AN OPEN SOURCE WEB APPLICATION FOR HISTORIC AIR PHOTO DISPLAY AND DISTRIBUTION IN WISCONSIN

H. Veregin [a], P. C. Gorman [b], J. Stoltenberg [c], AJ Wortley [a], M. Bricknell [a]

[a] Wisconsin State Cartographer's Office, Dept. of Geography, University of Wisconsin-Madison, USA - (veregin, lwortley, bricknell)@wisc.edu
[b] Digital Collections Center, University of Wisconsin-Madison, USA - pgorman@library.wisc.edu
[c] Robinson Map Library, Dept. of Geography, University of Wisconsin-Madison, USA - jstoltenberg@wisc.edu

KEY WORDS: History, Internet/Web, Imagery, Digitisation, Visualization, Interoperability, Retrieval

ABSTRACT:

This paper describes an effort to build a digital archive and web-based retrieval system for historic air photos for the state of Wisconsin. The University of Wisconsin-Madison has an extensive collection of historic air photos, including a rare 1937-1941 collection that covers almost the entire state. This collection is in constant demand, being the oldest systematic aerial survey of the state, and serves as a critical baseline dataset for understanding changes to Wisconsin's landscape over the last 70 years. However, since the physical media are fragile and demand is so high, access to the photos can be difficult. To solve these problems we embarked on a three-year project funded through the Baldwin Endowment at UW-Madison. Teams in the Department of Geography and Library Digital Collections collaborated to scan, catalog, and develop web-based access for the collection. In the course of the project we scanned and processed over 38,000 photos, and collected metadata for each image to provide search and cataloging capabilities. We developed automated pattern recognition techniques to extract digital flight lines and photo centers from paper index maps for map display and search functions. We then developed a web interface based on an open source framework for display and distribution of the imagery. The interface includes a commercial tiled basemap on which we superimposed additional map reference layers and photo center points and footprints. Users can interact with the system to select, preview and download digital images from any desktop connected to the web.

1. INTRODUCTION

This paper describes the development of a digital archive of historic air photos for the state of Wisconsin. The project involves digitizing, indexing, cataloging, and creating a web-based interface for search and retrieval of an air photo collection originally acquired by the United States Department of Agriculture (USDA) in the 1930s covering nearly all of Wisconsin. Project collaborators include the Robinson Map Library and the Wisconsin State Cartographer's Office – both units within the Department of Geography at the University of Wisconsin-Madison – and the University of Wisconsin Digital Collections Center (UWDCC). The air photo collection digitized in this project is owned by the Robinson Map Library and the Wisconsin Department of Transportation, and is the oldest and rarest in the Robinson Map Library's holdings. The collection is heavily used by a wide range of library patrons because it is the earliest systematic aerial survey of the state and has become a baseline for understanding important physical, cultural, social, and economic changes in Wisconsin over the past seventy years.

The digital archive initiative addresses two key limitations of the physical collection. First, the physical media are fragile and thus do not circulate from the library. Second, since access to these photos is difficult, demand to use them significantly exceeds their availability. To overcome these limitations we focused our efforts on building a system that allows users to easily discover, preview, and download imagery from any computer connected to the Web. We developed an extensible system and flexible workflows to enable the incorporation of additional years of photography and to facilitate interoperability with the UWDCC repository to allow access to other digital collections.

In the course of the project we scanned and processed over 38,000 photos, and collected metadata for each image to provide search and cataloging capabilities. We developed automated pattern recognition techniques to extract digital flight lines and approximate photo centers from paper index maps for map display and search functions. We developed a web interface based on the open source GeoMoose/OpenLayers application framework for display and distribution of the imagery. This system includes a commercial tiled basemap on which we superimposed additional map layers and photo center points and footprints. Users can interact with the system to select, preview and download digital images.

Funding for the project is through the Ira and Ineva Reilly Baldwin Wisconsin Idea Endowment at the University of Wisconsin-Madison. The purpose of the endowment is to engage university faculty, staff and students with off-campus audiences in the tradition of the "Wisconsin Idea" – the notion that resources and expertise at the University of

Wisconsin can benefit the citizens of the state by generating tangible and practical solutions to clearly defined problems (http://www.provost.wisc.edu/baldwin/call.html). As such the audience for the digital repository is broad, and includes not only land records and GIS professionals, but also the general public and other researchers who currently make use of the physical collection.

