Building a Geographic Information System to Represent Historic Archaeological Resources

By Ty Summerville

Abstract

In an effort to assist the cultural resource management (CRM) activities of the Texas Department of Transportation (TxDOT), a geographic information system (GIS) database was developed to locate historic resources. The area selected for the project was TxDOT's Houston District. Maps dating from 1871 to 1952 were georeferenced, historic resources identified, locations digitized and information pertaining to the structures and map sources added to the database. The greatest challenge of the project was accurately georeferencing maps that predate the adoption of the North American Datum of 1927. To deal with this issue, customized datum transformations were developed to transform maps into a modern coordinate system. Several quality control measures were also used to assess the accuracy of the generated GIS data including the use of root mean square error (RMSE) for positional accuracy.

Introduction

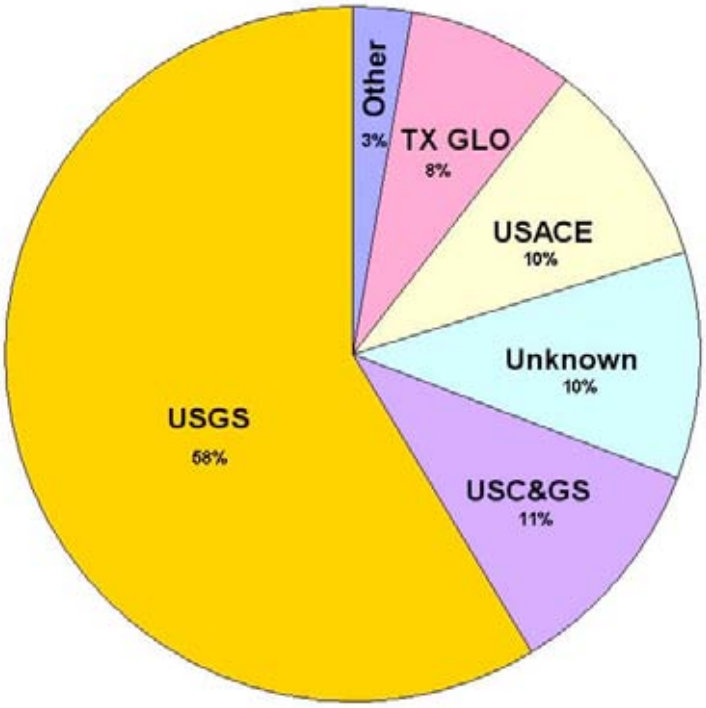
Management of the state highway system in the state of Texas is the responsibility of the Texas Department of Transportation (TxDOT). This highway system runs through more than 79,000 of miles of land in Texas. Many different environments are affected by the highway system and the constant demands made on it. One of these environments affected is our own man-made environment. Several Federal and State cultural preservation laws protect historically and archaeologically significant human occupation and habitation sites known as cultural resources. TxDOT's Environmental Affairs Division manages cultural resources affected by highway construction and maintenance. In an effort to more effectively manage cultural resources, archeologists and historians within TxDOT's Environmental Affairs Division approached PBS&J with a project concept involving the creation of a historic archaeological resource geographic information system (GIS). The source information for creating the GIS was to be cultural resources site locations derived from historic era maps. When completed, the GIS would provide important and readily available information needed for planning transportation improvements and maintenance projects.

Project Background

The project area selected for cultural resource GIS coverage was the TxDOT's Houston District. The Houston District contains 6 Texas counties: Harris, Galveston, Brazoria, Fort Bend, Waller and Montgomery. The project was designed to include 2 phases of data collection and analysis. The first phase consisted of cataloging pre-1930 maps from 15 selected repositories and then analyzing these source maps to assess their potential for providing relevant cultural resource site locations within the Houston District. Of the 400 maps identified, 97 were selected as source maps for further analysis and site-specific

data capture. The second phase included the acquisition, georeferencing and digitization required to generate the cultural resource GIS files. The second phase also included a series of scoping meetings designed to develop the criteria used for selecting and identifying the cultural resources to be captured from the source maps. From these meetings, 82 types of map features were selected and classified into 12 categories of cultural resources. A third phase was also issued to widen the coverage extents of the resource GIS by selecting post-1930 maps.

