Final Performance Report
Grant Number: HK-230973-15
Pleiades 3
Thomas R. Elliott
New York University
29 November 2019

Narrative Description

For readers unfamiliar with the *Pleiades* gazetteer of ancient places (https://pleiades.stoa.org), a history of the project and a description of its content and scholarly mission has been provided in Appendix 1.

In my 2015 proposal for this project, I responded to then-new provisions in the *Implementation Grant Guidelines* addressing "substantive changes to the design, technical architecture, and dissemination and preservation strategies" of established digital humanities projects. The Foundation's generous response empowered NYU's Institute for the Study of the Ancient World and its partners not only to improve the responsiveness and reliability of the hardware and software that powers the *Pleiades* community's editorial and dissemination processes, but also to add new features requested by community members in furtherance of the gazetteer's mission. *Pleiades* now benefits from reliable, scalable software more closely tailored to its users' needs. We are also better positioned to address future technological and scholarly requirements, so that we can continue to deliver innovation and value for years to come. Although some specific approaches outlined in the original proposal were changed in light of further analysis and on-going technological change, the full suite of anticipated outcomes has been achieved.

Project Activities

Major Activities

The proposal addressed project activities on pages 3-5 under the rubric "Innovations: Methods and Digital Technology". Additional activities were enumerated in the "Data Management Plan" (pp. D1-D2).

The bulk of project activity was devoted to software upgrade, development, and testing aimed at securing two key objectives. First, we sought to address egregious and growing performance issues that hampered not only our content creation and review process but also data dissemination and archiving. Second, we collaborated with our world-wide user community to identify, prioritize, and implement software improvements and new features aimed at streamlining editorial process, tailoring our data model to community research needs, and facilitating reuse and remixing of *Pleiades* in third-party systems and projects.[1] Other project work (administrative and travel/meeting) supported the primary activity.

Omissions and Changes in Project Activities

Unnamed contractor: By the time the project was awarded, I had identified my first choice for the "unnamed developer" whom I had intended should work "at a constant 25% level of effort for 25 months beginning in January 2016." I had called out the identification and hiring of this

[1] Work was managed using *GitHub.com*'s online, free issue-tracking application. Two-hundred fifty-nine discrete issues were addressed over the course of this project and may be reviewed here: https://github.com/isawnyu/pleiades-gazetteer/issues?q=is%3Aclosed+updated%3A%3E2015-08-31.

individual as one of the project's major risks and, indeed, by project start the person I wished to hire had moved from doing freelance work to a full-time position at Duke University's Collaborative for Classics Computing (DC3). Fortunately, DC3's mission and scheduling were compatible with taking on a revised version of the work originally proposed, and therefore NYU requested and received an award modification from the Foundation that permitted us to issue a subaward to Duke and transfer some other software development tasks to me along with a modest increase in effort. This arrangement worked well. Unspent subaward funds have been returned to the NEH.[2]

User Focus Sessions: I proposed a series of invited focus sessions with key users to be held alongside major disciplinary conferences during the period of performance. Funds were awarded to support these meetings, whose purpose was "talking over successes and challenges, evaluating the effects of recent software changes, and setting priorities for the next round of development." I chaired the first such meeting was held in San Francisco in January 2016, concurrently with the Annual Meetings of the Society for Classical Studies and the Archaeological Institute of America. Due to weather disruptions, and incompatible schedules on the part of desired participants, subsequent events could not be organized. I compensated by consulting with users electronically, both individually and in groups. I also organized meetings alongside other, unrelated travel, to meet with key users and groups as detailed in our biannual reports. Unspent travel funds have been returned to the NEH.

System Performance Upgrades: My proposal sketched a multi-step process for retooling the *Pleiades* web application (which uses the free and open-licensed *Plone* content management system, *Apache* web server, and *Varnish* reverse proxy cache) in order to address poor system performance resulting from overloaded, out-of-date software and intensive user demand.[3] I envisioned a two-part process aimed at realizing a more fine-tuned, scalable, and distributed system architecture capable of delivering significantly enhanced and consistent response to user interaction. First, I planned to upgrade the Plone CMS that supported both the consumer portions and the contributor/editor portions of the user interface. This software, which first "went live" in 2006 and received its last major upgrade in 2012, enables the *Pleiades* community to create, modify, and disseminate geographical information about ancient places and spaces. Then, in order to further reduce the load on the CMS, I intended to create a new, lightweight search-and-browse interface for read-only users. My hope was that this bipartite architecture – to be hosted on new hardware and tuned for performance using modifications to our *Apache* and *Varnish* Configurations – would do away with slow responses and system failures that had rendered both contribution and intensive use of the system nearly impossible.

After the project's contract software developers (*Jazkarta, Inc.*) completed initial upgrade steps for the *Plone* web application and migrated it to a new, better equipped server prepared by our long-time hosting provider *Tummy.com*, they instrumented the web application using the *New Relic* service in accordance with the plan outlined in the original proposal. As hoped, this instrumentation provided data about the portions of the web application that were placing the

[2] See separately submitted Final Financial Report.
[3] *Plone* Enterprise Content Management System: https://plone.org/. *Apache* HTTP Server Project: https://httpd.apache.org/. *Varnish* HTTP Cache: http://varnish-cache.org/.

heaviest loads on system resources like the central processors, memory, and disk drives. Once we started resolving these performance bottlenecks, it became clear that by downscaling the second proposed step to only replace the *Pleiades* "front page" and its search mechanisms, we could further improve the experience for our entire audience without complicating future maintenance and upgrade tasks by introducing a second, largely duplicative, code base.

During this process, it became clear that a scripted, repeatable mechanism for configuring a new *Pleiades* test or production server and deploying the myriad web application and middleware components would be valuable. We selected the *Ansible* provisioning tool for this purpose. Jazkarta modified and extended a widely used *Ansible* "playbook" for deploying Plone-based sites in order to manage every aspect of configuring and installing a production *Pleiades* server. They successfully used this playbook in April 2019 to migrate to another new server, this one equipped with solid-state drives. This move, which went smoothly and quickly thanks to our Ansible playbook, secured even more performance gains, especially for database-intensive tasks like complex user queries constructed through our Advanced Search form.

Accomplishments

System Performance: Both quantitative and qualitative assessments demonstrate the success of our efforts to produce a more reliably responsive system for anonymous, read-only users and authenticated contributors, reviewers, and editors alike. *Google Analytics* metrics demonstrate improvements in three key measures for user experience as demonstrated in Table 1.

Table 1: Page Speed Metrics from Google Analytics[4]

Measure[5]	Before upgrade	After upgrade	Improvement[6]
Average Server Response Time	4.2	0.97	78%
Average Page Download Time	0.21	0.12	43%
Average Page Load Time	8.48	3.18	63%

It should be noted that these performance improvements have been achieved while simultaneously removing restrictions on programmatic access to our data that I had put in place

[4] Actual improvements for the vast majority of users may be even better than asserted here. See the "Evaluation" section for caveats and limitations regarding *Google Analytics* data in the context of this project. In this table, "Before Upgrade" is the period spanning September 2015 – May 2016 and "After Upgrade" is the period spanning June 2016 – April 2019.

