Interim project report

Grant number: PW-50557-10
Title of project: Pleiades: Content and community for ancient geography
Name of project director: Dr. Thomas Robert Elliott
Name of grantee institution: New York University
Date report is submitted: 15 December 2011 (due: 30 November 2011)

Narrative description

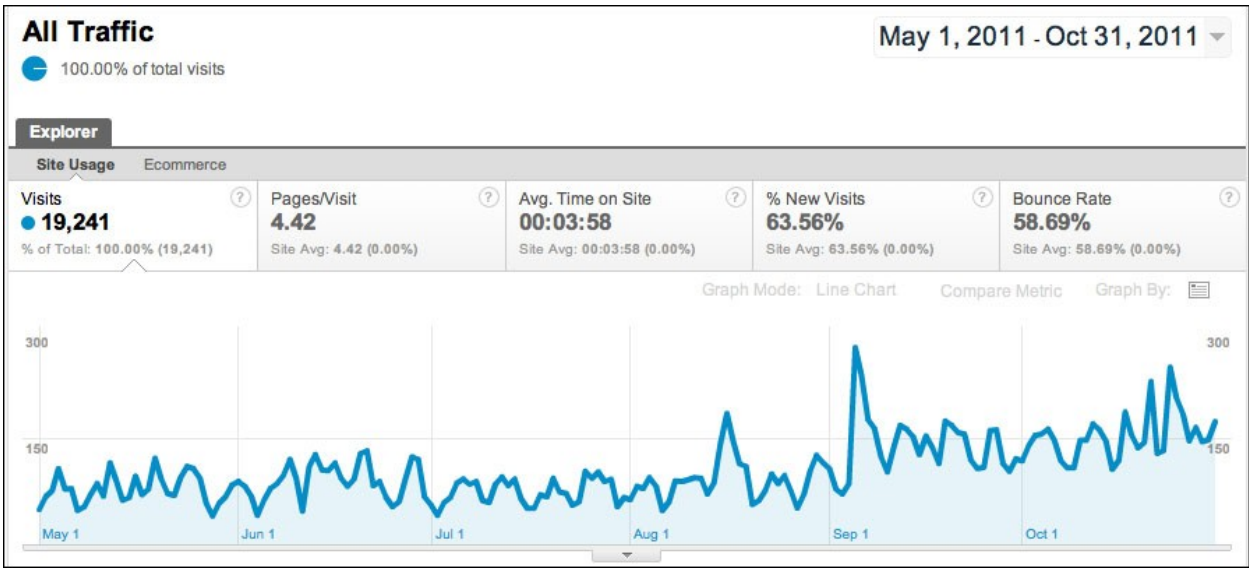
The Pleiades development and editorial team continues to be guided by the needs and interests of its user community in conceiving, prioritizing, and implementing bug fixes, new features, and content modifications. The following report summarizes the broad range of activities completed during the reporting period.

The following sections illustrate specific accomplishments, reflecting strong progress in all areas of the work plan. They are grouped thematically, rather than slavishly following the list of deliverables, since our understanding of deliverable groupings and impact has evolved and matured over time in response to community needs and technology changes.

Outreach and community building:

Patterns of use and community growth

During the reporting period, Pleiades saw over 19,000 discrete visits to http://pleiades.stoa.org. For the first three months of the period, visits per day averaged about 80; since August 2011 we have seen a slow-but-steady increase to an average of approximately 130 per day. Average pages-per-visit and time-on-site-per-visit metrics have remained roughly consistent, holding at 4.2 pages/visit and 3:58 minutes/visit respectively. The following chart, from Google Analytics, illustrates some of the key metrics and the trend in visit count.



