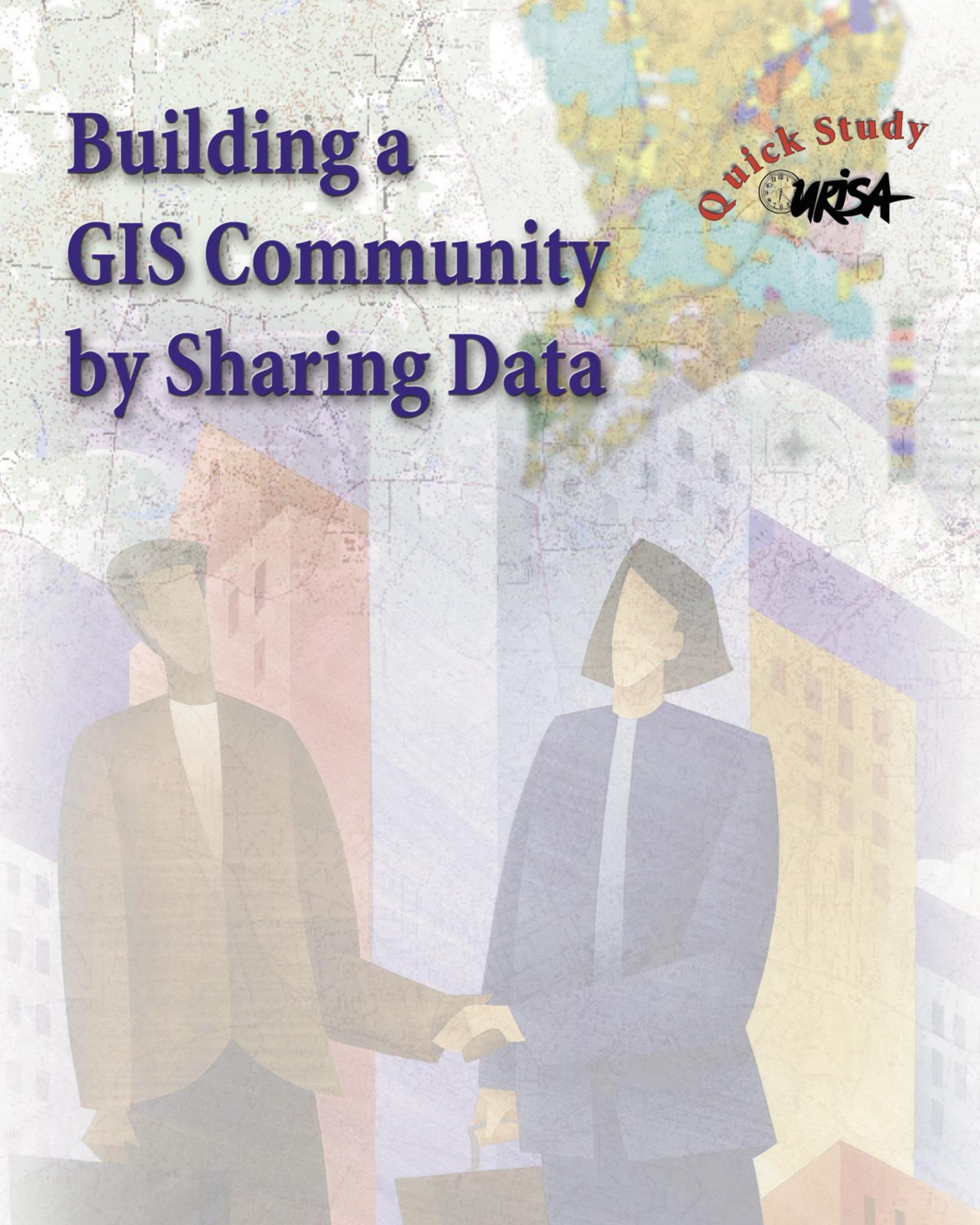
Building a GIS Community by Sharing Data

Quick Study

URISA



BUILDING A GIS COMMUNITY BY SHARING DATA

Paula Gee Davis, P.E.
Mapping Manager
Engineering Information Services
University of North Carolina at Chapel Hill
Phone (919) 962-9038
FAX (919) 962-9103
Email: paula@fac.unc.edu

Janet Jackson
GIS Coordinator
Orange Water and Sewer Authority
400 Jones Ferry Road
PO Box 366
Carrboro, NC 27516-0366
Phone (919) 968-4421 (Ext. 276)
FAX (919) 968-4464
Email: jjackson@owasa.org

Quick Study URISA

The Urban and Regional Information Systems Association
1460 Renaissance Drive, Suite 305
Park Ridge, IL 60068

Printed in the United States
ISBN #: 0-916848-39-6

TABLE OF CONTENTS

ABSTRACT

This Quick Study Guide tells the story of our Geographic Information System (GIS) Community. Within our community, we have a successful and rewarding process for sharing local data. Collaboration enhances our GIS programs and conserves time and resources at multiple levels: the county, towns, university, and public utility. Practical examples demonstrate that individual and collective productivity increases because we share ideas, methods, and, most importantly, data. We explain methods to find available local data, to search for data through the Web, and to use data from long-range planning studies. We offer ideas about updating and storing shared data. We discuss data standards and explore metadata as an extraordinary resource. Our data sharing creates opportunities for sound decision making and for planning community services across multiple organizations and jurisdictions.

