

Final Performance Report

PA-51873-06

Pleiades: Creating an Interactive Internet Archive for Ancient Geography

Richard J.A. Talbert

University of North Carolina at Chapel Hill (UNC acct # 5-41218)

September 29, 2008

Introduction

Pleiades (<http://pleiades.stoa.org>) was envisioned as a means of giving scholars, students and enthusiasts worldwide the ability to use, create and share historical geographic information about the Greek and Roman World in digital form. It was initiated by the Ancient World Mapping Center at the University of North Carolina, Chapel Hill, to consolidate, update, preserve and expand the achievements of the Classical Atlas Project, a 12-year, NEH-funded effort originally organized by the American Philological Association to produce a comprehensive Classical atlas in print. That important goal had been achieved in 2000, with the publication of the *Barrington Atlas of the Greek and Roman World* (R. Talbert, ed., Princeton); however, scholars and students alike began almost immediately to request coordinated digital products (e.g. maps and geographic datasets), but these could not be provided readily due to the hybrid conventional/digital production process used for the Atlas (its compilation began a few years ahead of the geographic computing revolution now so evident). Technological and methodological change, already noted during the final years of the Classical Atlas Project, continued to quicken. Developments in desktop GIS, digital libraries and other areas opened up new research possibilities. It was also clear that the passage of time was threatening the completeness and currency of the information itself. These concerns, coupled with the realization that a small Center at a single state-funded institution could not hope to keep pace with all relevant developments and publications across the entire spatial, linguistic and methodological footprint of the *Barrington Atlas*, motivated Tom Elliott (then AWMC Director) and Richard Talbert (BAtlas editor and Pleiades project PI) to examine emerging modes of broad, web-based collaboration.

Project Activities and Accomplishments

The Research and Development proposal for the work reported here promised “the development, testing and evaluation of a spatially-enabled, multilingual community support system enabling interested parties worldwide to participate in the maintenance, diversification and beneficial reuse” of the Classical Atlas Project data, an unspecified subset of which was to be digitized as part of the initial effort. The work plan called for a two-year period of iterative design, development and testing (February 2006 – January 2008) that would climax in the final 6 months with the recruitment of new participants and the opening of the collaborative content management system to initial public use. A no-cost extension stretched the period of performance through June 2008.

Significant research, planning and preparation of data preceded the funded project. Time estimates and initial designs for major aspects of the work were based on the best forecasts of project staff and our expert Steering Committee. For the project to be completed on schedule and in the manner originally proposed, the following steps had to be completed without difficulty:

- Continued availability of key personnel
- Prompt identification and hiring of the software developer
- Creation of web mapping and gazetteer services, export functions for archival purposes, and Federal Geospatial Data Committee (FGDC) metadata standards
- Realization of the geographic data model in a spatially-enabled relational database, linking it with Plone and with appropriate visualization and analysis component plugins in Plone
- Porting of the AWMC bibliographic database from Microsoft Access to Plone
- Effective implementation of the candidate editorial workflow (based on the procedures of the Classical Atlas Project) in Plone

Shortly after the grant award was announced, it emerged that, for domestic reasons, Tom Elliott (AWMC Director) would be moving away from the Chapel Hill area permanently, and therefore could no longer continue as Director. The proposal had envisioned Elliott devoting 66% of his time to the Pleiades project in order to provide both project management and software design/development leadership. Given the key role Elliott had played in envisioning and preparing for Pleiades, the Principal Investigator (Richard Talbert) undertook to reorganize the project with Elliott as full-time Project Manager, a position he would exercise remotely as a "teleworking" employee of UNC-CH.

Informal identification of potential candidates for the software developer position began early; however, Sean Gillies came to Elliott's attention through a Steering Committee member only days before the award was publicly announced. By virtue of his unique blend of relevant skills and experience, Gillies soon emerged as the leading applicant for the position once it was authorized and could be formally advertised, but his existing consulting contracts precluded a hiring date prior to June 2006. Talbert and Elliott made the decision to wait until Gillies was available to fill the position, aware that an extension of the grant period would be required in consequence, as well as additional support to enable Elliott to continue working on the project through June 2008. Time has demonstrated the wisdom of this decision, as swift-moving, unpredictable changes in geospatial and information technology quickly undermined many of the specific implementation decisions that had been proposed. Gillies' imagination, technical acumen and connections to many aspects of the geospatial industry proved invaluable in successfully navigating this rapidly changing landscape.

