

## Jump-starting the next level of online geospatial collaboration: Lessons from AfricaMap

Benjamin Lewis & Weihe Guan

*Center for Geographic Analysis, Harvard University, Cambridge, Massachusetts, USA*

**ABSTRACT:** This article discusses an opportunity to engage in new forms of collaborative research made possible by widespread access to the web and mapping technologies. While throughout history there have been references of various types to locations on the earth, very little has been organized in a way to support even simple spatial search or visualization on maps, let alone collaboration. Factors which hinder such geospatial collaboration will be suggested along with thoughts for moving forward. AfricaMap, a public online mapping project under development at Harvard, will be explored for ideas and lessons. Finally, a series of new tools is envisioned to push online collaboration to a more productive state of practice for researchers in all disciplines.

*Keywords:* AfricaMap, geospatial collaboration, mapping technologies, GIS, open source

### 1 DEFINITION OF THE PROBLEM

“Whatever occurs, occurs in space and time” (Wegener 2000). It follows that historic documents should include many references to geography (and time) that could be displayed dynamically on a map. Indeed, a study by the University of California of 5 million library catalog records found that half contain place name references (Petras 2004). Other studies which examine the contents of individual documents have confirmed the existence of an even larger percentage of place references.

Consider the potential significance of place references in light of the first law of geography which posits everything is related to everything else, but that near things are more related than distant things (Tobler 1970). If place references are converted to numbers corresponding to latitude and longitude a computer can understand, all documents become spatial datasets which can be discovered, visualized, and analyzed using geospatial technologies. To tackle the difficult problems of recognizing and then making productive use of geographic references within documents, a combination of online human collaboration and automated tool development is required.

Online collaboration occurs whenever someone adds new information to the web and shares it with others. Until recently technical skills were required to add content to the web. Now only literacy is required. The common and simple activity of writing on the web is changing the way we communicate and, in so doing, changing the way we do research. To give a sense of the breadth of this phenomenon, consider that it includes such diverse activities as creating a book review on Amazon, adding new code to an open source software project, tagging a photo in Flickr, editing an article on a wiki, loading a video to YouTube, adding a comment to a Facebook wall, as well as blogging, and tweeting. Even using a search engine writes information (not generally made public) which is used to shape future search results.

The use of a geographic reference is optional in most of these activities. In the past, few such references were added to information on the web because it was difficult, time consuming, and

costly. With the advent of free tools the cost is now close to zero, and the volume of geospatial writing is increasing. A document describing Google Earth's collaborative geospatial format KML stated, "the current geo-mass market operating environment consists of millions of users ... using tens of millions of existing and indexed KML files and resources ... which is growing rapidly" (Wilson 2009).

Collaborative Geographic Information Systems have the potential to change the way we organize and share scholarly materials. By using latitude and longitude in addition to traditional search methods such as keyword, title and author, new kinds of questions can be asked such as: "Show me all items of type 'basket' from the Peabody collection which were found within 100 miles of Lake Tanganika" or "Display all ethnographic regions which intersect the travel routes of Livingston." The Ushahidi project built an application using the Google Maps Application Programming Interface (API) that allows users to post information on a map associated with the violence that followed the 2008 elections in Kenya. The application allows users from anywhere in the world to build a detailed picture of the timing and distribution of events (Zuckerman 2008).

"For true media innovation to have human impact it must effect the imagination" (Scharl 2007). Displaying information visually and referencing it to a familiar landscape may help people to better interpret information. For example, a photograph of an insect that causes crop failure is more useful when accompanied by a latitude/longitude coordinate. The process of adding geographic coordinates to information on the web represents a new form of geospatial communication, which is a particular instance case of the broad phenomenon of online collaboration.

Until recently the tools which enable map-based collaboration were not practical because most users did not have access to bandwidth necessary to display maps or large images quickly. By contrast, ASCII text-based collaboration had been available since the early days of the Internet. In 1988 the Internet opened to the public, and it became possible for anyone to obtain an internet connection and set up a server. In those days ASCII naturally carried the bulk of communication content because of its expressiveness and small memory requirements. By the mid-90's, as the cost of memory and bandwidth decreased, the traffic volume of other forms of media requiring more memory, such as images, sound, and video, increased (Hobbes 2010).

In 1996 MapQuest was released and quickly became a popular web tool for finding driving directions. MapQuest was not collaborative as it did not allow people to add their own information to the map, and one could not use it to define new geographic coordinates. Requiring less bandwidth and memory, text-based collaboration was able to start earlier. In 1999 the first blogs were created and mass-scale, text-based web collaboration began (Blood 2000). Before blogs, one had to be able to write HTML to publish simple text messages on the web. It would not be until 6 years later in 2005 with the introduction of the Google Maps API, that map-based collaboration would be positioned to reach a mass audience. As with blogging, the uptake was rapid and exponential, jumping from a few geospatial web users to millions in a couple years. To the surprise of many in the geospatial field, the commercially funded Google Maps platform fulfilled many elements of the geospatial collaboration vision GIS professionals had been prototyping for years (Miller 2006).

Important improvements in web infrastructure occurred between the appearance of MapQuest and the arrival of Google Maps, which made the success of the latter possible. First, the speed of the average internet connection went up 50 fold during this period (Nielsen 2008), and processing power increased even more, doubling every 2 years as had been predicted by Moore's law (Wikipedia, Moore's Law 2010). At the same time browsers became more sophisticated application deployment platforms, able to process XML that, when combined with increases in bandwidth, made possible the first global base map delivery service. Google took advantage of infrastructural improvements to remove a fundamental barrier to geospatial activities on the web, specifically, access to fast, global, high quality mapping.

Google Maps represented the first widespread use of AJAX, a new combination of the existing technologies XML and JavaScript. The mashup was born, and along with it,

new forms of data sharing. By this time, the mass use of Global Positioning Systems (GPS) in conjunction with collaborative geospatial platforms like OpenStreetMaps (OSM) began to demonstrate the power of crowdsourcing for organizing geospatial knowledge. In a similar way, but starting earlier because of its text orientation, Wikipedia had begun to enable large scale organization of knowledge in the form of text and images. A significant difference between OSM and Wikipedia is that Wikipedia includes a large historical dimension. OSM and much of the geospatial web does not—yet.

The introduction of Google Maps and Google Earth in 2005 established many aspects of the technical foundation that exist today. Elements of Google's architecture such as the use of the Web Mercator projection and pre-cached map tiles were adopted by competitors. From a user perspective, the new elements were: 1) A detailed, current base map of roads and satellite imagery for the globe. 2) Free, fast delivery of base maps via an open Application Programming Interface (API) which can be incorporated into custom applications. 3) Tools for annotating the base maps. KML files created in Google Earth are one example of such tools. Mashups made with the Microsoft Bing Maps API are another.

Much has been built on top of this framework since 2005; however, most developments have been incremental. An important characteristic of Google's approach is the implicit distinction made between the two primary geospatial data structures, raster and vector. Raster formats (often JPEG, PNG, or GIF) requiring large amounts of storage space are used to display static satellite and roads base maps, while vector formats (primarily flavors of XML) requiring little storage space, support the addition of new geographic information to the web and serve as annotation of the base maps. The base maps, which if downloaded would comprise many terabytes, are delivered via a web service. The fundamentally necessary, always on, current, raster base map became the geospatial dial tone of the web. Against this reference system anything else can be referenced, comments can be made, new data created by creating XML annotations in the form of hyperlinks, geotags, GeoRSS, GeoJSON, KML, etc. For the first time, people with no professional background in geography or GIS began actively engaging with these technologies. The term VGI (Volunteered Geographic Information) was coined to describe the new kinds of data being created by this emerging class of geographers (Goodchild 2007).

Google's contributions, and the various efforts and innovations it spurred on the part of Microsoft, Yahoo, MapQuest, and others, moved geospatial collaboration forward. From the vantage point of a few years later, we can get a better sense of the size of the geospatial collaboration iceberg. We understand that most geospatial collaboration involves the creation of annotations of base maps that depicts current conditions. Little georeferenced historic information is available on the web, with the consequence that most annotations lack historical depth. Today makes sense as the place to start, but what about History? How should we handle the past geospatial annotations people have created in the form of cartographic (mostly paper) maps? Interestingly, people have been geospatially annotating the world for thousands of years (Harley 1987), often with more accuracy and consistency than web-born annotations being created today (Fig. 1).

Modern colonial powers have been drawing detailed maps of the territories they control or wish to control for hundreds of years. Since the development of the theodolite in 1790, a significant portion of historic maps are accurate enough to reference well to modern base map. (Wikipedia Ramsden 2010). Historical maps of lesser accuracy can also be profitably georeferenced. Examples of these are the David Rumsey maps prior to 1800 available in Google Earth.

