# **DHQ: Digital Humanities Quarterly**

Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure 2009

Volume 3 Number 1

## **Digital Geography and Classics**

Tom Elliott <tom\_dot\_elliott\_at\_nyu\_dot\_edu>, New York University Sean Gillies <sean\_dot\_gillies\_at\_gmail\_dot\_com>, New York University

### Abstract

The authors open by imagining one possible use of digital geographic techniques in the context of humanities research in 2017. They then outline the background to this vision, from early engagements in web-based mapping for the Classics to recent, fast-paced developments in web-based, collaborative geography. The article concludes with a description of their own Pleiades Project (http://pleiades.stoa.org), which gives scholars, students and enthusiasts worldwide the opportunity to use, create and share historical geographic information about the Greek and Roman World in digital form.

### The View From 2017

As I settle into my chair, a second cup of morning coffee in my hand, an expansive view of the eastern Mediterranean fades in to cover the blank wall in front of me. It's one of my favorite perspectives: from a viewpoint a thousand kilometers above the Red Sea, you can look north and west across an expanse that encompasses Jordan, Egypt and Libya in the foreground and Tunisia, Calabria and the Crimea along the distorting arc of the horizon. A simple voice command clears some of my default overlays: current precipitation and cloud cover, overnight news hotspots, and a handful of icons that represent colleagues whose profiles indicate they're currently at work. Now I see a new overlay of colored symbols associated with my current work: various research projects, two articles I'm peer-reviewing and various other bits of analysis, coding, writing and reading. These fade slowly to gray, but for two. Both of them are sprawling, irregular splatters and clumps of dots, lines and polygonal shapes.

1

2

3

4

The view pivots and zooms to frame these two symbol groups. The pink batch indicates the footprint of one of the review pieces, a survey article comparing Greek, Roman and Arabic land and sea itineraries. The other group, in pale blue, corresponds to my own never-ending "Roman boundary disputes" project. In both cases, I've previously tailored the display to map findspots of all texts and documents cited or included, as well as all places named in notes or the cited modern texts and ancient sources. These two sets of symbols remain highlighted because potentially relevant information has recently appeared.

I suspect the "hits" for the boundary disputes are just articles in the latest (nearly last) digitized Supplementband of the Pauly-Wissowa *Real-encyclopädie* to be delivered to subscribing libraries...nothing new from the old print version. I'll double-check that later. The more interesting results are illustrated when I switch focus to the itineraries article I'm reviewing. A new collection of colored symbols appear on the landscape, their intensity automatically varied to indicate likely similarity to the selected article. One particular subset jumps out at me: a series of bright red dots paralleling the southern bank of the Danube. They are connected by line segments and join up with a network of other points and lines that fan out northward into the space once occupied by the Roman province of Dacia. A new and extensive Roman-era itinerary has been discovered in Romania and published.

I can't resist making some quick explorations. It's easy enough to select the new publication and ask for an extract of the itinerary's geography, as well as the place names in their original orthographies. I follow up with a request for the

results of two correlation queries against all of the itineraries cited in the review article: how similar are their geographies and how similar are their place names? My programmatic research assistant reports back quickly: there's a strong correlation between the sequential locations in a portion of the new document and an overlapping portion of our only surviving copy of a Roman world map, the so-called "Peutinger Map", but the forms of several place names differ in some particulars.<sup>[1]</sup> Another intriguing result shows that the distances between nodal points in the new itinerary are statistically consistent with those recorded in the Claudian-era inscribed itinerary from Patara in Turkey, the so-called *Miliarium Lyciae*.<sup>[2]</sup> I bundle up these results and the new article, and add them to the review project. I can attach a short note to the author about it later.

But now it's time to teach. Another voice command stores the research stuff and clears the projected display, which reforms in two parts: another virtual globe draped with the day's lecture materials, and a largely empty grid destined to be slowly populated over the next half-hour by the various icons, avatars and video images of my students. I double-check that I'm set to do-not-disturb and lean back to review my lecture notes. Today, Caesar will cross the Rubicon.

5

6

7

8

9

### The View, Explained (and what we have left out)

We've taken an equally imagined, but less frenetic, view of the scholarly future a decade from now as that recently sketched for freelance business consultants by [Sterling 2007]. Nonetheless, we share with Sterling a number of common assumptions about the ways in which location will alter our next-decade information experiences, both at work and in private life. We envision a 2017 in which the geo-computing revolution, now underway, has intersected with other computational and societal trends to effect major changes in the way humanist scholars work, publish and teach. The rapid developments occurring at the intersection of geographic computing and web-based information technology cannot be identified with any single label, nor are they effectively described by any single body of academic literature. A variety of terms are in use for one or another aspect of this domain, including "web mapping", "neogeography", "social cartography", "the geoweb", "webGIS" and "volunteered geographic information". Those unfamiliar with this fast-moving field – one in which speculation and innovation are widespread – will gain some appreciation of its breadth and heterogeneity if they start their reading with [Turner 2006], [Goodchild 2007], [Boll 2008] and [Hudson-Smith 2008].<sup>[3]</sup>

For humanists, general research tasks will remain largely unchanged: the discovery, organization and analysis of primary and secondary materials with the goal of communicating and disseminating results and information for the use and education of others.<sup>[4]</sup> But we expect to see a more broadly collaborative regime in which a far greater percentage of work time is spent in analysis and professional communication, all underpinned by a pervasive, always-on network. Much of the tedious and solitary work of text mining, bibliographic research and information management will be handled by computational agents, but we will become more responsible for the quality and effectiveness of that work by virtue of how we publish our research results. Ubiquitous, less-obtrusive software will respond to our specific queries, and to the interests implied in our past searches, stored documents, cited literature and recently used datasets. The information offered us in return will be drawn from a global pastiche of digital repositories and publication mechanisms, surfacing virtually all new academic publication, as well as digital proxies for much of the printed, graphic and audio works now for sale, in circulation or on exhibit in one or more first-world, brick-and-mortar bookstores, libraries or museums.

It would be impossible in an article-length treatment to produce a comprehensive history of geography and the classics, digital or otherwise. The non-digital, cartographic achievements of the field to 1990 were surveyed in detail by [Talbert 1992]. Many of the projects discussed then have continued, diversified or gone digital, but there has not been to our knowledge a more recent survey. Developments on the complementary methodological axis of "Historical GIS" have been addressed recently in [Gregory 2002], [Knowles 2002] and [Knowles 2008].<sup>[5]</sup> No conference series, journal or research center has yet been able to establish the necessary international reporting linkages to provide a regular review of classical geographical research (digital or otherwise), an increasingly urgent need in a rapidly growing subfield.

Readers will detect this gap (and the compounding effects of limited space and time) in three regrettable categories of omission. Firstly, we have been unable to address many important comparanda for other periods, cultures and

disciplines, for example the China Historical GIS (http://www.fas.harvard.edu/~chgis/), the AfricaMap Project (http:// isites.harvard.edu/icb/icb.do?keyword=k28501), both spearheaded by Harvard's Center for Geographic Analysis. Secondly, our examples here are heavily weighted toward the English-speaking world; we have not been able to address such seminal efforts as the *Türkiye Arkeolojik Yerleşmeleri* (TAY) Projesi (Archaeological Settlements of Turkey Project: http://tayproject.org/). Finally, we cannot possibly grapple in this article with the burgeoning practice of GIS and remote sensing for large-scale site archaeology and regional survey, though we can point to examples of projects seeking to facilitate aggregation, preservation and discovery of related data, including FastiOnline (http://www.fastionline.org/), the Mediterranean Archaeological GIS (http://cgma.depauw.edu/MAGIS/) and OpenContext (http://opencontext.org/).

On the other hand, the initial impact on historical work of the publication in 2000 of the *Barrington Atlas of the Greek and Roman World* (R. Talbert, ed., Princeton) has been assessed by a number of reviews in major journals, and significant works have since appeared that take it as the reference basis for classical geographic features (e.g., [Hansen 2004]). Planning for a digital, user-friendly version of the Atlas continues as an area of active discussion between the Ancient World Mapping Center at the University of North Carolina in Chapel Hill (http://www.unc.edu/awmc) and other parties. Meanwhile, the analog and digital materials that underpinned publication of the Atlas are the present focus of the authors' Pleiades project (see further, below).

## The Primacy of Location: A Recent Example Drawn from Google

The ability to identify and locate relevant items of interest is of prime importance to users of scholarly information systems, and geography is a core axis of interest. As Buckland and Lancaster put it in their 2004 D-Lib Magazine article:

The most fundamental advantage of the emerging networked knowledge environment is that it provides a much-improved technological basis for sharing resources of all sorts from all sources. This situation increases the importance of effective access to information. Place, along with time, topic, and creator, is one of the fundamental components in how we define things and search for them. [Buckland 2004]

Indeed, to realize the vision we have outlined above, spatial information must become as easily searched — within, between and outside individual digital libraries — as HTML web pages are today. The Googlebot began crawling and indexing web-posted documents encoded in the Keyhole Markup Language (KML; http://en.wikipedia.org/wiki/Keyhole\_Markup\_Language), as well as web feeds containing GeoRSS markup (http://en.wikipedia.org/wiki/Georss) in late 2006 [Ohazama 2007]; [Schutzberg 2007]. Microsoft added similar support to its Local Search and Virtual Earth services in October 2007 (GeoWeb Exploration 2007). It is already safe to say that this is now a routine function for viable web search services. Henceforth, KML and GeoRSS resources on the Web will be found and indexed if they are surfaced via stable, discoverable URLs.