The origins and entry paths of our visitors demonstrate that we continue to be effective in extending the reach and online discoverability of Pleiades content by providing useful information in useful formats, tuning our site map, exposing our data in a granular and open fashion at stable URLs, and linking our resources to pivotal external reference resources like Wikipedia. Visitors during the reporting period came from 120 countries. Of these, the following accounted for over 500 visits each (in descending order): United States, United Kingdom, France, Germany, Italy, Netherlands, Turkey, and Spain. Over 50% of visits came by way of search engines (especially Google), which we consider a particular gratifying result, validating

our commitment to the open web and open data. Nearly 28% of our visitors were referred via links from 334 other sites. Of these, the following accounted for over 100 visits each (in descending order):

- Institute for the Study of the Ancient World (http://isaw.nyu.edu)
- Ancient World Mapping Center (http://www.unc.edu/awmc)
- Omnes Viae Roman Route Planner (http://omnesviae.org)
- Sean Gillies' Blog (http://sgillies.net)
- Facebook
- Barrington Atlas feature identifiers and other resources on our own http://www.atlantides.org
- Chuck Jones' Ancient World Online (AWOL) blog (http://ancientworldonline.blogspot.com)
- Nomisma.org, which aims to provide stable digital representations of numismatic concepts and entities (http://nomisma.org)
- Wikipedia (http://en.wikipedia.org)
- The American Numismatic Society online database (http://numismatics.org)

The core Pleiades team has been aggressive in using social media to promote the project. The project news blog at http://pleiades.stoa.org/news/blog has been used heavily to announce new features, content upgrades, and community activities. These blog posts are echoed to the Pleiades Community Listserv, various blog aggregators, Facebook, LinkedIn and Twitter. Reblogging and Facebook seem to be the most effective channels in terms of driving traffic back to the site; Twitter is surprisingly ineffective. We believe this is because of varying demographics: enthusiast and humanist users of Pleiades are more likely to be engaged with Facebook than with Twitter, which (for the humanities at least) is more the province of digital humanist practitioners than humanists who use digital resources.[1] LinkedIn is even more ineffective, but this doesn't surprise us given its demographics. Despite Academia.edu's large footprint among humanistic scholars, we have been unable to make effective use of it as a channel for outreach because the site provides no mechanisms for institutional data flows or external automation integration. Queries concerning such matters have gone unacknowledged.

Active participation by users as contributors has been steadily increasing. At present, we have 6 users who are making repeated, substantive updates to Pleiades content. Although we collect data on their activities, we do not at present have a simple means to aggregate, analyze, and report that information. We hope to include in our next interim report a section addressing this issue in more detail. Based on inquiries and community discussion, we expect the number of contributing users to grow significantly in the next 6 months.