From the perspective of our current geospatial collaboration paradigm in which current satellite and road map services are (arguably) the culturally neutral, interpretation-thin base against which other information is collated and discussed, existing paper maps are a treasure trove of interpretation-laden, historic commentary about places on the planet, yet to be brought into the light of the web and made available for the general user or professional researcher to explore and use. Historical maps constitute a rich source of historic places and

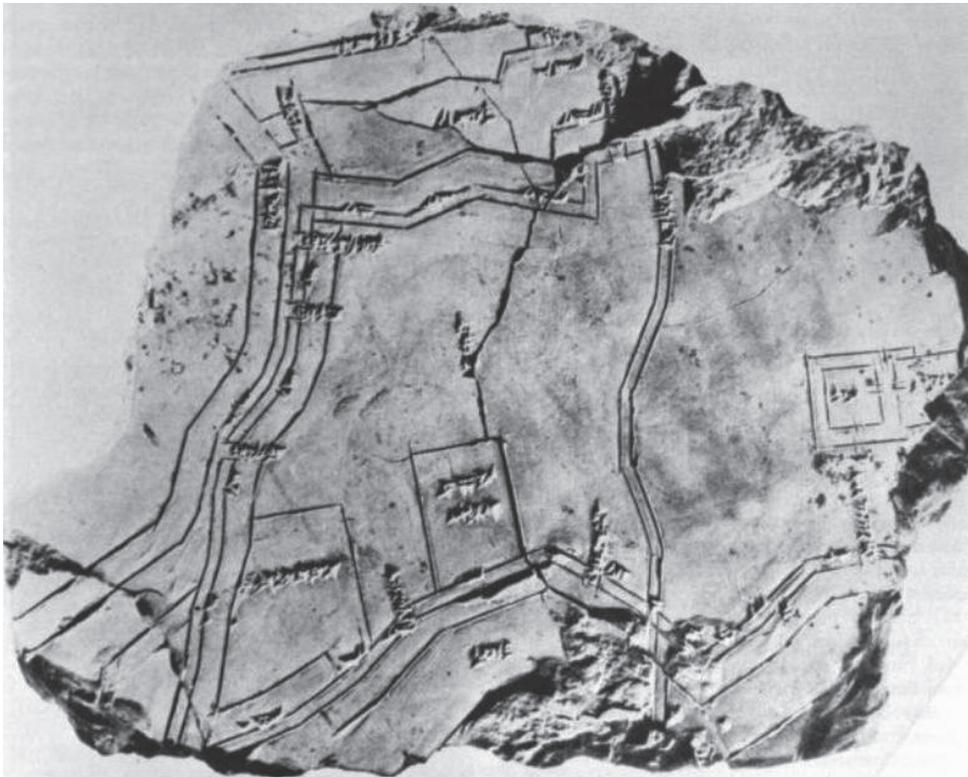


FIG. 6.7. PLAN OF NIPPUR, CA. 1500 B.C. Possibly the earliest town plan drawn to scale, this shows the temple of Enlil in its enclosure on the right edge, city walls, canals, storehouses, and a park. Size of the original: 18 X 21 cm, By permission of the Hilprecht Collection, Friedrich-Schiller-Universitat, Jena.

Figure 1. A detailed plan drawn to scale depicting a portion of the ancient city of Nippur with temple, city walls, canals, storehouses, and a park that was created 3500 years ago (Harley 1987).

names against which texts, photographs, documents, datasets, and video can be manually and automatically referenced.

There are important unaddressed issues which hinder the widespread adoption of online mapping tools to support historical research. One problem is that most existing historical materials have not yet been digitized or brought online, let alone georeferenced. Another is that there is no central place to find such information. Eventually search engines may be that place, but that is not the present case. If everyone used KML to tag web pages containing spatial data we might have the beginnings of a solution (Schutsberg 2007).

Google has current base maps and few historical maps. Google's search engine does not offer a good way to find old maps, especially ones which are georeferenced. (An exception to this is the excellent Google Earth interface for finding historic satellite imagery.) Other historical materials online are scattered across map libraries and private collections. Very few are georeferenced. Though the capabilities of online mapping systems have grown tremendously, it is still difficult for researchers to discover even the little historic spatial data there is on the web.

Barriers to spatial collaboration exist in the United States, but the situation is more severe elsewhere in the world. The problem is multifaceted, and solutions in this area must address a wide variety of areas including data storage, archiving, tool development, data integration, georeferencing, data distribution, interoperability, access to software tools, and cost (Mannak 2008). For many parts of the developing world, even basic scientific datasets are not available to researchers (Ogodo 2009). Based on our recent experience developing the

AfricaMap system, we will attempt to dissect the problem and put forward some ideas for moving the nascent domain of collaborative mapping forward.

The AfricaMap project developed at the Center for Geographic Analysis represents an initial attempt to address a part of the collaboration problem by building a framework to make some high quality data available on the web in a way that researchers can use across disciplines. AfricaMap allows users to search against large amounts of data from across disciplines and display result on a map. Based on our experience with AfricaMap, we will attempt to define the factors which help or hinder collaboration; we will discuss in detail, placing the factors in the context of current and developing technologies and initiatives.

We perceive three basic limiting factors which hinder online, spatially oriented collaboration:

1. The availability of historical or current mapping other than current satellite and roads for most parts of the world.
2. The availability of a public host to which organizations can load materials they wish to share and make materials available to others permanently for discovery and visualization.
3. The availability of tools to allow interested users to collaborate, mark up maps, hold discussions, and create workspaces.

These areas are presented roughly in order of importance. Without (1) basic mapping, there would be little incentive for users to add materials to the system and store it in (2) permanent storage. Without dependable storage it would not be possible to build (3) tools for collaboration to support researchers. Taken together, these limitations make it difficult to even imagine the possibilities in the area of collaborative map building. It is similar to trying to imagine a collaboratively built encyclopedia before Wikipedia.

The AfricaMap project starts to address problem (1), but it still lacks many features. Based on our experiences with AfricaMap and feedback from users, it is clear that there is a strong need for solutions to problems (2) and (3). These three core areas contain sub-topics. We will unpack and examine them in detail. Finally, will look at the current AfricaMap project for lessons learned and attempt to describe a way forward.

## 2 AFRICAMAP BACKGROUND

AfricaMap supports research and teaching on Africa that serves students and faculty across African studies at Harvard University and increasingly internationally. Much public data, current and historical, exists for Africa, but they are difficult to discover, let alone obtain; many researchers with projects on Africa spend critical time and funds gathering cartographic data from scratch. People in Africa have an even harder time accessing mapping of their own regions. When researchers gather data it is sometimes lost again because there is no place to store and reference it. The AfricaMap project represents a framework for organizing African data, allowing it to be retrievable, map-able, comparable, and downloadable. The data is served live from various systems inside and outside of Harvard over the Internet, and brought together within a web browser to form an integrated map for the user. Data that is stored on the AfricaMap servers are made available to other applications outside AfricaMap as map services.

At its core, AfricaMap consists of a set of public digital base maps of the continent, viewable dynamically at a range of scales and composed of the best cartographic mapping publicly available. Behind the scenes a gazetteer provides rapid navigation to specific locations. As more detailed mapping becomes available it is added to the system. Because of its decentralized architecture, there is in theory no hardware or software limitation on the amount of data that can be incorporated.

The idea for AfricaMap was developed in the Center for Geographic Analysis (CGA) at Harvard (see Acknowledgement). The project is overseen jointly by professors Suzanne Preston Blier and Peter Bol, and designed and managed by Ben Lewis. The CGA has the dual aim of supporting Harvard research that involves mapping and location-based investigation,

as well as making data created in the course of research available to others. In November of 2008 the Phase I release (Beta version) of AfricaMap was launched.

Basic AfricaMap features attempt to serve broad needs of researchers in multiple disciplines interested in Africa. These features include:

- A common web accessible set of current and historic maps for Africa. One of the greatest problems facing researchers is the lack of a commonly available base map for the continent upon which one can build one's own research materials and share them with others. Wherever possible AfricaMap is making base map data available without copyright restrictions.
- A comprehensive (and continually growing) gazetteer for African place names. Currently there are over a million names, in an array of spelling alternatives and languages in addition to English (French, Arabic, Chinese, Japanese and Russian etc.). This is not important to international users. Furthermore, by overlaying maps at different resolutions and showing different features, users can potentially "translate" place terms into these different languages.
- A repository for spatial datasets, both those compiled from other sources and new datasets created by researchers.
- A clearinghouse for research projects on Africa, allowing users to discover what regions are studied in current and past projects. Users can view the data or metadata relating to those projects and contact the researchers involved.