Obviously, much of the information that the web search providers seek to index is not in one of these geospatial formats. To determine the geographic relevance of web-facing content in HTML, MSWord and PDF files and the like, a process called geoparsing must be employed cf. [Hill 2006, 220]. Geoparsing entails first a series of steps, often called collectively "named entity recognition", in which proper nouns in unstructured or semi-structured texts are disambiguated from other words and then associated with known or notional entities of interest (e.g., places, persons, concepts). The potential of course exists for false identifications, ambiguous matches (i.e., one name for many discrete entities) and complete failures (e.g., due to the obscurity of a name or a referenced entity). Assuming some level of success in disambiguation and identification of geographic entities in a text, the next step is to associate coordinates with each named entity. Unless the textual source is itself a geographic gazetteer containing map coordinates that can be parsed accurately, the coordinates must be supplied from an external reference dataset that ideally holds the exact name variants occurring in the text as well as the geospatial coordinates; we must recognize that it will often be the case, especially with regard to primary sources and academic literature on historical topics, that such a reference dataset does not exist. Nonetheless, once coordinates are assigned to as many entities as possible, and the results are

11

12

stored in an index with a link to the original document, mapping, geosearch and other geographic computations can be performed.<sup>[6]</sup>

13

14

Despite clear challenges and the real possibility of incomplete results for many texts (we shall consider an example below), we believe that geoparsing-based services will mature and diversify rapidly. By 2017, we expect that all web-facing textual resources will be parsed (rightly or wrongly) for geographic content. Audio and video works, where these can be machine-transcribed, will receive similar treatment, whether they are accompanied by extensive metadata or not. Some texts, especially those produced by academics and specialist communities will have coordinates encoded into the digital texts in a way that facilitates automatic extraction, making their geographic indexing relatively simple (e.g., KML, GeoRSS or the like). The majority of texts, however, will still have only geographic names and geographic description, most with neither special tagging nor authority control. The search engines will automatically identify these components and add associated entries to their geographic indexes by matching them with coordinate data in reference datasets. We believe that these datasets will include information drawn from KML and GeoRSS-enabled documents on the web, as well as previously geo-parsed texts (especially those whose genre, topic and regional nature are easily identified using automated techniques). Topic- and genre-aware geoparsing may well prove essential to accurate disambiguation of non-unique place names.

The first fruits of geographic search are already apparent in information sets being surfaced to the web. Consider, for example, Google Books. A copy of J.S. Watson's 1854 translation of Xenophon's *Anabasis* was digitized at Harvard in June 2006. Bundled into the original print work by the publisher was W.F. Ainsworth's geographical commentary (pp. 263-338). Google's "About this book" page demonstrates the range of connections the search company's software is so far able to make between the book and other resources.<sup>[7]</sup> Basic bibliographic metadata is surfaced, along with direct links to a downloadable PDF file and the Google book reader application (and thence to the optically-recognized text itself). A variety of links to other services are also provided (e.g., Amazon and OCLC WorldCat). Of chief interest in the present discussion are the geographic components introduced under the heading "Places mentioned in this book." These consist at present of a map and an index of select places, as illustrated in Figure 1.



It would appear that most of the "places" currently highlighted in the Google Books presentation (evidently a maximum of 10) have been parsed from Ainsworth's commentary or the unattributed "Geographical Index", which we assume to be Watson's (pp. 339 - 348). As is the case with many Google search results, we are given no information about how the indexing or correlation with geographic coordinates (i.e., the geo-parsing) was accomplished. Some of the results are mapped correctly, but others have clearly gone awry. Consider, for example, the placement of Olympia, Syracuse and Tempe (see Figure 2).



The placement of these ancient features in North America illustrates one of the most obvious failure modes of geoparsing algorithms. If the algorithm succeeds in finding a match at all, it may find more than one match or a single, but incorrect match. Given the information publicly available, we cannot know for certain which of these two circumstances produced the results in question here. It seems probable, however, that at least the Olympia and Syracuse errors result from selecting the wrong result of several, as both of these well-known sites appear in many of the standard gazetteers and geographic handbooks.

There are published methods for resolving such ambiguities on the basis of proximity to other locations more reliably geo-parsed from the work, but these presuppose knowledge of a coherent geographic footprint for the work under consideration as well as a relatively comprehensive and highly relevant reference dataset that can provide at least a single match for most geographic names ([Smith 2001] and [Smith 2002a]). Google has not revealed, to our knowledge, whether such self-calibrating algorithms are at work, but we can be certain that they do not have access to a comprehensive, highly relevant reference dataset for the ancient Greek world. Many datasets — including the *Getty Thesaurus of Geographic Names*, the *Alexandria Digital Library Gazetteer* and the open-content *GeoNames database* [<sup>8]</sup> — contain significant numbers of historical Greek and Roman names; however, the authors' experience confirms that many ancient sites and regions are unaccounted for or incorrectly identified in these gazetteers, and many name variants and original-script orthographies are omitted. A complete and reliable dataset simply does not yet exist in a useful form.

It seems reasonable to assume that Google and its competitors will continue to seek or construct more comprehensive reference datasets to support geo-parsing improvements. Indeed, Google's embrace of VGI can be seen as a deliberate strategy for this purpose. The Google Earth Community Bulletin Board (http://bbs.keyhole.com) provides a venue for the sharing of thematic sets of labeled geographic coordinates. The expansion of the Googlebot's competencies to include KML and GeoRSS expand the harvest potential to the entire web. This circumstance highlights a class of research and publication work of critical importance for humanists and geographers over the next decade: the creation of open, structured, web-facing geo-historical reference works that can be used for a variety of purposes, including the training of geo-parsing tools and the population of geographic indexes (see [Elliott 2006]).

The automated features outlined above were added to Google Books in January 2007. The stated motivation was cast in terms of helping users "plan your next trip, research an area for academic purposes or visualize the haunts of your favorite fictional characters" [Petrou 2007]. In August 2007, Google made a further announcement of new features, which refined that motivation to echo the company's vision statement [Badger 2007], emphasis ours; compare the Google Company Overview, first sentence:

Earlier this year, we announced a first step toward *geomapping the world's literary information* by starting to integrate information from Google Book Search into Google Maps. Today, [we] announce the next step: a new layer in Google Earth that allows you to explore locations through the lens of the world's books...a dynamic and interesting way to explore the world's literature...a whole new way to visualize the written history of your hometown as well as your favorite books.

<sup>[9]</sup> If the "Places in this book" component of Google Books is intended to help users explore the geography of the book, the Google Books layer in Google Earth offers users the opportunity to discover books by way of the geography. If, by chance, we should find ourselves browsing northeastern Turkey in Google Earth, the Google Books layer offers us a book icon in the vicinity of the town of İspir. This icon represents Watson's 1854 translation of the Anabasis. Links on the popup description kite will take us back to specific pages of the scanned Harvard copy we discussed above. Another link provides us with a way to "Search for all books referencing Ispir." On 30 September 2007, we were offered information about 706 books. By refining the search to include the keyword "ancient," we narrowed the result set to 167 books. We judged these results to be highly relevant to our interests. We then selected a further suggested refinement from a list of links at the bottom of the page: "Geography, Ancient." The resulting 7 matches are unsurprisingly spotty and largely out-of-date, given the current state of Google's book digitization project, copyright restrictions and the degree of publisher text-sharing partnerships. Taking this into account, we again judged the results highly relevant. The results included [Kiepert 1878], [Kiepert 1881] and [Kiepert 1910], [Bunburry 1883], [Fabricius 1888] and [Syme 1995].

18

17

## Prelude to Geographic Search: Web-based Mapping

We should not be surprised that web-wide geo-search is with us already in embryo. It follows naturally, as a user requirement, once you have assembled a large index of web content and developed the ability to parse that content and discriminate elements of geographic significance.