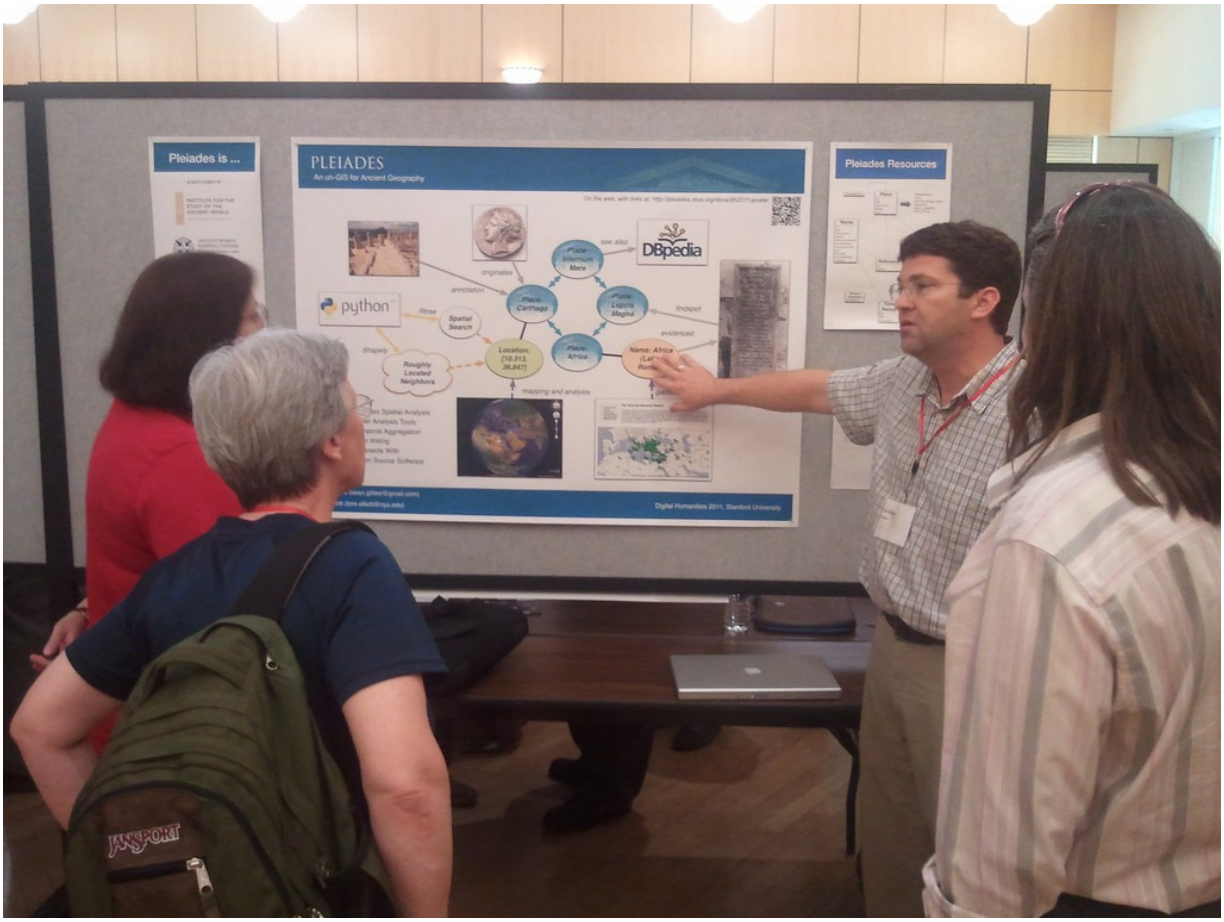
Outreach: Pleiades at AAG 2011

Sean Gillies represented Pleiades in a special session on gazetteers at the 2011 Annual Meeting of the Association of American Geographers (12-16 April 2011; Seattle, Washington). Our slides are online at http://atlantides.org/docs/slides/Pleiades-AAG2011 and attached to http://pleiades.stoa.org/docs/aag-2011-slides.

[1] This is an assumption based on anecdote and personal experience. We have no data to support it quantitatively.

Outreach: Pleiades at Digital Humanities 2011

Sean Gillies and Tom Elliott represented Pleiades at Digital Humanities 2011 (19-22 June 2011; Stanford University).



Our poster was well received. The trip also provided opportunity for informal conversations with many other projects and digital humanities practitioners. A clickable imagemap version of our poster, with links to Pleiades and other resources on the web, is maintained at http://pleiades.stoa.org/docs/dh2011-poster.

Outreach: Community meetup via Google+ hangout

Interested in trying another medium to facilitate conversation between the core Pleiades team and our globally distributed users, we scheduled a community meetup using the relatively new Google+ “Hangouts” feature. This event was held on 19 October 2011 and was not an unqualified success. Although substantive conversation was had amongst 4 of the participants, many others were unable to participate due to technical challenges associated with out-of-date browser plugins, software installation restrictions on work and university library computers, and slow network connections. The video call itself was interrupted on several occasions by involuntary dropouts that affected some members. The consensus opinion is that future meetups should go back to using group text chat via the #pleiades Internet Relay Chat channel.

Outreach: Content Improvement projects and volunteer opportunities

In an effort to help Pleiades users coordinate their efforts at content improvement with other members of the community, we have added a new section of the website to enable community members to suggest, organize and volunteer for projects to improve or expand Pleiades content. There is now a Pleiades Content Development page at http://pleiades.stoa.org/docs/content-development-projects where users can propose new projects, volunteer for existing ones, and coordinate work between ad-hoc teams. There is also a new “Using our Data” section, which highlights some of the ways in which external, partner projects are reusing and remixing Pleiades content: http://pleiades.stoa.org/docs/using-our-data.