### 3 FACTORS WHICH HINDER COLLABORATION

During the inception, design, development and support of AfricaMap, the authors have been exposed to user demands for spatially oriented collaboration tools, as well as the technical and institutional obstacles which exist in meeting these demands. The continued development of AfricaMap is aimed at overcoming these obstacles. In this section we will examine key elements based on our experience which hinder collaboration and propose ways of addressing them.

#### 3.1 *Availability of mapping materials*

##### 3.1.1 *Availability of historic base mapping*

Historical base mapping provides orientation for many geographic investigations. At a minimum base maps generally contain roads, town names, major geographic features, and political boundaries. In many cases historical mapping made in the past 50 years or so contains much more detailed information about a given place than one can identify on a Google satellite or roads base map, especially for rural parts of the world (Figs 2a-b). Historical base mapping, when coupled with Google's current satellite imagery, provides a powerful, hitherto underutilized way to visualize change.

Determining globally how much historical mapping is missing from our current online geospatial environment is difficult. Since the early 19th century, major world powers including Britain, France, Spain, Portugal, Russia and the U.S. have systematically mapped most of the habitable earth multiple times. Most of these maps exist in paper form in library drawers. Many are in private collections. A few libraries are in the process of scanning their maps, and a few of these are engaged in georeferencing them. East View Cartographics, a major private map seller which has a sizable collection of colonial maps, claims to have over 500,000 maps. Books are also a large source of valuable maps. John Orwant, the manager of the Google Books metadata team, estimates the number of book titles in existence as 168,178,719 from more than 100 metadata sources (Orwant 2009). If we assume just 1% of these books each contain 5 maps, that would be 8 million additional maps in books. By contrast, the number of historical georeferenced paper maps online which can be downloaded

or incorporated into an application via web services today must certainly be less than 1000, and there is no easy way to find them. The latter estimate is based on extensive searches by the authors.

Making such maps available in a web-based mapping system is more of an organizational and financial challenge than a technical one, though technical breakthroughs will be as important. For example, a free online georeferencing tool for anyone to use could have a major positive impact by reducing the cost of this work and allowing the work to be shared. Such a tool could help crowdsource the work of georeferencing historic materials. An online tool has an advantage over a desktop tool because as the work is done, the newly georeferenced maps and other materials are in turn available as a reference against which additional materials can be referenced. Such an approach could allow users to comment on and repeatedly georeference the same object until the community feels it is finished. The versioning history of a given map object could be made available to the community in wiki fashion.

### 3.1.2 *Availability of georeferenced materials*

Historical base mapping is only one type of georeferenced material. If one considers that nearly all materials in the corpus of human knowledge, including maps, books, manuscripts, letters, photos of objects, photos of people, sound recordings, and video recordings, either describe places on the surface of the earth or are tied to a location in some manner, it becomes apparent that only a small portion of historical knowledge is currently available online, let alone georeferenced.

The universe of potential materials to be included in a collaborative mapping system may be divided into two broad areas: materials such as satellite imagery and GPS data, which are

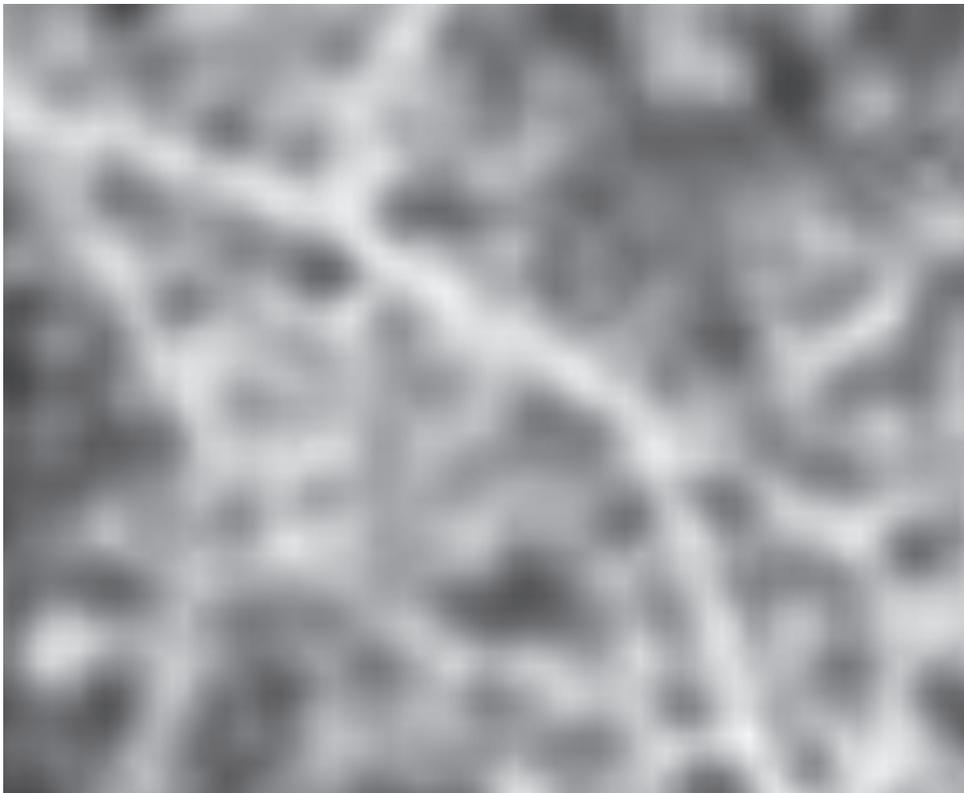


Figure 2a. Image from Google for most of Africa (Ife, Nigeria).



Figure 2b. Base mapping available in AfricaMap for the same location.

born digital and spatial, and information which must be converted to digital form and then georeferenced. Much of the material that is born digital is currently well-handled: Google Maps and other commercial providers are providing imagery via its Google Maps API, and Open Street Maps is providing a crowdsourcing platform for uploading GPS tracks to build high value datasets such as roads for all parts of the globe. AfricaMap can easily incorporate these types of materials.

The material which offers perhaps the greatest opportunity for expanding the nature of geospatial collaboration because of its both breadth of subject area and its temporal depth are paper maps in map collections and books, as well as references to places in books, datasets, images, and documents. As we have stated, much of the recorded history over the past 5 millennia describes events that happened somewhere on the surface of the earth; it is information that can be organized spatially in addition to being organized by subject and author.

Georeferencing is the establishment of relationships between information and locations using names or times or events, in order to improve searches against and visualizations of collections (Hill 2006). One can think about georeferencing as an additional method of organization that can be applied to just about any type of information. Geographic element definitions are included in many metadata formats such as Dublin Core (DCMI), providing opportunities for map collections. In many cases these fields have not been populated; that may change with the improved availability of tools for generating accurate coordinates.

Perhaps the important question for our purposes is not whether a given piece of information has a spatial component, but how much it costs to extract the spatial information and represent it in a meaningful way. To answer this we need to know what we wish to extract and how we want to represent it on a map. The choice made will vary according to information types: an ancient document might be best represented by a point while a travel account in a

book might require multiple lines. Irrespective of spatial information is extracted from an old map, a text document, or an old photograph, there are many considerations which must be taken into account, including what precisely it is that is being located in space and how the spatial information is likely to be used. Is the place the document describes to be mapped, or is it the place where the document was found to be mapped, or both?

### 3.1.3 *Integration of data with public commercial layers such as Google's*

Google and other commercial providers are making very detailed datasets available to the world and the terms are liberal. Data which a few years ago would have cost millions of dollars to access are available for free and can be incorporated into non-commercial, web-based systems easily. These datasets have immediately become, for many parts of the world, the best geographic information available to the public. But the commercial layers lack the historical depth and detailed place name information that are available on historic topographic maps. Because the commercial and historic maps each have their strengths and weaknesses, combining the two is logical. By doing so, suddenly it is possible to see change over time, and annotations on the historic mapping fill in for the missing labels on the commercial satellite mapping.

Commercial datasets have additional value because they are constantly being improved, are always on, and are well hosted on fast servers. Access never seems to be a problem, though that could change. For this reason it is important not to be dependent on the mapping of one commercial provider alone. There is a danger along with opportunity. These data may vanish at some point or become very expensive. The glib solution is to assume that they will disappear in a few years and therefore use them to the maximum degree to georeference other materials while they are still accessible. The other solution is to be aware of and support free data sources such as Open Street Maps and public domain satellite data which will always be available.

### 3.1.4 *The use of gazetteers*

A gazetteer is a list of place names with associated geographic coordinates which helps a user navigate a large spatial domain. Gazetteers can be used in a variety of ways to make content more accessible and comprehensible across space and across representations in multiple languages and historic variations. A gazetteer is very useful when combined with a historic base map from scanned paper maps. The scanned paper map is generally not searchable by place name, while the gazetteer is searchable but provides no cartographic information. As with the combination of satellite and scanned maps, the two combined are greater than the sum of their parts.