Quick and accurate geographic visualization had long been one of the holy grails of the web. The now-defunct Xerox PARC Map Viewer [Putz 1994] demonstrated to early users the potential of dynamic, web-based mapping, even if it was a stand-alone application. A variety of services have come and gone since, some concentrating on screen display, others attempting to provide dynamic maps suitable for print. Some sites have provided scans of paper maps (like the David Rumsey Historical Map Collection: http://www.davidrumsey.com/); others have served out digital remotely sensed imagery or readily printed static maps in PDF and other formats (the AWMC Maps for Students collection, for example: http://www.unc.edu/awmc/mapsforstudents.html). Flash and Shockwave have also been popular ways of providing animation and interactivity on the web (for example, some maps in [Mohr 2006]). Until recently, however, the dominant paradigms for dynamic on-line maps comprised iterative, server-side map image generation, sometimes mediated by a client-side Java applet, in response to discrete mouse clicks. To one degree or another most such map tools emulate the look and feel of desktop GIS programs, perpetuating the thematic layering inherited from analog film-based cartographic composition. This paradigm continues in active use (consider the map interface for [Foss 2007], for example).

Recent innovations in on-line cartographic visualization are revolutionary. We have only just grown accustomed to digital globes (like NASA WorldWind and Google Earth) and continuous-panning, "slippy" browser-side maps. Google, Yahoo, Microsoft Live Search and MapQuest have all gone this route, and there is now an excellent, open-source toolkit for building this sort of client-side maps: OpenLayers (http://openlayers.org). Many of these applications are doing more than giving us compelling new visual environments. Many of them are also breaking down traditional divisions between browse and search, thematic layers, web content, spatial processing and geographic datasets, not least through the mechanisms known generically as "mashups" (web applications that dynamically combine data and services from multiple, other web applications to provide a customized service or data product). These tools remain an area of intense, active development, so we should expect more pleasant surprises. Moreover, as the costs of projectors, large-format LCD displays and graphics cards fall, and as digital television standards encourage the replacement of older televisions, display sizes and resolutions will at last improve. Efficient techniques for the simple mosaicing of multiple display units and for the ad-hoc interfacing of mobile devices and on-demand displays provided by third parties expand the horizon still further. It will not be long before a broader canvas opens for cartographic display.<sup>[10]</sup>

There is not space in this article to treat other forms of geographic visualization, especially 3D modeling, in detail. It must suffice for us to say that this too is an area of vigorous change. From the earliest QuickTime VR panoramas (classicists will remember Bruce Hartzler's *Metis*, still hosted today by the Stoa: http://www.stoa.org/metis/) to such contemporary services as Google StreetView (http://en.wikipedia.org/wiki/Google\_Street\_View), 3D modeling and its integration with geographic and cartographic visualization tools proceeds apace.

## Web-mapping the Geographic Content of Texts: Example of the Perseus Atlas

Geo-mapping of texts has been with the field of classics since the late 1990s, a development that paralleled interdisciplinary innovations in digital library research and development. Although its technology has now been eclipsed by rapid and recent developments in web-based cartographic visualization, the Perseus Atlas stands as a pivotal early exemplar of the power of this approach.<sup>[11]</sup>

The first version of the Perseus Atlas was rolled out in June 1998, shortly after Rob Chavez joined the Perseus team 24 with the atlas as his primary task.<sup>[12]</sup> Its coverage was limited to the "Greek world," and it was designed solely to provide geographic context (illustrative maps) for discrete content items in the Perseus collection, e.g., individual texts (or portions of texts) or object records. Perseus was already in control of a significant amount of geographical information

22

21

drawn from its texts through geo-parsing.<sup>[13]</sup> This was further augmented from metadata associated with its coin and pottery databases. In the main, this dataset consisted of place names. A rudimentary digital gazetteer had been developed through computational matching of this names list with gazetteer records in various publicly-available US Government geographical datasets (especially the resource now known as the GEONet Names Server of the National Geospatial Intelligence Agency), which provided coordinates. After some initial experimentation, the Perseus team selected an early version of MapServer to provide dynamic mapping, and by way of Perl and CGI, this was integrated with the existing Perseus tooling. Geographic data was stored in a plain PostgreSQL database, the PostGIS spatial data extension for PostgreSQL having not yet been developed. Vector linework and polygons from the Digital Chart of the World were added to provide geographic context.

The Perseus Atlas saw two upgrades subsequent to its 1998 debut.<sup>[14]</sup> The 2000 upgrade rolled out a rewritten and expanded Atlas that boasted several new features: global geographic coverage, a relief base layer, and tighter integration between the Atlas and the Perseus Lookup Tool. This last feature — a collaboration between Chavez and David Smith — transformed the Atlas from a cartographic illustration mechanism to an alternative interface for the entire collection, capable of both object-specific and cross-object cartographic visualization and query. A fork of the Atlas, designed to support the Historic London Collection, also appeared during this period.<sup>[15]</sup> It offered essentially the same features, but employed a separate geographic datum, higher-resolution imagery and rudimentary temporal filtering capabilities by way of user-selectable, dated map layers. The 2002 upgrade improved the place name lookup and navigation tools in both Atlas branches, added an option for saving map views and augmented the London Atlas with links to QuickTime VR panoramas for select streets.

### The Geo-Library, the Web and Geographic Search

It is clear from even a cursory review of contemporaneous developments that the Perseus team was in the vanguard of an important revolution in web-based cartographic visualization. When Perseus first fielded its atlas in 1998, The Alexandria Digital Library (ADL) Project team had been working for four years to realize Larry Carver's 1983 vision of geographic search in a digital maps catalog [Hill 2006, 49]. At about the same time the Perseus team was conceptualizing its atlas as an interface to their entire, heterogeneous collection, the ADL vision was expanding to encompass geographic browsing and search in an entire digital library containing materials of all kinds, not just maps [Goodchild 2004]. A year after the Perseus Atlas appeared, the ADL vision of the "geo-library" was endorsed by the National Research Council Mapping Science Committee (National Research Council Mapping Science Committee 1999; see further [Hill 2006, 11–17]). The 2000 upgrade to the Perseus Atlas represented the first realization of this vision for the humanities outside ADL, and one of the first anywhere.

It is clear that, despite significant challenges, geo-library functionality like that pioneered by ADL and Perseus is rapidly maturing into one of the standard modes for browsing and searching all manner of digital information on the web.<sup>[16]</sup> Simple spatial queries (within bounding boxes, by proximity to points, and by proximity to other features) are becoming commonplace.<sup>[17]</sup> We expect that such "geo-aware" searches will occur more frequently outside special map-oriented interfaces and services, with the target coordinates being provided automatically and transparently by search software and other computational proxies.

These developments are being fueled by now-familiar business models.<sup>[18]</sup> The proximity of a retail outlet or entertainment venue to a customer (or the customer's preferred transportation method) is of interest to both the consumer and the vendor. Likewise, the location of a potential customer — or the locations of places she frequents, or that hold some sentimental or work-related interest for her — all have monetizable value. Location-aware ads and free-to-user planning, shopping and information-finding tools will be paid for by the businesses that stand to gain from the trade that results. Competition between search providers will open up more free search tools, some visible to users only by virtue of their results; however, it seems likely that industrial-strength GIS access to the underlying spatial indexes and georeferenced user profiles will remain a paid subscription service.

Spatial search will help manage more seamlessly the inevitable heterogeneity of the 2017 web. Formats, delivery 29

26

27

28

mechanisms, cost models and access challenges will surely continue to proliferate. We assume that some resources will be little advanced from the average 2007 vintage web page or blog post, whereas other works will exhibit the full range of meaningful structures and linkages now being worked out under the rubrics "semantic web" and "linked data".

<sup>[19]</sup> Some works will be posted to basic websites. Copies of other works will reside in institutional repositories, federated archive networks, publishers' portal sites or massive grid environments. Whether stand-alone or integrated in a collection, some resources will have appeared (or will be augmented subsequently) with a range of metadata and rich, relevant links to other resources and repositories. Others will constitute little more than plain text.

We expect current distinctions between geographic content and other digital works will continue to erode over the next decade. As the geographic content in most documents and datasets is identified, surfaced and exploited, "born geographic" works will be fashioned in increasingly flexible and interoperable ways. To some extent, specific industries and institutional consortia will surely continue to use specialized protocols, service-oriented architectures and specialist formats to provide lossless interchange and rich contextualization of critical data. Smaller players, and big players wishing to share data with them, will increasingly use RESTful models, simple URLs and widely understood open formats for basic information publishing and exchange, even when these methods are lossy.<sup>[20]</sup> In some cases, these documents will serve as proxies for the unadulterated content, encoded in whatever format is necessary to express its creators' intent, even if it is totally idiosyncratic. The value of universal, geo-aware search and manipulation will trump concerns about surfacing "incomplete" data. Even now, a KML file with appropriate descriptive content and hyperlinks can function as a table of contents, geographic index or abstract for a document collection coded in TEI or a custom web application running in Plone. We expect that extensible web feeds (like the recently codified Atom Syndication Format, RFC 4287: http://www.atompub.org/rfc4287.html) will form a simple, web-wide infrastructure for notification and metadata exchange, alongside the more complex and difficult-to-implement special protocols that are already linking federated digital repositories and grid systems.