Forthcoming outreach: Planning a Pleiades “hack day”

A virtual “hack day” is in the process of being scheduled for November 2011. It will provide users (and prospective users) an opportunity to work together to improve titles and descriptions for our place resources. On-the-fly training of new users is also anticipated.

User interface upgrades

User interface upgrade: New home page

In an effort to make the home page more clear and more engaging, we stepped back and rewrote it from scratch. Users had suggested that first visitors to the site had difficulty grasping the project’s aims and scope -- and essential geographic nature -- from the the old home page. The new page was built with these concerns in mind. Two paragraphs of text explain what and who (on the left) and how and why (on the right). Above that is a prominent search form. The centerpiece is a slide show of random places from Pleiades, highlighting the geographic nature of the project. The page is lightweight and fast: only about 10-20% of the latency of the old home page. The old home page continues at http://pleiades.stoa.org/home and users of the site will encounter it often. It is linked from the first of the portal tabs, the one labeled "Home", and is the root of the links in every content item's breadcrumb trail.

(see illustration on next page)

PLEIADES

A community-built gazetteer and graph of ancient places

Search the 34,068 Places

Internum Mare Go

Springing from the Classical Atlas Project and the *Barrington Atlas of the Greek and Roman World*, Pleiades is a historical gazetteer and more. It associates names and locations in time and provides structured information about the quality and provenance of these entities. There is also a graph in Pleiades: names and locations are collected within places and these collections are associated with other geographically connected places. Pleiades also serves as a vocabulary for talking about the geography of the ancient world within Linked Data sets and is referenced by research projects such as Google Ancient Places and PELAGIOS.

More about the data ...

Pleiades is one hundred percent open source, one hundred percent openly licensed and one hundred percent editable. Our open source software gets used and improved by a diverse crowd of programmers in the humanities and beyond. Our openly licensed resources have no surprises or encumberances, and their editability allows Pleiades users to correct or enrich resources – thereby tailoring the gazetteer and the entire linked data vocabulary to better meet their own particular needs. Pleiades welcomes your contribution, no matter how small, and we have a number of internal projects to suggest.