### 3.1.5 *Intellectual property considerations*

Collaboration requires the sharing of information. Often researchers do not want to give away their data until they have published their results. But sometimes it is more valuable to get feedback and find potential collaborators than to keep a dataset private. With the right tools, sometimes a researcher is willing to share data publically during research in controlled ways. Today such tools exist but they are limited in their power. One of the best is Google MyMaps but it is constrained in the number of features which can be displayed simultaneously. As there is no easy way in the geospatial world to make materials available to others and control accessibility on a day-to-day basis (as people are accustomed to doing with Facebook and other collaborative tools) many researchers default to not sharing.

Another reason that materials are not more frequently shared is that there may be no apparent, simple way of defining and enforcing the type of sharing that the user would like to implement. Perhaps a user would like to make materials available for only non-commercial use and would prefer to require attribution if data is used in a publication. The Creative Commons project (<http://creativecommons.org/>) provides legally defensible sharing-oriented licenses which cover the basic areas of attribution, modification, and commercial use and can be adapted to fit most requirements.

### 3.2 *Availability of a place to put mapping materials*

#### 3.2.1 *The cost of serving data*

It is expensive to set up a GIS server, which at the very least costs thousands of dollars in hardware, software, bandwidth, and technical expertise to put together. Providing a low cost or free hosting solution for scholarly spatial materials would be an important step towards making more of such data available. Universities have an interest in supporting initiatives which increase the access of their own scholars to materials and encourage collaboration between their scholars and those of other institutions. Many universities already spend considerable resources to serve non-spatial digital collections to the world.

#### 3.2.2 *The use of web services*

In addition to traditional web-based viewing and direct download of data, the use of web services, particularly open ones, makes many new kinds of data sharing possible. Web services allow interactive maps which reside on one system to be embedded in another organization's web or desktop application. Google Maps is such a service and can display on the Google Maps site or can be incorporated into other applications. Given the large file size of GIS downloads and the cost of hardware and software necessary to host a system, the use of web services promises to be an important part of geospatial collaboration.

#### 3.2.3 *Permanence and citations of materials*

Researchers are often willing to share their materials provided the cost is not high, their materials are easily discovered, and credit or citation goes to the owner or author. In academia, citations are the coin of the realm, and improved citeability could be used as an incentive to scholars to make their materials available. There are no map-oriented systems which offer facilities to both store geospatial information and make them permanently citable. There are systems which meet this need for tabular social science data however, such as the Dataverse Network (King 2007).

#### 3.2.4 *Data discovery, metadata and spatial indexing*

How can one find the best available online data for a particular purpose? This is currently very cumbersome when it comes to spatial materials and standard search engines. One reason is that search engines are oriented toward finding pages, not data, let alone the range of types of spatial data. Even within sites which specialize in data this task can be arduous. There is often no easy way to get a sense of the overall contents of a particular data warehouse. A user gets a list back, but if there are a large number of results in the list and the system does not have a good way to prioritize results, the chance of finding the materials searched for is small.

A possible solution is to provide a holistic view of the contents of a system using a map to show the footprints of all materials, a tag cloud view to show the concentration of keywords, or a time graph to allow one to find all materials which deal with a particular period in history. All of these approaches to some extent bypass the problem of discipline-based vocabularies or ontologies; instead they use discipline-independent terms such as space (lat/long), time (date range), and tag (user defined tag). Combining multiple such holistic views (map, tags, time) in a faceted search will provide a powerful way to sift through large collections without much a priori knowledge of the contents.

Many types of information describe a place on the earth and have an implied geographic footprint which defines the area of interest. Creating geographic footprints for non-spatial data is an important aspect of georeferencing. In AfricaMap, there is a Project Layer which is composed of the geographic footprints of research projects completed, on-going, or planned for in Africa. Such footprint layers are a spatial form of metadata, describing and linking to the project materials from the map and making non-spatial project information searchable spatially.

### 3.3 *Usability of the system*

#### 3.3.1 *Interface design*

Many GIS systems are cumbersome and difficult to use. There is often a large tool bar that the user is expected to learn. Because the intended users of such data sharing systems are often not professionals in information technology fields, the system's interface design should be simple and conform to de facto standards wherever possible. Advanced functionality can be present, but should be ideally behind the scenes. Simple design promises a system easier to implement across disciplines and across cultures and languages.

#### 3.3.2 *The importance of performance and scalability*

Many geospatial systems that are oriented toward scholars or developed by non-commercial efforts are slow. Slowness isn't just a problem of wasting a user's time; it is often a show stopper. People generally do not come back to unresponsive systems and are unlikely to recommend them to others. The problem of speed is related to its scalability. A system might be fast with a moderate amount of traffic but very slow when traffic increases beyond a certain point. In order to handle gracefully increasing amounts of traffic a system must be scalable.

A web services architecture supports scalability because it provides a way to decentralize computationally intensive functions. The same technologies can also support collaboration by allowing various organizations to provide data services which can be organized for the end user as a single mapping system.

In addition to technical approaches which support scalability, software cost is an important factor for allowing a system to scale in order to meet demand. When additional servers are needed to help with performance, the cost of software to support more servers can be prohibitive. Open Source GIS software is highly scalable in terms of cost; however there are many factors that must be considered when calculating cost, such as the skill set of the organization.

#### 3.3.3 *Interoperability with other systems*

Web services, open standards, open source software, open data formats, open metadata formats, web access, operating system platform independence—all these are important elements of interoperability and are important to consider when designing systems that attempt to reduce barriers to data access and improve the chances of successful collaboration between scholars.

### 3.4 *Availability of a place to collaborate with mapping materials*

#### 3.4.1 *Research audience should be defined broadly*

We are interested in improving the ability of academic researchers to collaborate, but we are also interested in building systems which bring a wide variety of users together, whether from government, non-profits, private sector, or the general public. The problems and opportunities are too great and crucial for interested people not to be able to work together.

The research audience differs from but overlaps with a general audience, and the success of Wikipedia is a vivid example of this. A system that only aims to serve the academic community but does not build common cause with other groups interested in mapping knowledge is missing an opportunity to create a more significant system with greater longevity.

#### 3.4.2 *Multiple language support*

For both the research audience and the general public, multiple language support is essential for collaboration. Most of the open source components used in the AfricaMap framework are able to support multiple languages without further development. For example, one can search by place names in any language and will be able to find the place if it is recorded in that language in the gazetteer. However, some challenges exist in making the system compatible with a variety of client operating systems internationally; these are issues that which go beyond language support.

### 3.4.3 Collaboration tools

True collaborative research requires more than data searching, viewing, uploading and downloading. It should also include tools for defining and inviting collaborators, controlling access to some materials, commenting on materials, tracking changes to editable layers, and more. AfricaMap currently lacks many of the tools we deem to be important. Those we envision are described in the next section.

## 4 AFRICAMAP LESSONS LEARNED AND SOLUTIONS RECOMMENDED

### 4.1 Data currently in the system

The project started with publicly available continental scale data, and used that as a framework against which to start to bring in more detailed materials. In addition to Google's satellite and physical maps, AfricaMap includes continent-wide layers to support place-based search and materials georeferencing for even the most remote regions of the continent. Other important layers in AfricaMap include the Operational Navigation Charts (ONC), Tactical Pilotage Charts (TPC), Joint Operations Graphics (JOG), and Army Map Service (AMS) sets of mapping from the U.S. National Imagery and Mapping Agency (NIMA). In addition, a full set of seamless Soviet mapping at 1:500,000 scale for the continent was acquired. Other important thematic layers in AfricaMap useful for georeferencing include current and historical administrative boundaries, historic maps, ethnographic boundaries, and language boundaries. There are additional critical datasets that are gradually being added to AfricaMap such as Russian 200,000 scale mapping, census data, and health data. Table 1a and 1b list the major data layers in the current AfricaMap release.

The gazetteer (place name) layer in AfricaMap was derived from GeoNames (<http://geonames.org>) which itself is composed of several gazetteers. The main source for GeoNames is the GEOnet Names Server (GNS) maintained by the U.S. National Geospatial Intelligence Agency (NGA). The vocabulary of place types in GeoNames follows the NGA's place types, and GeoNames has added a few new types. Of the various public global gazetteers, none are complete for all feature types, and there is no widely agreed upon vocabulary for place types. GeoNames is perhaps the closest to providing a de facto standard for gazetteers, and because it is being improved wiki fashion by a global user community, seems most likely to become the most complete public gazetteer. One weakness of GeoNames is the lack of historical depth. Even though it records place name aliases, they are not marked by the time when the alias was in use.

