Spring 2011 - Desktop GIS: Building a County-Wide Address Points Database

By Dan Castaneda and Randy Knippel

The Office of GIS is currently working with all cities and townships in Dakota County to develop a county-wide address point database. The goal of the project is to create a single, official source for every address in Dakota County, conforming to database standards developed by MetroGIS, which is based on a proposed federal standard. The database will be available for distribution through MetroGIS, and will be used in various GIS applications, as well as other systems.

The immediate purpose for this database is to support a new computer aided dispatch (CAD) system, used by the Dakota Communications Center. The new CAD system is being hosted by LOGIS (Local Government Information Systems), a consortium in Golden Valley. This CAD system, and others like it, has been updated to take full advantage of GIS data and will be much more effective at directing emergency responders by using these address points. Similarly, the next generation of E911 systems (NG911) will use the same address points when that technology is deployed in the next few years.

Dakota County has a strong foundation for addresses based on our parcel data, which will typically give us one address per parcel. However, there are areas where multiple addresses exist for a single parcel, typically in commercial and industrial areas and in higher density residential areas such as apartment complexes. These areas have gaps in addresses that need to be filled. To help fill these gaps, Dakota County was provided additional address points by Dakota Electric Association. The Office of GIS also incorporated address from a business database, which had been built from multiple sources. Finally, additional addresses from LOGIS that were already in the existing CAD system were added as well.

Local knowledge and data can help further refine the gaps in this area, and Dakota County is engaging the cities to accomplish this. City staffs typically are much more familiar with addresses and assign new addresses as necessary. Also, many cities maintain extensive lists of addresses for a variety of purposes including economic development, building inspections, and utility billing. Any addresses in these lists that don't already exist in the address points database will be appended using a systematic approach.

Several address locators using the previously mentioned datasets were used to geocode the address lists. The first locators used in the process run against the existing address points database. This generates a list of addresses which do not match an existing address in the database. These addresses are then run against locators representing other geocoding methods based on streets and address ranges to generate new address points, which are then added to the database. Finally, some addresses that don't match any of the geocoding methods are returned to the city so they can determine whether or not the address is valid, or if it contains a mistake that can be corrected. Once the lists are checked over and returned, the process is repeated to determine if more matches exist. If there are any unmatched addresses, manual verification may be needed, but will be used as a last resort. Throughout this entire process, a coding system is used to keep track of the source of the address, and to track which dataset each point matched.

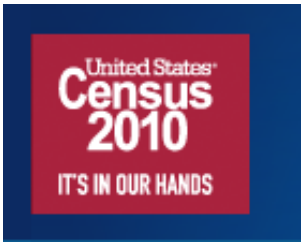
Through the geocoding process, addresses are located using various methods representing various levels of accuracy and precision. One final step automatically locates many of the points directly on the building itself. The automated geocoding and refinement creates locations for most addresses in the county. However, additional interactive refinement is necessary to place the address on the front door, which is considered the most accurate

location.

Once all of the geocoding, review, and refinement steps are completed, the address points database will be the most comprehensive and accurate set of address locations ever available to County and city staff. This database will undoubtedly prove to be a tremendously valuable asset to departments throughout the County and cities.

Spring 2011 - GIS 101: The 2010 Census and the American Community Survey - The Evolution of Census Data

By Scott Laursen



There have been a lot of news reports featuring data released by the Census Bureau in recent months. I've seen headlines about the release of the U.S. House of Representatives re-apportionment data, breakdowns of the diversity of the various cities in the Metro area, and an article about how some rural counties are getting older. The source of this information is frequently simply given as "data from the Census Bureau". In most people's minds, that means "data from the 2010 Census". While it would have been safe to assume that such data came from the most recent Census in the past, that is no longer true. The data used in these front-page articles was actually generated from two different sources –the 2010 Census and the American Community Survey. While the Census Bureau produces both of these datasets, they are quite different from each other, and it is important to know what those differences are when looking at data from them.

For the 2000 Census, the Census Bureau sent out two forms for citizens to fill out. Most households received the "short form", which contained only eight questions about race, ethnicity, age, gender, and home ownership. Alternately, about one out of every 40 households instead received the "long form" which, in addition to the questions found on the "short form", also asked a large number of detailed questions about the household, gathering information on income, poverty, disability, country of origin, etc. The results of the additional "long form" questions were then statistically manipulated to produce estimates of what the data would have looked like had these questions been asked of the entire population.