Within months after work on the project began (and just as Gillies joined the team), Google Earth was released. Although many researchers familiar with the development and use of Geographic Information Systems had been aware of Keyhole Software and its virtual globe application, Google's purchase of the company and free distribution of an upgraded version of its software, backed by streaming distribution of high resolution earth imagery, served to revolutionize expectations about the display and manipulation of geographic information on the web. The advent of Google Earth is just one example of a major paradigm shift – not only in online mapping, but also in the nature, origins and uses of digital spatial information – that is still unfolding. A range of free mapping and spatial data services are now available, and these are increasingly easy to use in combination with information drawn from other sources on the web.

These developments led us to change the project's emphasis. Instead of implementing the suite of mainstream, enterprise-level GIS service protocols outlined in the original proposal, we elected to support the lightweight, easy-to-use formats associated with Google Earth and other emerging "neo-geographical" applications. This decision reflected our assessment that these applications would quickly achieve widespread use, which would then drive standardization and regularization of the associated file formats. Our estimation has been validated, not only by the near-ubiquity of Google Earth and online mapping tools, but by the recent adoption of the Keyhole Markup Language (KML; the encoding format used by Google Earth) as a recognized standard by the Open GIS Consortium (OGC), the same body that had elaborated the formats and protocols identified in our original proposal. To date, the vast majority of prospective users of Pleiades have been interested in the neo-geographical approach.

At the beginning of the project, Elliott had made the identification of a "static file format" for Pleiades data a significant priority. The goal was to adopt a standard encoding scheme or format that could be used to encode every aspect of the Pleiades content set completely and unambiguously. The resulting static files would furnish users with an easy way to move large subsets (or even the entire published collection) of Pleiades content into a desktop GIS or other system. Moreover, such a format would be readily amenable to time-stamped release, and to deposit into multiple digital archives as a preservation method. Preference was for an open, widely used standard. Consultation with spatial and archaeological data archivists during the grant period and subsequently has so far failed to identify a fully satisfactory

solution for a static export format, although a slowly growing number of archives are taking various GIS formats, at least under bit-preservation terms. The original proposal anticipated that the Alexandria Digital Library (ADL) Gazetteer format would provide such a vehicle, and it went so far as to propose that Pleiades would support the full ADL Gazetteer Protocol for the exchange of such data.

Unfortunately, development of this putative standard stalled with the retirement of its primary research leader (a Pleiades Steering Committee member), and it has not seen significant adoption outside its original application in the Alexandria Digital Library. Its complex and deeply hierarchical XML file structure lost ground in competition with the Open GIS (OGC) Consortium's detailed Geography Markup Language (in use by government and commercial GIS entities) and the lightweight KML format used by Google Earth. Moreover, many potential Pleiades users requested the well-known Shapefile format, which we would need to package with multiple data files containing tabular attribute data in comma-separated value (CSV) format. Ultimately, we elected to defer a final decision, and implementation of a solution, to a later phase of the project in hopes that a more appropriate standard approach would emerge.

After careful consideration, we departed from the original proposal in another technical area. Rather than implementing a hybrid system with content in Plone and geo-location data in another relational database, we chose to geo-enable the Zope Object Database (ZODB), which Plone incorporates. This decision made it much easier to test the geospatial aspects of our system: relatively low level development is easier to test than is integration of two different databases. This approach also brought us into alignment with a small but dynamic group of Plone and Python developers who were moving in a similar direction. Their adoption of our `zgeo.spatialindex` package has greatly helped us through contributions of code and robust testing.

We elected to defer full implementation of FGDC metadata to a later phase of the project, subject to recommendations by expert members of the community. The FGDC guidelines are more concerned with integrated datasets than granular data. Pleiades allows users to create their own datasets by freely aggregating the vetted locations, names, and places. It is not clear how to apply the FGDC metadata model comprehensively to this situation. Insofar as FGDC standards of description apply to specific aspects of our content (for example, horizontal accuracy and precision of location coordinates), we have incorporated them into our data model. We are also exploring visualization conventions that take metadata values into account, surfacing them to the user in intuitive ways (again, in light of published FGDC guidance). In particular, we are considering an upgrade of the KML we produce to include not only a coordinate pair for a point location, but also a circle whose radius corresponds to the nominal horizontal error in those coordinates, based on the precision and accuracy of the underlying data.

The migration plan for the bibliographic database was re-imagined after a series of attempts to model it effectively in Plone. One NEH evaluator of our proposal had commented on the complexity of the relational structure for the database supplied in the proposal (a database that was already in use at AWMC, but required porting to a more robust database system), and suggested that the eXtensible Markup Language might provide a more natural way to store bibliographic information. The object-oriented approach in Plone proved equally complicated, and we eventually elected a different approach entirely: this combines a simple model for bibliographic citations in Plone and a collection of web-facing bibliographic records authored in XML (using the Metadata Object Description Schema, or MODS, standard promulgated by the Library of Congress).