# Big Science, Repositories, Neo-geography and Volunteered Geographic Information

Many commercial innovations for online search, retrieval and visualization will be of benefit to humanist scholars and their students. Such benefits will depend, to a significant degree, on the willingness and ability of humanists, their employers and their publishers to adapt their practices to fit the new regime.

To date, humanists have taken one of two approaches in preparing and disseminating geo-historical information on the web. Large projects with significant funding have tended to follow the lead of the larger geo-science and commercial GIS communities in placing emphasis on the elaboration of extensive metadata describing their datasets, thereby creating a basis for their discovery and inclusion in digital repositories. In some countries (notably the U.K.) emphasis by national funding bodies has encouraged such behavior. At the other end of the spectrum, hand-crafted datasets and web pages are posted to the web with minimal metadata or distributed informally within particular research communities. The neo-geographical revolution, and in particular the explosive interest in Google Earth, has recently provided a much richer and easier mechanism for the dissemination of some types of work in this latter class (e.g., [Scaife 2006]).

In the United States, e-science approaches have long been dominated by ideas expressed in an executive order issued by President Clinton in 1994 and amended by President Bush in 2003 [Executive Order 12906] and [Executive Order 13286]. These orders chartered a National Spatial Data Infrastructure, encompassing a federally directed set of initiatives aimed at the creation of a distributed National Geospatial Data Clearinghouse (http:// clearinghouse1.fgdc.gov/); standards for documentation, collection and exchange; and policies and procedures for data dissemination, especially by government agencies. More recently, the Library of Congress constructed, through the National Digital Information Infrastructure and Preservation Program, a National Geospatial Digital Archive (http://www.ngda.org/), but this entity so far has been of little use to classicists, as it has focused collection development on "content relating to the east and west coasts of the United States" (http://www.ngda.org/research.php#CD).<sup>[21]</sup>

30

The "national repository" approach has similarly found a coherent advocate for academic geospatial data in the UK by way of EDINA, the national academic data center at the University of Edinburgh (http://edina.ac.uk/); however, the recent governmental decision to de-fund the UK Arts and Humanities Data Service has introduced a significant level of uncertainty and chaos into sustainability planning for digital humanities projects there (details at http://ahds.ac.uk/). More broadly, the Global Spatial Data Infrastructure Association seeks to bring together representatives of government, industry and academia to encourage and promote work in this area (http://gsdi.org), but to date it seems to have focused exclusively on spatial data as it relates to governance, commerce and humanitarian activities.

The top-down, national repository model (and its cousin, the institutional repository) contrasts sharply with the neogeographical methods now rapidly proliferating on the web. To a certain degree, the differences reflect alternative modes of production. The repository model was born in a period when spatial datasets were created and used almost exclusively by teams of experts wielding specialized software. Concerns about efficiency, duplication of effort and preservation were key in provoking state interest, and consequently these issues informed repository specifications. VGI, however, has only recently been enabled on the large scale by innovations in web applications.<sup>[22]</sup> Many such efforts – from early innovators like the Degree Confluence Project (1996; http://www.confluence.org/), to more recent undertakings like Wikimapia (2006; http://wikimapia.org/) – reflect the interests of communities whose membership is mostly or almost entirely composed of non-specialists. They tend to be more broadly collaborative, iterative and open than traditional geospatial data creation efforts, but both commercial and institutional entities are also increasingly engaged in collaborative development or open dissemination of geographic information. VGI datasets are published via a wide range of web applications; recent standardization of formats and accommodation by commercial search engines has opened the field to "webby" distribution. For example, KML and GeoRSS-tagged web feeds can now be posted and managed just like other basic web content.<sup>[23]</sup>

### The Electronic Cultural Atlas Initiative

In 1997, The Electronic Cultural Atlas Initiative was chartered in Berkeley to ease and promote cross-project sharing of spatio-temporal data in the humanities [Buckland 2004]; http://www.ecai.org. Among its initiatives was a metadata clearinghouse for registered projects (http://ecaimaps.berkeley.edu/clearinghouse/). Scholars wishing to register their projects arrange for the published data to be posted to the web, and create "ECAI metadata" (a Dublin Core derivative), which is then added to the clearinghouse. Geographic and temporal components of the search are facilitated by a map-and-timeline interface realized with server-side TimeMap software (see http://www.timemap.net). This tool may be seen, on one level, as a reflection of the dominant web search mode of the day, the directory. But it also anticipated the future of VGI, appreciating that many humanities projects with spatial components would not be able to fit their content into the formats supported by the big repositories. Instead, it assumed that these projects would arrange for web hosting themselves, and take voluntary steps to share it with others. ECAI's Clearinghouse was also innovative in providing for the discovery of indexed projects through a combination of search and browse methods that exploited both spatial and temporal information in intuitive ways. ECAI remains a vibrant international community, organizing regular conferences and maintaining linkages across a wide range of projects and disciplines (see http://ecai.org/Activities/conferences.asp).

# The Stoa Waypoint Database and the Register of Ancient Geographic Entities

The first attempt at full-fledged VGI in Classics was the Stoa Waypoint Database, a joint initiative of Robert Chavez, then with the Perseus Project, and the late Ross Scaife, on behalf of the Stoa Consortium for Electronic Publication in the Humanities (http://www.stoa.org). In its public unveiling, [Scaife 1999] cast the resource as "an archive [and] freely accessible source of geographic data...for archaeologists...students...digital map makers, or anyone else engaged in study and research". At initial publication, the dataset comprised slightly over 2,000 point features (settlements, sites and river mouths), drawn from work Chavez and Maria Daniels had done for the Perseus Atlas and personal research projects. The points included both GPS coordinates and coordinates drawn from various public domain (mostly US government) gazetteers and data sources. Chavez and Scaife also invited contributions of new data, especially encouraging the donation of GPS waypoints and tracks gathered in the field. A set of Guidelines for Recording

1

35

34

36

Handheld GPS Waypoints were promulgated to support this work [Chavez 1999]. The original application for download of the database was retired from the Stoa server some time ago; however, the data set has recently been reposted by the Ancient World Mapping Center in KML format (http://www.unc.edu/awmc/pleiades/data/stoagnd/). Despite limited success in soliciting outside contributions, the idea of the Stoa Waypoint Database had a formative influence on the early conceptualization of the authors' Pleiades project.

The Stoa's interest in GPS waypoints reflected a worldwide trend: in April 1995, the NAVSTAR Global Positioning System constellation had reached full operational capability, marking one of the pivotal moments in the geospatial revolution we are now experiencing. Originally conceived as a military aid to navigation, GPS quickly became indispensable to both civilian navigation and map-making, as well as the widest imaginable range of recreational and scientific uses. In May 2000, an executive order eliminated the intentional degradation of publicly accessible GPS signals, known as selective availability [Office 2000]. This decision effectively placed unprecedentedly accurate georeferencing and navigating capabilities in the hands of the average citizen (~15m horizontal accuracy).

Work on the Stoa Waypoints Database was clearly informed by the Perseus Atlas Project, and also by another 39 concurrent project at the Stoa: the Suda Online (SOL; http://www.stoa.org/sol/). This all-volunteer, collaborative effort undertook the English translation of a major Byzantine encyclopedia of significance for Classicists and Byzantinists alike. The Suda provides information about many places and peoples. An obvious enhancement to the supporting web application would have been a mapping system like the Perseus Atlas. In exploring the possibility, Scaife and Chavez realized that the variant place names in the Suda and in the Perseus Atlas were a significant impediment to implementation. The solution they envisioned in collaboration with Neel Smith was dubbed the Register of Geographic Entities (RAGE). They imagined an inventory of conceptual spatial units and a set of associated web services that would store project-specific identifiers for geographic features, together with associated names. This index would provide for cross-project lookup of names, and dynamic mapping. Some development work was done subsequently at the University of Kentucky [Mohammed 2002], and the most current version may be had in the Registry XML format currently under development by Smith and colleagues at the Center for Hellenic Studies via the CHS Registry Browser (cf., [Smith 2005]). At present, it contains just over 3,500 entries drawn almost entirely from Ptolemy's works and thus has so far not seen wide use as a general dataset for the classical world.<sup>[24]</sup> The concern that informed the RAGE initiative remains valid: geographic interoperation between existing classics-related digital publications will require the collation of disparate, project-level gazetteers. It is our hope that, as Pleiades content is published, its open licensing and comprehensive coverage will catalyze a geo-webby solution.

### The Pleiades Project

Our own project, Pleiades (http://pleiades.stoa.org), is heavily influenced by both by the scholarly practices of our predecessors and by on-going developments in web-enabled geography. We are producing a standard reference dataset for Classical geography, together with associated services for interoperability. Combining VGI approaches with academic-style editorial review, Pleiades will enable (from September 2008) anyone — from university professors to casual students of antiquity — to suggest updates to geographic names, descriptive essays, bibliographic references and geographic coordinates. Once vetted for accuracy and pertinence, these suggestions will become a permanent, author-attributed part of future publications and data services. The project was initiated by the Ancient World Mapping Center at the University of North Carolina, with development and design collaboration and resources provided by the Stoa. In February 2008, the Institute for the Study of the Ancient World joined the project as a partner and it is there that development efforts for the project are now directed.

