Interim Performance Report
Grant Number: HK-230973-15
Project Title: Pleiades 3
Project Director: Tom Elliott
Grantee Institution: New York University
Date Submitted: 30 September 2016

Performance

As anticipated in our last interim performance report (March 2016), the upgrade of the *Pleiades* web application (from Plone 3 to Plone 4) was completed and deployed in May 2016. As anticipated, we saw immediate performance improvements for both anonymous, read-only users and authenticated contributors, reviewers, and editors: the prime deliverable of the current project. Subsequent interventions in key code components, identified and prioritized by our developers (Jazkarta, Inc.) using code review, reports from the *New Relic* application monitoring service, and their engineering judgment, yielded even greater boosts. In the 9 months prior to the upgrade, *Google Analytics* calculated our average server response time to be 4.2 seconds and our average page load time to be 8.48 seconds. As pointed out in the proposal, these numbers were unacceptably poor. Since the upgrade, we are averaging 0.9 seconds on server response and 2.1 seconds on page load time. Fully 26% of pages are fully loaded in less than one second. Another 37% in between 1-2 seconds. This improvement in responsiveness is illustrated in Figure 1, which was generated with *Google Analytics*.[1]

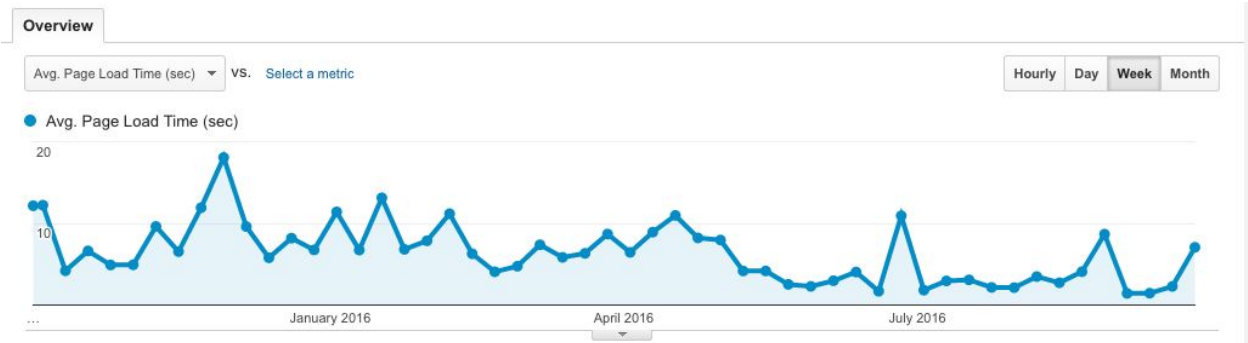

Figure 1: Pleiades page load timing (October 1, 2015 - September 28, 2016)[2]

It is important to note that the pre-upgrade timings were measured during a period when bot activity on the site (search engine indexers and third-party data users) was severely restricted through server configuration. Since the upgrade, these restrictions have been removed (although demonstrated bad actors are still sometimes banned). One measure of this change may be had by viewing error reports in the *Google Search Console*. Prior to the upgrade, the GoogleBot was encountering upwards of 40,000 crawl errors per month on the *Pleiades* website (a combination of server errors due to overloading and of "access denied" errors resulting from restrictions). Immediately after the upgrade, this number dwindled to 7,000, and is now down to

---

[1] *Pleiades* uses a customized version of the open-source *Plone* content management system (https://plone.org/). Jazkarta: https://jazkarta.com/. *New Relic*: https://newrelic.com/. *Google Analytics:* https://analytics.google.com.

[2] Bot traffic is not included in these numbers. Nor are requests for resources in formats other than HTML. Three post-upgrade spikes in page load time may be observed in the data: one in July, one in August, and one in late September. The first two correspond to periods of degraded system performance caused by difficulties with subsequent feature upgrades, now resolved. The most recent (which post-dates the current reporting period) is under investigation at the time of this writing.

zero, apart from intentional "access denied" cases that we use to keep bots from making multiple passes through indexes and search components of the site. So, the page load and server response figures reported for the post-upgrade period above should be understood to correspond to a period in which search bots supporting Google, Bing, and other entities -- as well as third-party users of our data like *Dickinson College Commentaries* -- have been given free rein to view and collect data from *Pleiades* as frequently and quickly as they like. We hope to measure bot traffic systematically on the site in future; it appears from a cursory review of our server logs that we handle several thousand bot requests each day.[3]

We anticipate that our users will see further performance improvements in coming months when we have re-written the most time-costly component of our web appication at present: the "Photos" portlet that queries *Flickr.com* for related images on every place page.

