The Travelers in the Middle East Archive (TIMEA) Project: Researching and Exploring the Middle East Through Online Maps

German Diaz and Eva Garza
Fondren Library MS 225
Rice University
P.O. Box 1892
Houston, TX USA 77251-1892
gdiaz@rice.edu
evagarza@rice.edu

Abstract. Performing scholarly and historical research has always been a tedious and scattered process. However, with the advent of on-line interactive mapping, there is potential to revamp the way we access information for large geographic areas. Through a project called Travelers in the Middle East Archive, the GIS/Data Center at Fondren Library has been developing various on-line maps and GIS data sets of the Middle East region. These maps would allow for research in a geographic context using maps as interface to documents, books, or commentaries. In this paper we will discuss the challenges and objectives of the GIS module of the project, the data sets to be developed, and the opportunities for the research community and the public to explore the Middle East at a whole new level.

Keywords: ArcIMS maps, Middle East, Digital Archive, Egypt, Cyprus

Introduction

The Travelers in the Middle East Archive is a digital archive of narratives documenting travel to the Middle East published between the 18^{th} and early 20^{th} centuries. During this time period many Europeans traveled to the Middle East and published a wealth of literature. Such literature represents invaluable resources for scholars and researchers in numerous disciplines like religion, history, women's studies, and archaeology. TIMEA attempts to facilitate access to these materials as a one-stop research portal (www.timea.rice.edu) for electronic texts, books, postcards, images, and maps of the region. This project aims to have a greater impact on education by serving as a teaching tool for middle, high school, and college students. The project is a close collaboration between scholars, library staff, and technical developers. The archive is divided into four basic areas: electronic texts, images, research modules, and web maps. The technical work behind the archive involves selection of materials, scanning, metadata creation, production of electronic texts, and development of ArcIMS maps. This paper will discuss the GIS data sets used to build the ArcIMS maps, the technical processes developed, and the challenges faced while developing the current GIS interface. We will also touch on the issue of public access to the materials and opportunities for partnerships with Middle East institutions.

GIS Data Sets

TIMEA is an innovative project implementing the most current mapping technologies. The GIS/Data Center at Fondren Library is responsible for developing the maps that give the archive its geospatial component. Having a GIS interface not only makes the collection visually attractive, but also opens doors for a more comprehensive study of the Middle East. The maps are intended to provide access to searchable pieces of the document library through map links. Maps will give users a complete view of a region, provide search/download tools for GIS data sets, and act as access points for electronic texts within the archive. Also, one of the funding requirements of the project is to make GIS data available for public use. Map searching and data download capabilities will facilitate this data exchange requirement, and make the archive even more beneficial to the GIS community. The maps will be linked to books and materials within the archive's database, using ancient and modern place names as the link between the map and the database. Basically, researchers will be able to open a map, search for a place and display a detailed map of that location along with a list of related documents, points of interest, books, postcards, and pictures. In the same way, researchers will be able to link to maps from the materials' content.

The following present day nations will be part of TIMEA: Azerbaijan, Bahrain, Cyprus, Egypt, Georgia, Iran, Iraq, Israel, Jordan, Lebanon, Oman, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates, and Yemen. For display purposes and to provide a sense of continuity and completeness to the region the project will also include Armenia, Eritrea, Ethiopia, Kuwait, Libya, Sudan, the Gaza Strip, and the West Bank. Please see Figure 1 for a detailed map of the project area.

Figure 1. Map of Project Area

Currently the pilot project centers on Cyprus. However, data sets are also being created for Egypt, and later the rest of the countries will be completed and added. Cyprus was selected as the pilot country because of its manageable territorial size, and because one of the scholars involved with the project, Dr. Michael Decker, has expertise in this area. This has proven to be an effective approach since more trial-and-error data processing can be done in a smaller project. It also helps in assessing timing and resources for the entire project.

The present Cyprus base map located at http://aries.rice.edu/website/cyprus/ is made up of elevation, satellite imagery, roads, Roman roads, olive oil mills, place names, railroads, administrative subdivisions, and drainage system layers. The first step in building the initial base map was to search and select available data sources. The Digital Chart the World (DCW) dataset [1] was the most logical place to start since the data was already available at the center and to the public. The DCW files provided country and state boundaries, canals, drainage lines, inland water bodies, roads, and railroads. However, the data set only provided an initial overview of the areas, and higher resolution data sets were going to be needed for features like boundaries and drainage systems. The major problem was the difference in resolution and scale between the vector and raster files. When overlaid, boundaries and roads barely matched against satellite and elevation files. The team had to find ways to get more detailed layers without having to purchase expensive data products. Table 1 lists the data sets used in the pilot project, along with software packages applied in processing. More information on each layer will soon be displayed as a metadata link on the web maps.