Table 1a. Major raster data layers in AfricaMap phase 1 beta release.

Series name	# Map sheets	Coverage	Scale	Year	Size on disk GB
U.S. ONC	47	95% continent	1:1,000,000	1979–1998	13
U.S. TPC	142	75% continent	1:500,000	1974–2002	40
U.S. JOG	1039	50% continent	1:250,000	1965–2005	58
U.S. Burundi	40	100% continent	1:50,000	1994	3
Soviet	550	100% continent	1:500,000	1962–2003	6
Soviet	3250	80% continent	1:200,000	1962–2004	40
Nigeria	1039	50% country	1:100,000, 1:50,000	1960–1978	58
French 1898	62	100% continent	1:2,000,000	1898	12
Sierra leone	190	100% country	1:50,000	1964–1973	3
Freetown	108	100% city	1:2,500	1941–1965	5
Historical maps	10	100% continent	varies	1612–1930	2
Health maps	4	100% continent	1:10,000,000	2000	1
<b>TOTAL</b>	<b>6477</b>				<b>240</b>

Table 1b. Major vector data layers in AfricaMap phase 1 beta release.

Map name	# Features	Coverage
HRAF ethnographic atlas	847	100% continent
People's ethnographic atlas	1927	100% continent
Scholarly maps on history	150	100% continent
Harvard map collection index	260	100% continent
Projects index	45	NA
Place name gazetteer	~1,000,000	100% continent
Admin boundaries	17,155	100% continent
Population surface	1 km grid	100% continent
Population centroids	109,177	100% continent
Lakes and rivers	150,572	100% continent
Soils great groups	4,909	100% continent
Surficial geology	11,977	100% continent
Land cover	5 km grid	100% continent
Power plants	1637	100% continent
<b>TOTAL</b>	<b>1,298,506</b>	

## 4.2 Current AfricaMap system architecture

### 4.2.1 Hardware

AfricaMap currently runs on a single Linux server running the Red Hat Enterprise 5 operating system and is hosted at the Harvard-MIT Data Center (HMDC) with a one gigabyte per second connection to the Internet. AfricaMap currently accesses about half a terabyte of raster and vector data on disk.

### 4.2.2 Software

AfricaMap uses the Open Source PostgreSQL relational database for storage of vector (point, line polygon) spatial features. Geodatabase functionality is enabled within PostgreSQL by the PostGIS spatial libraries. PostGIS supports a wide variety of spatial SQL operators and is a platform for querying, analyzing, extracting, and converting spatial data. For rendering and symbolizing map data out of PostGIS, MapServer is used. MapServer connects directly to PostGIS and provides an Open Geospatial Consortium (OGC) compliant Web Map Service (WMS) interface for AfricaMap (Fig. 3).

AfricaMap uses a JavaScript-based map client to display map tiles from a local tile cache as well as tiles from remote servers such as Google's. All map data is rendered as tile caches including the gazetteer (place name database), political boundaries, and historic maps. Tile generation occurs based on WMS (Open Geospatial Consortium Web Map Service) requests.

Caches are generated from the original GIS data by map renderers MapServer and GeoServer. All queries against vector (point/line/polygon) GIS datasets are handled by PostGIS, a spatial relational database. Tiles are stored in an open tile format called TMS (Tile Map Service).

Data can be stored in any source projection and re-projected on-the-fly to Spherical Mercator for overlay with commercial map services such as Google Maps. Google Earth versions of the raster layers are also made available as KMZ files using superoverlays.

### 4.2.3 Web services

Using WMS for map rendering within AfricaMap makes it possible to provide web service access to AfricaMap data from applications anywhere on the web. The maps in AfricaMap can be used by systems other than AfricaMap without their needing to download and configure the (often very large) datasets. Since AfricaMap consumes open services it is also able to consume map services hosted by other organizations.

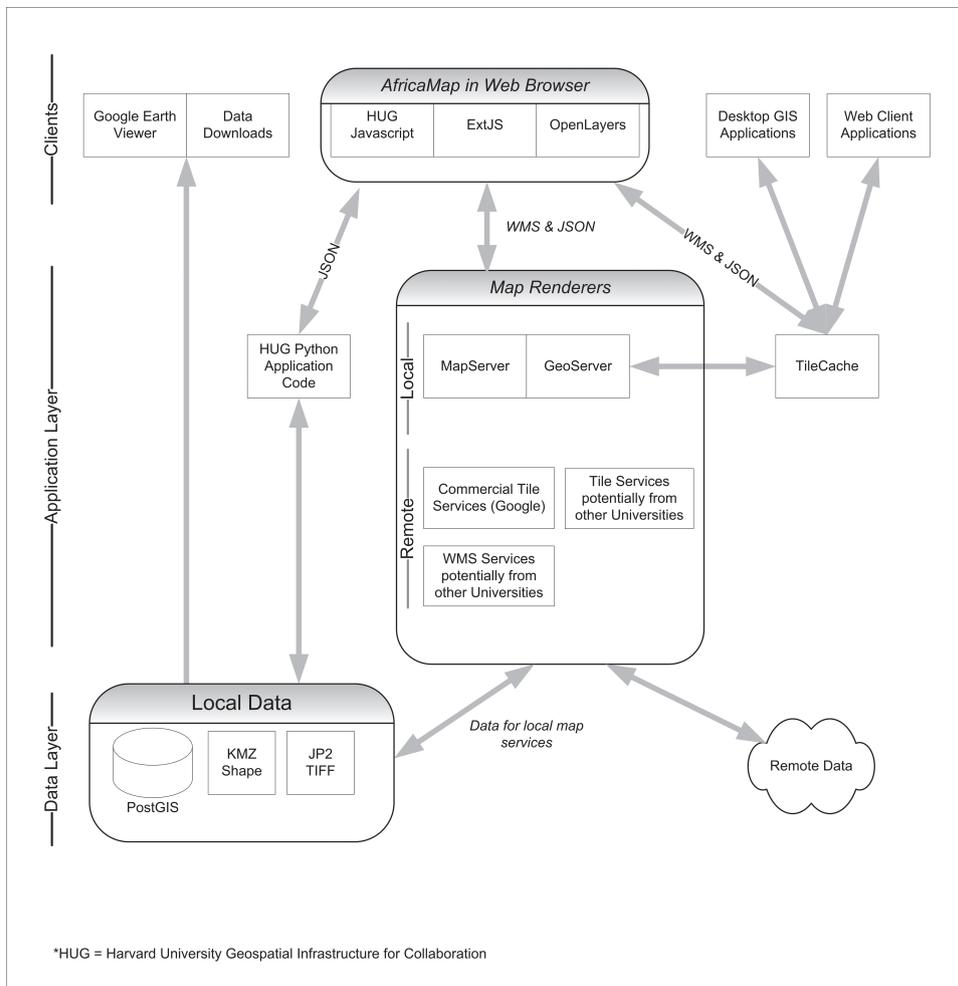


Figure 3. Current AfricaMap phase I architecture.

To make AfricaMap's map services scalable, TileCache is used which generates an Open Source Geospatial Foundation (OSGeo) compliant Tile Map Service (TMS) from a WMS. In order to integrate TMS layers and Google Maps layers within a common application, AfricaMap uses the OpenLayers JavaScript library. OpenLayers also contains connectors to proprietary mapping services such as Google Maps and Microsoft Virtual Earth. Unlike Google Maps mashups, OpenLayers-based platforms are not dependent on Google Maps, but can use many different commercial or non-commercial services.

To support rapid query and map display against datasets with millions of features such as the AfricaMap place name database, full text indexing using the Tsearch2 PostgreSQL extension is used. This approach allows a user to search multiple layers at once and display the results of all matches from all layers on the map. Unlike many web GIS applications, there is no browser-side limit to the number of query results that can be displayed. This is accomplished by using an image-based service instead of an XML or vector-base approach for rendering features.

AfricaMap uses open standards, open web service protocols, and open data formats. Open standards are the basis of the web, and AfricaMap is a web-centric infrastructure which

benefits greatly from the technical and social benefits accruing from open standards used in the Open Source software from which it is built.

#### 4.2.4 *Data formats*

GIS image files used by AfricaMap conform to the open GeoTIFF standard for spatial imagery. Vector data conforms to the open Shapefile specification and Well Known Text (WKT), and large images are made available in the open JPEG 2000 format and KMZ format. All of these file types represent leading open geospatial data formats. The core GIS libraries used are the Open Source OGR Simple Features Library and the Geospatial Abstraction Library (GDAL), which are capable of reading and writing a wide variety of formats.