User Interface and Interactions

Another project deliverable identified in the proposal was the creation and deployment of a new graphical user interface. This deliverable reflected, in large part, our intention to replace the use of the Plone content management system and our heavily-modified version of its native web application for non-authenticated, read-only users of the site. We argued in the proposal that this architectural change would be necessary to achieve desired performance improvements. Once we had analyzed the performance of the upgraded Plone web application, it became clear that a better course would be to continue performance-tuning the upgraded web application while focusing time and attention on improvements to the existing user interface and to the underlying data model and database structure.

In addition to basic improvements in layout, icons, and map interfaces, we have made a number of enhancements designed to help contributors and editors streamline the process of content creation, review, and publication. Upon viewing a place page, an authenticated contributor now sees lists not only of the published names, locations, and connections associated with that place, but also of any such content that is still in draft or pending-review state. Previously, only published content or content belonging to the contributor in question was visible. This change reduces duplication of effort, which previously had to be detected and reconciled by the editorial college. It is complemented by another new feature that helps contributors recover from inadvertently creating doublets of existing content or making other mistakes: authenticated contributors can now delete their own drafts (previously a task that required a Managing Editor's intervention). Contributor requests for other features have also been honored, including more flexibility in working with the citation lists and the ability to import polyline and polygon features from *OpenStreetMap* (previously only points were supported). The editorial college now also benefits from a speedy and comprehensive review list (the prior version was so slow to load as

[3] *Google Search Console* (formerly *Google Webmaster Tools*): https://www.google.com/webmasters/tools/. *Dickinson College Commentaries:* http://dcc.dickinson.edu/.

to be useless) and a one-click mechanism for converting existing location resources into complete place resources.[4]

This last feature facilitates the implementation of a long-sought change in editorial policy: *Pleiades* now encourages the treatment of individual buildings, monuments, and similarly scaled features as full-fledged place resources. Originally only sites that were large or historically significant enough to be places on a 1:500,000-scale map were designated "places", with their own Uniform Resource Identifiers (URIs). But Pleiades is not constrained by the limitations of paper-based map publishing, and so smaller features can be treated with the same depth and sophistication as larger ones. Consequently, places like monuments and buildings can be documented and cited online as easily as settlements, islands, provinces, and the like. This change clarifies and enriches *Pleiades* content and increases its impact as a reference resource across other digital projects in the field. It has been complemented by the addition of a new "location type" vocabulary, which lets the contributors and editors indicate the function of each location resource and the character of its relationship with the associated place as defined in the *Editorial Guidelines*. The editors and reviewers are now working through existing content, identifying locations that need to be converted to places in order to conform to the new policy.[5]

Anonymous users of *Pleiades* have also benefited from interface changes requested by members of the community in our prior focus sessions and via online communication. The coordinates of the "representative point" for each place (an automatically calculated centroid of all associated locations) are now prominently displayed on each place page and shown clearly on the accompanying map. An associated copy-to-clipboard button has been added to facilitate easy collection of the coordinates by users who do not wish to use more programmatic methods to extract them from one of our downloadable export or online serialization formats. Similar buttons have also been added for the "canonical URI", "suggested citation", and *Wikipedia* citation template fields on place pages (see Figure 2). Quick search results via the search box on each page have been streamlined to include only matching places, as users found the default behavior (intermixing of places, names, locations, connections, and other content items) to be at odds with their intent in using this affordance. Full-site search remains possible through the "advanced search" form.