Our timetable for implementation of workflow in Plone incurred delays when we decided to wait for a major release of Plone that was slated to offer easier site customization, workflow definition and content versioning. This latter feature was particularly important because we have always viewed the ability to produce a change history for individual content items as an essential prerequisite for scholarly citation and verification. In the proposal, we had identified a Plone plugin that was then under active

development; we had intended to use this to support versioning. Subsequent to the start of the project, the Plone development community elected to cease work on this plugin, and instead it promoted versioning as a major feature in the 3.0 release. As we waited for this release, which finally appeared in late 2007, we put extra labor into low-level modifications to the Plone indexing and content management functions that will enable enhanced spatial search and spatial feature correlation in a second phase of Pleiades. We also followed the guidance of the Plone community in preparing our existing software for the upgrade to Plone 3. This work paid off: the transition from Plone 2.5 (our initial starting point) to Plone 3 was not difficult.

We committed significantly more time to refining and experimenting with our proposed editorial workflow than expected. The project proposal described a complex process involving multiple content states, hierarchical roles and regimented review steps. It reflected an idealized version of the mature Classical Atlas Project editorial workflow, adjusted for a then-hypothetical digital environment. However, as work on Pleiades advanced, it emerged that this candidate workflow could be fully implemented in Plone, but that it was surely too rigid and too complicated to serve the full range of needs of our contributors and editors. It would obstruct, rather than encourage, rapid piecemeal improvement of content, and overly burden editors with a series of mandatory, computationally mediated checklists. Through a series of prototyping exercises conducted in consultation with key Steering Committee members and Technical Observers, we devised a simpler editorial workflow that preserves the essential safeguards inherent in the Classical Atlas Project working methods, while providing the Pleiades Community with a fluid process for quickly making and vetting suggestions and additions. The Pleiades workflow of 2008 is radically simplified, consisting of the fewest number of content states, transitions, and user roles that can possibly work. Development cost to implement is reduced. Community and collaborative interactions are emphasized.

The “Dissemination” section of the proposal promised the preparation of two “white papers” in addition to the project final report. The first such paper was to detail the customizations and extensions to Plone necessary to achieve the desired workflow and content objects. The second was to evaluate the achievements of the project and the effectiveness of its approaches. We also committed the project leadership to discussion and presentation of the project at appropriate disciplinary conferences. We have satisfied the first commitment by continuously disseminating information about our technical approach and software releases via the project wiki and our personal web logs (see the Pleiades News Aggregator: <http://planet.atlantides.org/pleiades/>). We are currently satisfying the second commitment by posting a modified version of this report on the project wiki. Project staff have delivered papers on Pleiades in several venues (see appendices).

Community, Continuation and Impact

Despite a number of personal and institutional changes, the Pleiades Project team remains strong, and is maturing into a community of practice. During 2008, with the recruitment of Elliott to be Associate Director for Digital Programs at NYU’s Institute for the Study of the Ancient World (ISAW), the locus of Pleiades technical development has shifted there. ISAW is better equipped administratively and financially to ensure the full realization and long-term vibrancy of this critical initiative. The Institute expended internal funds so that Elliott could devote 20% of his time during the period February – April 2008 to the project. Pleiades’ chief software engineer (Sean Gillies) is also now an ISAW employee (from July 1, 2008). Elliott and Gillies are both presently engaged in work on a joint NEH/JISC-funded project with colleagues at King’s College, London, that is exploring mechanisms for web-based interoperability between Pleiades and several digital publications of epigraphic and papyrological texts (<http://concordia.atlantides.org>). For its part, AWMC remains deeply engaged in the design and realization of the Pleiades project. In writing a letter of support for a second-phase Pleiades proposal (see

below), Richard Talbert affirmed his commitment to serving as Pleiades Co-Senior-Editor (with Roger Bagnall, ISAW Director) and to the inclusion of Co-Managing-Editor duties in the portfolio of the AWMC Director. The untimely death of Ross Scaife occurred only 3 months before the end of the grant period. As editor of the Stoa Consortium at the University of Kentucky, Scaife had fostered pre-proposal planning for Pleiades, and had provided the development server used for the Pleiades R&D grant period. We had already been working with Scaife to plan for the transition of Pleiades from development to production, and so a transition from the Stoa development server to a more powerful, commercially hosted machine was effected smoothly with assistance from Scaife's Kentucky colleagues and students. AWMC is defraying the cost of this server through December 2008. Thereafter, ISAW will assume responsibility for Pleiades hosting costs. Recently, Elliott has begun polling members of the Pleiades Steering Committee and Technical Advisers Group to gauge their interest in continued affiliation with the project as its inaugural content creators and editors. The vast majority of those contacted so far have expressed enthusiastic support for the ongoing work.