In AfricaMap all map data are projected to a common spatial reference system, Spherical Mercator, in order to overlay with commercial layers such as Google Satellite. Because Spherical Mercator is not commonly used in desktop GIS systems, data made available for download are provided in the more commonly used Geographic space WGS84 space. Web services in AfricaMap support both Spherical Mercator and Geographic. AfricaMap uses the OGC Well Known Text (WKT) standard for describing and communicating projection information in web services and GIS data.

#### 4.3 *Functionality of the system and what's missing*

AfricaMap supports an initial implementation of a solution to some of the problems identified in this section. The areas for which AfricaMap does not provide solutions are critical and need to be addressed in order to jump start the next level of online geospatial collaboration envisioned here.

The AfricaMap project has made a small start toward addressing the problem of availability of basic mapping and other kinds of spatial data; AfricaMap presents some large datasets to users via a simple interface that supports intuitive comparisons between mapping layers. However as groups and organizations have been exposed to and become users of AfricaMap, many have expressed the desire for tools that would allow them to load their own materials to the system as well as tools to support various forms of online collaboration.

In addition, many groups have asked for versions of the system which focus on other geographic regions, and these groups of course need data loaded for those regions. In fact there have been so many requests for additional data that we have not been able to keep up with the requests and are working on a tool to allow users to upload their own materials. There have also been many requests for the ability to control access to datasets. Because everything in AfricaMap is public at this point, there are many datasets which we have not been able to add to the system because they cannot be made completely public.

Furthermore, there have been requests from various groups for tools to use to georeference collections of materials such as map sheets and digital objects representing objects in museums. The availability of georeferencing tools will be key to the creation of an environment in which a range of scholarly materials can be searched and resulting distributions visualized on a map.

## 5 PROPOSED ENHANCEMENTS TO AFRICAMAP

Based on our experience, the following are the key areas that must be addressed in order to move the AfricaMap framework forward in the direction of a general purpose collaborative platform:

1. Assemble fundamental sharable geographic materials such as major map series for globe developed by the United States and Soviet Union at a minimum. Include integration with commercial providers, gazetteers, and best global layers such as the LandScan population distribution surface and GADM administrative boundaries.

2. Provide a place (ideally a central index to materials on various servers) to store materials, control access, and make map materials usable. Discovery of any portion of the data should be possible via search engines as well as from a central mapping application.
3. Tools for georeferencing materials of various types need to be provided for anyone to use.
4. Tools for collaborating and engaging in scholarly exchanges with regard to map materials should be provided.

The current Beta release of AfricaMap delivered only a small portion of these functions; however, its architectural framework is designed to accommodate further enhancements in this direction.

### 5.1 *Proposed collaboration tools*

Users need to be able to create their own logins to the system and upload, georeference, and control access to their materials. One simple type of georeferencing tool would involve the referencing of URLs. Being able to define a geographic location and basic metadata for a persistent URL which points to valuable information is a potentially useful enhancement. Such URLs could point to digital objects in collections or to valuable online resources which describe a given place. The geographic “footprint” of the URL might consist of a single or multiple points or lines or polygons.

A bulk version of this tool could allow the user to upload a table of URL information and specify fields in their table likely to contain references to places. These could be automatically parsed and matched against a gazetteer to determine preliminary lat/long coordinates. After performing a preliminary automated coordinate lookup, the owner would be able to check and correct the coordinates assigned within an interactive mapping environment. Coordinate values will be stored in the database in the OGC standard Well Known Text format which could then be output to any spatial format. It is important that users be able to export out any materials they create for use in other systems.

A critical feature of this proposed system is the ability for users to control in a fine-grained way how their materials area shared. Users must be able set permissions such that their materials are visible to only themselves, visible to those they choose to share access with, or made public. Users should be able to control whether others can edit their materials or only view them. Such an approach supports the typical research materials life-cycle in which sources at early points in a project must be confidential; after publishing, the materials can be made public. There may even be an active desire on the part of the researcher to make materials available to support a published article.

For a map-centric collaboration platform, a tool to georeference scanned map images is critical. Historic maps are a key resource in this system which will be used to georeference other historical materials. The system should allow the user to upload a scanned map image, and then georeference the scanned image by identifying common points between the uploaded image and already projected mapping in the system. Once a sufficient number of common points are defined, the system would reference the scanned map to geographic space. The system could be designed to allow users to perform this task as frequently as desired to obtain a good fit between the map and the “real” world. Users would be able to comment on one another’s referencing work and even revise referencing performed by others. The creator would be able to choose whether to accept a corrected version of their material.

A tool to facilitate scholarly exchanges and conversations relating to spatial objects could allow such a system to become a platform for communication at various levels of geographic resolution. Users will be able to comment on mapped materials by placing notes in the form of point, line, or polygon features on the map and annotating them. Such a system should automatically track the user, date, location, and map reference for each comment, and the user should be able to email their map markup to colleagues. A key element here could be a URL, automatically generated, and included in the email along with the comment. Upon clicking on the URL, the email recipient would be taken into AfricaMap and the comment

displayed in the context of any maps that were displayed when the email sender created the comment.

Using this approach any comment created by a user could be syndicated using the syndication standard for blogs, RSS. Users would be able to choose to make any comment appear in their RSS feed which anyone would be able to subscribe to. This type of feature could make map comments made in AfricaMap usable by any blog in the world. As with the emailed feed, a subscriber reading a feed will have a URL that will take them directly into the AfricaMap system to view the comment in the map context in which it was made. We envision that syndication could be an important means of dissemination and could bring new users into the site where they would be positioned to contribute to and continue a discussion by adding comments.

Links pointing to content on other servers are subject to breakage whenever the linked-to site changes the URL or disappears. An important feature of a system would be an administrative tool that could be set to check periodically all links in the system and generate a report for the administrator. After a defined number of failed attempts to access a given link, the system would notify the creator of the link through email.

In addition to latitude and longitude, dates are a powerful metadata type that cut across disciplines; by using maps to represent time, time-enabled materials can be presented visually. To attain these ends, users need to be able to precisely define time ranges (start date and end date) for any feature in the system. To support interoperability with other systems ISO 8601, the international time standard, could be used to store date ranges in the database and in data output from the system. ISO 8601 is compatible with Dublin Core used by OAI and with KML used by Google Earth.

A simple way to enable AfricaMap to support the visualization of time enabled materials would be to output a version of the features in KML format for exploration using the Google Earth time bar. In addition, the query interface could support searches using start date and end date, and a time graph view could display the temporal distribution of all materials in the system for which time ranges are defined.

In addition to the standard query result list, we propose three interactive graphic views be available to help the user characterize a collection or the results of a query: map view; time graph view, and word cloud view. The map view will display a heat map showing the distribution of materials across the geographic extent of the selected set of materials. Areas with greater concentrations will show up in a different heat map color. The time graph will display a graph of the materials distributed across the time frame they represent. The word cloud will use the approach of a standard tag cloud except that it will be possible to choose certain fields such as Language and Author to create field-based, focused clouds in addition to the more traditional unstructured tag clouds.

The relationship between the four pieces (results list, map, time graph, word cloud) should be dynamic. Zooming in on the map to show a subset of the results on the map would filter the rest of the views, and the results will change to reflect the new spatial filter. Changing the time range will filter the other three views. Selecting a word from the word cloud will cause the other views to adapt to the refined query. In this way a user will be able to explore large collections of georeferenced materials with a better a priori sense of the distribution of materials in the collection.

## 5.2 *Proposed new architecture of AfricaMap*

Figure 4 presents the existing architecture in black and the proposed in blue. Highlighted areas are of significant software development.

The proposed architecture builds from the current AfricaMap and adds some significant new features to support ongoing collaboration and time-enabled GIS dataset development between geographically distributed participants.