[4] *OpenStreetMap*: https://www.openstreetmap.org/

[5] About *Pleiades* URIs: https://pleiades.stoa.org/help/what-are-pleiades-uris. *Pleiades Editorial Guidelines*: https://pleiades.stoa.org/help/editorial-guidelines

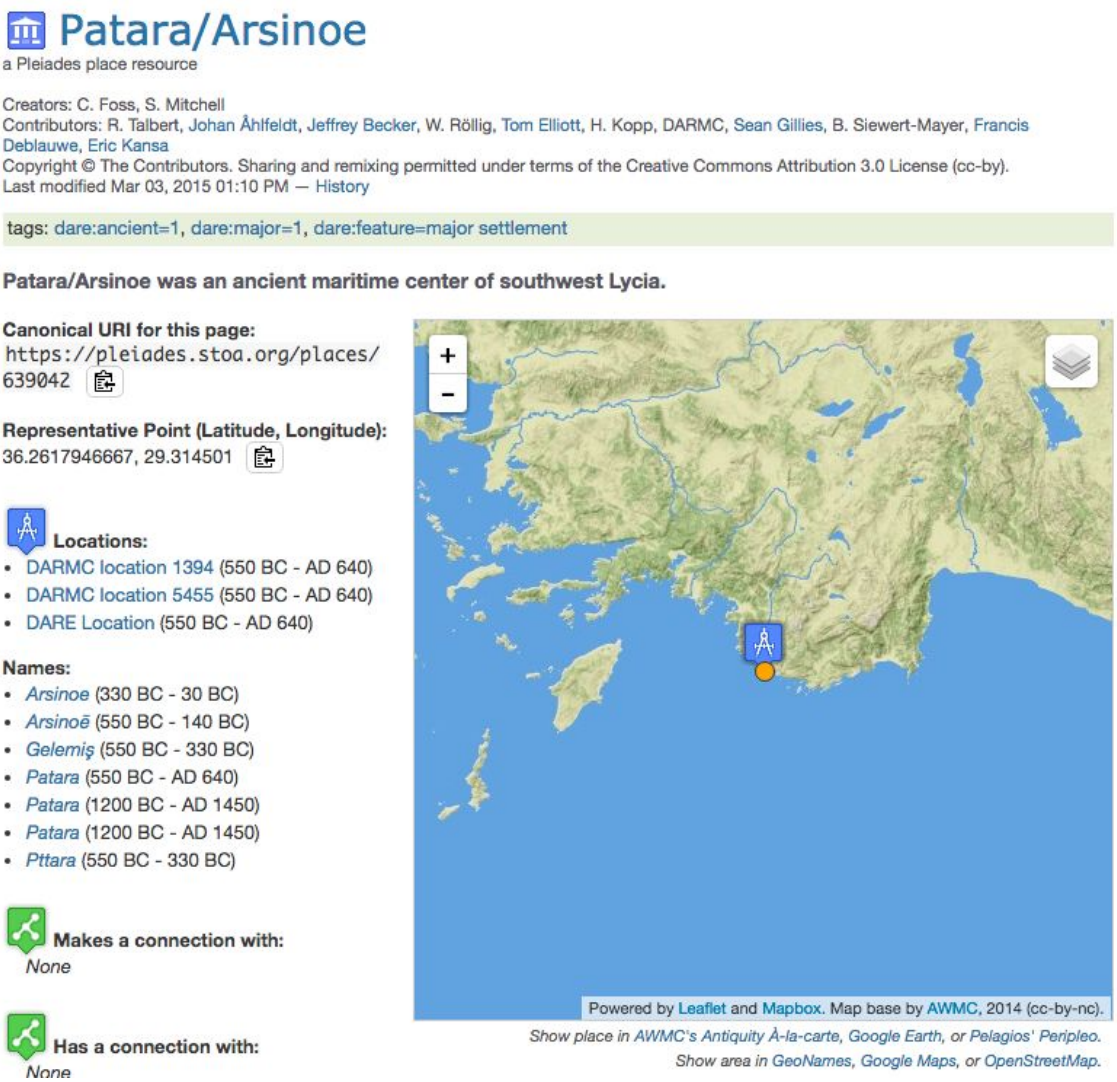

Figure 2: Screen capture illustrating some of the interface improvements discussed herein

Citation and Bibliographic Management

Citation and reference management for *Pleiades* has been more thoroughly revamped than one might imagine from reading about the interface changes mentioned above. As promised in the proposal, *Pleiades* has been integrated with the open *Zotero* citation management system. This initiative moves *Pleiades* and its users beyond the tedious nineteenth-century cul-de-sac of manual bibliographic citation processes while maintaining and enhancing the essential scholarly functions of documentation, argumentative support, repetition, and resource discovery.[6]

For individuals who wish to use *Zotero* to capture structured bibliographic data about *Pleiades* content without (retyping, cutting, pasting, and reformatting), we have upgraded all *Pleiades* pages to include rich metadata that can be parsed automatically by *Zotero* browser plugins.