As indicated in the preceding paragraph, a proposal for a second phase of Pleiades was submitted to NEH's Division of Preservation and Access in July 2008 under the Humanities Collections and Resources rubric. This proposal seeks two years' additional support to transform Pleiades from a promising research and development effort into an active component of the emerging cyber-infrastructure for the humanities. Accordingly, we mean to shift our focus to the maturation and growth of our international community of collaborators by accelerating the addition of legacy Classical Atlas Project information to the gazetteer; enhancing Pleiades features and data durability in response to community priorities; facilitating usability through documentation development and hands-on training; and planning for the sustainability and future growth of the resource. The proposal reflects not only the continued commitment of AWMC (outlined above), but also a significant cost-sharing from ISAW, which commits its own funds to the digitization of over half of the remaining Classical Atlas Project compilation materials as well as other project costs.

In the Pleiades proposal we highlighted an urgent need, widely acknowledged as crucial to the future of humanities scholarship: new ways to build and sustain up-to-date reference works. We argued that Pleiades, if successful, would provide a digital geographic reference work for classicists that met high standards of currency and scholarship. Moreover, we offered Pleiades as an innovative model for addressing the problem of durability in reference works across the breadth of the humanities. We held up the Barrington Atlas as a potential example of the threat: within only a few years of its appearance, published research findings had already begun to undermine its currency and completeness. This circumstance should certainly not be seen as a criticism of the value of the Atlas, or of the extraordinary collaborative effort and expertise that went into its development. Rather, the nature of scholarly publication causes the difficulty: new work builds upon and supplants earlier work. Comprehensive reference works like the Atlas have unique value as infrastructural components for entire disciplines, yet they are routinely so expensive to create in print that – for many fields in the humanities – revised editions cannot be produced at regular intervals. Unfortunately, even the majority of digital reference works in the humanities labor under similar constraints. They typically employ small core staffs of dedicated researchers and administrators or, at best, engage a limited group of widely dispersed collaborators, whose expertise must often be compensated. Recommendations for sustaining such projects often invoke “business models” and “revenue channels” as potential solutions to the problem, but subscription regimes, gated resources and restrictive license terms can prove awkward to implement effectively for specialist resources whose greatest value to scholarship may only surface when provided free to scholars and students; many potential users (especially at second-tier institutions and in developing countries) simply lack the resources to pay.

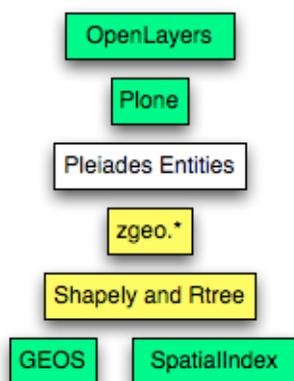
An alternative model that aims to drive down, or even eliminate, core content-related costs brings together elements of “open access” licensing and “long-tail” or “social” authoring approaches. The use of

Internet technologies to establish and maintain geographically and linguistically distributed communities – with membership open to a heterogeneous public – can significantly increase the workforce available to develop important content. Enthusiasts and students can be exposed to the approaches of trained scholars, and even work alongside them. Scholars in such communities benefit from the enthusiasm and collaboration of non-academically credentialed collaborators, who often adopt such valuable roles as translator, journal monitor, grammar checker or provocative questioner. Prompt publication of individual articles circumvents delays endemic to conventional publication projects; the latter must either await the completion of all content, or must adopt a serial publication approach (for example, in separate, alphabetically organized volumes). In the alternative model, the ease with which individual components can be updated further helps to limit the threat of obsolescence. As leading institutions begin to step up to the challenge of providing a humanities-oriented technical infrastructure, the use of common tools and services by such projects may further help to control costs and promote stability.

We still believe these assertions to be true, and appreciate NEH support to develop this new alternative model. We look forward to building on the solid foundation established by Pleiades in 2006-2008.