After the 2000 Census, the Census Bureau decided to change the way they did things. People who used Census data had long complained that the 10-year gap between Censuses was too long, that the data became out-of-date and irrelevant after three or four years. After listening to these complaints, the Bureau decided to eliminate the decennial "long form" and instead produce estimates of those data on a yearly basis. The "long form" was replaced in 2005 by the American Community Survey, or ACS. The ACS provides what's called "continuous measurement" of social, economic and housing data by sending out monthly surveys to approximately 250,000 randomly selected households across the country. These surveys are combined and statistically manipulated to produce yearly estimates.

This solution goes a long ways toward solving the problem of having out-of-date data, but it isn't perfect. Places with smaller populations, such as small cities and rural areas, often don't receive enough surveys during a year for their results to be considered statistically significant. As a result, the Bureau has to aggregate their data results over time to be able to produce acceptable estimates. In general, if an area has a population of greater than 65,000, the ACS is able to provide yearly estimates. If an area's population is between 20,000 and 65,000, the ACS results show data gathered over a three-year period. Finally, if an area has a population of less than 20,000, the Census Bureau has to wait five years to collect enough data to provide accurate estimates.

The initial five-year data collection period has now passed, and the first batch of five-year ACS data was released late in 2010. This is the data that generated many of the news reports mentioned earlier. However, it is important to understand the limits of this data. The economic and social characteristics of a community can change in three years, and it can change a great deal in five years. This needs to be kept in mind when reviewing ACS data.

What about the 2010 Census? The "short form" was the only questionnaire used in 2010; it was sent to every household in the country. It is the source of the actual population counts that were used to determine which states gained or lost seats in the U.S. House of Representatives. This data will also be used in the redistricting

efforts that will take place early in 2012.

The 2010 Census data is considered to be more accurate than the ACS data, because it is as close as we can come to an actual count of the population. The ACS data, while considered accurate, is only an estimate. This is why it is important to know which type of “Census data” is being used. Right now, data from the 2010 Census is more accurate than the ACS data. Five years from now, though, the ACS estimates will still be coming out on a yearly basis, while the 2010 Census data will be considerably out-of-date, so at that point the ACS information will be considered to be more accurate.

More information about the 2010 Census and the ACS can be found on the Census Bureau’s website – http://www.census.gov.

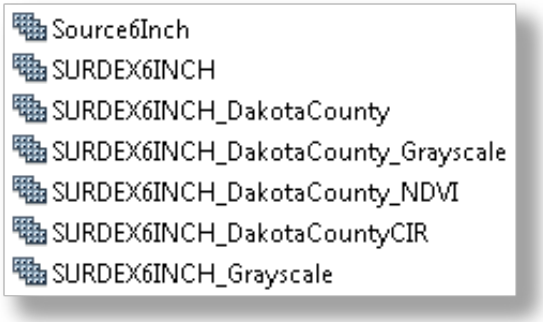
Spring 2011 - Tech Talk: Mosaic Dataset: Faster Raster Delivery to the Desktop

By Joe Sapletal, GISP

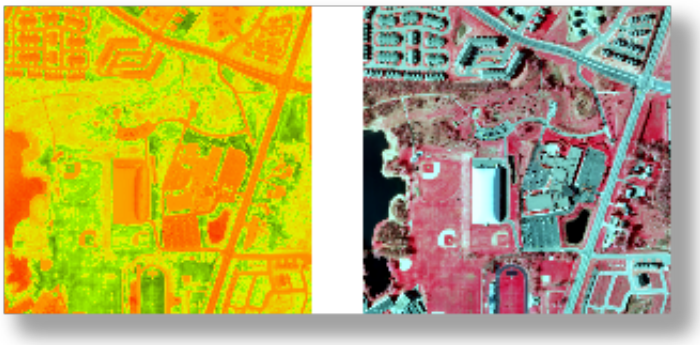
There's nothing worse than knowing the new imagery has arrived and only a select few people get a peek at it before the GIS Database Administrator spends days loading it into the database for everyone else to access. At ArcGIS 10 there is a solution for getting imagery into the hands of the users quickly, and with a pretty good display time, too - the Mosaic Dataset.

"A mosaic dataset allows you to store, manage, view, and query small to vast collections of raster and image data. It is a data model within the geodatabase used to manage a collection of raster datasets (images) stored as a catalog and viewed as a mosaicked image." - ESRI Desktop Help. The store and manage items are key to making imagery available as soon as possible upon its arrival.

By using the Create Mosaic Dataset tool and the Add Rasters to Mosaic Dataset tool, in a matter of moments we were able to create a mosaic dataset of our new imagery. Essentially all it is are pointers to where actual image tiles are stored on a network drive. After using the Build Overviews tool the imagery displayed quite quickly. The imagery was ready to be viewed, but we didn't stop at just being able to display the color photography.