In addition, the digital archive will serve as a foundational historic research tool for scholars in many disciplines, including archaeology, history, geography, geology, historic preservation, landscape architecture, environmental science, and education. Digital geospatial technology is increasingly becoming incorporated into these and other fields, in large part due to growing awareness of this technology and its utility for visualizing and analyzing spatial patterns (Goodchild, 2004; Goodchild and Janelle, 2010). In particular, the availability of web-mapping APIs (Application Programming Interfaces) has allowed many new users to develop mashups combining spatially-referenced data from a variety of different sources. Growing awareness of the power of place and the ability to visualize spatial patterns has led to important new digital research initiatives and suggests to some a new framework for interdisciplinary collaboration (Bodenhamer et al., 2010). This project is situated within this new multi-disciplinary research area. The ultimate goal of our project is to help open doors for expanded research and public participation in understanding cultural and social phenomena in Wisconsin over time.

2. DESIGN REQUIREMENTS

As noted above, the primary rationale for our work is to provide broader access to historic aerial imagery, and to preserve a widely used and quickly deteriorating collection. The use cases and user requirements for our interface evolved from our experiences interacting with and assisting users with the hard copy collection. We built our interface with the needs of this group in mind, and based on an understanding of their typical tasks and needs.

Our interface design is also informed by an evaluation of selected state-based historic air photo viewers. A summary of the features and characteristics of these viewers is provided in Table 1. Some viewers are predominantly GIS-based and thus require a steep learning curve for public consumers. Others are tailored more to traditional library searches and thus lack the advantages of spatial selection and visualization. Based on an analysis of these interfaces, we chose to adopt an approach that balances the use of an innovative map-based navigation, selection and preview interface tailored for non-professional users with the use of library-based repository storage, delivery and additional viewing options. To achieve this balance, an open source software approach was chosen as this offered the most flexibility in terms of customization and design.

2.1 User Community Considerations

Our user community is broad, and our application is designed to meet the diverse needs of this community. For novice users, such as members of the general public, simple navigation that mirrors other online mapping applications is key. For more advanced users, enhanced functionality is important, including multiple search avenues, the ability to change basemaps and turn layers on and off to customize the interface, and the ability to interact with the photos to compare change.

2.2 Photo Identification

One goal of our application is to emulate the experience users have when manually searching for air photos on paper index maps, and to improve upon that experience wherever possible.

A primary requirement is to design the interface to emulate the look of paper index maps, by superimposing photo center locations on top of a basemap. This provides a natural entryway for users familiar with the use of paper index maps. In addition, we wanted to improve the ability to identify photos and reduce user effort associated with this task. Paper index maps do not provide photo footprints, such that users cannot easily identify the set of overlapping photos associated with their location of interest. Some trial and error is required. A requirement for the interface is for the user to be able to click on any point on the basemap and have the system return information about all photos that contain that point. In addition, the system needs to respond with feedback showing the user the footprints of these photos.

		Alabama	Colorado	Connecticut	Florida	Illinois	Indiana	Minnesota	New York	Pennsylvania	Wisconsin
Technology	ArcGIS					●	●				
	OpenLayers										●
	Adobe Flash	●									
	Google API			●	●			●			●
	Bing API									●	
	Other	●	●						●		
Navigation	Zoom to county						●				●
	Zoom to PLSS										●
	Geocode address						●	●		●	●
	Photo extents				●						●
	Photo center points					●		●	●	●	●
	Flight lines		●								
	Map full screen						●				●
	Overlay/transparency			●							●
	Photo mosaic			●			●				
Access	Photo downloads				●	●		●		●	●
	PDF printable			●							
	Preview viewer				●				●		●
	Free imagery		●	●	●	●	●	●	●	●	●

URLs	*Alabama* http://alabamamaps.ua.edu/aerials/index.html *Colorado* http://ucblibraries.colorado.edu/aerialphotos/home.asp *Connecticut* http://www.econmap.com *Florida* http://ufdc.uflib.ufl.edu/ufdc/?c=flap *Illinois* http://www.isgs.illinois.edu/nsdihome/webdocs/ilhap *Indiana* http://129.79.145.7/arcims/IHAPI/index.html *Minnesota* http://map.lib.umn.edu/mhapo *New York* http://test.aerial-ny.library.cornell.edu *Pennsylvania* http://www.pennpilot.psu.edu *Wisconsin* http://maps.sco.wisc.edu/APfinder/index.html (Sites accessed 10/9/2010.)