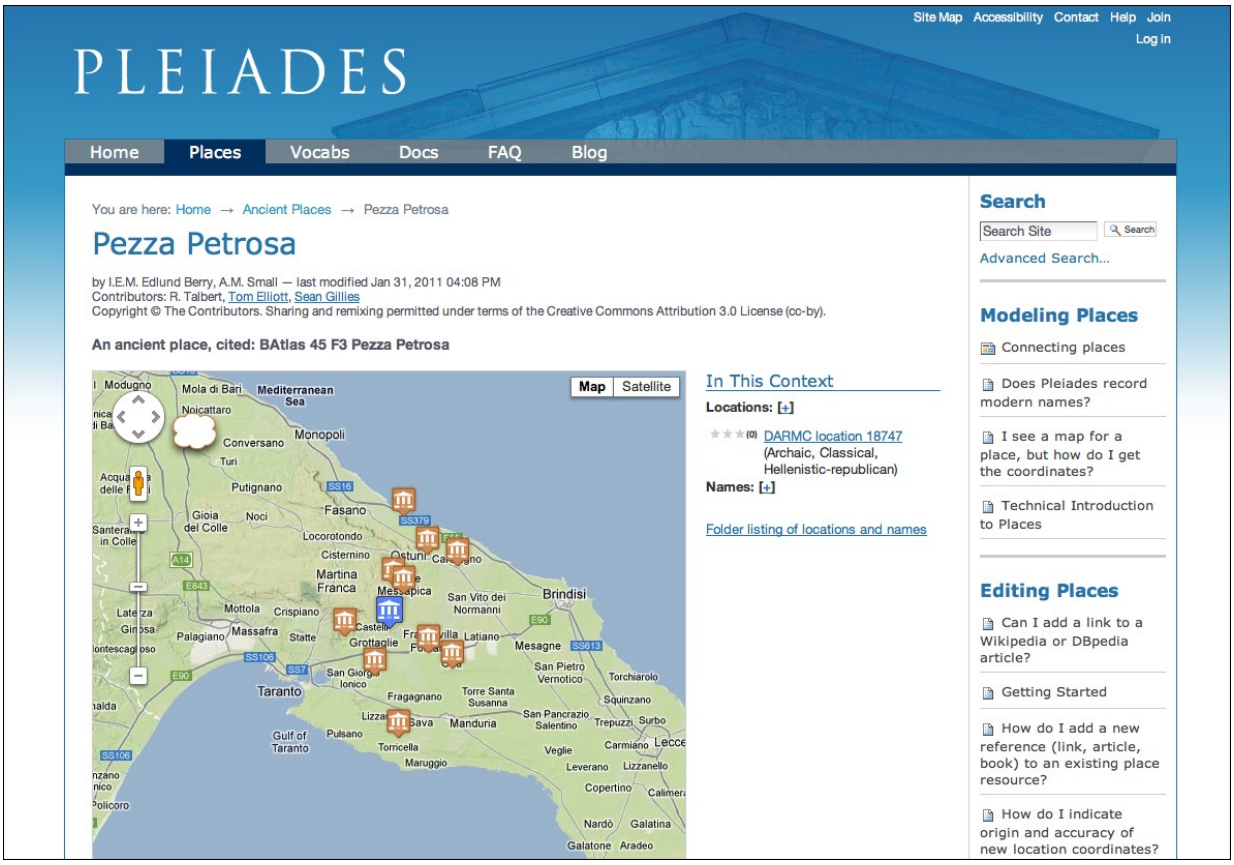
More about contributing ...

Pleiades is supported by:

National Endowment for the Humanities

User interface upgrade: Maps on place resource pages

Key improvements to the map and user interface are now in production.



We've consolidated the map and listing of locations and names for each place and made the map larger. Maps are fully functional for unpublished places; for the first time you get a view of the published places in the neighborhood of your work in progress. Aggregations of roughly located places formerly shown as orange rectangles are now represented by a cloud marker that hovers near the place of interest. Mousing over the cloud reveals the associated box, and clicking on the cloud reveals a listing of the roughly located places.

User interface upgrade: New top-level tabs

The "portal tab" elements formerly appearing as "Home, Browse, Download, Contribute, Follow" are now replaced by "Home, Places, Vocabs, Docs, FAQ, Blog". The benefit for users is consistent highlighting of the proper tab when browsing the site and more direct access to documentation and pages about places.

User interface upgrade: New landing page for places

A new and improved landing page for the folder of places and a technical introduction have been added at http://pleiades.stoa.org/places. Users report finding the new introduction to the Pleiades model of place and the list of starting points for getting into the data very helpful.

According to our Google Analytics account, the bounce rate for /places/ since the change is 34%, down from 41% for the preceding 6 months. More visitors to that page are going on to find more information than before.

Performance enhancements

Performance enhancements: the open-source Shapely package

We have tried harder to communicate to our community exactly what Pleiades has gained from using and producing open source software. One of Pleiades' key software spin-offs is the Python package named Shapely (https://github.com/sgillies/shapely). Shapely's 1.2.10 release, which went into production for Pleiades during this period, included significant improvements in speed and functionality. The Pleiades team didn't author these improvements, but got them by virtue of having carved a general purpose open source GIS package out of the Pleiades software stack. We've cultured a growing open source software community around Shapely that multiplies the original investment in Pleiades funding. We were pleased to recently see Shapely used and cited by programmers from high-profile, innovative companies like Development Seed (http://developmentseed.org/) and SimpleGeo (https://simplegeo.com/).