[5] Online help in *Google Analytics* defines "average server response time" as "the average amount of time (in seconds) your server takes to respond to a user request, including the network time from user's location to your server;" "average page download time" as "the average amount of time (in seconds) to download this page;" and "average page load time" as "the average amount of time (in seconds) it takes for pages from the sample set to load, from initiation of the pageview (e.g., click on a page link) to load completion in the browser."

[6] "Improvement" is calculated here as the difference between the later and earlier values, divided by the earlier value and presented as a percentage.

prior to this project in order to protect system performance for human users. *Pleiades* is currently responding to between 500 and 2,000 "bot" requests per hour, sometimes clustered in short, intense periods during which multiple bots are requesting multiple resources nearly contemporaneously. Of these, many are likely intended to keep search engine indexes up-to-date (I can identify, for example, up to 50% of these requests as originating from the GoogleBot, which updates Google's search index). According to *Google Analytics* data gathered between 2016 and 2019, over 60% of *Pleiades*' human site visitors come to the site by way of an external search engine result, demonstrating the importance of making and keeping the site responsive not only to the human visitors but to the bots that provide the data that helps them get there in the first place. The rest of our bot traffic comes from other search engines; from content access or harvesting programs written by external digital humanities researchers and projects; and from software whose purposes are more obscure.

Data Discovery and Reuse: In addition to *Pleiades*' enhanced capacity to support search engine indexing as discussed in the preceding paragraph, new tools to facilitate on-site discovery and reuse of *Pleiades* data were also produced by the project. Under subaward, our collaborators at Duke University's Classics Collaboratory (DC3) developed both a new search interface for our front page and a reconciliation service for *Pleiades* users. The new search form uses client-side code to provide users with fast place-name searches and intuitive result lists, all without placing any load on the *Pleiades* CMS or other middleware. Subsequent work by Jazkarta developers during the period of performance improved the overview and result-set maps on this page, providing even faster response, a more modern look, and lower pricing from our basemap provider, *Mapbox.com*. The reconciliation tool (dubbed "Geocollider") permits individual users or third-party software code to upload a file containing placenames and/or longitude/latitude coordinates, set matching parameters, and get back a file of possible matches. It also provides an online Application Programming Interface (API) compatible with the *OpenRefine* data-cleaning tool, which is widely used in digital humanities and digital library circles.[7]

Software Enhancements: Several user-requested enhancements have been made to the *Pleiades* content model and to associated database, middleware, and user interface software. Chief among these is the introduction of a new class of *Pleiades* resource: the "Connection." Connections join Places, Locations, and Names in the *Pleiades* data model.[8] They are designed to record typed relationships between places (e.g., spatial relationships like abutment or direction of river flow; administrative relationships between provinces and cities; and more), together with temporal qualifiers and citation of related primary and secondary sources. Apart from the introduction of Connections, numerous small refinements to the user interface and to workflow management processes – too varied to be listed here – have also been implemented.

Data Export and Preservation: Regular deposit of a comprehensive, open data export from *Pleiades* into third-party archives and into the hands of other digital projects is a major deliverable outlined in the project proposal. In February 2018, the PI completed work on a "*Pleiades* Datasets" package for this purpose, building on capabilities established earlier in the period of performance. This package of data files and documentation is publicly available via

[7] https://openrefine.org/.
[8] See further, Appendix 2: The *Pleiades* Data Model.

GitHub, where its contents are updated frequently based on the latest publications in *Pleiades*. Formal releases of this data are made periodically and deposited at *Zenodo.org* (CERN's public Open Data repository), at the *Internet Archive*, and at NYU's *Faculty Digital Archive*.[9] Our aim is to establish a quarterly archiving schedule.

Additionally, we refined our data serialization routines so that both daily bulk exports and on-demand data serializations of individual place resources contain every attribute field on every published object (places, names, locations, and connections). In the proposal, we identified the JSON serialization, as well as the bulk-export-only CSV serialization, as the formats we would make comprehensive.[10] After consultation with frequent users of our export data, we elected to pass over the CSV and concentrate on the JSON, because of its ability to reflect complex, nested and interrelated data structures in a way that would be extremely cumbersome to achieve in CSV. The daily bulk export can be had via the *Pleiades* downloads page. JSON serializations of individual place pages are accessed by appending the string "/json" to any *Pleiades* place URI.[11]

Bulk Ingest and Other Editorial Process Tools: A bulk upload script has been created and a series of software tools have been (and continue to be) refined by the PI and other editors for ingesting, checking, and preparing external data for incorporation into *Pleiades*. An early version of this toolkit was used to add citations of pages on the *ToposText* website to over 2,000 *Pleiades* place resources in the later part of 2018.[12] Numerous other data ingest projects are now under discussion or in preparation.

Audiences

Pleiades performs for the ancient Mediterranean historical-cultural sphere (and, increasingly, beyond) all the functions of a traditional scholarly gazetteer, identifying and describing ancient places and spaces for the benefit of scholars, students, and the general public. But it does so on an unprecedented scale, continuously enabling and drawing upon the work of individuals, groups, and their computational agents as the hub of a growing international scholarly communications network.

We know anecdotally that *Pleiades* is used with undergraduates by instructors at a number of institutions for individual and class projects. Sometimes it is used as an information source (either as a springboard to relevant literature or as a source for spatial coordinates for a mapping project). Other instructors sometimes set students the task of researching and preparing new content for submission to *Pleiades*. Indeed, our Credits Page lists some individuals whose

[9] See, for example, *Pleiades Datasets* version 1.1: https://pleiades.stoa.org/news/blog/pleiades-datasets-version-1.1, with links.
[10] JSON: JavaScript Object Notation (https://www.loc.gov/preservation/digital/formats/fdd/fdd000381.shtml). CSV: Comma-Separate Values (https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml).
[11] Pleiades Downloads page: https://pleiades.stoa.org/downloads/. Example of a Pleiades JSON serialization: https://pleiades.stoa.org/places/395622/json. Pleiades "Place URIs": https://pleiades.stoa.org/help/what-are-pleiades-uris.
[12] https://topostext.org/.

contributions to *Pleiades* were made when they were undergrads (and, in at least one case, in high school).[13] Other contributors to *Pleiades* identify variously as "interested amateurs", non-professional enthusiasts, members of clergy, geospatial professionals, graduate students, post-doctoral researchers, independent scholars, high-school teachers, and university professors. At the time of this writing, *Pleiades* has over 600 registered contributors, of whom nearly 170 can boast at least one published resource in the gazetteer.

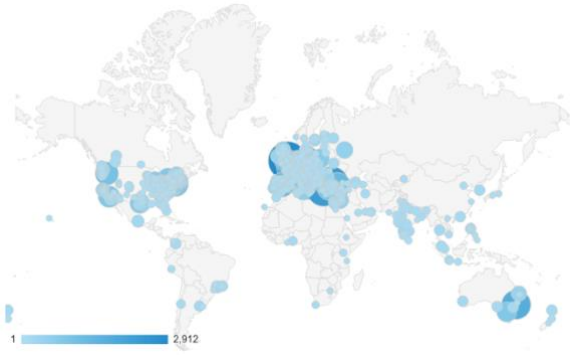


Figure 1: Density of Pleiades Users by City

During the period of performance, Google Analytics estimates that the HTML portions of the *Pleiades* website were visited by nearly 150,000 users who made nearly 230,000 visits to the site (sessions).[14] Twenty-seven percent of these users were based in the United States, 8.5% in the United Kingdom, 6% each in Italy and Turkey, 5% in Germany, 4% each in France and Spain, 3% each in Canada and Mexico, 19% in the rest of Europe, 8% in the rest of Asia, and 3% each in Africa and Oceania.