Table 1. Summary of Selected Historic Air Photo Viewers

2.3 Requirements for the Basemap

Another limitation of paper basemaps is their lack of detail, clarity, and currency. Often these maps are quite fuzzy, contain confusing text overstrikes, and are several years older than the date of the photos. Lack of currency means users may not have current roads, reservoirs, or city shapes to make visual comparisons with current information. Hence another design requirement for our system is that the basemap must be current, and must be of high cartographic quality to allow for better and easier search. In short, high-quality cartography is a priority in the design of the interface. Use of one of several available commercial basemap options (i.e., Google maps) allows for a lower maintenance burden in providing this requirement. Further analysis of this option indicated the need to integrate two additional map reference layers – county and PLSS (Public Land Survey System) boundaries and labels -- via OGC (Open Geospatial Consortium) WMS (Web Map Service) protocols to provide a seamless integrated basemap specific to our users' needs.

2.4 Search Functionality

Another requirement for our system relates to search functionality. For paper index maps, search is limited due to the non-interactive nature of the maps. Users looking for an unincorporated place, a small cemetery, or other local feature, may need to do research independently in advance. For our system, we included search capabilities using digital gazetteers and address matching (geocoding) functionality.

A large percentage of users looking for air photos are familiar with PLSS and often have a legal description of the property they are searching for. Paper photo index maps do not include such information, so in the library users must use plat books to determine the location of the feature based on the PLSS description. Giving the users the ability to zoom to a particular PLSS feature enhances their experience greatly. In addition to PLSS, the system allows for searches on county name, by address, and by feature name.

2.5 File Delivery

A further requirement of our system is the ability to download high-quality image scans. We wanted the system to emulate and enhance the library experience, in which users can walk out with a scanned photo.

An important aspect of the project relates to the fact that the interface offers seamless integration of services by administratively separate departments. Photos are processed and delivered to users by means of a repository API which allows one department (UWDCC) to host and curate image data, while a separate department (Geography) provides the user interface. Separation of storage and management functions from indexing and access applications ensures that all content can be preserved for future generations even as access technologies and user interfaces change over time.

3. DATA COLLECTION

3.1 Selection and Metadata

Our efforts focus on a collection of aerial photographs taken by the United States Department of Agriculture (USDA) in the 1930s. This collection of photographs is in constant demand by a wide range of users because it is the earliest systematic aerial survey of the state and has become a baseline for understanding important physical, cultural, social, and economic changes in Wisconsin over the past seventy years. The collection is rare, with physical copies residing in the Robinson Map Library (approximately 75% complete) and the Survey and Mapping Section of the Wisconsin Department of Transportation (WisDOT) (almost 100% complete).

The entire digitized collection of 1937-1941 photographs totals 38,906 images, from which are missing only 85 exposures. In addition, because these photographs were flown in stereo and have significant side-lap and end-lap, most of the missing 85 exposures do not result in actual gaps in geographic coverage. This digital archive represents the most complete collection of aerial images from the first USDA statewide flight in Wisconsin.

To assist in cataloging and search functions, we collected metadata for each scanned photograph. We devised a modified Dublin Core metadata schema (http://dublincore.org) consisting of key descriptive information (Table 2). Students working directly with the photograph collection at the Robinson Map Library created metadata records with the photographs at hand and entered these into a web interface coupled to an underlying database. A unique database identifier for each record was also transposed to each photograph at this time.

Field	Example
DC_Title Name of county, year of photo.	Dane County, 1937
DC_Creator Agency responsible for acquisition.	United States Department of Agriculture
DC_Contributor Company contracted for acquisition	Chicago Aerial Survey Co., Chicago IL
DC_Subject Subject terms from list.	Water
DC_Description Scale information.	1:20000 or 1" = 1667'
DC_Date Photo date.	1937-06-22
DC_Format Photo format, dimensions.	Black and white photograph, 7x9
DC_Identifier Reel, exposure, year.	1-3, 1937
DC_Identifier_LocalID Unique database ID.	Aerial00001
DC_Relation_IsPartOf Name of the collection.	The Changing Landscapes of Wisconsin: A digital archive of historic aerial photographs
DC_Coverage Location and/or relative temporal information.	Spatial: Dane County (Wisconsin)
DC_Rights Rights for the original object.	As a work of the United States government, this material is in the public domain.
DC_Rights_Ownership Copyright owner of original object versus image.	Original material owned by Arthur H. Robinson Map Library.
Note Note terms from list.	Few markings.
Submitted Submitting institution.	Arthur H. Robinson Map Library.
QCFlag Quality control level.	Final

Table 2. Modified Dublin Core Metadata

3.2 Scanning

The University of Wisconsin Digital Collections Center (UWDCC) took the lead on the digital imaging aspect of the project. UWDCC surveyed various user communities to identify preferred scanning specifications. To meet the needs of scholars intending to use the photography in geospatial analysis, a resolution of 600 dots per inch with 8 bit depth was ultimately adopted.