[6] *Zotero:* https://www.zotero.org/

Figure 3 illustrates the differences between two *Zotero* records for the same *Pleiades* place page, one captured before we added rich metadata support, and the other afterwards.

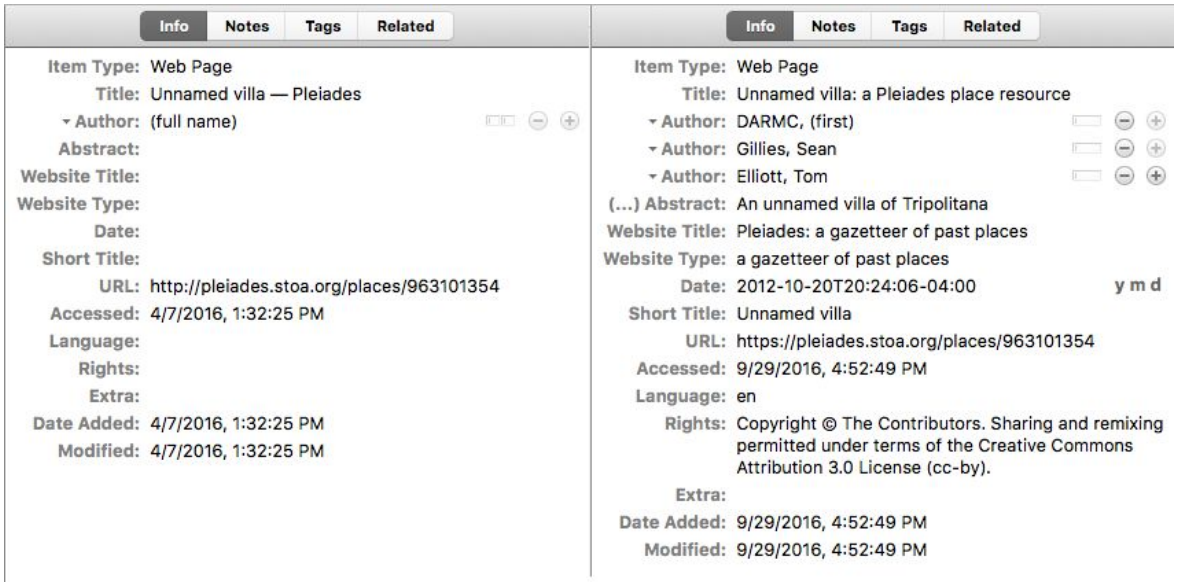

Figure 3: Zotero record for a Pleiades place (before and after rich metadata upgrade)

Contributors can now also draw upon *Zotero* records to create new citations in *Pleiades*. Previously, contributors were required to enter bibliographic information manually into *Pleiades* for each work they wished to cite. Editors then spent scarce time and effort checking not only the content and validity of the works referenced but also the conformance of the citation formatting to the *Pleiades Citation Guide*. These tedious, manual tasks are now a thing of the past. To cite a work in *Pleiades* today, the contributor simply pastes a link to any public *Zotero* record into a form and clicks the associated "Fetch Zotero" button. *Pleiades* uses the open *Zotero* Application Programming Interface (API) to copy information (including an automatically formatted citation in the editors' preferred style) into the fields in the *Pleiades* form. The contributor then needs only to select the "citation type" and fill in any "citation detail" (e.g., page numbers for the relevant portion of the work) and move on. Figure 4 illustrates a before-and-after sequence. Figure 5 illustrates the automatic use of this imported data in the resulting published place page.[7]