Figure 1: Percentages of Source Maps from Identified Sources



Georeferencing Historic Maps

To build a useful GIS, features must be located using a common coordinate system. Selected maps for this project were scanned and georeferenced in order for a common coordinate system to be used with them. Georeferencing is the process of registering a map to a common coordinate system. Geospatial data in a common coordinate system can be analyzed and plotted with other geospatial data such as proposed transportation corridors. Digitizing features on georeferenced maps produces higher positional accuracy when feature placements scales are standardized to be larger than the source map scale. Additional advantages of using georeferenced imagery for data generation are the lack of a need for a digitizing tablet and the ability to access accuracy of the georeferencing processes more readily by producing a visual product of the result.

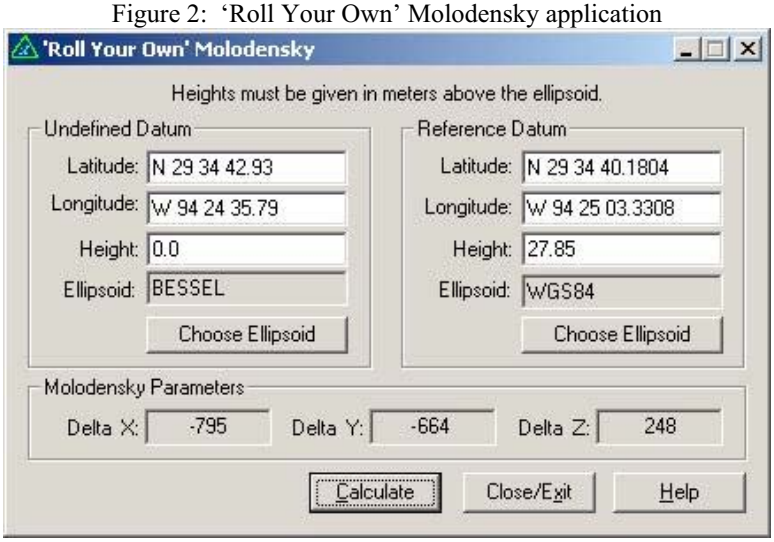
The process of georeferencing maps only requires some basic information about the maps being georeferenced if those maps already use a common coordinate system. A common coordinate system can be any coordinate system commonly available as input/output

coordinate system selections with commercially available coordinate transformation software. This type of software will only accurately transform coordinates for maps of the United States that are created with the North American Datum of 1927 (NAD27) or more recent horizontal datums. Even though NAD27 was adopted as the standard datum for use in the U. S. in 1927, some areas in the U. S. were mapped using earlier datums for a number of years. This was certainly the case in this project with a few maps being reprinted into the 1960s using a datum predating NAD27. Overall the map set selected for this GIS project included maps ranging in map publication dates from 1871 to 1969 (reprint of earlier map). Of the 97 source maps, 24 use NAD27 while 73 use another datum.

For the 73 source maps not using NAD27, additional cartographic information was needed in order to accomplish the needed transformation from an antiquated coordinate system to a common modern system. Based on research, the maps with antiquated coordinates were found to employ the North American Datum (1913), the U. S. Standard Datum (1901) or other obscure datums available prior to development of regional geodetic datums. The method used to generate a coordinate equivalency for these antiquated coordinate systems involved the development of a 3-parameter Molodensky transformation. Such a transformation enables coordinates using one datum to be transformed to coordinates using another datum. The process works by calculating the shift along the x, y and z-axes from one system to another. To calculate this shift, a freeware called 'Roll Your Own' Molodensky was used. The application was obtained free of charge from Mentor Software, Inc.