Pleiades may be seen from several angles. From the point of view of the Classical Atlas Project and its heir, the Ancient World Mapping Center, Pleiades is an innovative tool for the perpetual update and diversification of the dataset originally assembled to underpin the Barrington Atlas, which is being digitized and adapted for inclusion. From an editorial point of view, Pleiades is much like an academic journal, but with some important innovations. Instead of a thematic organization and primary subdivision into individually authored articles, Pleiades pushes discrete authoring and editing down to the fine level of structured reports on individual places and names, their relationships with each other and the scholarly rationale behind their content. In a real sense then Pleiades is also like an encyclopedic reference

40

41

work, but with the built-in assumption of on-going revision and iterative publishing of versions (an increasingly common model for digital academic references). From the point of view of "neo-geo" applications, Pleiades is a source of services and data to support a variety of needs: dynamic mapping, proximity query and authority for names and places.

Pleiades incorporates a data model that diverges from the structure of conventional GIS datasets. Complexity in the historical record, combined with varying uncertainty in our ability to interpret it, necessitate a flexible approach sensitive to inherent ambiguity and the likelihood of changing and divergent interpretations (cf. [Peuquet 2002, 262–281]). In particular, we have rejected both coordinates and toponyms as the primary organizing theme for geo-historical data. Instead, we have settled upon the concept of "place" understood as a bundle of associations between attested names and measured (or estimated) locations (including areas). We call these bundles "features". Individual features can be positioned in time, and the confidence of the scholar or analyst can be registered with respect to any feature using a limited vocabulary.

42

45

46

We intend for Pleiades content to be reused and remixed by others. For this reason, we release the content in multiple formats under the terms of a Creative Commons Attribution Share-Alike license (http://creativecommons.org/licenses/ by-sa/3.0/). The Pleiades website presents HTML versions of our content that provide users with the full complement of information recorded for each place, feature, name and location. In our web services, we employ proxies for our content (KML and GeoRSS-enhanced Atom feeds) so that users can visualize and exploit it in a variety of automated ways. In this way, we provide a computationally actionable bridge between a nuanced, scholarly publication and the geographic discovery and exploitation tools now emerging on the web. But for us, these formats are lossy: they cannot represent our data model in a structured way that preserves all nuance and detail and permits ready parsing and exploitation by software agents. Indeed, we have been unable to identify a standard XML-based data format that simply and losslessly supports the full expression of the Pleiades data model.<sup>[25]</sup>

In 2009, we plan to address users' need for a lossless export format by implementing code to produce file sets composed of ESRI Shapefiles and attribute tables in comma-separated value (CSV) format. The addition of a Shapefile+CSV export capability will facilitate a download-oriented dissemination method, as well as position us to deposit time-stamped versions of our data into the institutional repository at NYU and other archival contexts as appropriate. Indeed, this is the most common format requested from us after KML. Although the Shapefile format is proprietary, it is used around the world in a variety of commercial and open-source GIS systems and can be readily decoded by third-party and open-source software. It is, in our judgment, the most readily useful format for individual desktop GIS users and small projects, and because of its ubiquity has a high likelihood of translation into new formats in the context of long-term preservation repositories.

Historical time remains a problematic aspect of the web that cannot be divorced from geography. The use of virtual map layers to represent time periods remains a common metaphor. Timelines and animated maps are also not uncommon, but all these techniques must either remain tied to specific web applications or rely on cross-project data in standard formats that only handle the Gregorian calendar and do not provide mechanisms for the representation of uncertainty. Bruce Robertson's Historical Event Markup Language (HEML; http://heml.mta.ca) remains the most obvious candidate for providing the extra flexibility humanists need in this area, both for modeling and for expression in markup. It is our hope that it too will find modes of realization in the realm of web feeds and semantic interchange.

In its final report, the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences highlighted the importance of accessibility, openness, interoperability and public/private collaboration as some of the prerequisites for a culture of vigorous digital scholarship [Unsworth 2006, 28]. We believe that Pleiades exemplifies these characteristics, both on its own terms and as an emulator of prior efforts that helped identify and define them. We recommend our two-fold agenda to other digital projects in the Classics whose content includes or informs the geographic:

- 1. Publish works that help (in terms of content, structure, delivery and licensing) other humanists work and teach
- 2. Publish works that potentially improve the performance and accuracy of the (geo)web

### Conclusion

We view the history of computing, classics and geography as a rich and profitable dialogue between many disciplines and practitioners. It is tragic indeed that we lost much too soon our friend and mentor Ross Scaife, who emerges as a pivotal figure in our narrative of this history. It was he who invited Elliott to the University of Kentucky in 2001 to give the first public presentation about the proposed Pleiades Project at the Center for Computational Studies, and it was he who provided the development server that supported the first two years of Pleiades software development. We are confident that, were Ross with us today as he was at the workshop that occasioned the original version of this paper, he would still be motivating us with challenging examples and stimulating ideas, connecting us with new collaborators and encouraging us to push harder for the changes we wish to see in our discipline.

47

48

Despite the sense of loss that inevitably runs through the papers in this volume, and despite the challenges we face in a chaotic and interdisciplinary milieu, we also view the history of geographic computing and the classics as a hopeful omen for the future. We have a spectacular rising wave to ride: a wave of technological and societal change that may well help us conduct research and teach with more rigor and completeness than before while breaking down artificial boundaries between scholars in the academy and members of an increasingly educated and engaged public whose professional skills, public hobbies and personal interests coincide with ours at the fascinating intersection of humanity, space and time.

### **Notes**

[1] The most recent published work on the Peutinger map at the time of publication was [Elliott2008], which looks forward to Richard Talbert's forthcoming digital edition of this unique primary source for the study of Roman cartography. A number of digital offprints on related topics are available from the Ancient World Mapping Center at http://www.unc.edu/awmc/talbertarticles.html.

[2] The definitive publication of the so-called Miliarium Lyciae is [Şahin 2007].

[3] An interesting range of ideas and views surfaced in the context of the meeting in late 2007 that occasioned Goodchild's paper (see the position papers and presentations collected at http://www.ncgia.ucsb.edu/projects/vgi/), but it is instructive to note the discussion this meeting and paper generated beyond the academy: see the list thread beginning with http://geowanking.org/pipermail/geowanking\_geowanking.org/ 2008-January/016319.html, as well as [Fee 2008] (including comments). For clarity in this article, we adopt the phrase VGI throughout to denote the full range of open, community-oriented geographic collaborations.

[4] It is not our intent to imply, by this formulation, that we reject the notion that humanistic research will extend to address questions that cannot be formulated or investigated outside a computational context; rather we seek to emphasize certain underlying tasks that we believe will persist even if automated and applied to novel research goals. It is a formulation shared to one degree or other by (e.g.) [Unsworth 2000] and [Mostern 2008]. With regard to these, and the rest of the assumptions outlined in this paragraph, we cannot possibly predict the details of their manifestation, nor the order in which they will appear. Collaboration and automation of data finding, information management and scholarly communication/publication tasks are widely assumed (e.g., [Boonstra 2004, 91-99]; [Genoni 2005]; [Boast 2007]; [Unsworth 2006]; [Dunn 2008]; [Guthrie 2008, 16017]; [Pritchard 2008]; [Spiro 2008]). These assumptions spring from knowledge of work in a variety of fields. Web search customization, based on profiling, social bookmarking and other novel ranking schemes, has been an active area of research for over a decade, building on a venerable lineage of research in information retrieval and text mining. Hinrich Schütze's Information Retrieval Resources page provides an excellent entry point to the literature (via standard texts), as well as links to related professional societies, research centers, courses and conferences: http://www-csli.stanford.edu/~hinrich/information-retrieval.html. Emerging tools, frameworks and standard data formats for easy discovery, creation and social management and annotation of bibliographic information (e.g., WorldCat http://www.wordcat.org; Bibsonomy http://www.bibsonomy.org/; LibraryThing http://www.librarything.com/; The Open Library http://openlibrary.org/; and Zotero http://www.zotero.org/) lend credence to our expectation that the tedious, citation-management aspects of bibliographic work will become increasingly automated. The intersection of such systems with emerging digital archives and digital publications (both retrospective and borndigital) will create opportunity for a wide range of potential research support services for information discovery, document delivery and text mining. It is beyond the scope of this article to venture any authoritative summary or quantitative assessment of the progress and ultimate "success" of the vast range of digitization and digital publication efforts - commercial, public and consortial - now underway, but [Wilson 2007] and [Crawford 2008] provide recent assessments and access to related announcements and publications.

[5] A major conference on the topic, Historical GIS 2008, has just concluded (August 2008). It is hoped that proceedings from the meeting will be

forthcoming (see http://www.hgis.org.uk/HGIS\_conference/index.htm).