Table 1. List of Data Sets for Pilot

Data Set	Source	Processing Software
Political boundaries/outlines, railroads, roads, inland/coastal water bodies	Tobin Global Planner -Digital Chart of the World -Russian Topo Maps 1:500000 scale http://www.tobin.com/psGlobalPlanner.asp	ArcMap editing tools, WinTOPO, and AutoCAD.
Topography/elevation	GTOPO30 http://topex.ucsd.edu/cgi-bin/get_data.cgi Shuttle Radar Topographic Mission (SRTM) http://edcsns17.cr.usgs.gov/srtmbil/	GeoStatistical Analyst, ERDAS Imagine, Photoshop
Place Names	GEOnet Names Server http://earth-info.nga.mil/gns/html/	MS Access 2000
Satellite Imagery	Landsat Orthorectified, 1990 +/- 3yrs, 30m http://edc.usgs.gov/products/satellite/tm.html	ERDAS Imagine, Photoshop
Roman Roads Olive Oil Mills	Historical information by Dr. Michael Decker http://www.cas.usf.edu/history/fs/decker.html	ArcMap 8.x

Once completed, some maps will also have thematic layers that directly relate to the materials in the archive. For instance, olive oil mills and the Roman roads in Cyprian maps represent important thematic layers of the time period covered by the project. Another example of thematic layers would be historic maps, which can provide more valuable research information. A search is currently underway for relevant historic maps of both Cyprus and Egypt.

Technical Processes

Having to create large amounts of data requires careful and meticulous work to get the most accurate information. In many cases data was created a certain way, then discarded later when more effective ways of data development were found. The three most significant processes that had to be developed were the editing and extraction of administrative outlines, the creation of a place names database, and the development of shaded-relief files. More thorough technical information will be provided on the development of the database and shaded-relief files to benefit others doing similar work.

Administrative/Political Boundaries

Many trials were done to collect boundary layers. Initially the DCW files were going to be used throughout the project, but the scale did not conform to the scale of the rest of the datasets. Since the Russian topographic maps provided a better scale, we decided to digitize all coastal lines, and administrative/political boundaries from those maps. First, a free version of a software package called WinTopo was used to extract all the vector layers at once using a raster-to-vector function. This was useful in some areas, but required intensive cleaning to separate boundaries from roads and other line features. Manual digitizing ended up being the best option as it allowed for more quality control and gave the best results. Figure 2 shows before and after images of a segment of the Cyprus coastline. The image on the left shows the coastline (in red) from the DCW files, and the image on the right shows the edited coastline after they were digitized from the topographic maps. The boundary line from the DCW data set was noticeably off the actual country boundary. After digitized, the boundary file’s scale improved from 1:1000000 to 1:500000.

Figure 2. Boundary Files Before and After Editing

Place Names

Much of the linking functionality between GIS maps and electronic materials will be based on place features like cities, museums, and historic sites. For this reason, place layers have a significant role in the overall integration of GIS with the rest of the archive. The GEOnet Names Server (GNS) provides access to the National Geospatial-Intelligence Agency's (NGA) and the U.S. Board on Geographic Names' (US BGN) database of foreign geographic feature names. Available free of charge, this database contains place name files (in .txt format) for all countries. Moreover, an FTP website makes downloading easy with files organized by country codes.

Two tables were downloaded, one with designations codes and one with place names. The designation codes table contains hundreds of codes (DSGs) for features that are not significant to the project, and could cause overcrowding of the maps. Overcrowding of maps occurs when:

- The file contains features irrelevant for the project (airport, zoo).
- There are different features at the same lat/long location (a cemetery in a town).
- There are features with alternative names (Paphos and Bass).
- There are features with the same name, but in different location (Ayia, ruin at NE of Cyprus, and Ayia populated place SW of Cyprus).