Grant Products

The Pleiades Project has delivered its core product under this grant: a working set of software, capable of supporting collaborative maintenance, update and web publication of historical geographic information derived from the Classical Atlas Project compilation materials and other sources. As originally proposed, this software is entirely open source. We use some components developed by other communities ("external components"), but we have also developed further components to meet the specific needs of the project ("Pleiades components"). We have sometimes found it necessary to make improvements to external components; wherever possible, we have contributed these modifications back to the originating community. Conversely, we have found – as predicted in the proposal – that some of our own Pleiades components are of interest to other communities of practice; as a result, we have received contributions of code and testing from software developers supporting those communities.



The software "stack" is shown in the diagram at left. The external components to which we contribute are shown in green. Components created by the Pleiades team are in yellow and white; components to which other entities contribute are in yellow. OpenLayers (<http://openlayers.org>) is the leading open source web map toolkit. Pleiades has modestly enhanced its features, and employs it to provide contextual maps in its web application. Plone (<http://plone.org>) is a leading open source content management system. Pleiades has made modest improvements to its vocabulary manager and to its user interface framework, and contributed these code improvements back to the Plone code base.

Pleiades Entities is the implementation of the data model for use in Plone. zgeo.* is a suite of Python software packages including: zgeo.geographer, zgeo.spatialindex, zgeo.atom, and zgeo.kml. These packages provide support for the Pleiades Entities component, and enjoy contributions from programmers employed by The Open Planning Project and Makina Corpus SA. Shapely and Rtree are general-purpose Python GIS software elements that support the zgeo.* packages. Shapely enjoys contributions from programmers and researchers employed by Campocamp SA, the University of California, and the National Oceanographic and Atmospheric Administration. GEOS and SpatialIndex are low-level libraries for geometry and spatial indexing

computing. Users include the PostGIS project and Autodesk. Pleiades has made modest contributions to each, and has helped SpatialIndex to become an openly developed project.

Pointers to (and information about) the components of the Pleiades software stack are available from the Pleiades Software Wiki page: <http://www.atlantides.org/trac/pleiades/wiki/PleiadesSoftware>.

Appendix: List of Papers and Presentations

In addition to an ongoing series of on-line blog postings (collected at <http://planet.atlantides.org/pleiades>), the Pleiades team has presented the project in the following contexts:

- Elliott 2006
T. Elliott, "Pleiades: Beyond the Barrington Atlas," paper delivered to the *London Ancient History Seminar*, 27 October 2006:
<http://www.atlantides.org/trac/pleiades/wiki/ElliottLondonPaper>
- Elliott 2007
T. Elliott, "Beyond the Barrington Atlas: An Online Workspace for Ancient Geography," paper delivered to the panel *Ancient Geography: New Discoveries and Perspectives* at the Annual Meeting of the American Philological Association, San Diego, 6 January 2007:
<http://www.atlantides.org/trac/pleiades/wiki/ElliottAPAPaper>
- Elliott 2007a
T. Elliott, "Heterogeneity," a session keynote paper delivered to *Space and Time: Methods in Geospatial Computing for Mapping the Past*, E-Science Institute, Edinburgh, 23-24 July 2007
- Talbert 2007
R. Talbert and S. Gillies, "Pleiades: Advancing Ancient Geography Online," paper delivered at *Using New Technologies to Explore Cultural Heritage*, a conference sponsored jointly by the National Endowment for the Humanities and the Consiglio Nazionale delle Ricerche (Italy), Washington, D.C., 5 October 2007:
http://www.neh.gov/digitalhumanities/Conference_07Oct/DH_Conference.html
- T. Elliott participated in the *Digital Gazetteer Research and Practice Workshop*, Santa Barbara, CA, 7-9 December 2007: <http://www.ncgia.ucsb.edu/projects/nga/ncgia.html>
- Elliott 2008
T. Elliott, "pleiades.stoa.org: open collaboration and the enhancement of historical geodata," a paper delivered to the *Workshop on Web-based research tools for Mediterranean archaeology*, 2008 Annual Meeting of the Archaeological Institute of America, Chicago, IL, 3-6 January 2008:
<http://pleiades.stoa.org/Members/thomase/my-news/slides-for-pleiades-talk-at-aia-2008>
- Elliott 2008a
T. Elliott, "Cyrenaican archaeology, epigraphy and geography: an approach for webfeed-based interoperability," a paper delivered remotely to *Inscriptions of Roman Cyrenaica: First International Workshop*, British School at Rome, 28-29 February 2008:
<http://horothesia.blogspot.com/2008/02/atomgeorss-for-interoperability.html>
- Elliott forthcoming
T. Elliott and S. Gillies, "Digital Geography and Classics," *Digital Humanities Quarterly*, forthcoming.