[6] The Perseus Digital Library team at Tufts University was an early leader in this field (see references at note 13), and continues to push forward research on named entity work ([Crane 2006], [Crane 2006a]). The recently released, open-licensed Perseus text files include the name tagging inserted by their system (http://www.perseus.tufts.edu/hopper/opensource); cf. [Crane 2006b]. Outside the Perseus milieu, geoparsing has become increasingly important in both academic and commercial contexts. Some recent academic papers on name entity recognition, with additional bibliography, include [Kozareva 2006], [Nadeau 2006], [babeu 2008], [Lucarelli 2007] and [Shaalan 2008]. In 2006, the commercial firm Metacarta (http://www.metacarta.com/) unveiled Gutenkarte (http://gutenkarte.org/) as a demonstration of its geoparsing techniques using public domain texts from Project Gutenberg. [Bell 2008] provides a recent, concise description of geoparsing techniques from the point of view of Yahoo!'s Geo Technologies division. Google also is a major player in this space: we treat some of their services and approaches in detail below.

[7] Google Books' "About this Book" page for Xenophon, *The Anabasis, or Expedition of Cyrus, and the Memorabilia of Socrates*, J.S. Watson (trans.), London: 1854 (copy scanned from Harvard University Library on 21 June 2006): http://books.google.com/books? id=wqcKoOn76aMC&dq=editions:0GclhKHLkS0MeN2kDT76mh/; accessed 30 September 2007.

[8] Getty Thesaurus of Geographic Names: http://www.getty.edu/research/conducting\_research/vocabularies/tgn/; Alexandria Digital Library Gazetteer: http://middleware.alexandria.ucsb.edu/client/gaz/adl/index.jsp/; Geonames: http://www.geonames.org/

[9] As it appeared in September 2007 at http://www.google.com/corporate/.

[10] The U.S. Congress has legislated 17 February 2009 as the date by which all licensed, full-power television stations in the United States stop broadcasting in analog format and broadcast only in digital format (http://www.fcc.gov/cgb/consumerfacts/digitaltv.html). Recent technological advances treating large displays, multi-projector displays and on-demand displays for mobile devices are treated in [Czerwinski 2006], [Bhasker 2007] and [Stødle 2007]. Our assumption about the increasing ubiquity of pixels is shared by Tim Berners-Lee, who last year expanded on his vision of "pixels everywhere" in the context of mobile computing (see [Berlind 2007]).

[11] To our knowledge, there is no comprehensive published account of the development and workings of the Atlas; what follows therefore derives from papers published by the Perseus team on aspects of their named entity recognition and temporal-spatial metadata work, as well as the authors' recalled experience, short announcements and help documents on the Perseus website and a telephone interview with Rob Chavez (17 July 2007).

[12] Announcing the New Perseus Web Atlas (12 June 1998): http://www.perseus.tufts.edu/atlas.ann.html.

[13] See [Rydberg-Cox 2000]; Crane 2001, pp. 6,8 and figures 3-5, 8-11; [Smith 2001]; [Smith 2002a]; [Smith 2002b].

[14] Perseus Web Atlas Update (28 September 2000): http://www.perseus.tufts.edu/PR/atlas2000.ann.html; An Upgrade to the Perseus Interactive Web Atlas (26 March 2002): http://www.perseus.tufts.edu/PR/atlas.ann.html.

[15] New Collections, New Navigation (17 November 2000): http://www.perseus.tufts.edu/PR/frontpage.ann.html. This development, of course, depended heavily on Smith's named entity recognition work (see references in note 13).

[16] [Asadi 2007] assesses some aspects of the state-of-the-art and provides recent bibliography on techniques and problems.

[17] Google Earth provides this functionality in its "Find Businesses" widget: you enter (e.g.) "Pizza" in the "What" text box; the "Where" text box is prepopulated with the default value "Current View" (effectively a bounding box). MapQuest provides geographic proximity search through its "Search Nearby" feature (http://help.mapquest.com/jive/entry.jspa?externalID=349&categoryID=6). Yahoo! Local provides similar capability (http://local.yahoo.com/). Other examples and experiments abound; their enumeration here would overly belabor the point.

[18] Our own casual review of U.S. patents and patent applications, via online patent search services, reinforced our long-standing perception of widespread commercial interest in "location-based" services and information retrieval; although not rigorously screened, we believe the number of patents and applications in this field over the past decade may easily number in the thousands. A web search for a phrase like "location-based" or "location-aware search" will retrieve a significant number of relevant blog postings, articles and business websites.

[19] It would be naïve for us to assert that, in 10 years, the Semantic Web as presently imagined (see [Shadbolt 2006]; [Miller 2007]) will constitute the web itself. Rather, we expect to see a range of practices develop from the creative intersection of current patterns with proposed Semantic Web approaches and with other proposals and innovations. Our own recent experiments with linked data (see http://linkeddata.org/) have encouraged us to imagine that an increasing number of resources – especially those useful for academic research and reference – will be

exposed to the web in ways that provide stable, discrete URIs and multiple standard formats for each potentially cited or remixed element of interest. Such a development, combined with methods for the virtual aggregation and linking of resources in a manner compatible with basic web architecture and indexing (e.g., OAI/ORE in Atom: http://www.openarchives.org/ore/0.9/atom-implementation.html), would be of great service to humanities scholars (see further [Cohen 2008], [Elliott 2008a]).

[20] REpresentational State Transfer (REST) is an approach to the development of web services that seeks to exploit essential modes of web interaction to provide simpler, stabler and more open systems than traditional service-oriented architectures (see [Tilkov 2007] for a good introduction). It shares some principles with the linked data movement now championed by Tim Berners-Lee and others (see note 16 for references). The uneven uptake of REST in the geo-computing community was highlighted by links and comments in [Gillies 2008]; note also this recent press release from ESRI, a leading developer of commercial GIS software: http://www.esri.com/news/arcnews/spring08articles/ attendees-get.html.

[21] In private correspondence with Elliott, NGDA principals have expressed interest in eventual accession of Pleiades data, but to date no concrete requirements or processes have been articulated.

#### [22] See above, and note 3.

[23] The publicly open and collaborative aspects of VGI projects has sometimes led to the impression that use of the resulting data for serious geospatial work is irresponsible, not least because of perceptions of disregard for traditional forms of metadata (e.g., in a discussion thread on the closed-archive Maps-L discussion list in April 2007; subject line "Query: map and GIS user needs"). It may be that the inherently public aspect of VGI efforts makes it far easier to discover such oversights; however, we are not aware of any authoritative study addressing the issue in quantitative terms. There can be no doubt that many VGI datasets, and many of the web applications that enable their creation and use, fail to address issues as fundamental as precision, accuracy and data origin. It is particularly frustrating, for example, that content from Google's geo-aware photo sharing service, Panoramio (http://www.panoramio.com/), is surfaced in Google Earth with no reflection of its spatial accuracy. This despite the fact that Panoramio content includes both images that were geo-tagged manually using an arbitrarily scaled map interface and images whose EXIF headers originally contained high-resolution coordinates and other metadata provided automatically by GPS-enabled digital cameras. On the other hand, an increasing number of VGI projects whose technical staff include academic or professional geographers are addressing these concerns and finding innovative ways to solve them in the new media.

[24] Of course the primary source of delay for RAGE was the lack of availability of a comprehensive set of coordinates, identifiers and names around which to array the content of other gazetteers. The Barrington Atlas constitutes such a resource, but has not been available in an appropriate form, nor under a suitable license that would permit such reuse. Elliott and Gillies, in collaboration with others (including Smith) and with funding from the National Endowment for the Humanities and the U.K. Joint Information Systems Committee have recently devised and promulgated a list of identifiers and name strings for every discrete geographic feature in the Barrington Atlas (see [Elliott 2008b]). Projects and publishers are encouraged to use these identifiers in their digital datasets; the Pleiades project aims to establish a web service for linking these identifiers to its online coordinate and name data as it is digitized. Linking or incorporation of the Ptolemaic information now in the RAGE registry will thereafter be a straightforward matter.

[25] The original Pleiades proposal, written in 2005, anticipated that the Alexandria Digital Library (ADL) Gazetteer format (http:// www.alexandria.ucsb.edu/gazetteer/ContentStandard/version3.2/GCS3.2-guide.htm) would provide an export and archiving vehicle (Elliott had supplied comments to the joint ADL and ECAI teams that developed the gazetteer format with this use in mind). We 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 seems to have stalled with the retirement of its primary research leader (a Pleiades Steering Committee member), and to our knowledge it has not seen significant adoption outside its original application in the Alexandria Digital Library and a few closely collaborating historical-cultural academic projects. In our view, 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 introduced by Google Earth. Recently introduced guidance for the description of "places" (in addition to "place names") by the Text Encoding Initiative might provide another "lossless" alternative, but it seems unlikely to see wide adoption outside scholarly humanities circles (see TEI Consortium 2008, sub "13. Names, Dates, Peoples and Places", http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html).