Once downloaded, the tables were imported into MS Access. The designation code table was cleaned and queried to contain only a selected set of DGSs. This new table was then used to query and obtain another table with only selected places names. See Figure 3 for more details.

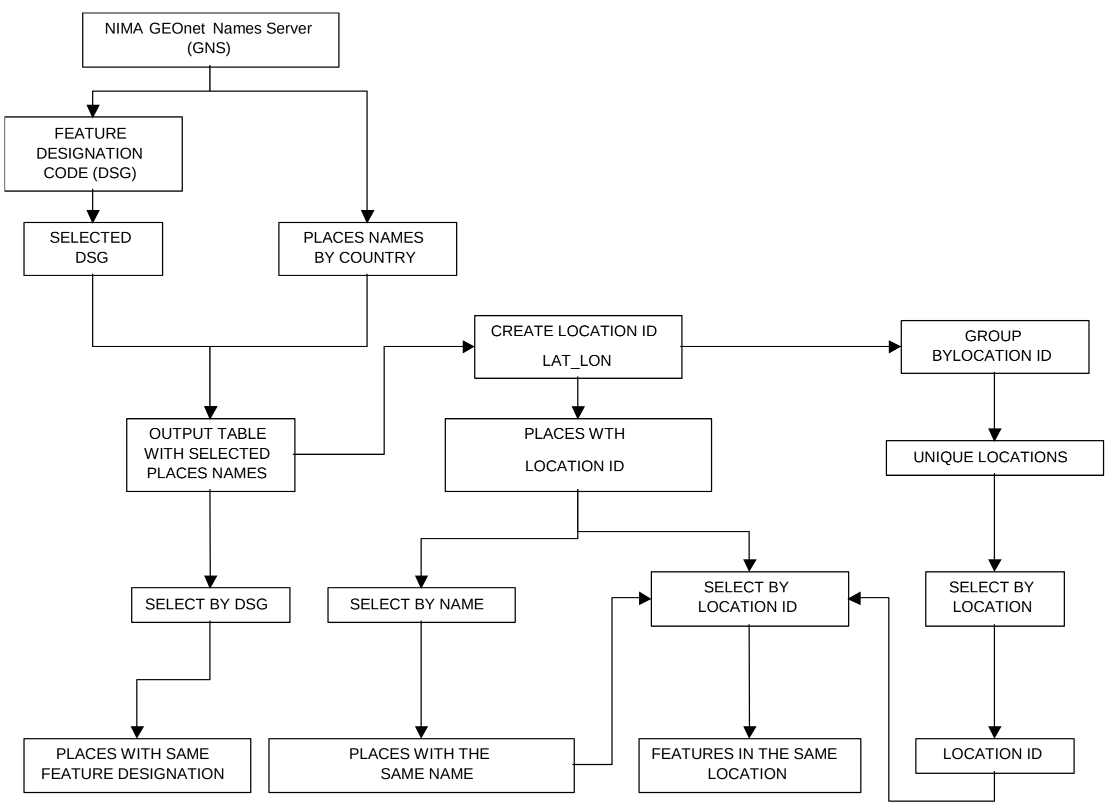

Figure 3. Steps to Process Place Names Files

To try to minimize feature overlapping caused by having places in the same location or with alternative names, the LAT and LON fields were used to create a location ID (LOC_ID = LAT &"_"&LON). This ID allowed the creation of two new tables, one with the complete list of places including the Location ID, and another with the list of unique locations (one record per location).

With the resulting three tables and with the aid of a relational database structure, it was possible to eliminate irrelevant features, create map of places based on classifications (churches or ancient places), search for a place by name or keyword, select by location (by clicking or entering the lat/long), and from a specified place identify features in the same location (using alternative names or different features in the same location).

Elevation Data

The project conducted many trials to create elevation layers for Cyprus and Egypt. The GTOPO30 files were processed using ESRI's GeoStatistical Analyst to create the initial Cyprus elevation layer. However, elevation files were later replaced once SRTM's files became available. This dataset is publicly accessible in Digital Elevation Model (DEM) format in files covering a 1 by 1 degree area with a resolution of 30 meters. The challenge was to design a process to correct and generate a color graded hill shaded relief file derived from the DEM files and then apply this process first to the files covering Cyprus and Egypt, and later to the rest of the countries. Figure 4 shows a rough flowchart of the steps involved in creating the finished elevation files.