Performance enhancements: site search improvements

Users reported being tripped up by the difference in live search results in the small search form and the listing returned after clicking its "Search" button. We've remedied the situation so that now the live results match the results from a form-submitted search. Furthermore, the * glob character applied automatically in the live search is propagated to the form and links in the search result page. This has the automatic effect of causing both search contexts to perform a substring search, rather than allowing the form-submitted search to default to a full-word-match search. Users report finding this arrangement more intuitive and effective.

Performance enhancements: more memory

We doubled the RAM on our server from 4 to 8 gigabytes. Pages load faster, which pleases users, but this improvement has also helped Pleiades places be almost entirely indexed by Google, which previously seems to have left off indexing Pleiades resources because of slow page loading speeds.

New features

New feature: comments (with OpenID support)

Support for commenting was added to Pleiades over the course of the reporting period in a series of graduated steps. Users are now able to make comments on news items (i.e., blog posts), as well as name, place, and location resources. These comments are queued for moderation by the editors prior to publication. The motivation for adding this feature was to

lower the barrier to participation and encourage a broadening of the community. Key to the implementation was the decision to allow anyone with an Open ID (see http://openid.net/) to get a limited role in Pleiades. They can create a profile and they can comment on resources (as can users with full-fledged Pleiades accounts). This makes it much easier for a casual visitor who notices an omission or opportunity for improvement to register same with the editors without waiting for approval of a formal account request while still protecting the site and its moderators from having to cope with spam bots. The Pleiades login page at http://pleiades.stoa.org/login_form has been modified to reflect and explain these two options, with a link to the registration instructions.

The RSS feed and listing of recent comments are linked from http://pleiades.stoa.org/news/comments. Reviewers can subscribe to the RSS feed to stay abreast of incoming comments. A listing of resources ordered by their number of contents, the most commented on at the top, can be had from the following search URL: http://pleiades.stoa.org/search?total_comments:int=1&total_comments_usage=range:min&sort_on=total_comments&sort_order=descending. The total comments index can also be used in collections criteria as shown in http://pleiades.stoa.org/Members/sgillies/topic-testing/comments-on-roman-period-resources.

New feature: connecting places

Geographically connected place resources can now be linked together even if their exact coordinates are unknown. This new feature is best explained by way of an example.

Pleiades does not currently have precise GIS-ready coordinates for the aqueduct that supplied *Nemausus* and CAD-grade representations of it that would permit architectural visualizations or hydraulic analysis are beyond the scope of this project. What Pleiades can express, starting now, is that the *Nîmes* aqueduct is geographically connected with the spring at *Ura Fons*, with the *Pont du Gard*, and ultimately with the city of *Nemausus*. Similarly, Pleiades can express geographic connections between rivers, ports, and unbounded bodies of water like the *Ledus fl., Lattara,* the *Taurus/Volcarum Stagna*, and the *Gallicum Mare*. As roads are imported into Pleiades, we will be using what structured information we have to create the same kind of connections in an automated way, forming networks and opening up some interesting avenues of discovery and exploration in Pleiades. "Connects with" does not mean only touching or adjoining, it also encompasses intersection, overlap, and containment. It is a very coarse-grained relationship.

The home page for any place lists the other places it makes a connection with and, separately, the other places that make a connection with it. For places that only just connect and have no overlap or part/whole relationship, the directionality is not important. However, evolving custom is that "small" places make connections to "large" places: a bridge connects with a river, a river connects with a sea. For overlapping places, hierarchy can be suggested by exploiting directionality of the connection: a pass should make a connection with a mountain range, a tribe's chief town should make a connection with the tribe's territory.

New feature: Ratings for locations and names

In order to further lower barriers to participation, we have made it possible for users to promote or demote names and locations within the context of their place. Pleiades is designed to allow for any number of names and locations in the context of a place, but has been lacking a

measure of relative significance or value that might be usefully conveyed to researchers or the public. Some features of the site itself stand to benefit from a weighing of objects: if a place were to be mapped or labeled using just one location or name, which would it be? The answer, we think, is the one most preferred by the Pleiades user community. There are ways to infer preference, but first and foremost we will actively solicit preferences using a new user interface element: a three star rating for each location or name. To accommodate ratings we've changed the layout of place pages, moving the map to the left and location and name listings to the right.