To use the 'Roll Your Own' Molodensky application, two pair of geographic coordinates for a single position are required. One pair of the coordinates is the geographic location of a point expressed using the undefined datum. The undefined datum in this case is the datum used on the historic maps. The second pair of the coordinates required is the geographic location of the same point expressed using a known datum. By supplying both sets of coordinates to the application, the program will calculate the shift needed to convert between the two datums.

Figure 2: 'Roll Your Own' Molodensky application



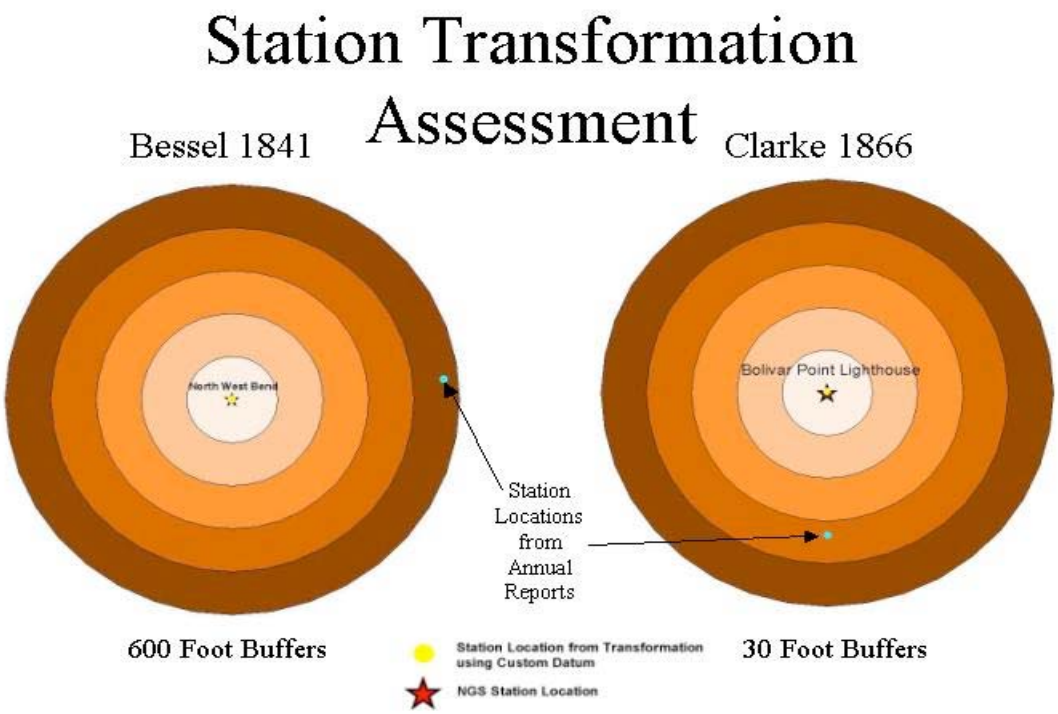
Finding coordinates that could be used with the 'Roll Your Own' Molodensky application involved a substantial research effort. Two pair of coordinates were needed for each datum used by the historic maps that predated the use of NAD27. Since the selected historic maps for the project employed two such datums, 2 sets of coordinates were needed for each of these datums. Each set required that coordinates of a single feature be used. This feature had to exist during the time that the maps were drawn and a coordinate location determination must have been recorded using the location technology of that era. Coordinates for the same feature were also required using a known modern datum. Archival research to find such information yielded a perfect source for the coordinate information needed. The source used was the U.S. Coast Survey, later known as the U.S. Coast and Geodetic Survey, which is now called the National Geodetic Survey (NGS), a part of the National Oceanic and Atmospheric Administration. This agency has, since its beginning in 1807, been the preeminent branch of the U.S. government involved with geodetic surveys of the nation. In 1851 the U.S. Coast Survey first published, in its Annual Report to Congress, the geographic positions of triangulation stations used in geodetic surveys. This report and subsequent annual reports by the U.S. Coast Survey provided half of the information needed to calculate a 3-parameter Molodensky transformation. The other half of the information would be obtained from the National Geodetic Survey. It was found that a handful of triangulation stations used in the 19th century and early 20th century are still being maintained by the NGS in and around the Houston District. The NGS publishes information concerning maintained triangulation stations and other survey control monuments using what is called NGS datasheets. These datasheets provide geographic coordinates for the stations using know modern datums and a detailed history of the stations that includes the date that each station was monumented. By using the geographic coordinates listed for a triangulation station from both sources, the 'Roll Your Own' Molodensky application was used to generate a 3-parameter Molodensky transformation for each of the two antiquated datums used by the selected historic maps.