Each aerial photograph was scanned on flat-bed scanner. After scanning, images were cropped, color corrected, and saved in both TIFF and JPEG2000 formats. The JPEG2000 images were processed based on parameters developed for the Library of Congress's National Digital Newspaper Program (Buckley and Roger, 2006). JPEG2000 images offer dynamic scaling in the user interface as well as zoom-and-pan high-resolution viewing. Users are able to download either image format.

All digital content and metadata has been ingested into the UWDCC's repository, built upon the Fedora Commons platform. This provides flexible, application-independent access to digital objects through a robust API, while also providing a stable archival platform for long-term management. By separating storage and management functions from indexing and access applications, Fedora Commons will ensure that all content can be preserved for future generations even as access technologies and user interfaces change over time.

To ensure long-term preservation of the material, the original TIFF images have been archived and stored at separate locations. Content in the Fedora repository is periodically verified against archived data to ensure consistency, and files with inconsistencies are backed up from tape. The original physical copies of the photos have been returned to the Robinson Map Library where they will remain publically accessible onsite for users who would like to take advantage of the stereo pairs, or simply prefer to handle the original material versus the digital images.

3.3 Indexing

Index maps of air photos provide the basis for navigation and search functions. For the 1930s USDA photos these index maps are from two to three feet in size. A common practice is for a library patron to use an index map to identify the photograph center point symbols, marked with a roll and exposure number, closest to the feature or area of interest. This tabular index is then used to locate the photos in a filing cabinet.

Successful deployment of web-based analogs to this search capability required extraction and digitization of geographic information from the hard-copy index maps. With the help of custom-written pattern recognition software and standard GIS tools, we extracted approximate photograph center point coordinates (Figure 1) and created an associated database populated with latitude, longitude, roll and exposure numbers, county identifier, PLSS coverage, and photograph date.

The process involves automated photo center capture followed by manually tagging with flight line and exposure number attributes. On a county-by-county basis, index points were reconciled against metadata, and missing or duplicate index points were corrected.

Each county-based index map produced a spatial index of between two hundred and one thousand points, which were then aggregated into a statewide database for all seventy-two counties in Wisconsin (Figure 2). This statewide database was then integrated into the web-based interface to facilitate search and discovery of photos.

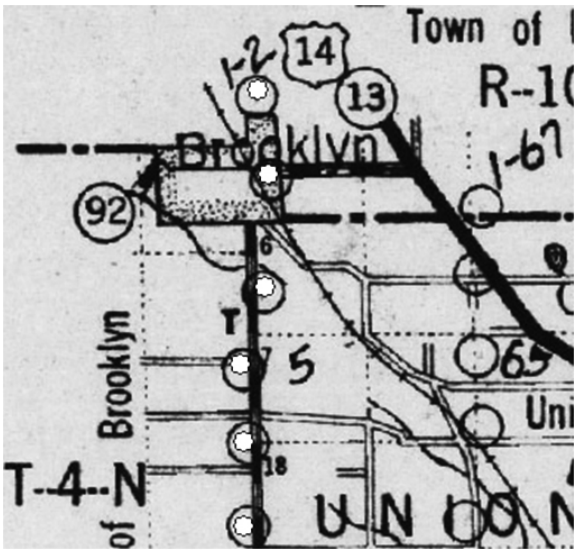

Figure 1. Extraction of Center Points from Index Map

Figure 2. Extracted Photo Center Points for Entire State

4. WEB INTERFACE

4.1 Interface Design

To allow patrons to explore and access the collection we developed a web-map interface based on an open source application framework. The interface incorporates a Google base map with additional custom visual reference layers, including county boundaries and names, and PLSS features. Users can search and navigate by address, geographic coordinate, place name, or PLSS description. Additionally, users can navigate manually using standard web-map control tools. The intelligent spatial index described above is also superimposed over the basemap, allowing users to visually navigate to an area of interest and identify all relevant historic photographs. Once identified, photographs can be previewed with varying transparency on a reference map or aerial imagery basemap to validate their location. Users can also examine the photographs in a high-resolution viewer prior to deciding to download images for later use.