Evaluation

The project did not involve a formal, third-party evaluation component; however, a number of quantitative and qualitative means have helped me to assess its effectiveness.

Web Performance Measures*:* For each interim reporting period I used Google Analytics, backed up by feedback from our user community, to assess our progress in improving system performance. I focused primarily on two *Google Analytics* metrics: average server response time and average page load time. Significant overall improvement in these metrics, as well as a third (average page download time), over the period of performance is summarized in Table 1Table 1 on page 3. Nonetheless, it bears noting that the reliability of these metrics is limited in some unquantifiable ways, including:

- The small sample size of page-timing data received by *Google Analytics* (less than 1.2% of total pageviews);
- The filtering out of likely bot traffic from the totals, which *Google Analytics* does automatically, but which we seek to serve alongside our browser-using human audience;

[13] https://pleiades.stoa.org/credits.
[14] HTML: Hypertext Markup Language (used to create web pages for viewing in browsers by human users): https://en.wikipedia.org/wiki/HTML.

- The absence of a mechanism for assessing use of web resources other than HTML pages, such as our JSON and KML data serializations;[15]
- The growing use of browser-based privacy tools and "blocker" software designed to prevent Google from tracking web use; and
- The tendency of users to avoid site use during periods of system instability and unresponsiveness.

Given all these factors, I have tried to avoid becoming exercised when one or two website visits take so long as to affect negatively the Google Analytics metrics for an entire day, week, or month. Graphs showing the effects of such events on the "Server Response Time" metric are provided in Appendix 3: *Google Analytics* Data.[16]

Productivity Assessment: The productivity of *Pleiades* content creators, reviewers, and editors is another lens through which to view the effectiveness of the current project. Appendix 4: Editorial History Data provides a series of graphs illustrating the productivity of the *Pleiades* community over time with respect to total published content, "creation events" (in which a new place, name, location, or connection object is created), and "modification events" (in which a new object is corrected or enhanced). Three major periods in *Pleiades* content history may be discerned:

- The Golden Age of Ingest during which *Pleiades* was launched and its initial body of content created via data import *en masse* (2nd Quarter 2009 – 3rd Quarter 2013);
- A Renaissance that followed the first tranche of changes funded by this grant (3rd Quarter 2016 – 2nd Quarter 2019); and
- The Dark Times that fell between these periods: a dismal epoch during which poor site performance inhibited collaborative work and publication (4th Quarter 2013 – 2nd Quarter 2016).

Table 2 enumerates the quarterly average of "creation" and "modification" events that occurred during each of these three periods. It also presents the percentage improvement in each type of event from the Dark Times before the grant to the Renaissance it has brought forth.

[15] KML: Keyhole Markup Language (used by Google Earth): https://en.wikipedia.org/wiki/Keyhole_Markup_Language.

[16] NB: we removed all *Google Analytics* code from the *Pleiades* website in April 2019 in order to further improve system performance and to respect the privacy of our users, an increasing number of whom either expressed concerns about our use of *Google Analytics* or made use of browser-plugins that prevented *Google Analytics* from harvesting data about their visits.

Table 2: Curatorial Productivity

Period	Dates	Creation Events per Quarter	Creation Improvement[17]	Modification Events per Quarter	Modification Improvement
Ingest	2Q09-3Q13	5805		13542	
Dark Times	4Q13-2Q16	279		1797	
Renaissance	3Q16-2Q19	985	253%	2729	52%

User Feedback: I have not gathered user feedback in any sort of formal way, therefore assertions here should be regarded as anecdotal. Nonetheless, informal feedback from members of the Editorial College, as well as other frequent users, has been uniformly positive and encouraging. Users report increased confidence in the site and its future prospects, and are further encouraged both to use and to contribute content by our data export and archiving arrangements.

Continuation of the Project

New York University's Institute for the Study of the Ancient World continues to cover the annual costs of *Pleiades*' web hosting and to provide for a 10% research time commitment by the Associate Director for Digital Programs (Elliott), which he devotes to *Pleiades* management and programming work. The volunteer community of editors, contributors, and reviewers continues to grow and to become more active. External projects continue to reuse and remix *Pleiades* content, and to contact the Editorial College concerning potential collaborative projects, cross-citation, and data contribution. All of these factors reinforce our intent and expectation that the project should continue.

Long Term Impact

Pleiades already has an established reputation and function within the landscape of digital humanities projects and the community of digital practice surrounding ancient world studies and linked digital gazetteers. We believe that the statement of Humanities Significance contained in our proposal remains valid. The current project has reinforced perceptions of the reliability and flexibility of *Pleiades*, encouraging existing and new users alike to participate. Thanks to this investment in technological stability and publishing infrastructure, it is our expectation that funders will see in future multiple proposals for *Pleiades* content and analysis work focused on specific research questions and information needs at the intersections of ancient geography, history, and archaeology.

[17] See note 6.

Award Products

Open-licensed code and data related to grant products mentioned above are published on-line as follows:

- Content Management System, Middleware, and Content Model: https://github.com/isawnyu/pleiades-gazetteer and other repositories references there, especially:
  - https://github.com/isawnyu/pleiades3-buildout
  - https://github.com/isawnyu/PleiadesEntity
- Front Page and Simple Search Interface: https://github.com/isawnyu/pleiades-frontpage
- Exports and Serializations:
  - Code
    - CSV: https://github.com/isawnyu/pleiades-dump
    - JSON: https://github.com/isawnyu/pleiades-json
    - KML: https://github.com/isawnyu/pleiades-kml
    - RDF: https://github.com/isawnyu/pleiades-rdf
  - Data (in bulk) via https://pleiades.stoa.org/downloads/
- Bulk Ingest: https://github.com/isawnyu/pleiades3-buildout/blob/jazkarta-plone4/scripts/batch_update.py
- Automated Deployment with Ansible: https://github.com/isawnyu/pleiades-ansible
- Reconciliation:
  - https://pleiades.stoa.org/news/blog/introducing-geocollider
  - https://www.paregorios.org/posts/2017/10/using-openrefine-with-pleiades/
  - https://github.com/ryanfb/geocollider
  - https://github.com/ryanfb/geocollider-sinatra

Appendices

Appendix 1. Description of *Pleiades*

Like any good gazetteer, *Pleiades* (http://pleiades.stoa.org) is an organized spatial reference work for use in research, publication, and teaching. Jointly operated by the Institute for the Study of the Ancient World at New York University and by the Ancient World Mapping Center at the University of North Carolina at Chapel Hill, it constitutes the most accurate and comprehensive geographic dataset for the ancient Mediterranean world, identifying and describing nearly 35,000 places, spaces, territories, and regions.[18] It serves scholars and students alike as a ready reference and a guide to associated bibliography for places encountered in primary sources and secondary literature. As a database fronted by a web application, *Pleiades* transcends the familiar structure of print gazetteers (alphabetical lists of placenames), to provide its users with multiple modes of discovery: thematic browsing, hyperlinks, and a search engine. *Pleiades* publishes not just for individual human users, but also for search engines and for the burgeoning array of computational research and visualization tools that support work in fields like computational linguistics, digital text encoding, computational text analysis, natural language processing, named entity recognition, and Geographic Information Systems (GIS).