INTRODUCTION

In this Quick Study, we describe our Geographic Information System (GIS) Community. This group meets regularly to share experiences, ideas, and data. By sharing data, we are able to answer questions regarding planning and services across multiple organizations and jurisdictions quickly and easily. This study is not about sharing proprietary software, licensing agreements, or purchased data. It is about sharing results, methods, and common objectives to enhance production and give a more complete look for the small or big picture.

The idea of data sharing seems to be such a simple idea; how could it be necessary to write a Quick Study about it? However, data sharing has continued to be a popular presentation at local, regional, and national GIS conventions. So we thought that some of these ideas might help other communities get started, navigate through obstacles, and continue toward success.

The title, "Putting Your Reputation on the Line: Building a GIS Community by Sharing Data," implies a risk in sharing because data can take on a life of its own. However, by building a community of users who understand the local data and its limitations, sharing data is the most cost effective, quickest, and best way to acquire data for a successful GIS program.

CHAPTER 1

Invitation to the Community — Join, Join, Join

Decide that you want to share and then do it!

Our community interest in Geographic Information Systems began about 15 years ago. Clearly, our county, towns, university, and public utility could each use cadastral (lot lines used for taxation), zoning, fire hydrant locations, and jurisdictional boundary information. However, the expense of collection and update is too great for each agency to manage all of the data. The public is unmistakably served by agencies collecting data necessary for its unique responsibilities and managing redundant data as a shared resource. Shared data distributes the cost of collection and updates. So, the concept of sharing data was born of a necessity to reduce expenses and to keep projects moving toward success.

In order to share data and coordinate GIS services, representatives from our county, towns, university, and public utility met to evaluate and purchase the same hardware and software. Actually, this effort is an excellent example of how the best-laid plans can go wrong. Because of budgets, preferred vendors, and state contracts, each jurisdiction purchased different hardware and GIS software. Individual GIS programs were started that were, at best, loosely affiliated. So the dream of easily sharing data was put off for a while.

As was the case with many early GIS users, after the initial purchase of software and hardware, each agency began the process of data collection. This is not an easy task. The county engaged a consultant to enter cadastral data. Some data were digitized and some were surveyed. Some data were required by local ordinances and obtained through permitting processes at the town engineering or planning departments.

Now that Global Positioning System (GPS) has become a popular means of gathering data, all members in our User Group utilize some type of data that has been obtained by the use of GPS. For example, the Orange Water and Sewer Authority recently added 20,000 meter boxes to its GIS data.

Approximately 10 years ago, the engineer for the town of Chapel Hill, North Carolina, began monthly meetings so that area GIS Community professionals could meet, discuss, and exchange ideas. Not data – just ideas! (But it was a start.) Today, this group encompasses a diverse group of GIS users from the police, fire, 911, county land records, county tax assessor, county and town planning, town engineering, local water/sewer utility, and the university. One private consultant regularly attends as does a person from a private university who represents their land resource management. Although there is not extensive representation from private companies, they would be welcomed. By talking about the project goals and objectives, we began to get requests for our data. After a few successful pilot projects, the data sharing became routine.

At first, we shared cadastral data and limited data sets of existing features. Then jurisdictions, street centerlines, and building footprints were shared. Utilities, parks, and bus routes are now available. All of these are existing features that can be captured and confirmed. These features are collected and updated by different agencies and are displayed in conjunction with each other because we use the same coordinate system (more about this in Chapter 2).

Viewing existing features is powerful. However, in the last couple of years, we have seen proposed features become more accessible. The display of proposed data is very powerful because the display of data can shape the planning process. Imagine displaying a sewer replacement project from public utility data in conjunction with repaving of the street from the town data. The road paving could be scheduled immediately following the replacement of the sewer line. Knowledge of proposed data enables project planning among different jurisdictions.

GIS provides a tremendous opportunity to coordinate services with great savings to taxpayers and customers. GIS shows "connectivity," "adjacency," and "neighborhood." Brief examples of these concepts using the sewer line replacement follow.