4.2 Technical Implementation

In approaching the application architecture, we paid particular attention to creating a component-based application structure by maintaining, where possible, a separation between the photo repository, the spatial database of aerial index and reference map information, and the application interface logic. This separation provides for use of standard communication protocols between databases and the interface and allows for extensible introduction of additional components or replacement of an existing piece of the architecture. Where possible, standard protocols and APIs were used to communicate with the interface. Examples include the Google Maps API V.2, Google Geocoding API, OGC Web Map Services, and a custom API from the Fedora Commons repository at UWDCC (Figure 3).

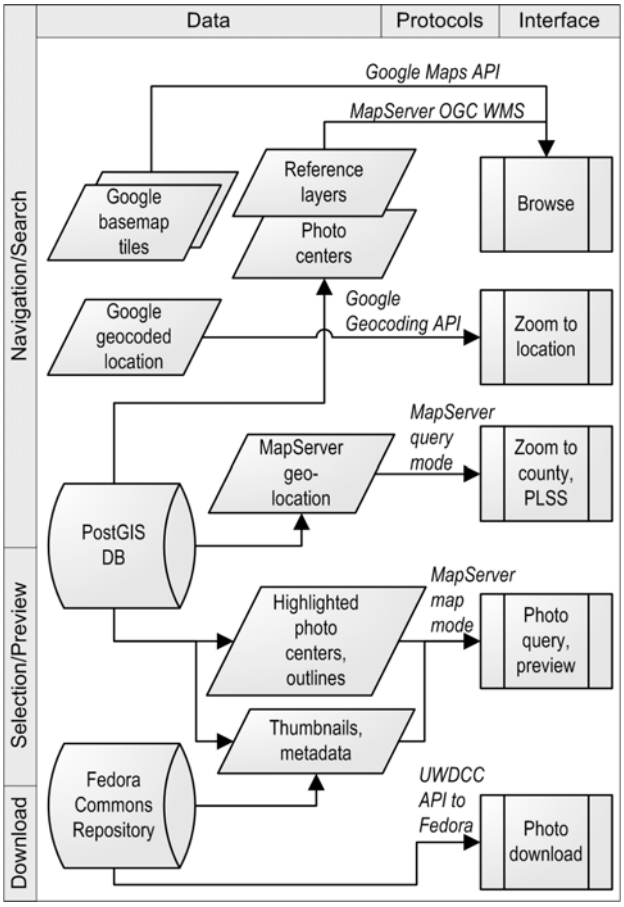

Figure 3. Simplified View of System Architecture

We utilized the open source web-mapping framework called GeoMoose (an extension upon the OpenLayers application framework) to develop the application interface. This framework provided a desirable base level of functionality along with clear support for scripted customizable functionality. The framework is written in PHP and Javascript and provides native support for Google Maps and Geocoding API services (among others), WMS integration, and documentation for adding custom "services" to the interface.

For storage of aerial photo center points and outlines as well as ancillary GIS reference layers, we chose a PostGIS/Postgresql spatial database implementation along with Mapserver software for WMS and spatial query support. UWDCC developed a custom API to provide seamless access to the Fedora Commons-based photo repository functions including thumbnails, previews, and download capabilities.

The following is a summary of system components:

- *Fedora Commons* is an open source enterprise library repository software, used for long-term storage, preservation, maintenance, and online access for digital air photo scans.
- *Postgresql/PostGIS* is an open source geospatial database used for storage of air photo center points and extents as well as spatial map reference data integrated with the basemap. It provides query capability for navigation and photo selection and functions, and serves as the back-end data source to Mapserver for map display.
- *MapServer* is an open source mapping engine providing WMS views of map reference layers, air photo center points, and extents and custom views of highlighted photo selections.
- *OpenLayers* is an open source Javascript web mapping framework commonly used for integrating commercial and WMS map views as well as providing various map query capabilities. It provides a standard interface to multiple commercial base map services.
- *GeoMoose* is a PHP web mapping framework built in Version 2 as extension upon the OpenLayers framework. It provides additional map application functionality and direct integration with the MapServer engine.
- *Google Maps API* is an application programming interface to Google for requesting Google basemap tiles.
- *Google Geocoding API* is an application programming interface to Google servers for resolving a location to coordinates for navigation to an area of interest.
- *UWDCC Fedora API* is a custom application programming interface authored by UWDCC for efficient access to scanned photo metadata, thumbnails, and images for download at multiple resolutions.
- *OGC WMS* is a web mapping service standard from the Open Geospatial Consortium for standard authoring of map layer views that may be integrated in a web mapping framework or other GIS client.