### **Works Cited**

Asadi 2007 Asadi, Saeid, Xiaofang Zhou, Hamid R. Jamali and Hossein Mofrad. "Location-Based Search Engines Tasks and Capabilities: A Comparative Study". *Webology* 4: 4 (2007). http://www.webology.ir/2007/v4n4/a48.html.

Badger 2007 Badger, Brandon. Google Book Search in Google Earth. Google, August 20, 2007. http://google-latlong.blogspot.com/2007/08/google-book-search-in-google-earth.html.

- Bell 2008 BellTyler. A Tale of Two Cities. The Yahoo! Geo Technologies Blog. 28 August 2008. http://www.ygeoblog.com/blog/2008/08/28/a-tale-of-two-cities/.
- Berlind 2007 Berlind, David. Web inventor Sir Tim Berners-Lee waxes on 'pixels everywhere.'. ZDNet, June 7, 2007. http://blogs.zdnet.com/BTL/?p=5318.
- **Bhasker 2007** Bhasker, Ezekiel S., Ray Juang and Aditi Majumder. "Advances Towards Next-Generation Flexible Multi-Projector Display Walls". Presented at *EDT 2007. Proceedings of the 2007 Workshop on Emerging Displays Technologies: Images and Beyond: The Future of Display and Interaction* (2007). http://portal.acm.org/citation.cfm? id=1278240.1278251&coll=GUIDE&dI=GUIDE&CFID=8598811&CFTOKEN=44119944.
- Boast 2007 Boast, Robin, Michael Bravo and Ramesh Srinivasan. "Return to Babel: Emergent Diversity, Digital Resources, and Local Knowledge". *The Information Society* 23 (2007), pp. 395-403.
- Boll 2008 Warning: Biblio formatting not applied. BollSusanne. JonesChristopher. KansaEric. KishorPuneet. NaamanMoor. PurvesRoss. ScharlArno. WildeEric. LOCWEB '08: Proceedings of the First International Workshop on Location and the Web. New York. ACM. 2008.
- Boonstra 2004 Boonstra, Onno, Leen Breure and Peter Doorn. "Past, Present and Future of Historical Information Science". *Historiche Sozialforschung* 29: 2 (2004).
- Buckland 2004 Buckland, Michael, and Lewis Lancaster. "Combining Place, Time and Topic: The Electronic Cultural Atlas Initiative". D-Lib Magazine 10: 5 (2004). http://www.dlib.org/dlib/may04/buckland/05buckland.html.
- **Bunburry 1883** Bunbury, Edward Herbert. A History of Ancient Geography Among the Greeks and Romans from the Earliest Ages till the Fall of the Roman Empire. 2nd edition. London: John Murray, 1883.
- **Carlson 2007a** Carlson, Nicholas. *John Hanke, GM, Google Earth*. InternetNews.com, July 6, 2007. http://www.internetnews.com/xSP/article.php/3687471.
- Chavez 1999 Chavez, Robert. Guidelines for Recording Handheld GPS Waypoints. 1999. http://www.stoa.org/guides/gps.shtml.
- **Cohen 2008** Cohen, Dan. *The Vision of ORE*. Dan Cohen's Digital Humanities Blog, 3 March 2008. http://www.dancohen.org/2008/03/03/the-vision-of-ore/.
- Crane 2001 Crane, Gregory, R.F. Chavez, A. Mahoney, T.L. Milbank, J.A. Rydberg-Cox, D.A. Smith and C.E. Wulfman. "Drudgery and Deep Thought: Designing a Digital Library for the Humanities". *Communications of the Association for Computing Machinery* 44: 5 (2001).
- Crane 2006 Crane, Gregory, and Alison Jones. *The Perseus American Collection 1.0.* 2006. http://dl.tufts.edu//view\_pdf.jsp?urn=tufts:facpubs:gcrane-2006.00001.
- Crane 2006a Crane, Gregory, and Alison Jones. "The Challenge of Virginia Banks: an Evaluation of Named Entity Analysis in a 19th Century Newspaper Collection". Presented at JDCL 2006. Proceedings of the 6th ACM/IEEE-CS joint conference on Digital Libraries (2006), pp. 31-40.
- **Crane 2006b** Crane, Gregory. "The Perseus Digital Library: New content and services for 19th century American documents". *D-Lib Magazine* 12: 3 (March 2006).
- Crawford 2008 Crawford, Walt. "Updating the Book Discovery Projects". *Cites and Insights: Crawford at Large* 8 (2008). http://citesandinsights.info/v8i9b.htm.
- Czerwinski 2006 Czerwinski, Mary. "Large Display Research Overview". Presented at CHI 2006. Extended Abstracts on Human Factors in Computing Systems Conference (2006), pp. 69-74.
- Dunn 2008 Dunn, S., and T. Blanke. "Next Steps for E-Science, the Textual Humanities and VREs: A Report on Text and Grid: Research Questions for the Humanities, Sciences and Industry, UK e-Science All Hands Meeting 2007". D-Lib Magazine 14 (2008).
- Elliott 2006 Elliott, Tom. "Pleiades: Beyond the Barrington Atlas". Presented at APA 2006. Proceedings of the American Philological Association (2006).
- Elliott 2008a Elliott, Tom. "Constructing a Digital Publication for the Peutinger Map". In R. Talbert and R. Unger, *Cartography in Antiquity and the Middle Ages: Fresh Perspectives, New Methods*. Leiden: 2008. pp. 99-110.
- Elliott 2008b Elliott, Tom. *The Hidden Web: Don't Love It, Leave It*. Horothesia, 1 August 2008. http://horothesia.blogspot.com/2008/08/hidden-web-dont-love-it-leave-it.html.

- Elliott2008 Elliott, Tom. Barrington Atlas IDs. Horothesia, 10 July 2008. http://horothesia.blogspot.com/2008/07/barringtonatlas-ids.html.
- Executive Order 12906 Warning: Biblio formatting not applied. *Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure*. 11 April 1994. http://www.archives.gov/federal-register/executive-orders/pdf/12906.pdf.
- Executive Order 13286 Warning: Biblio formatting not applied. Amendment of Executive Orders, and Other Actions, in Connection With the Transfer of Certain Functions to the Secretary of Homeland Security . 28 February 2003. http://edocket.access.gpo.gov/2003/pdf/03-5343.pdf.
- **Fabricius 1888** Fabricius, Wilhelm. *Theophanes von Mytilene und Quintus Dellius als Quellen des Geographie des Strabon*. Strassburg: Heitz und Mündel, 1888.
- Fee 2008 Fee, J. VGI Meh. Spatially Adjusted, 13 March 2008. http://www.spatiallyadjusted.com/2008/03/13/vgi-meh/.
- Foss 2007 Foss, Pedar, and Rebecca Schindler, eds. *MAGIS: Mediterranean Archaeological GIS*. The Collaboratory for GIS and Mediterranean Archaeology, 2007. http://cgma.depauw.edu/MAGIS/.
- Genoni 2005 Genoni, Paul, Helen Merrick and Michele Wilson. "The Use of the Internet to Activate Latent Ties in Scholarly Communities". *First Monday* 10 (2005). http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1301/1221.
- **GeoWeb Exploration 2007** GeoWeb Exploration: Mining Information from Collections and KML. Virtual Earth / Live Maps Blog . 19 October 2007. http://virtualearth.spaces.live.com/blog/cns!2BBC66E99FDCDB98!9381.entry.
- Gillies 2008 Gillies, Sean. Entries: Category [REST]: Representational State Transfer. Sean Gillies Blog, 2008. http://sgillies.net/blog.
- Goodchild 2004 Goodchild, Michael F. "The Alexandria Digital Library Project: Review, Assessment, and Prospects". *D-Lib Magazine* 10: 5 (May 2004).
- **Goodchild 2007** Goodchild, Michael F. *Citizens as Sensors: The World of Volunteered Geography*. 18 September 2007. http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Goodchild\\_VGI2007.pdf.
- Goodchild 2007a Goodchild, Michael F., and R. Gupta. Workshop on Volunteered Geographic Information. 2007. http://www.ncgia.ucsb.edu/projects/vgi/.
- **Gregory 2002** Gregory, Ian. A Place in History: A Guide to Using GIS in Historical Research. AHDS, 2002. http://hds.essex.ac.uk/g2gp/gis/index.asp.
- **Guthrie 2008** Guthrie, Kevin, Rebecca Griffiths and Nancy Maron. *Sustainability and Revenue Models for Online Academic Resources: An Ithaka Report*. Ithaka, 2008. http://sca.jiscinvolve.org/files/2008/06/sca\_ithaka\_sustainability\_report-final.pdf.
- Hansen 2004 Hansen, M., and T. Nielsen. An Inventory of Archaic and Classical Poleis. Oxford: 2004.
- Hartzler 1998 HartzlerB. Metis QTVR. The Stoa Consortium. http://www.stoa.org/metis.
- Hill 2006 Hill, L. Georeferencing: The Geographic Associations of Information. Cambridge, MA: MIT Press, 2006.
- Hudson-Smith 2008 Hudson-Smith, Andrew, and Andrew Crooks. "The Renaissance of Geographic Information: Neogeography, Gaming and Second". *Centre for Advanced Spatial Analysis* 142 (2008). http://www.bartlett.ucl.ac.uk/casa/publications/working-paper-142.
- Kiepert 1878 Kiepert, H. Lehrbuch der Alten Geographie. Berlin: D. Reimer, 1878.
- Kiepert 1881 Kiepert, H., and G.A. Macmillan. A Manual of Ancient Geography. London: Macmillan and Co, 1881.
- Kiepert 1910 Kiepert, H. Atlas Antiquus. Berlin: D. Reimer, 1902.
- Knowles 2002 Knowles, A., ed. Past Time, Past Place: GIS for History. Redlands, CA: 2002.
- Knowles 2008 Knowles, A., and A. Hillier, eds. *Placing History: How Maps, Spatial Data and GIS are Changing Historical Scholarship*. Redlands, CA: 2008.
- Kozareva 2006 Kozareva, Zornita. "Bootstrapping Named Entity Recognition with Automatically Generated Gazetteer Lists". Presented at ACL 2006. Proceedings of the Eleventh Conference of the European Chapter of the Association for