The oldest datum used by the selected historic maps was the Bessel 1841 ellipsoid. Prior to the development of regional geodetic datums, ellipsoids were used with a geographic coordinate for an origin point contained in the area being mapped. These input parameters are called a local datum. The Bessel 1841 local datum was used on 5 of the selected historic U.S. Coast Survey maps. The 3-parameter Molodensky transformation developed for the Bessel 1841 local datum used the North West Bend triangulation station, first monumented in 1861. This station was chosen for its proximity to the Houston District and its monumentation date. The only station listed in the 1864 annual Coast Survey report that is still being reported currently by NGS is the North West Bend station. The results from the 'Roll Your Own' Molodensky application were calibrated by creating a custom datum using Blue Marble's Geographic Calculator and simply transforming 1864 geographic coordinates for North West Bend to geographic coordinates using the North American Datum of 1983 (NAD83). NAD83 coordinates are used on NGS datasheets. The resulting NAD83 coordinates were then compared to the modern NGS NAD83 coordinates on the North West Bend datasheet. Several iterations of this process were conducted to find a 3-parameter Molodensky transformation that

produced coordinates closely matching those found on the NGS datasheet for North West Bend.

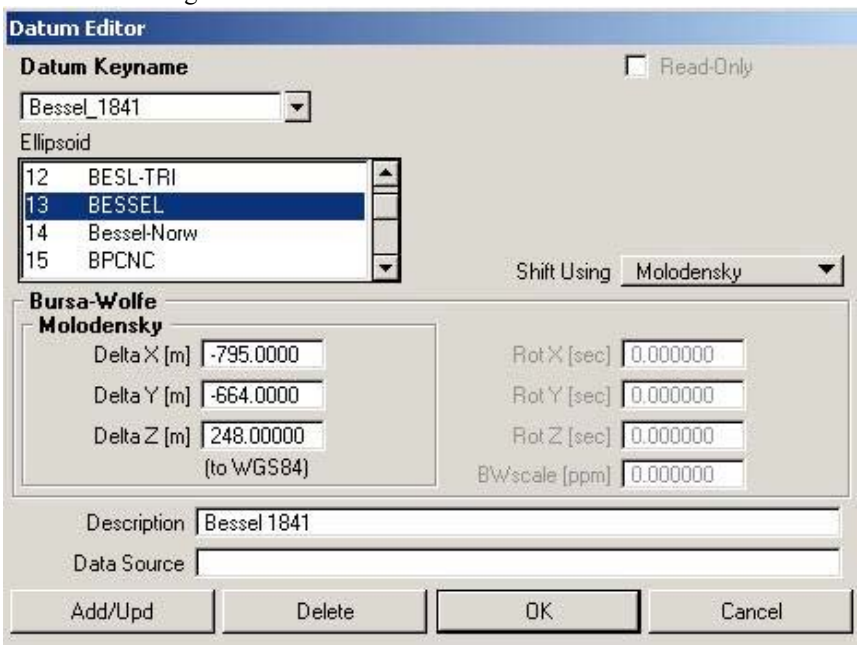
The other antiquated datum used by the selected historic maps is actually a range of datums all using the Clarke 1866 ellipsoid. In 1870, the U.S. Coast Survey adopted the Clarke 1866 ellipsoid for use for geodetic purposes including mapping. This ellipsoid was used as a local datum until the adoption of the first geodetic datum of the U.S., the U.S. Standard Datum of 1901. In 1913, the U.S. Standard Datum was renamed to the North American Datum (NAD) with the adoption of its use by Canada and Mexico. Continual refinement of geodetic information after 1870 eventually lead to NAD and NAD27. Due to this progression, a single 3-parameter Molodensky transformation was generated for maps using a datum based on the Clarke 1866 ellipsoid. Once again U.S. Coast Survey and NGS geographic positions were used. For this datum, the Bolivar Point Lighthouse Station, monumented in 1900, was used. This station was chosen because it is located within the study area and is depicted on several source maps. The results from the 'Roll Your Own' Molodensky application were calibrated by comparing the results of the 3-parameter Molodensky transformation used as a custom datum with Geographic Calculator with the NGS datasheet for the Bolivar Point Lighthouse Station. Once again, several iterations were conducted for this process before an accurate transformation could be found.

