinga, 'Extracting historical time periods from the Web', *Journal of the American Society for Information Science and Technology*, 61.9 (2010) 1888–1908. doi:10.1002/asi.21378.

Doerr, Martin, Athina Kritsotaki, & Stephen Stead, 'Which Period Is It? A Methodology to Create Thesauri of Historical Periods', in *Beyond the Artifact. Digital Interpretation of the Past. Proceedings of CAA2004, Prato 13-17 April 2004*, ed. by Franco Niccolucci and Sorin Hermon (Budapest: Archaeolingua, 2010) 70–75. Retrieved from http://proceedings.caaconference.org/files/2004/10_Doerr_et_al_CAA_2004.pdf

Heath, Tom and Christian Bizer, *Linked Data: Evolving the Web into a global data space*, 1st edn (Morgan & Claypool, 2011). doi:10.2200/S00334ED1V01Y201102WBE001

'Linked Data - W3C', *W3C* (2015). Retrieved from https://www.w3.org/standards/semanticweb/data

Maria Mouroutsou, Stella Markantonatou, & Vasilis Papavasiliou, 'The development of vocabularies of historical period names from web acquired corpora7', *Mediterranean Archaeology & Archaeometry*, 14 (2014) 165–74.