The present project has its origins in the Classical Atlas Project (CAP), a 12-year effort that produced the *Barrington Atlas of the Greek and Roman World*. In acknowledgement of this heritage, *Pleiades* is named after the daughters of Atlas in Greek mythology. CAP, which had received funding from the NEH and other sources, redressed a critical gap in research tools for the Classics that had been identified in the early 1980s: historical cartography was so neglected as a subdiscipline of ancient studies that there existed no comprehensive, up-to-date scholarly reference atlas for Greek and Roman civilization.[19] An NEH Preservation and Access Research and Development grant (PA-51873-06, PI: Talbert) launched *Pleiades* in Chapel Hill in 2006.[20] It culminated in 2008 having funded conceptual and experimental work to develop the data model now used, proving our hypothesis that off-the-shelf GIS data models were inadequate for the task and that a custom database was required.[21] From 2008-2010, a JISC/NEH Transatlantic Digitization Collaboration Grant (PX-50003-08; PI Bagnall) with a team at King's College, London used the *Pleiades* prototype to explore modes of cross-project geographic linking. Its findings, adapted to the more suitable LOD model, formed the conceptual foundation for the Pelagios Project.[22] From 2010 to 2014, *Pleiades* moved into its implementation phase, supported by a combination of institutional funds and an NEH Preservation and Access Humanities Collections and Reference Resources grant (PW-50557-10, PI: Elliott). This grant facilitated the

[18] At the time of writing, *Pleiades* has published 36,865 records for places and spaces that include information about 32,784 associated toponyms and 39,985 associated spatial location geometries.
[19] Roger S Bagnall, ed. *Research Tools for the Classics: The Report of the American Philological Association's Ad Hoc Committee on Basic Research Tools*. Chico, Calif.: Scholars Press. 1980.
[20] Final reports from all referenced grants available at: http://pleiades.stoa.org/docs/reports/.
[21] Sean Gillies. "What's an Un-GIS?" *And a Laser in My Shoe*. November 1, 2010. https://web.archive.org/web/20101109124826/https://sgillies.net/blog/1055/whats-an-un-gis/.
[22] LOD: Linked Open Data: https://en.wikipedia.org/wiki/Linked_data#Linked_open_data.

complete digitization of the *Barrington Atlas* materials, incorporating work already done with separate funds by Prof. Michael McCormick and his students for Harvard's *Digital Atlas of Roman and Medieval Civilization*.[23]

Pleiades content is updated collaboratively by volunteers around the world who work under the supervision of a volunteer Editorial College, filling a gap in the scholarly communications fabric for the typically small, highly technical notes and bits of geographic data that are otherwise unlikely to see the light of day in conventional, narrative-focused publications.[24] Through *Pleiades* these volunteers also provide authority control for ancient places. Just as the Library of Congress issues "authority files" to help improve discoverability in library catalogs by standardizing subject terms, names, and titles, *Pleiades* assigns a stable identifying number to each place and to each record about a place (a "Place ID" or PID). The use of numbers, rather than "standard" or "preferred" placenames, permits *Pleiades* to itemize places that never had a formal name in antiquity (e.g., a particular courtyard, road segment, or bridge) or whose ancient name has been lost from our sources over the intervening centuries. The PIDs are incorporated into the page address in a uniform way that is easy to remember and to encode, thereby forming for each place a Uniform Resource Identifier (URI), a standard mechanism for identifying documents and objects on the world-wide web.[25] So, for example, *Pleiades* assigns the PID 29573 to an ancient region known as "Gedrosia". The URI for the *Pleiades* page for this Gedrosia is http://pleiades.stoa.org/places/29573. This practice facilitates citation and linking for individual users and establishes a simple, permanent reference number for each place that can be used anywhere such references are valuable.

In fact, by establishing an authoritative gazetteer and associated open information services, *Pleiades* helped start a revolution: a constellation of online publications treating the histories, languages, texts, and artifacts of antiquity that make use of *Pleiades* to contextualize their holdings interoperably across the Web. Operating under the aegis of the UK-based Pelagios Commons, over 40 teams from 8 different countries have incorporated *Pleiades* URIs into their datasets, thereby identifying places mentioned in texts and the origins and findspots of artifacts.[26] Several of these projects use *Pleiades* data to provide their users with links, dynamic maps, and other services. All of them publish the pairwise matches between their own records and *Pleiades* place records (over a million so far) as open data for other parties to use and other systems to harvest. *Pleiades* uses this data too, creating lists of related content on every place page by comparing the URI for the place in question against the data published by Pelagios partners.

[23] http://darmc.harvard.edu/icb/icb.do.
[24] Editorial policy: http://pleiades.stoa.org/docs/editorial-guidelines.
[25] More specifically, Pleiades uses HTTP URIs: Leo Sauermann and Richard Cyganiak, eds. *Cool URIs for the Semantic Web*. W3C Interest Group Note. 2008. World-Wide Web Consortium (W3C). http://www.w3.org/TR/cooluris/.
[26] Pelagios partners and their datasets: http://pelagios.dme.ait.ac.at/api/datasets.

Appendix 2. The *Pleiades* Data Model

Pleiades is a gazetteer of ancient places. Most content in the gazetteer is represented using four classes of information structure: **places**, **locations**, **names**, and **connections**.

Pleiades **places** are the primary organizational construct of the gazetteer. They are conceptual entities: the term "place" applies to any locus of human attention, material or intellectual, in a real-world geographic context. A settlement mentioned in an ancient text is a **place**, whether or not it can now be located; an archaeological site is a **place**; a modern city located atop an ancient settlement is a **place**. Basically, any spatial feature that is connected to the pre-modern past and that a human being has noticed and discussed as such between the past and the present is a **place**.

Places in *Pleiades* can therefore represent:

- areas of fairly intensive human activity like settlements and sanctuaries;
- large-scale geological features known in antiquity like mountains, rivers, lakes;
- political, social, or cultural constructs like provinces and mining districts; and
- individual structures, when they have been referred to individually by ancient sources or modern scholars (e.g., the Parthenon, the Queen's Megaron at Knossos, the Basilica Iulia, the House of the Faun).
- Spatial extents or thematic groupings of places defined by modern scholars or administrative entities for purposes of analysis, description, reference, or heritage management

Pleiades recognizes a variety of place categories or types; new categories can be added as needed by the Editorial College.

Places are entirely abstract, conceptual entities. They are objects of thought, speech, or writing, not tangible, mappable points on the earth's surface. They have no spatial or temporal attributes of their own. A **place** can exist in name only in an ancient source, without any material correlate; conversely, an archaeological site can exist as a **place** without an ancient name.