An important enhancement is the creation of a mechanism for leveraging the wealth of historic geographic information in the world's paper map collections by making it easier for

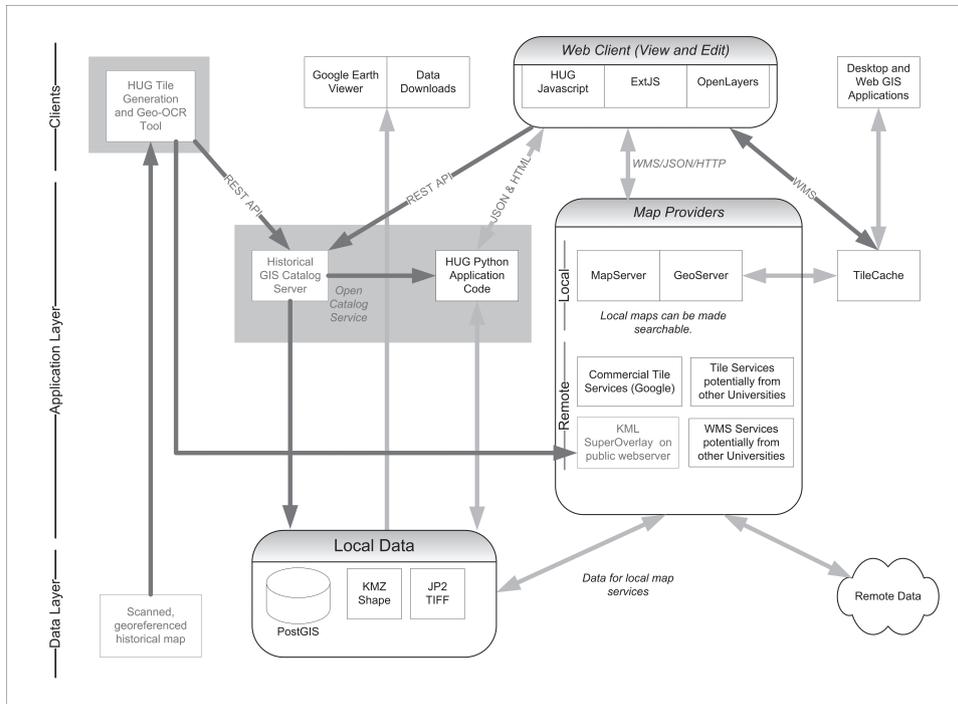


Figure 4. Proposed AfricaMap phase II architecture.

organizations to publish their maps in a useful way. This approach uses tile caches as in the current AfricaMap but with some twists. First, the list of available tile caches is put on the web in the form of a database (GIS Catalog Server) which can be updated by those who wish to contribute new caches to it, and which can be queried by users searching for maps. These caches do not need to reside on a central server, but can reside anywhere, including on the server of the organization that owns the original paper map. No special software is required in order for the organization to serve them efficiently. The caches just need to be stored in a web accessible space.

The next phase of the AfricaMap project proposes the development of an open source tool for these organizations to use to create tile caches from their georeferenced maps and register metadata for the caches along with the cache location URL on the central Catalog Server. In order to be able to process large numbers of tile caches the central server should farm out the tile generating jobs using a grid technology like BOINC (BOINC 2010). Once a set of tiles has been create for a given georeferenced map and metadata defined for it, the tile cache should be uploaded to the data owner's web space and registered on the central catalog server index. The Catalog Server will make its contents searchable via an open API so that others will be able to build a client that can query the catalog and retrieve tiles from wherever the tiles reside.

In addition to a framework for distributed sharing of historic map data, tools will be provided (AfricaMap Python Application Code in diagram) for collaboratively editing GIS layers based on the tile cache backdrops. Tools will allow users to add features to existing layers with defined schemas. A key layer will be the time-enabled gazetteer (place name database). This layer will evolve and improve as more and more historic maps are uncovered and more place name information is extracted from them.

An important feature of the editing tool will be provenance maintenance. For any new feature created, the reference information for the base map used to derive the new feature will be stored and associated with the new feature.

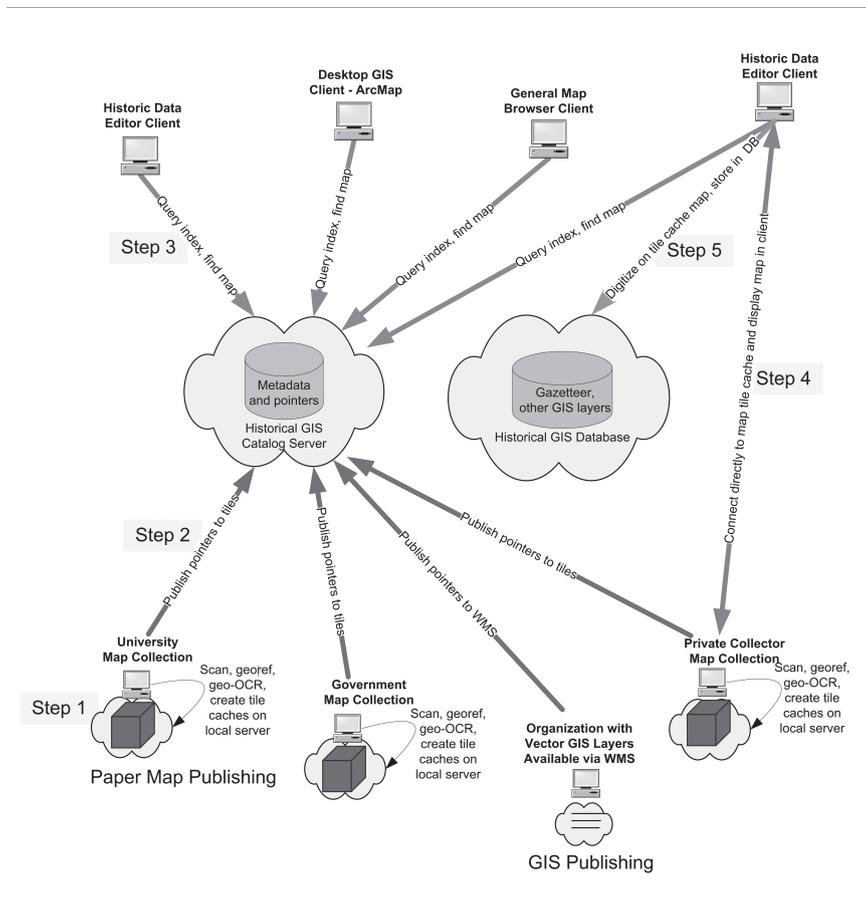


Figure 5. High level views of stages in collaboration.

Step 1 – Librarian at an organization with paper map collection scans map, (optionally) georeferences, (optionally) extracts place name features, creates tile cache, and puts cache on web accessible space. An online tool for georeferencing scanned maps will be provided.

Step 2 – Librarian notifies the Catalog Server of the existence of the tile cache by URL, fills in basic metadata regarding source, date, georeferencing technique used, etc.

Step 3 – Researcher looking for historic maps to support work in a given region queries the Catalog server using the system's web client and finds a list of matching datasets.

Step 4 – Researcher turns on desired tile cache and a request is made directly to the tile cache server.

Step 5 – The researcher uses the tile cache as a backdrop and digitizes new place name features to add to an historic layer.

### 5.3 Proposed collaboration workflows

There are several ways in which historical map images can be made available to the AfricaMap system for use in building new GIS datasets. As shown in Figure 5, the source data providers are on the bottom, and the map viewing and editing clients are on the top.

Figure 6 illustrates the proposed paper map publishing data flow in the next release of AfricaMap.

In addition to the manual extraction of place name features from historic maps, we also propose the development of an innovative means of gazetteer population: a specialized version of OCR (Optical Character Recognition) that is adapted to support the extraction of place names and locations (lat long) from scanned, georeferenced historic maps. We are calling this

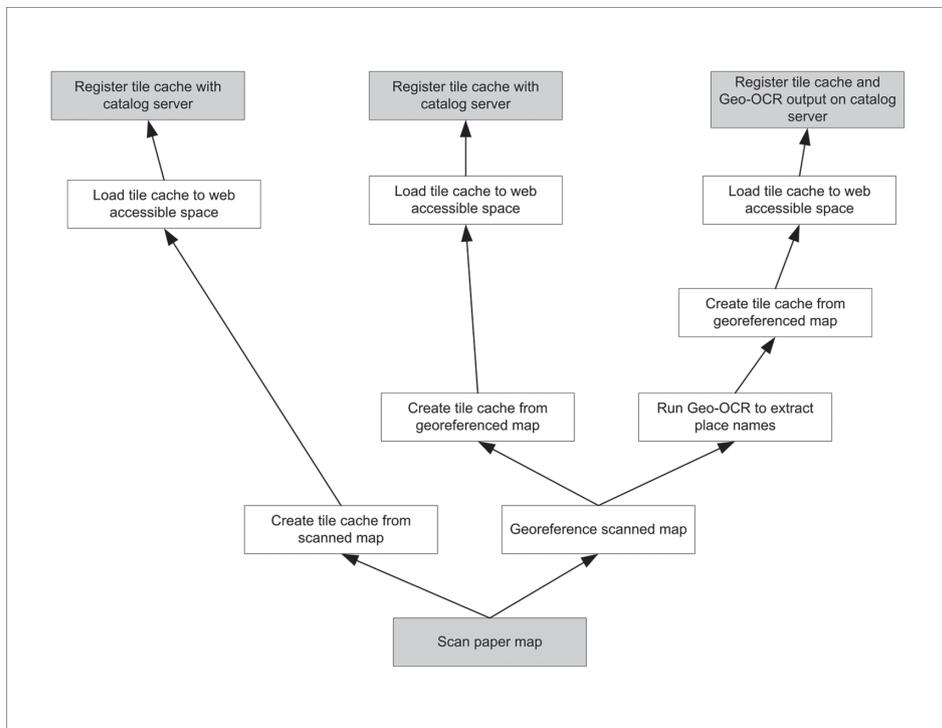


Figure 6. Paper map publishing data flows.

technology Geo-OCR or Map-OCR. Such an approach, in a collaborative environment such as this one, does not have to be perfect to be useful. Even if it starts with a relatively low accuracy rate, it will provide an automated “guess” as a jump start; iterative corrections as more reference maps are made available through the system will gradually improve the accuracy, thus making significant contributions to an ever-growing historical gazetteer.