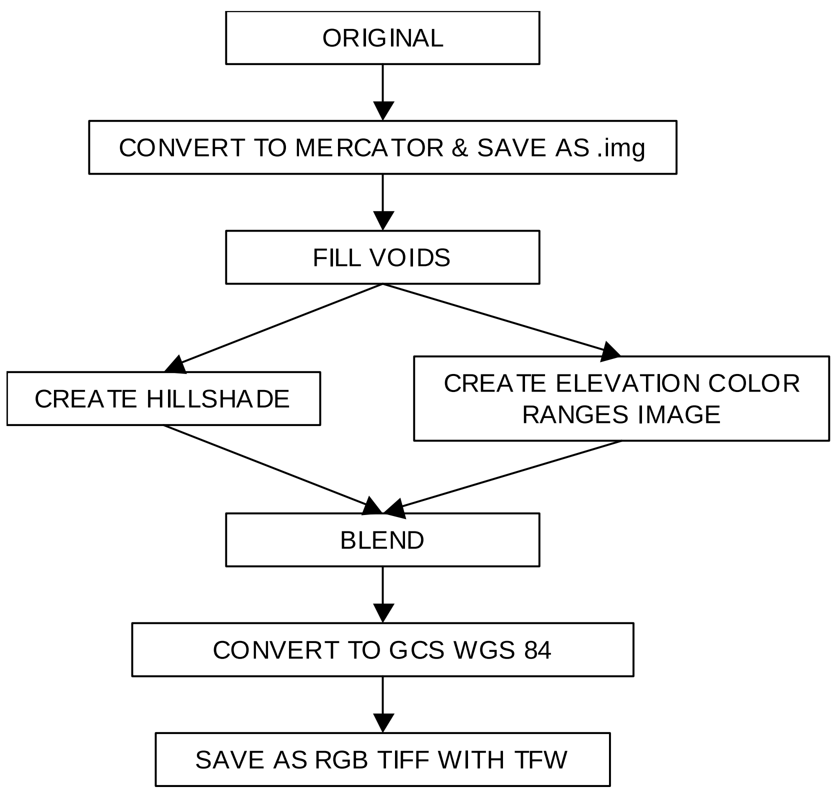

Figure 4. Steps to Process Elevation Files

The first step in the process was to convert the files to World Mercator projection and save them as ERDAS Imagine .img files. ERDAS Imagine takes the original file and reprojects it to World Mercator to have the X, Y, & Z values in meters (originally the X and Y coordinates are in decimal degrees, and Z value in meters). The program requires a consistent set of value for all XY & Z.

Due to the sensitive interaction of the radar with the ground during data collection [2], some DEM areas had inconsistencies recorded in the DEM as voids with a pixel value of –32767. To correct this situation, a method was designed to interpolate the voids using an ERDAS Imagine script. The script filled the voids by using the values of the pixels surrounding the void. The script scans the image with a 3x3 matrix, centering it in the void, and assigns to the void the majority of the values of the surrounding pixels containing real information. In cases where the void is just an isolated pixel, this method is highly accurate. For areas where the group of voids is bigger than a 3x3 matrix, the script makes multiple passes, filling the voids more completely each time, thereby creating a smoother, continuous surface at every pass. Figure 5 shows the basic logic of the script.