The spatial aspects of *Pleiades* **places** (i.e., latitude and longitude coordinates in space), as well as their ancient and modern names, are addressed through two other conceptual entities: **locations** and **names**. **Connections** are used to express and document relationships between different places. Temporal characteristics are recorded at the **name, location,** and **connection** levels as appropriate.

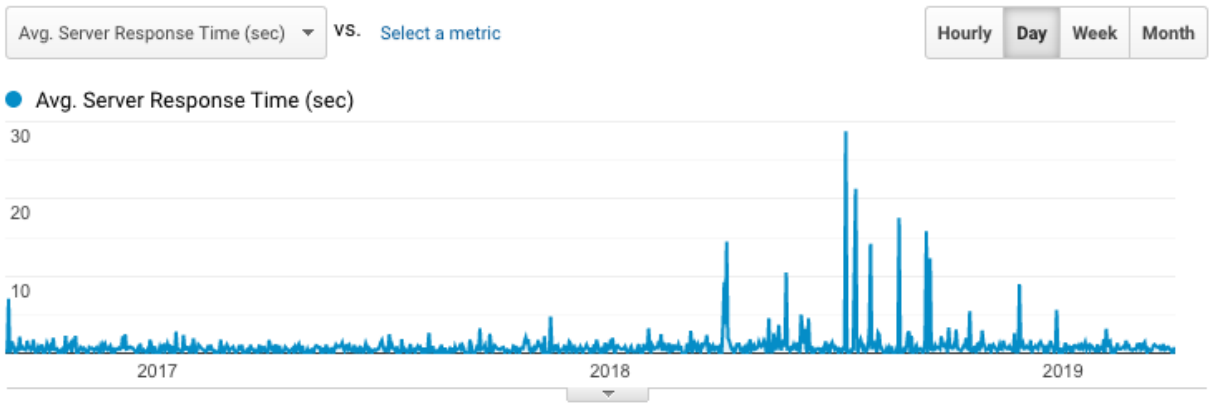
Locations in *Pleiades* connect **places** to coordinates in space. A **location** identifies a specific area of interest on the earth's surface that is associated with a **place** during a particular date range. A **place** can contain multiple **locations**. A **location**, on the other hand, is associated with one and only one **place**. Depending on the state of the evidence, the association between **location** and **place** may vary in certainty; some **places**, attested by name in ancient sources, may have no associated **location** at all because modern scholarship cannot pinpoint reliably the ancient site or area in question.

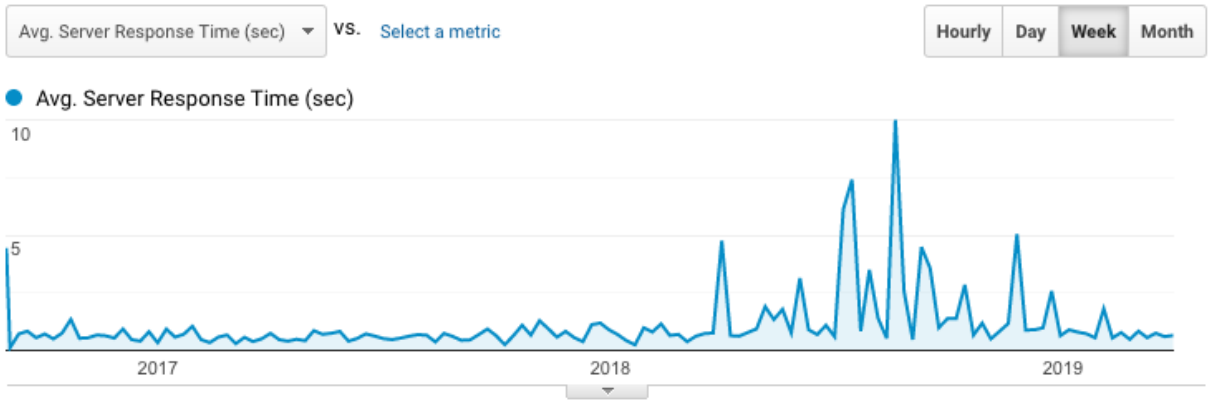
Names in *Pleiades* are also connected with **places**. A **name** reflects the identity of a **place** in human language, not its physical location in the landscape. **Names** have no spatial coordinates, but they are always annotated with the time period(s) of the textual source(s) in which they are attested. As with **locations**, a single **place** can have multiple **names**, but an individual **name** can be associated with one and only one **place**. This is true even if the same sequence of characters is also attested as a **name** for another **place**; *Pleiades* treats these "identical" **names** as separate entities.

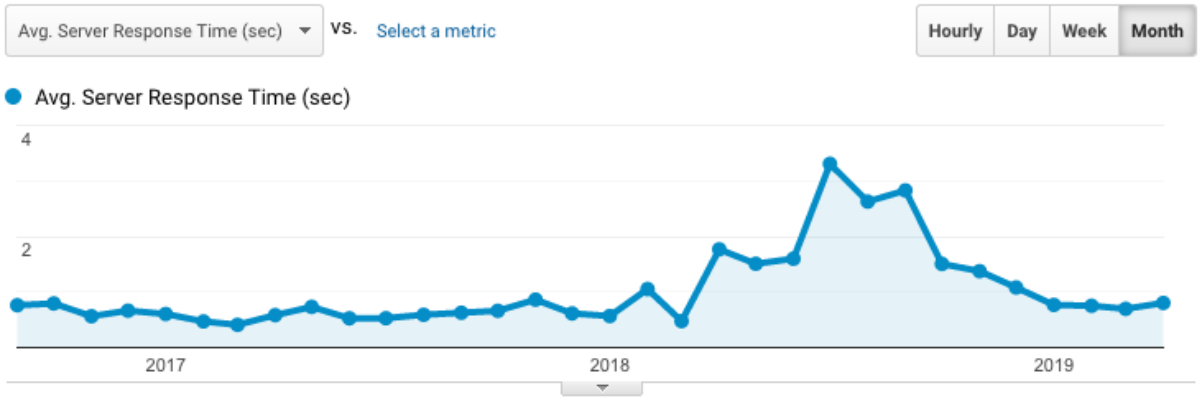
Connections are direct place-to-place relationships allowing the expression and documentation of geographic hierarchies, networks, and linkages. Like **names** and **locations**, they can be bounded in time and justified or explained by reference to ancient evidence and modern scholarly argument. Partitive, flow, and proximity connections are preferred over bounding boxes, convex hulls, and the like as mechanisms for creating spatial footprints for otherwise unlocatable **places** and for **places** with uncertain extents or unmappable boundaries. Temporal, political, economic, and analytical **connections** can also be expressed. The *Pleiades* editorial college maintains a growing vocabulary of connection types that provides standard terms and definitions for the classes of relationships that *Pleiades* contributors are entering into the gazetteer.

Appendix 3. *Google Analytics* Data

Figures in this section illustrate the outsize effects of a small number of unusually lengthy user interactions (often due to a slow metered or rural overseas Internet connection) on Google Analytics estimates of a key performance metric. They are relevant to the discussion on page 6 sub "Web Performance Measures."