Geo-OCR combined with a gazetteer could even provide the beginnings of an automated georeferencing system. The Holy Grail would be a completely automated georeferencing process for scanned maps uploaded to the system followed by a fairly complete place name and feature extraction. It is just a matter of time before this problem is solved. Our estimate is that it will take at least 5 years for a generic, Open Source solution to be developed that will work for most paper maps in most languages, that is, unless Google gets serious about it first.

## 6 CONCLUSIONS

Today a vast amount of human cultural knowledge is available online, providing more and more materials and sources for research in all disciplines. In most cases these materials must be accessed on an individual basis by going to the provider’s web site and searching the holdings using the particular site’s search interface. Not only are these searching methods often limited, but also there is no viable way of addressing these data comparatively within one system or across multiple systems. In short, unique challenges exist for accessing and effectively comparing and integrating these data. GIS (Geographic Information System) technologies offer a means of starting to address some of the issues of accessibility and comparability. Maps provide a common language for understanding the world.

The mapping revolution made possible by the web provides new means for integrating, analyzing, presenting, and sharing this information. We see web-based GIS as a promising

technology for expanding the range of search techniques available to users, making possible more precise searches of materials otherwise isolated by discipline category; this in turn and thus inspires and spurs individuals and institutions to offer their data to the web. GIS provides a uniform geospatial cataloguing format, and because it allows users to view multiple layers comparatively, it encourages a multi-disciplinary approach to understanding events and the related artifacts. GIS technology, used to view historical, environmental, socio-cultural, economic, and political spatial data layers in conjunction with multi-media data (images, video, audio, 3d models) will allow for the exploration and imagining of the earth's past and future in new perspectives.

AfricaMap begins to address also the problem of interdisciplinary geospatial data availability for Africa. The system provides some of the highest resolution mapping publicly available for the African continent, alongside an array of other data. AfricaMap is currently good at display, but it does not allow users to add comments or new materials. As potentially useful as AfricaMap is in its present state, to break new ground in improving access to materials and facilitate collaborative research, it needs to be enhanced in several ways.

Based on feedback from users, the improvements should include enabling users to upload, georeference, time-enable, tag, and control access to their materials. Also important will be a method by which users can comment on materials and email or syndicate their comments. To take advantage and motivate the creation of enhanced metadata, we propose creating a new kind of faceted search and visualization environment that supports visualization of search results by map, time graph, and tag cloud. The authors believe that the lessons learned from AfricaMap, once implemented as enhancements to the system, could help jump-start the next level of online geospatial collaboration.

## ACKNOWLEDGEMENTS

The authors thank Professors Suzanne Blier and Peter Bol, co-principle investigators of the AfricaMap project, for their vision and guidance in conceptualizing, developing and promoting AfricaMap. The authors also thank Dr. Frank Chang for his review comments on this manuscript. The AfricaMap project team at CGA includes the authors and David Siegel, Bo Zhao, Julia Finkelstein, Molly Groome, and Giovanni Zambotti.

The AfricaMap project was funded by the following funds and organizations of Harvard University:

- The Provost's Fund for Innovative Technology
- The William F. Milton Fund
- The Center for Geographic Analysis
- The W.E.B Du Bois Institute
- The Department of African and African-American Studies
- The Committee on African Studies

Collaborating Organizations in data, infrastructure and system development include the following:

- Organizations within Harvard University
  - The Center for Geographic Analysis
  - Harvard Map Collection
  - Harvard Geospatial Library
  - Harvard-MIT Data Center
  - The Dataverse Network Project
  - The Institute for Quantitative Social Science
- External contributors
  - East View Cartographic
  - Avencia
  - MetaCarta

For a complete list of the AfricaMap project steering committee members, please see the “About” tab on <http://africamap.harvard.edu>

## REFERENCES

- Blood, Rebecca. (2000) Weblogs: A History and Perspective. In Rebecca's Pocket. 07th September 2000. 25th October 2006. [Online] Available from: [http://www.rebeccablood.net/essays/weblog\\_history.html](http://www.rebeccablood.net/essays/weblog_history.html)
- BOINC. *Open Source software for volunteer computing and grid computing*, [Online] Available from: <http://boinc.berkeley.edu/>
- Giles, J. (2005) Special Report Internet Encyclopedias Go Head to Head. *Nature* 438, 900–901.
- Global Roads Workshop. (2008) *A Strategy for Developing an Improved Global Roads Data Set. 2.* Columbia University. [Online] Available from: [http://www.ciesin.columbia.edu/confluence/download/attachments/3407907/GlobalRoads\\_StrategyPaper\\_20feb08.pdf](http://www.ciesin.columbia.edu/confluence/download/attachments/3407907/GlobalRoads_StrategyPaper_20feb08.pdf)
- Goodchild, M. (2007) Citizens as Sensors: The World of Volunteered Geography: 1, 2. Santa Barbara: Position Paper for the Workshop on Volunteered Geographic Information. [Online] Available from: [http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Goodchild\\_VGI2007.pdf](http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Goodchild_VGI2007.pdf)
- Harley, J.B. & Woodward, D. (1987) *The History of Cartography: Cartography in prehistoric, ancient, and medieval Europe and the Mediterranean*: 112. Chicago: Humana Press.
- Hill, L.L. (2006) *Georeferencing: The Geographic Associations of Information*: 5. Cambridge, Massachusetts: The MIT Press.
- Hobbes, Z. (2010) *Hobbes Internet Timeline 10*. [Online] Available from: <http://www.zakon.org/robert/internet/timeline/>
- King, G. (2007). An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. 195. *Sociological Methods and Research*, Volume 36, No. 2. [Online] Available from: <http://gking.harvard.edu/files/dvn.pdf>
- Mannak, M. (2008) Development: Africa Still Hampered by Lack of Geographical Data. Inter-Press Service News Agency [Online] Available from: <http://ipsnews.net/news.asp?idnews=43685>
- Miller, C. (2006) A Beast in the Field: The Google Maps Mashup as GIS/2: 188. *Cartographica* volume 41, issue 3.
- Nielsen, J. (2008) Nielsen's Law of Internet Bandwidth. useit.com. [Online] Available from: <http://www.useit.com/alertbox/980405.html>
- Ogodo, O. (2009) Africa Facing Climate Data Shortage. 11th November 2009. *Science and Development Network*, [Online] Available from: <http://www.scidev.net/en/news/africa-facing-climate-data-shortage.html>
- Orwant, J. (2009) Comment in Language Log Blog. [Online] Available from: <http://languageblog ldc.upenn.edu/nll/?p=1701#comment-41758>
- Petras, V. (2004) Statistical Analysis of Geographic and Language clues in the MARC record. Technical Report for Going Places in the CatalogL Improved Geographical Access Project. School for Information Management and Systems, University of California Berkeley. [Online] Available from: <http://metadata.sims.berkeley.edu/papers/Marcplaces.pdf>
- Scharl, A. & Tochtermann, K. (eds.) (2007) *The Geospatial Web*. London: Springer-Verlag.
- Schutsberg, A. (2007) Google KML Search: What Does it Mean for Geospatial Professionals? [Online] Available from: [http://www.directionsmag.com/article.php?article\\_id=2409](http://www.directionsmag.com/article.php?article_id=2409)
- Tobler, W.R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 244–40.
- Wegener, M. & Fotheringham, S. (2000) *Spatial Models and GIS: New and Potential Models*, London: CRC.
- Wikipedia, (2010) *Moore's Law* [Online] Available from: [http://en.wikipedia.org/wiki/Moore%27\\_s\\_law](http://en.wikipedia.org/wiki/Moore%27_s_law)
- Wikipedia, (2010) *Ramsden Theodolite* [Online] Available from: [http://en.wikipedia.org/wiki/Ramsden\\_theodolite](http://en.wikipedia.org/wiki/Ramsden_theodolite)
- Wilson, T. & Burggraf, D. (ed.), (2009) *KML Standard Development Best Practices*: vii, Open Geospatial Consortium.
- Zuckerman, E. (2008). *my heart's in accra*: 1/18/08 Blog Posting. [Online] Available from: <http://www.ethanzuckerman.com/blog/2008/01/18/kenya-mapping-the-dark-and-the-light/>