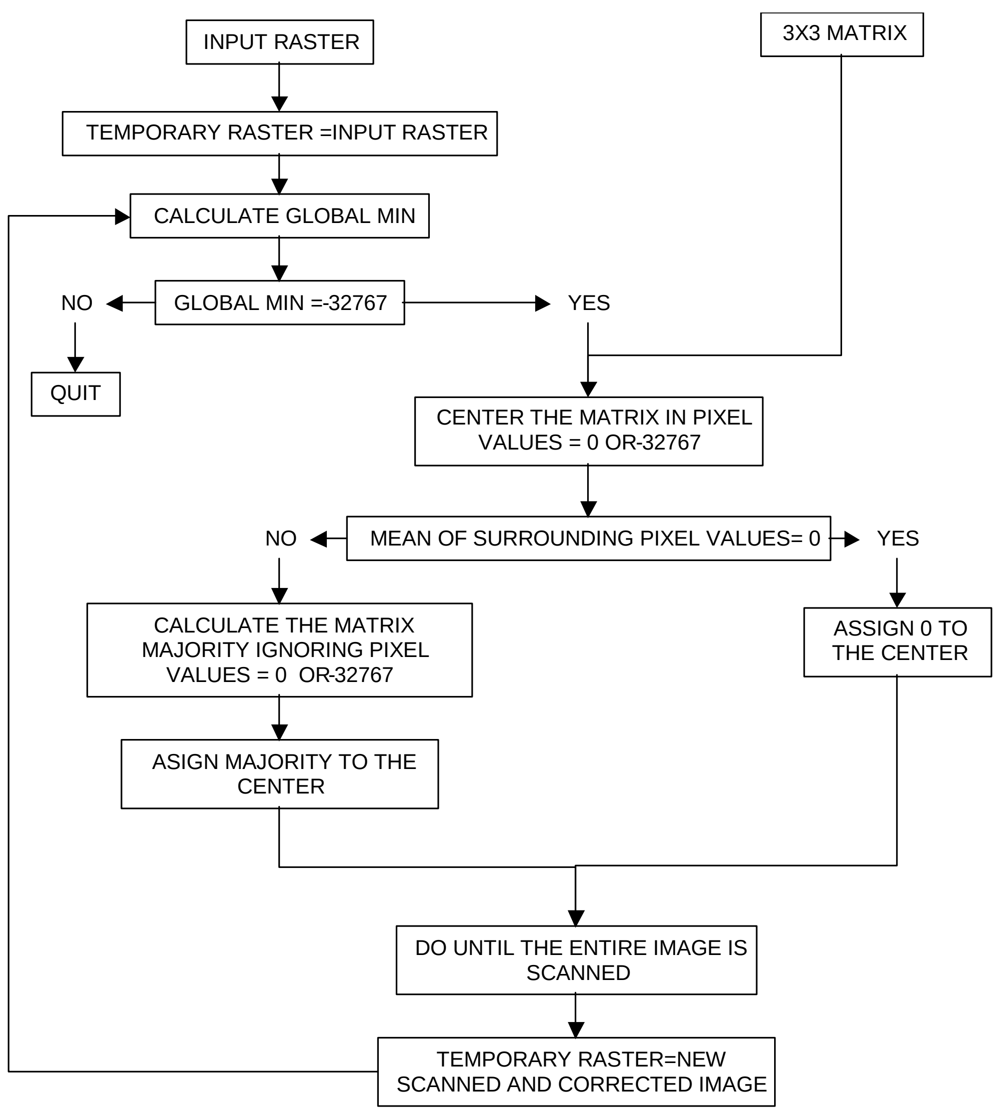

Figure 5. Logic of Fill Void Script

Two issues have to be taken in consideration while using this process. The first is that ERDAS Imagine recognizes the values outside the image as 0 values. When the array scans the pixels located at the edges of the image, part of the array falls outside of the image, computing those values as 0 instead of null values. To address this problem the script had to ignore the values 0 and –32767 when interpolating. This point brings us to the next issue, what happens when all the surrounding pixel values in the matrix are 0? In this case the script does not make any calculations and the script can loop forever. To resolve this problem, when all the surrounding pixels of a void area are equal to zero, the void will automatically be filled with zero values.

Once all the voids of the image were filled, a script was used to create the hill shaded relief model as a new temporary image file. To create the color-graded model, a series of elevation ranges were defined assigning some elevation breaks. Because most of the area of study falls within an elevation range of 0 to 1000 meters, more breaks were included in this range to highlight more topographic features.

Using the ArcGIS standard elevation color scheme in Figure 6, a specific color for each break was defined in red, green, and blue values. A script evaluated the pixel value, determined in which range it fell, and calculated its respective RGB value according to the formula in Figure 7. As a result, three temporary images were created in red, green, and blue values respectively. The next step was to combine all these images together using a "stack" function in ERDAS Imagine. Then the shaded relief file was blended with the new RGB colored image using the calculation: *Blended image = (Shaded Relief Image*255)+RGB Image –255*