Appendix 4. Editorial History Data

Figures in this appendix were constructed using data extracted from the *Pleiades* content history data included in our comprehensive JSON export serialization.

Figure 2: Content Growth (Lifetime)

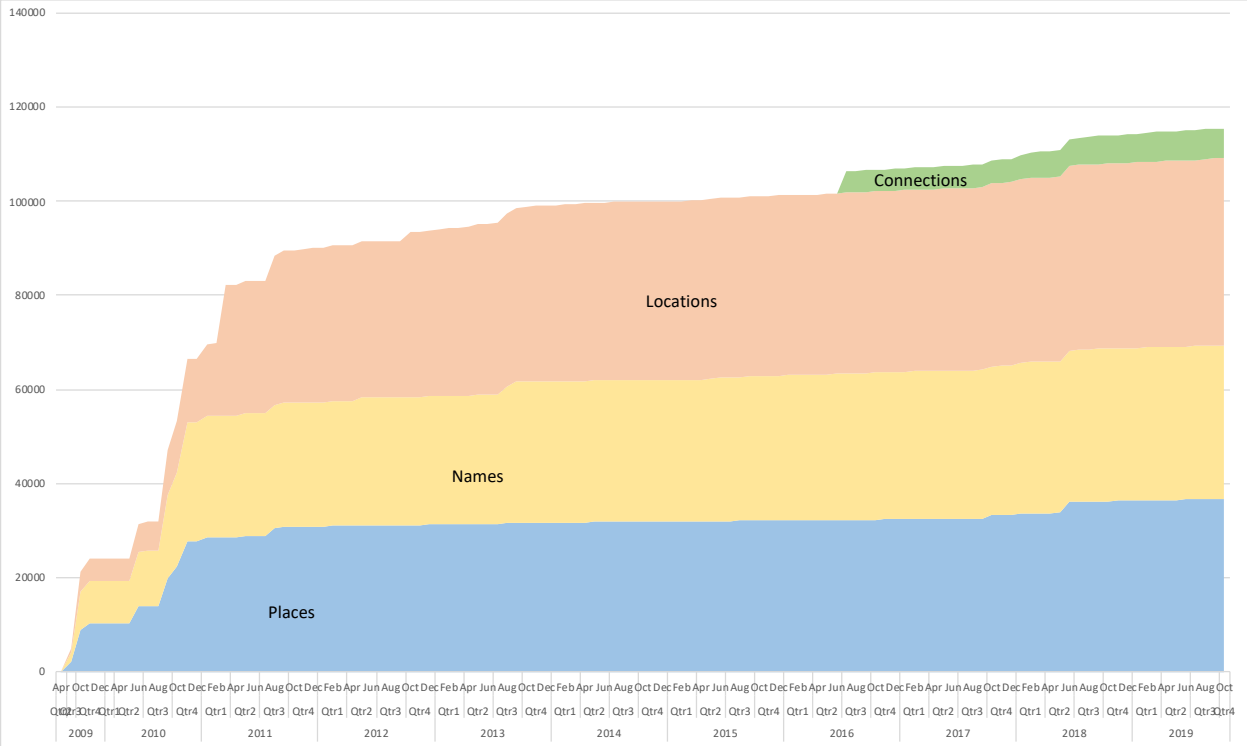


Figure 2 illustrates the growth of *Pleiades* content over the life of the gazetteer to date. Initial publications of place and name resources during the years 2009-2010 reflect the programmatic ingest of *Map-by-Map Directory* content from the *Barrington Atlas of the Greek and Roman World* (R. Talbert, ed., Princeton, 2000). Location resources added during later 2010 and early 2011 reflect the accession of coordinate data derived from the *Barrington Atlas* maps by the *Digital Atlas of Roman and Medieval Civilization* project at Harvard University, as well as data developed by staff and scholars at the Ancient World Mapping Center at the University of North Carolina at Chapel Hill. More modest concurrent and subsequent increases in content numbers reflect the work of individual contributors, reviewers, and editors working through the *Pleiades* website, with four notable upticks corresponding to programmatic interventions. Import of better coordinate data for some places developed by the *Digital Atlas of the Roman Empire* (DARE) project produced a jump in total locations in October 2012[27]. Addition of cross-references to DARE and to other resources catalogued by DARE resulted in a contemporaneous spike in place resource modifications. In mid-2013, a significant number of names and locations for sites in the Ancient Near East were added thanks to the work of the NEH-funded Gazetteer of the Ancient Near East (GANE) project, carried out by the Alexandria Archive Institute.[28] These major moments in *Pleiades* history are also evident in the subsequent 3 figures, which analyze types of editorial activity, rather than their end result.

27 http://dare.ht.lu.se/.
28 https://alexandriaarchive.org/gane/.

Figure 3: Editorial Overview (Lifetime)

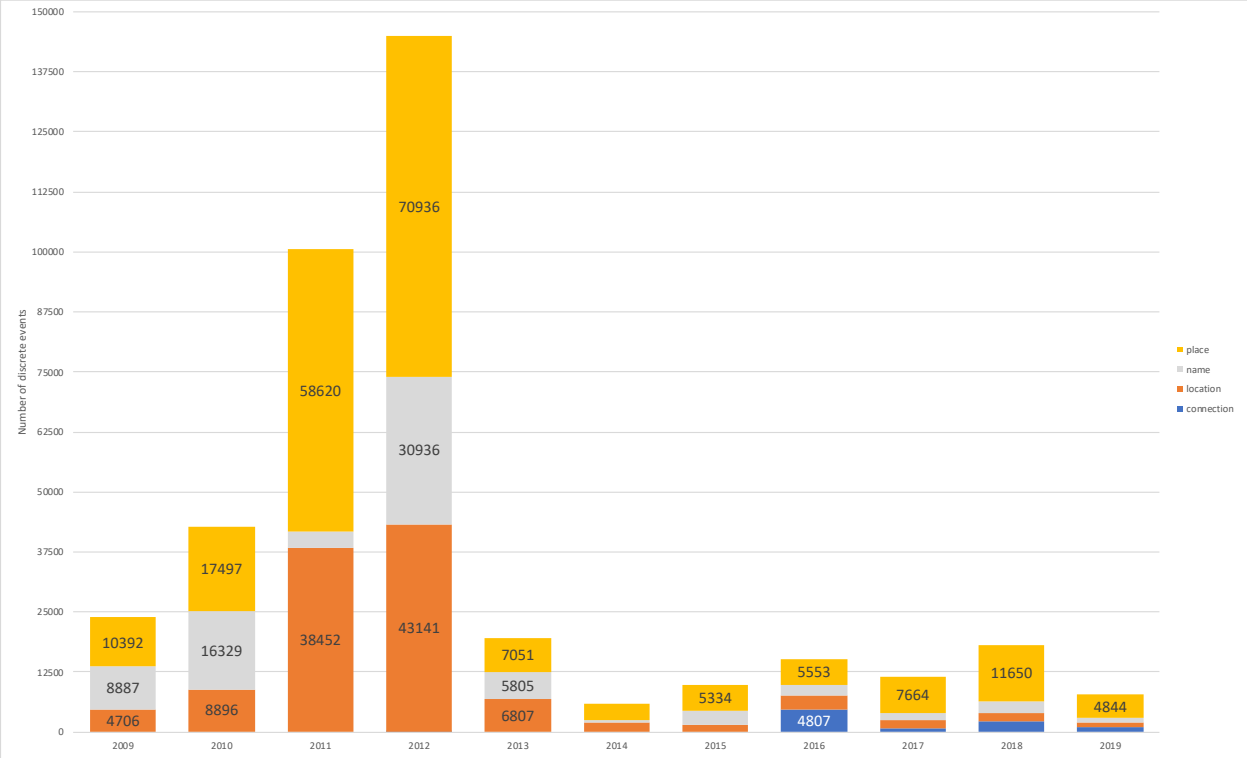


Figure 4: Content Creation (Lifetime)