4.3 Usability Testing

In developing the interface, particular attention was paid to aspects of usability in an effort to provide access to users from diverse disciplines and with different skills. Our work on interface design is informed by standards and best practices associated with human-computer interaction and user-centered design. Human-computer interaction (HCI) is concerned with the design, implementation and use of interactive computer-based systems and with understanding the human-computer interface (http://www.sigchi.org). User-centered design (UCD) is a design philosophy aligned to the user-centered focus of HCI. UCD puts the user at the center of the design process by incorporating user requirements into the process from the beginning in order to produce effective systems that conform to user needs (Hacklay, 2010). The benefits of designing systems in this way include higher productivity, greater user satisfaction, increased demand and use, and improved return on investment.

Standards related to HCI and UCD for interactive web interfaces have been developed by the International Organization for Standardization (ISO) and published as ISO 9241-151 ("Guidance on World Wide Web user interfaces") and ISO 9241-210 ("Human-centred design for interactive systems"). These publications provide a framework for improving web interface usability, including standards for interface design, content, and navigation and search functions, and methodologies for planning, implementing, and evaluating UCD projects. In general UCD requires early and continued attention to user needs, empirical measurement and testing to evaluate usability, and iterative refinement resulting from the testing process.

We plan to evaluate the interface through user-based testing methods to obtain structured, quantitative data about the effectiveness of the interface. Our plan includes outreach to institutions throughout the state to incorporate a varied pool of users and work environments. Our targeted user community is diverse, covering the spectrum from experienced geospatial professionals to casual users with little knowledge of geospatial data. This diversity necessitates segmentation of testing groups and strategies to accurately ascertain usability concerns.

The objective of our tests is to provide empirical evidence about how the application is being used, how well requirements are being met, and what problems need to be addressed. At this time we are less concerned with such quantitative aspects as speed and efficiency, and more concerned with basic usability. We are looking for affirmation that basic tasks can be carried out and that nothing is missing from the interface.

We plan to conduct testing within real work environments, including libraries, in order to ensure validity of test scenarios. Our approach is informal, yet structured. Users will be asked to perform certain tasks and information will be collected about performance. Both direct observation/interviews and post-performance questionnaires will be used to maximize the number of subjects given time and budget constraints. Tasks will be randomized to reduce learning-bias. Groups will be stratified based on experience levels.

5. CONCLUSIONS

Future work on this project includes seeking continued funding to expand the photo collections. The Robinson Map Library has an additional 77,000 air photos from the 1950s, 1960s and 1970s that could be included. In addition we hope to build relationships with counties or agencies that have additional print collections that could expand the archive. Incorporation of additional years of photography provides the ability to view a time-series of photos for a particular place, which is potentially a rich research tool. We are also interested in geo-enabling other, non-spatial digital collections and incorporating them into the archive. The ability to visualize the geographic landscape, cross-reference disparate thematic content for the same geographic area, and analyze changing spatial patterns over time has the power to greatly facilitate understanding of cultural phenomena and their interrelationships.

REFERENCES

Bodenhamer, D.J., J. Corrigan and R. M. Harris (Eds.), 2010. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington: Indiana University Press.

Buckley, R. and S. Roger, 2006. JPEG 2000 Profile for the National Digital Newspaper Program. Washington, DC: Library of Congress Office of Strategic Initiatives. http://www.loc.gov/ndnp/pdf/NDNP_JP2HistNewsProfile.pdf

Goodchild, M. F., 2004. The Alexandria Digital Library, Review, Assessment, and Prospects. *DLIB Magazine (10)5*. http://www.dlib.org/dlib/may04/goodchild/05goodchild.html

Goodchild, M. F. and D. G. Janelle, 2010. Toward Critical Spatial Thinking in the Social Sciences and Humanities. *GeoJournal* (75), pp. 3-12.

Haklay, M. (Ed.), 2010. *Interacting with Geospatial Technologies*. Chichester: Wiley-Blackwell.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the Ira and Ineva Reilly Baldwin Wisconsin Idea Endowment at the University of Wisconsin-Madison for its funding of this project. The project team also wishes to acknowledge Dr. Mark Harrower and Jeremy White for their contributions.