Figure 3: Positional accuracy assessment of each 3-parameter Molodensky transformation generated



The next step in the process of georeferencing the historic maps was to create the custom datums necessary in an application that would replicate the grid lines and ticks used on each historic map. The application selected to do this was Bentley's GeoCoordinator, which is now part of version 8 Geographics. The selection of GeoCoordinator was purely pragmatic based on the need of an application to generate grid lines using a custom datum based on the 3-parameter Molodensky transformations derived from the 'Roll Your Own' Molodensky application. Each of the antiquated datums were created in GeoCoordinator by assigning the appropriate ellipsoid definition and the 3-parameter Molodensky transformation calculated for it.

Figure 4: Custom datum editor in GeoCoordinator



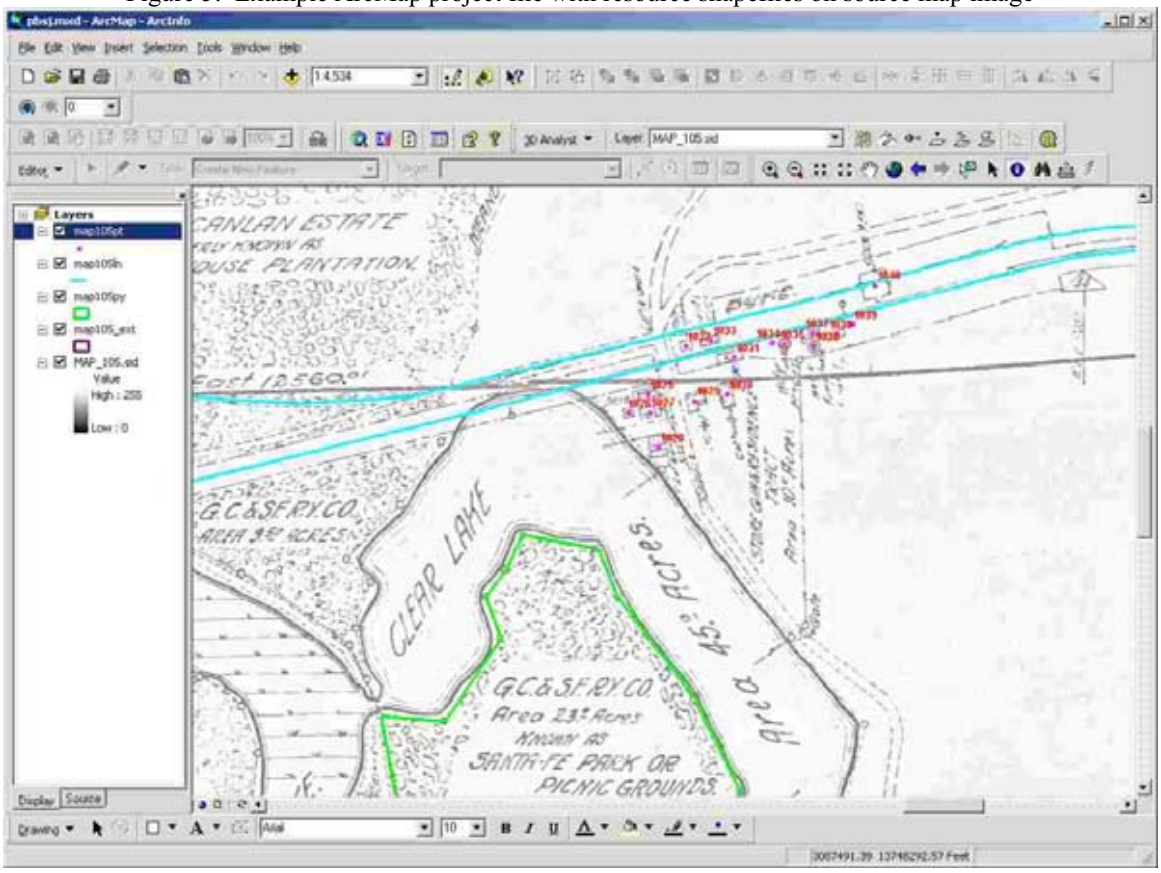
The final step in the process of georeferencing the historic maps was to resample each of the historic maps by assigning points in common (control points) between the registration grids and the grids used on the historic maps. To do this Bentley's Descartes was used. The choice to use Descartes was based solely on the dynamic warping functionality. Dynamic warping allows the user to view vector data on top of an uncorrected image by applying the coordinates of the control points selected as part of the registration model in reverse. Instead of matching control points and then resampling an image, the effects of control point placement can be viewed dynamically with the vector data shifting on the uncorrected image. This process was critical in georeferencing most of the historic maps. In some cases there were no grid ticks or grid lines on the historic maps. In these situations property lines, railroads, streets and bridges were used instead of registration grids to form the control points necessary to build a registration model for an image. The use of dynamic warping provided a real-time quality control mechanism by allowing the user to see first-hand the implication of selecting control points and the flexibility to delete and/or add additional control points before resampling an image. Once resampled, each of the selected historic map images were georeferenced to a common coordinate system.