We created a number of mosaic datasets from the 2010 6-inch orthophotography we received. We created a source mosaic dataset (called Source6Inch) which served as the source mosaic dataset for all the others. In ArcCatalog if you preview this dataset you will see the records that represent the 1,150 image tiles, as well as the overviews that were created. Each of the other datasets we made from the source simply have one record referencing Source6Inch. Each of those was created by using the Create Mosaic Dataset tool, and instead of referencing the source imagery, it references the original mosaic dataset we made. Hopefully if there are any images that need to be removed or added to the datasets, that can be done easier than if picking and choosing had to be done to each data set. The additional datasets were created on the fly by applying functions to them to extend the usage of the imagery. They appear to display just as quickly as the original mosaicked dataset.



We have a number of other mosaic datasets (some shown in the graphic), all from the same source, that were created using Functions. Functions allow you to create other datasets without actually creating new datasets. When you have 1,150 images that are 350MB each, being able to apply functions instead of creating entirely new datasets saves a lot of disk space. Functions can be as basic as a Clip to extract or exclude areas from the output dataset, to grayscale to turn a color image into a grayscale image. They can be as complex as NVDI (Normalized Vegetation Differential Index) or, if your raster is a surface, creating a shaded relief or hillshade without all the processing.

You can apply some functions to a raster, be it imagery or surface data, to do on the fly processing without the need for extra diskspace, without sacrificing display speed.

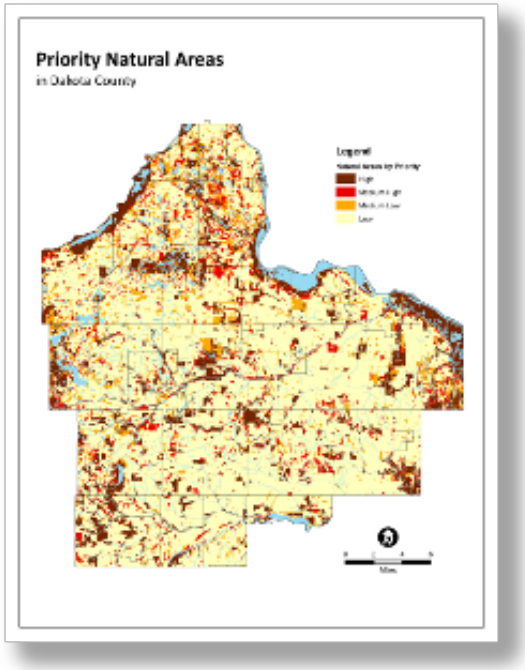
These are just a few of the many ways the mosaic dataset can help improve access to your imagery. They are relatively simple to set up, and have many options to fit your needs.

Spring 2011 - Department Spotlight: Dakota County Land Conservation Vision - Using GIS to Focus Conservation Investments

By Brendon Slotterback, Dakota County Office of Planning and Analysis, and Al Singer, Dakota County Parks and Open Space Department

In 2010, the County's Office and Planning and Analysis (OPA) began a project with the Park and Open Space Department to update the County's comprehensive land conservation vision. This work inventoried and reassessed priorities for all land conservation programs in the county, including Parks, Greenways, natural areas, farmland, and the buffer corridor initiative. A key goal of this work was to re-prioritize land for public protection and management.

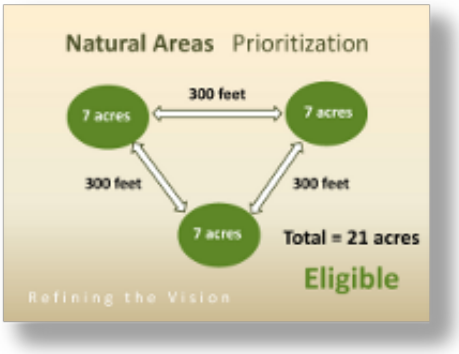
Previously, the scoring system for the Farmland and Natural Areas Program (FNAP) was weighted heavily towards landowner donation and some subjective criteria that was then used by a citizen advisory committee to evaluate projects. The goal of this joint project was to revise the criteria and scoring system through a more objective and updated analysis of natural resource quality. OPA staff determined that GIS was the appropriate tool to complete this work objectively on a county-wide basis. Accordingly, staff from the Dakota County Office of GIS assisted these efforts.