[7] *Pleiades Citation Guide:* https://pleiades.stoa.org/help/citation-guide

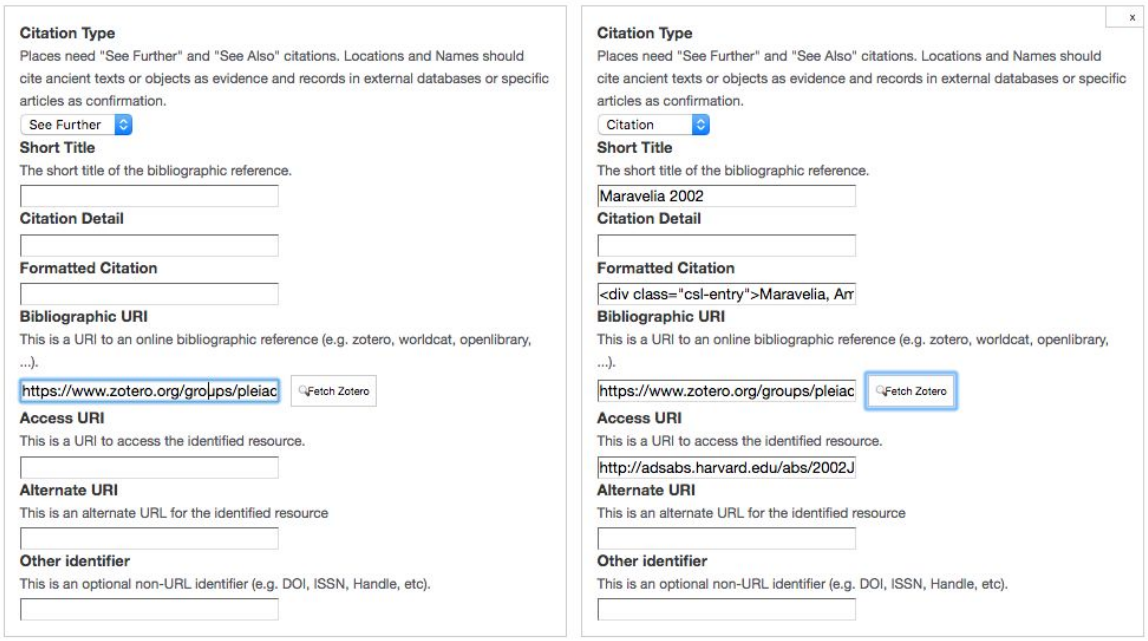

Figure 4: Zotero import, before and after

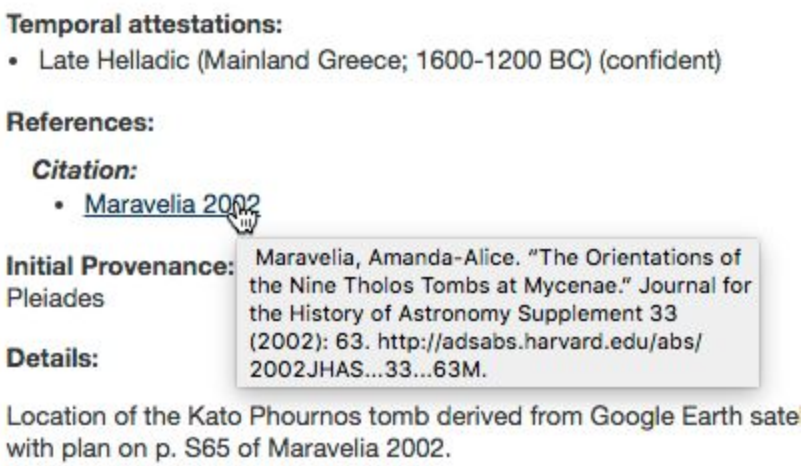

Figure 5: Imported Zotero data in use

Readers familiar with the original *Pleiades* bibliographic implementation will detect in Figure 4 the addition of new fields to the reference data model. We have tried to avoid implementing a complete, structured bibliographic data management tool in *Pleiades*, since *Zotero* provides this functionality in an open service. Instead, we have emphasized the immediate needs of our users: direct connections to resources available online and to discovery mechanisms for print-only titles, as well as the bibliographic enrichment of our Linked Data outputs. We also considered the ability to continue operating in the event of a temporary *Zotero* outage. In particular, we have added the ability to record not only a primary Uniform Resource Identifier (URI) for access to the work referenced (for digital and digitized works), but also an alternate

URI (to use for an archival copy in the *Internet Archive* or an institutional repository). Other additions include: a "formatted citation" field (for display to users), a "bibliographic URI" (so we can point to the source *Zotero* record), and an "other identifier" field in which to store International Standard Book Numbers (ISBNs) or Digital Object Identifiers (DOIs) for the work if appropriate.[8]