Recording Historic Resources

Creation of the historic resource GIS was approached in a distributed manner. Different analysts conducted resource identification, database entry of resource information and digitization of resource locations separately. A resource identification analyst recorded information describing each of the identified historic resources in Microsoft Access tables. Data entry forms were developed to record a UID number, map identification number, name, comment and resource type for each historic resource.

Locations of historic resources were digitized as either Microstation DGN files or ESRI personal geodatabases or shapefiles. Heads up digitizing was conducted in both softwares using the georeferenced imagery. A map scale of at least one-half the published scale was used as the standard viewing scale for digitizing historic resources. The only attribution initially attached to the historic resource graphics was a unique identification number (UID). In the Microstation environment this attribution was acquired using tags which are directly convertible to table attribution in ArcView or ArcMap. Data in the Access database tables were joined to the shapefile versions of all digitized historic resources. Shapefile format was specified as the GIS file format for the delivery of the data to TxDOT.

Figure 5: Example ArcMap project file with resource shapefiles on source map image



Assessing Accuracy

The final stage of the project consisted of assessing the quality of the data that had been generated. To assess the positional accuracy of the historic resources, a root mean square error (RMSE) evaluation was conducted. A RMSE calculation measures the spatial variability of coordinates for generated data being tested for accuracy when compared to matching data of higher positional accuracy. Map features in common between the historic maps and maps meeting National Accuracy Standards or matching published NGS datasheet coordinates were called sample control points (SCP). Features selected as SCP's were, in order of preference, survey control monuments, railroad intersections, road intersections and landmark point locations such as lighthouses and churches. Less than one-third of the SCP's matched station coordinates on NGS datasheets; but the SCP's that did match provided the most accurate assessments. In all cases, a point shapefile was created using images of a modern U. S. Geological Survey (USGS) 7.5' quadrangle and a selected historic map. For most maps, at least 5 sample control points were identified and used to evaluate the positional accuracy of individual source maps. The RMSE was calculated using the XY coordinate variability of the sample control points. The results of the RMSE calculation for each map were recorded in the database containing the detailed information regarding the selected historic maps.