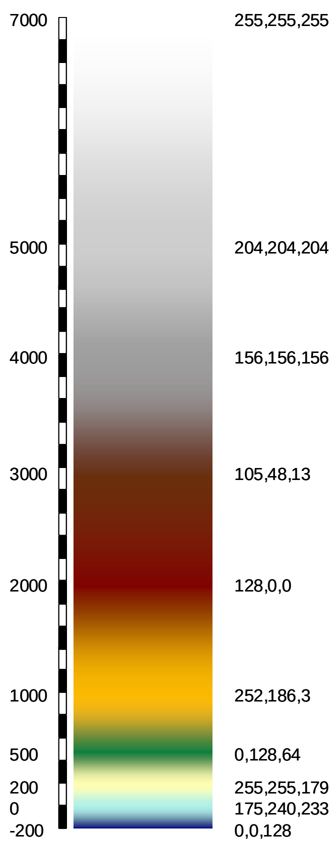

Figure 6. Elevation Color Scheme

$$R = \left(\frac{R\max - R\min}{RanMax - RanMin} \times \left(Pvalue - RanMin \right) \right) + R\min$$

$$G = \left(\frac{G\max - G\min}{RanMax - RanMin} \times \left(Pvalue - RanMin \right) \right) + G\min$$

$$B = \left(\frac{B\max - B\min}{RanMax - RanMin} \times \left(Pvalue - RanMin \right) \right) + B\min$$

Where:
Rmax=Maximum red value in the range
Rmin= Minimum red value in the range
Gmax=Maximum green value in the range
Gmin= Minimum green value in the range
Bmax=Maximum blue value in the range
Bmin= Minimum blue value in the range
RanMax= Top of the range
RanMin= Bottom of the range
Pvalue = Pixel Value

Figure 7. Formula for Assigning Pixel Values

Lastly, the blended image was reprojected to the "GCS WGS 84" coordinate system to be consistent with the coordinate system of the project. Since ArcIMS only reads RGB and not multi-channel files, the image also had to be exported as an RGB .tiff file with its corresponding world file (.tfw).

The Challenges

Data collection, staffing, and data processing have been the biggest challenges for a project as ambitious as TIMEA. Data collection is difficult because of the lack of extensive and affordable GIS data resources for the Middle East and Africa. In areas where no data is available, features have to be gathered, digitized or recreated almost entirely. Also, the variation in scale and resolution between the data sets calls for more clean-up work to make data sets match spatially. Some layers take longer to edit than others, depending on the amount of error and how badly they align with the rest of the data. This requires more work and staff, which brings us to the next challenge, staffing. The GIS team is composed of three people, two GIS professionals and one student worker. Only 10-20% of the teams' time is dedicated to TIMEA, which dictates how fast or slow the project moves. Also, since students only work temporarily, the workflow often slows down between semesters, as new students have to be hired and trained. Hiring GIS experienced students minimizes work interruption. Students typically perform scanning, digitizing, and data organization. The two GIS analysts manage the work progress, and produce methods for data development and conversions. This separation of duties helps with the workflow.

Data processing is one of the biggest challenges because of the vast amount of data that has to be processed with limited hardware resources. The elevation data set for Cyprus and Egypt for example contained hundreds of files. During processing, these files were multiplied creating three or four times the initial amount of data. To diminish this issue, data processing was often done in smaller steps taking into consideration staff's time and available computers.

Closing

One of the main goals of TIMEA is to serve as a teaching tool for all levels of education, thus it is being developed as a public resource. Also, as licensing permits, most of the data layers of the maps will be available for download once the maps and data sets are fully functional. During this stage of the project, the team is open to collaboration with other institutions and companies that may want to participate and contribute content or resources to the project. The TIMEA team has made great progress but a lot more remains to be done. Some of the current tasks include further development of "linking" functions between maps and books, addition of scanned historic maps, and data gathering.

The integration of GIS with non-traditional resources such as cultural heritage archives truly makes this project exceptional and a model for others. Being able to geographically search for research materials opens many doors for exploration, and even allows higher levels of understanding by showing relationships between places, books, and historic events.

References

1. Digital Chart of the World (Data Dictionary). Environmental Systems Research Institute Inc., 380 New York Street, Redlands, California 92373-8100, USA, 1993.

2. Trina Kuuskivi, Jennifer Lock, Xiaopeng Li, Steve Dowding, Bryan Mercer (2005). Void Fill of SRTM Elevation Data: Performance Evaluations. In: *Geospatial Goes Global: From Your Neighborhood to the Whole Planet*, ASPRS 2005 Annual Conference, March 7-11, Baltimore, Maryland, USA.