To begin the assessment, a qualitative natural area scoring system developed by the Dakota County Soil and Water Conservation District was used. This information layer provided a score based on a range of land cover types, proximity to habitat, and other features. Project staff modified this scoring system slightly to meet the needs of the land conservation study. The resulting natural resource scores/quality ranking would provide the basis for selecting the highest priority areas in the county for protection through land conservation programs.

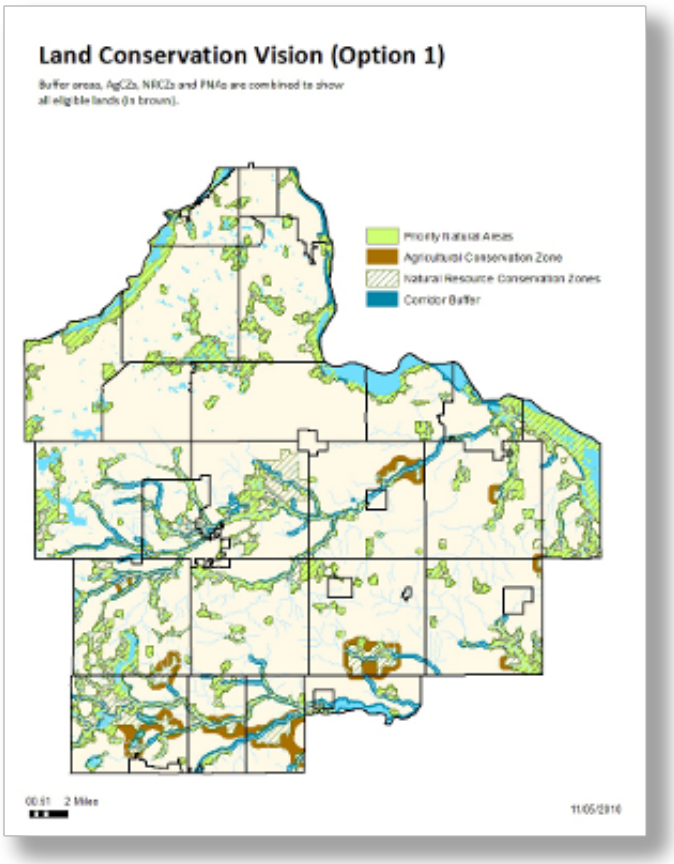
A number of factors based on natural resources science were used to identify the highest priority protection areas. These included:

- Natural area score
- Size of a contiguous natural area
- Proximity to other natural areas



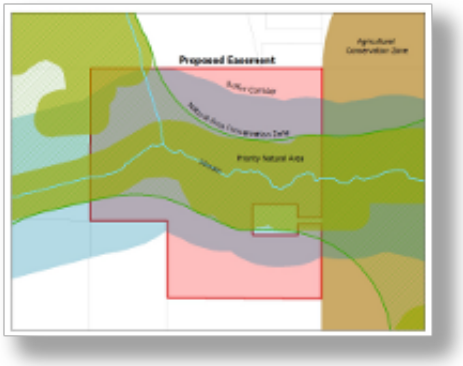
While a minimum area size was an important consideration for prioritization, small natural areas close to other small natural areas created groupings that may have higher natural resource value for wildlife habitat. Using GIS, staff was able to select natural areas that were either a minimum size, or whose area when combined with nearby features exceeded the minimum size.

These high Priority Natural Areas (PNAs) would be areas that typically received the highest score for a potential land conservation project. Realizing that it was also important to connect high quality natural areas with each other and with other protected land (federal, state, regional, county, local and private), GIS was used to create connecting corridors to form contiguous Natural Area Conservation Zones that would score well in the conservation programs.



The relative importance of water adjacency was increased to further enhance the priority of protecting "Buffer Corridors" along all streams, rivers and undeveloped water bodies.

Finally, areas near farmland that had already been protected through the FNAP were considered important in order to encourage the creation of farmland "blocks" that would have additional benefits for the landowner and the county. These areas, essentially buffers around existing protected farmland, are called Agricultural Conservation Zones.



This project resulted in the creation of many overlapping zones (Priority Natural Areas, Buffer Corridors, Natural Area Conservation Zones and Agricultural Conservation Zones), with each zone contributing varying amounts of points to potential land conservation projects. This new location-focused scoring system will allow staff to more objectively rate projects. Staff is also considering the possibility of developing a web map which will allow landowners who may be interested in land protection and management to quickly assess the score of their land prior to submitting any application information to land conservation staff. This could potentially attract more interest in the program, while at the same time, discourage landowners from submitting lower scoring projects.

The new criteria and scoring system has been reviewed and recommended by the FNAP Advisory Committee and will be reviewed by the County Board in May.