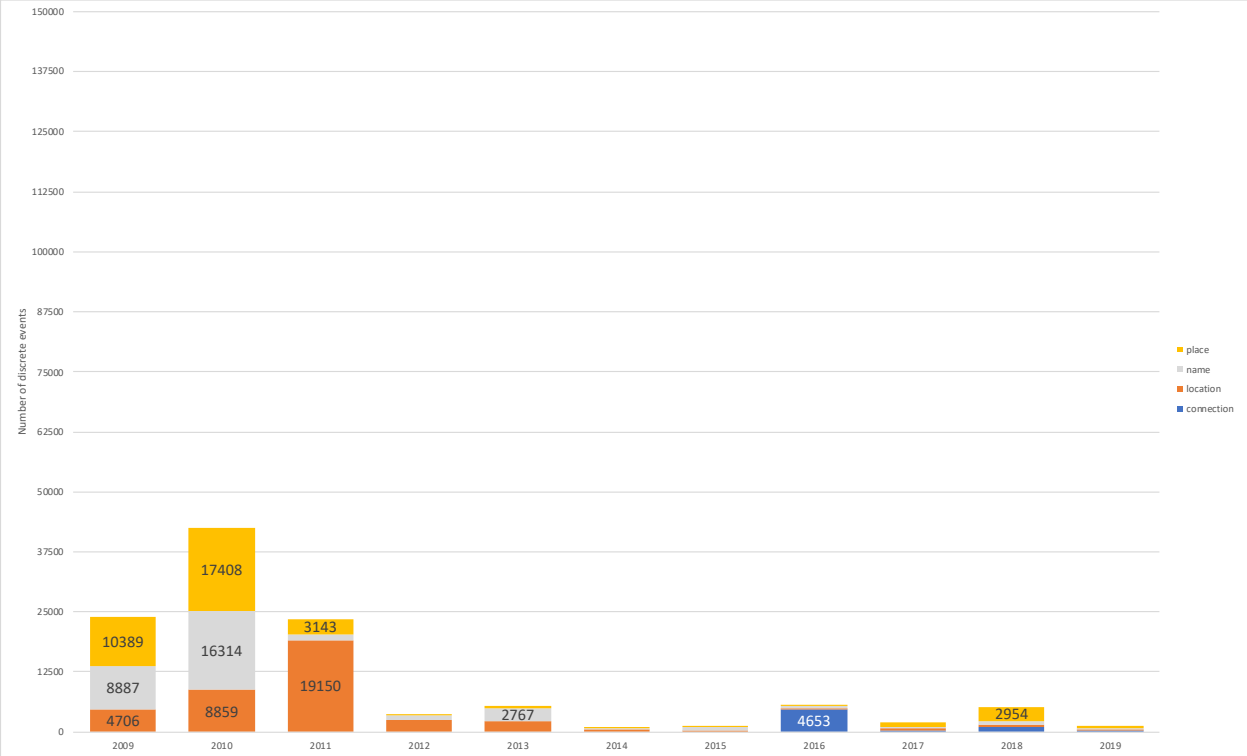
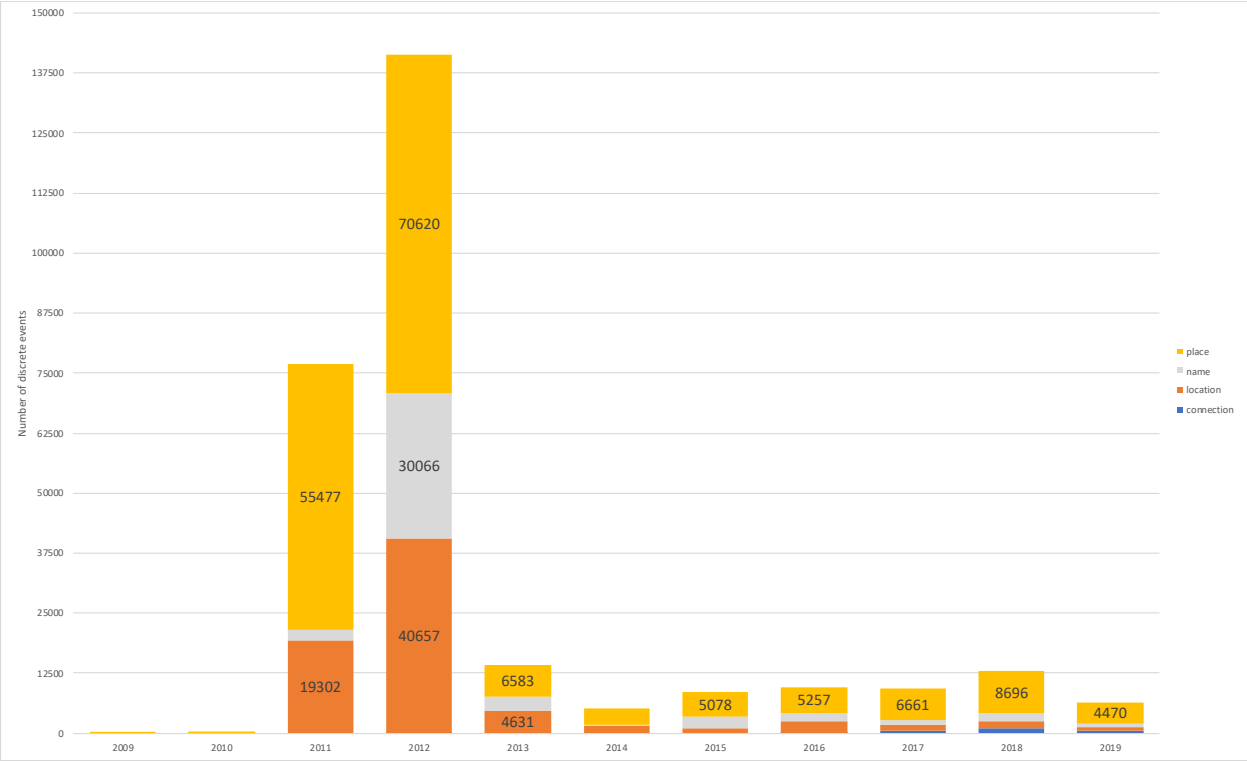


Figure 5: Content Modification (Lifetime)



The current project's period of performance began in September 2015. The sudden introduction in Figure 2 of a significant number of "Connections" in the third quarter of 2016 corresponds to grant-funded work that added the "Connection" type to the *Pleiades* content model (q.v.). Rollout of that modification included the migration of prior data held in individual place resources as new connection objects. Figure 8 highlights this moment. Subsequent increases in the creation of all content types reflect coordinated effort by the *Pleiades* community to bring a number of monuments and structures into the dataset, as well as general and varied content curation. A complementary surge in modifications to places in the second quarter of 2018 resulted from the ingest of data prepared and shared with *Pleiades* by the *ToposText* project (see especially Figure 9).