In addition to positional accuracy, the accuracy of the historic resource attribution was also assessed. This assessment was a 10% stratified random sample of historic resources. The sample was stratified by graphic feature type so that 10% of each class were reviewed. The evaluation consisted of a 7 criteria error matrix in which each resource was checked for correct UID number, map number, name, comment, resource type and location. Analysts who did not participate in the creation of the data conducted the review. Random numbers were generated using the data analysis tools in Microsoft Excel. 10% of the total number of potential resources in each graphic feature class comprised the number of random numbers generated for the selection of the samples. The random numbers matching a sequential listing of the resource UID numbers comprised the samples evaluated.

The completeness and consistency of the historic resource shapefiles were the last assessments conducted before the data was final. A PBS&J historian or archaeologist reviewed each selected historic map for completeness. This was accomplished by plotting each georeferenced historic map image along with the historic resource shapefiles labeled with the resource UID numbers. This review identified 17 resources (less than two-tenths of one percent of the total number of resources captured) that were missed during the identification and/or digitization process. Moreover, the consistency by which historic resources were attributed including the graphic feature classes used to represent the resources were also checked. Consistency checks were done using database field summaries for attribution consistency. Attribute values in these summary tables were verified or updated to reflect consistent values where appropriate. Graphic feature class consistency was also evaluated by comparing the types used in the shapefiles to the

resource type classifications selected during resource identification in Access database tables.

Findings and Conclusions

This project identified over 12,000 historic resources within TxDOT's Houston District from 97 historic source maps as source material. A variety of source maps were used, each with various levels of accuracy. All of the maps used during the first two phases of the project were obtained directly from the repositories visited during the first phase of the project. Maps used during the third phase were ordered directly from the historic map archive at the USGS. These maps are photographic reproductions of the original quadrangle maps. This acquisition method was found be the quickest, cheapest and most efficient method. While the USGS does not publish a database of historic maps within their archives they do take research requests and can provide the results of a database queries from the internal map database that is used by the USGS. Using this information, the third phase of the project was able to extend the coverage area of the historic resource GIS by 50%. The only drawback to using the photographic reproductions is a slight degradation of clarity or blurriness due to the photographic process. The reproductions are to scale and can easily be scanned and georeferenced

As should be no surprise, the most positionally accurate georeferenced maps in the source maps set were the most recently published maps using NAD27. Likewise, maps using NAD were found to have higher positional accuracy than maps using the U.S. Standard Datum or the Bessel 1841 local datum. Larger scale maps were generally more accurate than smaller scale maps. Even with the extensive research of the cartographic history of these maps and the techniques employed during the construction of these maps, several specific gaps in the information remain. Many maps contained no information about projections or datums used to create the map. Assumptions were mainly used based on what was determined to be the mapping techniques used by those who created the map. These assumptions obviously had some affect on the resulting positional accuracy of the data set. For higher levels of positional accuracy much more detailed and specific information would be required for each map used.

USGS quadrangles were found to contain 3 to 4 times the number of historic resources compared to other selected source maps. Moreover, most non-USGS maps from the selected source map set covered an area larger than 3 to 4 times the extents of a 7.5' quadrangle. Due to the density of historic resources on the USGS maps and the quality these maps they were by far the project's most valuable source.

Archaeologists and historians will use the historic resources shapefiles delivered to TxDOT as part of the cultural resource management conducted by the agency. Other uses obviously exist for transportation planners but it's possible that the data set has many uses yet to be discovered. TxDOT has 25 transportation districts within the state and the successful use of the data produced by this project could spur the development of similar datasets for other districts. The wealth of information that exists in various archives in the form of historic documents and maps can at times seem limitless. The

development of cultural resource GIS databases will undoubtedly serve the increasingly complicated needs of cultural resource managers in the future.

Ty Summerville
Senior GIS Analyst
PBS&J
206 Wild Basin Road, Suite 300
Austin, Texas 78746
512-329-8342 (x9661)
tdsummerville@pbsj.com