Existing structured bibliographic records for works cited by the *Barrington Atlas* and previously imported into *Pleiades* are being migrated to a *Pleiades Zotero Group* so that they can be readily reused in updating existing citations in *Pleiades* and in creating new ones.[9]

Enriching the Pleiades Linked Data Graph

As one of the earliest Open Linked Data projects in ancient studies, the *Pleiades* community is committed to maintaining its participation and leadership role in same. Two relevant initiatives highlighted in the proposal have been implemented during this reporting period: the addition of rich relationships to the *Pleiades* data model and the connection of key controlled vocabularies to relevant external vocabularies that are widely used in Linked Data.

During a previous grant-funded effort, *Pleiades* experimented with relationships between places by adding so-called "connections" and including them in our linked data serializations and exports. Each such connection was a simple, directed relationship between one place and another, and the community limited their use to express relationships of belonging (as a settlement relates to an administrative area or a geographical region) and of flow (as an aqueduct connecting to a settlement). Contributors and editors were eager to make more nuanced and robust connections between places than this simple model permitted, and so we have implemented rich relationships. A *Pleiades* connection remains a directed relation between two places, but it can now be attributed with temporal limits, bibliographic citations, and other annotations. Connections can also be typed, using a controlled vocabulary managed by the editorial college. Current values in the vocabulary include: "at", "on", "part of (administrative)", "part of (regional)", "near", "intersects", and "bounds". Additional terms will no doubt be added from time to time as needs are expressed by contributors and editors.

Post-upgrade timing studies of the *Pleiades* web application identified our controlled vocabulary management code as a top performance hog. This mechanism is deeply intertwined with every single resource on the site, providing management and lookup mechanisms for time periods, place types, and language and script codes, as well as measures of certainty, confidence, and the like. Its replacement with new, high-performance code gave us the opportunity to add some additional fields to each term in a vocabulary. The new fields allow us to specify equivalences between *Pleiades* vocabulary terms and terms in external vocabularies and to surface these equivalences in our Linked Data. At the time of this writing, we are already publishing "same as"

---

[8] *Internet Archive:* https://archive.org/

[9] *Barrington Atlas of the Greek and Roman World:* http://press.princeton.edu/titles/6773.html. *Pleiades Zotero Group:* https://www.zotero.org/groups/pleiades.

links between our time periods and the corresponding records in the *Perio.do* gazetteer of period definitions, which is funded by the NEH and the Institute of Museum and Library Services. We are currently working on crosswalking our "place type" vocabulary to the *Getty Art and Architecture Thesaurus*.[10]

Batch Create and Update

We are presently testing code that will constitute another key deliverable outlined in the proposal: a script that the Managing Editors can use to create and update *Pleiades* content *en masse.* We are working with real data provided by external partners (including the *Inscriptions of Israel/Palestine* project at Brown University and the *Cross-cultural After-Life of Classical Sites* project at Institute of Classical Studies, University of London). We anticipate reporting on completion and successful production use of this script in our next semi-annual report.[11]

Downloading, Archiving, Reusing

Pleiades now boasts a comprehensive, daily export of all its content: another key project deliverable. This "data dump" is provided (at the request of our users) in the JavaScript Object Notation (JSON) format, and may be had via the *Pleiades* Downloads page. Work is on-going to automate its quarterly deposit into the *NYU Digital Faculty Archive* and dissemination to other third-party archival contexts, as outlined in the Data Management Plan included with our proposal.[12]

---

[10] *Perio.do:* http://perio.do/ and the *Pleiades* time periods therein: http://n2t.net/ark:/99152/p03wskdnxfq. *Getty Art and Architecture Thesaurus:* http://www.getty.edu/research/tools/vocabularies/aat/.
[11] *Inscriptions of Israel/Palestine:* http://cds.library.brown.edu/projects/Inscriptions/. *Cross-cultural After-Life of Classical Sites:* http://commons.pelagios.org/2016/08/cross-cultural-after-life-of-classical-sites/.
[12] JSON: https://tools.ietf.org/html/rfc7159. NYU Faculty Digital Archive: http://archive.nyu.edu/.