Our first automated use for ratings will be in the representative latitude and longitude values written into the daily table of places (see http://pleiades.stoa.org/docs/using-our-data/how-to-get-pleiades-data). The representative point will be taken from the place's highest average rated location. In general, we'll be emphasizing the average ratings, though it's clear that rating open the door for some kinds of personalized user experience (map only locations you've rated 1 or higher, for example).

Additions to content and coverage

Content and coverage: Expanded time periods

Requests from Pleiades users and collaborating projects have led the editors to expand the temporal range of Pleiades content. This has been done by adding to the thesaurus of named time periods on which users can draw when creating or updating name and location resources. A general-purpose medieval/Byzantine time period was added to accommodate expressions of later evidence that are relevant to the study of antiquity. Basic sub-periods for the Byzantine empire have also been added. ISAW's own Ancient World Image Bank project has been disseminating open-access photos of ancient sites and we plan to eventually link these closely to Pleiades. Some of these photographs depict sites and features in Egypt that pre-date the period of Greek and Roman engagement there, and so we have followed the chronology that Thomas Schneider's developed for the NEH-funded UCLA Encyclopedia of Egyptology, in order to ensure consistency across the data sets. Discussions about needed periods with the Alexandria Archive Institute team that is working toward adding Ancient Near East sites and names to Pleiades are underway. We are also taking part in a larger discussion with members of the GeoDia, PELAGIOS, and OpenContext projects[1] aimed at defining and prototyping a better solution for developing and maintaining a thesaurus of time-period definitions that can be used across projects in linked data formats. A full list of the named time periods currently supported by Pleiades may be found at http://pleiades.stoa.org/vocabularies/time-periods.

Content and coverage: Ancient Near East

From September 2011 we have begun consultations with the Alexandria Archive Institute's Gazetteer of the Ancient Near East project (http://alexandriaarchive.org/projects/gane/), which is separately funded by an NEH Level II Digital Humanities Startup grant. GANE intends to publish its data via the Pleiades platform, thereby integrating it with the material already there. Discussions so far have centered on modifications necessary to support the historical and modern name information, time periods, and languages and scripts required by the project.

[1] GeoDia: http://geodia.laits.utexas.edu/; PELAGIOS: http://pelagios-project.blogspot.com/; OpenContext: http://opencontext.org/.

Content and coverage: Missing places from Barrington map 87 imported

When preparing *Barrington Atlas* directory tables for import we confused the map 87 inset directory with the map 87 directory (Cimmerius Bosphorus and Pontus-Phasis). The arrangement of these directories differed from all the others in the *Barrington*. The latter were not loaded, and the former were loaded into the Pleiades identifier range intended for the latter. We've now loaded the missing places and patched them in to the identifier space. Our master identifier mapping table at http://atlantides.org/trac/pleiades/wiki/IdentifierMapping has been updated accordingly.

Content and coverage: Updated categorization of places

Hundreds of rivers, islands, lakes, capes, temples, monasteries, and monuments whose place types were originally set to “named feature” have been precisely categorized. Taking advantage of standardized abbreviations in Pleiades resource titles inherited from the Classical Atlas Project ("fl." for river, "Ins(s)." for island(s), "L." for lakes, "Pr." for capes, "T." for temples, "Mon." for monasteries, and "Monu." for monuments), we've newly categorized hundreds of Pleiades places. Details, including links to the scripts are found at: http://atlantides.org/trac/pleiades/wiki/PlaceUpgradesAndMigrations. The lake category is new, with lakes collected at http://pleiades.stoa.org/collections/lake.

Content and coverage: More locations from DARMC/BAtlas

In the *Barrington Atlas* feature data digitized by DARMC and shared with Pleiades, we identified a number of coordinate pairs that the DARMC team had been unable to match with Pleiades places. All of the DARMC-matched content had already been loaded to Pleiades during the previous reporting period. Of the remaining features, 2,937 were named (i.e., had labels on the Barrington maps) and another 6,171 were unnamed (just point symbols without labels). For the named features, some of the match failures resulted from label transcription errors during digitization or from alterations to label text because of DARMC presentation preferences.