Figure 6: Editorial Overview (Period of Performance)

Figure 7: Content Creation and Modification (Period of Performance)

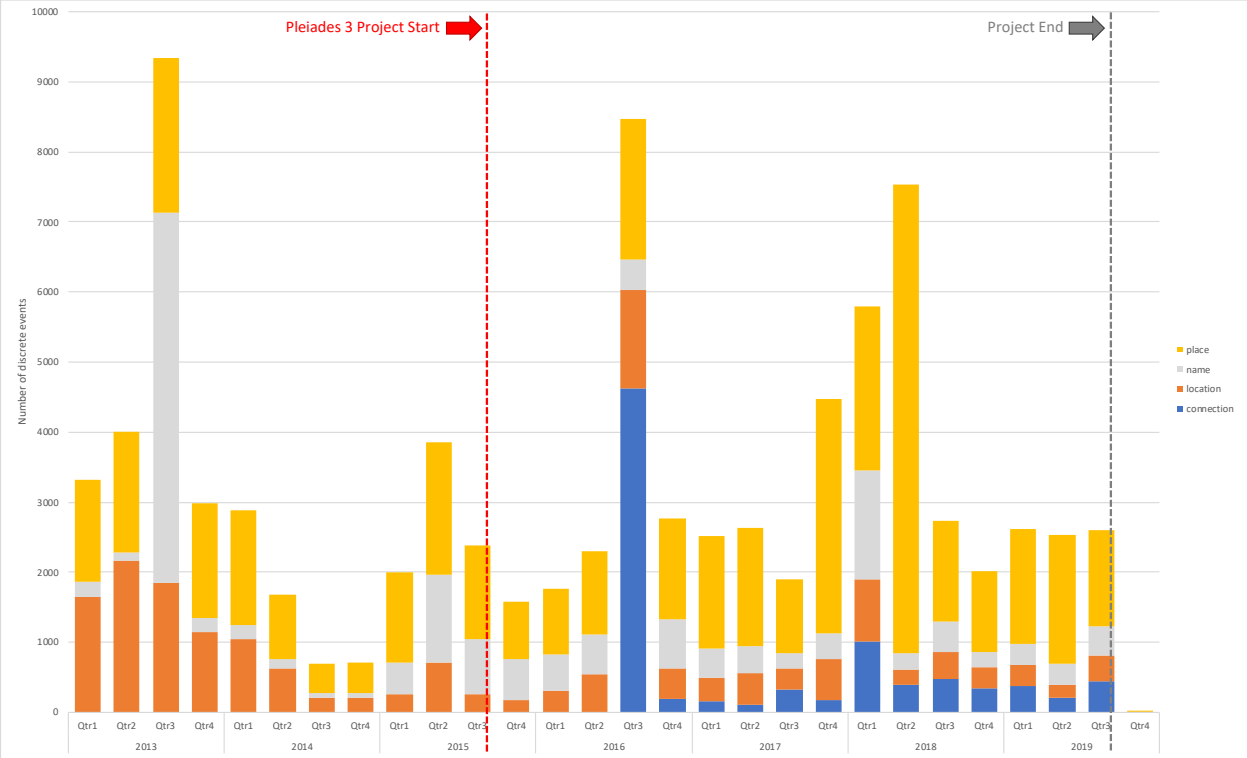


Figure 8: Content Creation (Period of Performance)

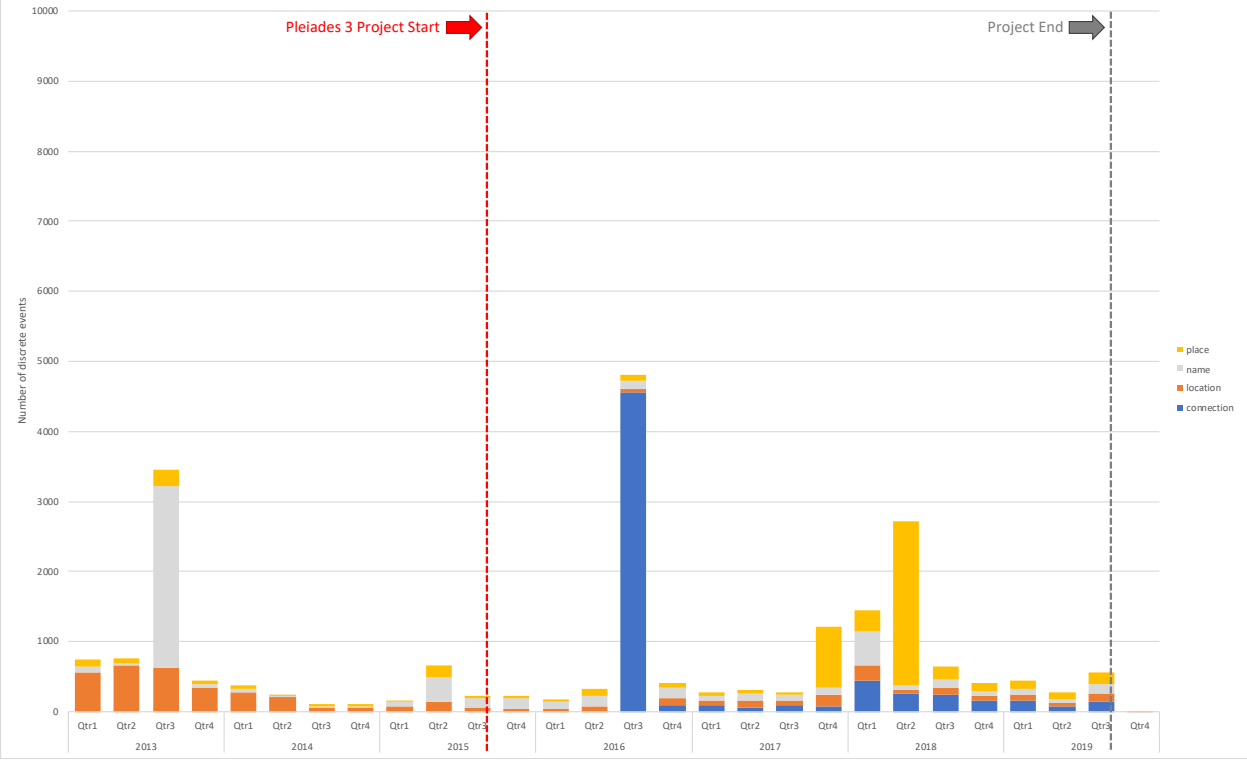


Figure 9: Content Modification (Period of Performance)