Computational Linguistics (2006), pp. 15-21.

- Lucarelli 2007 Lucarelli, Giorgio, Xenofon Vasilakos and Ion Androutsopoulos. "Named Entity Recognition in Greek Texts with an Ensemble of SVMs". International Journal on Artificial Intelligence Tools 16 (2007), pp. 1015-1045.
- Miller 2007 Miller, P. *The Semantic Web Is Everyone Confused*? Nodalities: From Semantic Web to Web of Data, 20 September 2007. http://blogs.talis.com/nodalities/2007/09/the\_semantic\_web\_is\_everyone\_c.php.
- **Mohammed 2002** Mohammed, S. *From SOL to RAGE: New Work on Collaborative Technologies for Classics*. 19 March 2002. https://www.ccs.uky.edu/ccs/profiles/Spring02/bbag/feb19.html.
- Mohr 2006 Mohr, J., J. Nicols, P. Funke and B. Stollberg-Rilinger, eds. *Mapping History*. 1997-2006. http://www.uoregon.edu/~atlas/.
- **Mostern 2008** Mostern, R. "Historical Gazetteers: An Experiential Perspective, with Examples from Chinese History". *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 41 (2008), pp. 39-46.
- Nadeau 2006 Nadeau, David, Peter D. Turney and Stan Matwin. "Unsupervised Named-Entity Recognition: Generating Gazetteers and Resolving Ambiguity". Presented at CAIAC 2006. Proceedings of the Canadian Conference on Artificial Intelligence (2006), pp. 266-277.
- Office 2000 Office of the Press Secretary, The White House. Statement by the President Regarding the United States' Decision to Stop Degrading Global Positioning System Accuracy. 1 May 2000. http://www.au.af.mil/au/awc/awcgate/space/gps\_clinton.htm.
- **Ohazama 2007** Ohazama, C. Search for KML in Google Earth. Google Maps API Blog, 14 February 2007. http://googlemapsapi.blogspot.com/2007/02/search-for-kml-in-google-earth.html.
- Petrou 2007 Petrou, D. *Books: Mapped*. Inside Google Book Search, 25 January 2007. http://booksearch.blogspot.com/2007/01/books-mapped.html.
- Peuquet 2002 Peuquet, Donna. Representations of Space and Time. New York: Guilford Press, 2002.
- Pritchard 2008 Pritchard, David. "Working Papers, Open Access, and Cyber-infrastructures in Classical Studies". *Literary and Linguistic Computing* 23: 2 (2008), pp. 149-162. http://ses.library.usyd.edu.au/handle/2123/2226.
- **Putz 1994** Putz, S. *Interactive Information Services Using World-Wide Web Hypertext*. 1994. http://www2.parc.com/istl/projects/www94/iisuwwwh.html.
- Ratliff 2007 Ratliff, E. "Google Maps Is Changing the Way We See The World". Wired Magazine 15: 7 (July 2007).
- Rohs 2007 Rohs, M. Camera Phones as Magic Lenses for Paper Maps. 3 September 2007. http://www.deutsche-telekomlaboratories.de/~rohs/wikeye/.
- **Rydberg-Cox 2000** Rydberg-Cox, J.A. "Knowledge Management in the Perseus Digital Library". *Ariadne* 25 (24 September 2000).
- Scaife 1999 Scaife, R. Stoa Waypoint Database (fwd). posting to the classics-I listserv. 5 October 1999. http://web.archive.org/web/20031206054959/http://omega.cohums.ohio-state.edu/mailing\\_lists/CLA-M/1999/0294.php.
- Scaife 2006 Scaife, R. Bronze Age Crete: Archaeological Sites of Minoan Crete and the Cyclades Associated with Descriptive Scholarly Resources. 22-31 May 2006. http://bbs.keyhole.com/ubb/showflat.php? Cat=0&Board=EarthHistory&Number=425243.
- Schutzberg 2007 Schutzberg, A. "Google KML Search: What Does It Mean for Geospatial Professionals?". *Directions Magazine* (16 February 2007).
- Shaalan 2008 Shaalan, Khaled F., and Hafsa Raza. "Arabic Named Entity Recognition from Diverse Text Types". Presented at *GoTAL 2008. Proceedings of the Sixth International Conference on Natural Language Processing* (2008). http://dblp.uni-trier.de/db/conf/tal/gotal2008.html#ShaalanR08.
- Shadbolt 2006 Shadbolt, N. "The Semantic Web Revisited". IEEE Intelligent Systems 2: 13 (2006), pp. 96-101.
- Smith 2001 Smith, David A., and Gregory Crane. "Disambiguating Geographic Names in a Historical Digital Library". Presented at ECDL 2001. Proceedings of the Fifth European Conference on Research and Advanced Technology for Digital Libraries (2001), pp. 127-136. http://perseus.mpiwg-berlin.mpg.de/Articles/geodl01.pdf.

Smith 2002a Smith, David A. "Detecting Events with Date and Place Information in Unstructured Text". Presented at JCDL

2002. Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries (2002), pp. 191-196.

- Smith 2002b Smith, David A. "Detecting and Browsing Events in Unstructured Text". Presented at SIGIR 2002. Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2002).
- Smith 2005 Smith, N. *Digital Incunabula pre-release development version*. 19 August 2005. http://chs75.harvard.edu/projects/diginc/techpub/registry.
- Spiro 2008 Spiro, L. Signs that Social Scholarship is Catching on in the Humanities. Digital Scholarship in the Humanities, 11 March 2008. http://digitalscholarship.wordpress.com/2008/03/11/signs-that-social-scholarship-is-catching-on-in-thehumanities/.
- Sterling 2007 Sterling, B. "Dispatches From the Hyperlocal Future". Wired Magazine 15: 7 (July 2007).
- **Stødle 2007** Stødle, Daniel, John Markus Bjørndalen and Otto J. Anshus. "The 22 Megapixel Laptop". Presented at *EDT* 2007. *Proceedings of the 2007 workshop on Emerging Displays Technologies* (2007). http://doi.acm.org/10.1145/1278240.1278248.
- Syme 1995 Syme, R. Anatolica: Studies in Strabo. Oxford: Oxford University Press, 1995.
- **Talbert 1992** Talbert, R. "Mapping the classical world: major atlases and map series 1872-1990". *Journal of Roman Archaeology* 5 (1992), pp. 5-38.
- Tilkov 2007 Tilkov, S. A Brief Introduction to REST. infoQ, December 2007. http://www.infoq.com/articles/rest-introduction.
- Turner 2006 Turner, A. Introduction to NeoGeography. O'Reilly media, 2006. http://oreilly.com/catalog/9780596529956/.
- **Unsworth 2000** Unsworth, J. Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? Symposium on Humanities Computing: formal methods, experimental practice sponsored by King's College, London, 13 May 2000. http://www.iath.virginia.edu/~jmu2m/Kings.5-00/primitives.html.
- **Unsworth 2006** Unsworth, John, et al. *Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences. The ACLS Report.* 2006. http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf.

Wilson 2007 Wilson, B. "News About Digitization Projects". D-Lib Magazine 13 (November 2007).

**babeu 2008** Babeu, Alison, David Bamman, Gregory Crane, Robert Kummer and Gabriel Weaver. "Named Entity Identification and Cyberinfrastructure". Presented at *ECDL 2007. Proceedings of the 11th European Conference for Research and Advanced Technology for Digital Libraries* (2007).

Şahin 2007 Sahin, S., and M. Adak, eds. Stadiasmus Patarensis Itinera Romana Provinciae Lyciae. Istanbul: 2007.



This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.