During the current reporting period, Sean Gillies developed an algorithm to facilitate supervised, fuzzy matching of the remaining data. As a result, nearly 2,000 of the named features and another 2,000 unnamed features have now been matched and loaded. 732 named features remain unmatched. These derive mostly from Barrington Atlas insets, which for technical reasons, do not submit to programmatic matching solutions. We plan to marshal community participation to help us match these by hand in coming months. Features sourced by DARMC from the *Tabula Imperii Byzantini* and other non-Barrington sources have not been added to Pleiades. More discussion of the algorithm and its results may be found at http://atlantides.org/trac/pleiades/wiki/DarmcUnmatchedGridReconciliation.

Data portability, extramural interoperation, and linked data

Representative latitude and longitude added to daily tables

Each morning, we dump our catalog data into 3 tables: one for Location resources, one for Name resources, and one for Place resources. Each of these tables has a column holding the

relative (to http://pleiades.stoa.org) URIs of the resources and the Location and Name tables have a pid (place identifier) column that affords joins with rows from the Place table. In September, we added a representative latitude/longitude column to each table to make it easier for users to make simple, small scale maps with third-party GIS and mapping tools.

The representative point for a Location is the easiest to describe: it's the centroid of the resource's geometry attribute. The representative point for a Place is the centroid of all geometries of published subordinate Locations. To compute the representative point for a Name, we take the centroid of all temporally co-located published sibling Locations. More date information was also added to the dump files.

Collaboration with the Google Ancient Places (GAP) Project

GAP (http://googleancientplaces.wordpress.com/) is a collaborative effort, funded through one of Google's digital humanities grants, to find ways to mine and exploit materials (like books) that Google has digitized in ways that benefit scholars and students of the ancient world. GAP is using Pleiades data in the following ways:

- Pleiades+ is a toponym extension to Pleiades, based on GeoNames. Pleiades+ attempts to match Pleiades and GeoNames identifiers together based on both location and toponym. When such matches are found, the additional toponyms from GeoNames are associated with the Pleiades URI. http://googleancientplaces.wordpress.com/pleiades/

- The GAP dataset, which is available for download and reuse, expresses machine-identified relationships between tokens in texts and place resources in Pleiades in the form of Resource Description Framework triples in the N3 format, following the PELAGIOS system of annotations. http://googleancientplaces.wordpress.com/project-data/

- GAPVis is an online interface for reading and visualizing GAP texts and their machine-identified relationships with Pleiades. You can read more about it in blog posts here, here and here. http://nrabinowitz.github.com/gapvis

Collaboration with the PELAGIOS Project

Funded by the Joint Information Systems Council in the United Kingdom, PELAGIOS[1] (http://pelagios-project.blogspot.com/) is an international consortium that is using Linked Open Data approaches to join up a variety of online resources that refer to places in the Ancient World. PELAGIOS partners use the Open Annotation ontology and Pleiades URIs to express references to ancient places within their information resources. This not only allows them to draw on the contextual information that Pleiades provides, such as coordinates and bibliographic references, but permits cross-browsing between resource sets. Pelagios does not centralize infrastructure, it simply facilitates those wishing to publish Linked Open Ancient World Data on the Web, using the Resource Description Framework (RDF).

More details about PELAGIOS and its use of Pleiades content can be found at http://pleiades.stoa.org/docs/using-our-data/pelagios.

[1] PELAGIOS stands for 'Pelagios: Enable Linked Ancient Geodata In Open Systems'

User documentation and on-line help

We continue to expand our help and documentation resources. Highlights this period include the addition of a “new contributor quick start guide” (http://pleiades.stoa.org/docs/getting-started) and additional FAQs at (http://pleiades.stoa.org/faq/), some with links to short how-to screencasts, hosted via YouTube.