- "Connectivity" can be understood as a network of interconnecting at-grade streets. When the sewer line is replaced in one street, the use of GIS will show how the traffic can be rerouted. Connectivity makes this rerouting possible either by knowledgeable professionals analyzing a map and selecting the streets that will take the traffic or by coding the streets and letting the computer route the traffic.
- "Adjacency" can be understood as the relationship of features to one another. GIS can show the pipe under the street or the fire hydrant on the north or south side of the street. Adjacency makes it possible to determine which sewer line will be replaced and, consequently, which street will be repaved following the replacement of the sewer line. Just like connectivity, adjacency can by shown on a map by overlaying the sewer lines with the streets. Also like connectivity, GIS programs can be used to define adjacency. A query might be: Which streets fall within the buffered areas of the replaced sewer line? Using this information, whether generated by technicians or a GIS query, the identified streets are repaved.
- "Neighborhood" allows characteristics of a near neighbor to determine values of an unknown. This is very useful in GIS for topographic calculations and other GIS applications that generate surfaces. For the sewer line replacement, the "neighborhood" concept would allow knowledgeable professionals to make assumptions about an unknown pipe size. If upstream and downstream pipes were both 12 inches, a good guess as to the unknown pipe size would also be 12 inches. In a domestic water system, the pressure in the loop can predict water pressure in the replaced domestic waterline or fire flow water for a hydrant.

Hopefully, the above samples show that a GIS does not have to be complicated to be useful. GIS displays data. "Connectivity," "adjacency," and "neighborhood" are concepts that are intuitive in the display of data. Police rerouting traffic is using connectivity. Technicians replacing a sewer line and then repaving a road are using adjacency. Engineers determining a 12-inch pipe lies between two existing 12-inch pipes are using neighborhood.

On the other hand, the above samples also show that a GIS can be more sophisticated. GIS can use coded street centerline data to determine a very complicated rerouting within minutes for an emergency response (connectivity). GIS can quickly give a list of the roads to be replaced once a series of sewer improvements are implemented with a time schedule and cost (adjacency). GIS can be used for complicated models to predict the water pressure anywhere in the system (neighborhood).

The GIS can display both existing and proposed data. Data displaying existing conditions and proposed improvements enables the opportunity for coordinating planning and services.

Our agencies have been successfully exchanging data for several years. We still get together and talk about maps we are making, projects we are tackling, and any information that may be helpful to the group. Generally, between 10 and 15 people attend the meetings – some are regular attendees and some attend when they can.

GIS has formal organizations that are based on the workplace as well as informal organizations that consist of GIS colleagues. Our managers encourage data sharing, so that they know they are seeing the same maps as everyone else. Confidence in the data strengthens confidence that their decisions will be the right ones. If we had a written vision statement in our GIS Community, it could be “sharing data, challenges, and solutions to strengthen our community.”

Another way in which we share data is by using the GIS Community to obtain data through the “economy of scale.” Approximately 5 years ago, the county coordinated aerial photography with planimetrics. The entire county was flown at 1” = 400’. The towns were flown at 1” = 100’. We worked together in proposing features that should be collected and the data dictionary. Since we share the same geographical area, it makes sense to share the same data (Figure 1).

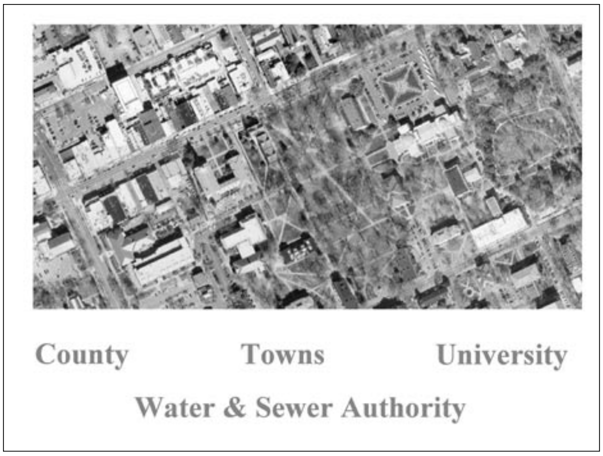


Figure 1. Since we share the same geographical area, it makes sense to share the same data.

This year, the town of Chapel Hill coordinated color aerial photography with planimetrics. This time, the county is receiving 1” = 200’, the towns are receiving 1” = 100’, and the university is receiving 1” = 50’. This coordinated effort saves money for taxpayers and customers.

So by beginning with the end in mind, our Community group has been a tremendous asset throughout the years as projects and objectives have expanded. During the GIS Community meetings, we discuss current projects and the others comment on data that they have that could be applicable. The atmosphere is very open and honest. It takes more than just shared data to make collaboration happen. It takes people, data, software/hardware, and methods all coming together. The following chapters outline how we have combined these five main areas to get the fullest benefit with the least amount of money.

CHAPTER 2

Finding Out Where Data are Available Locally/Globally — Meet, Meet, Meet

In our community GIS, the meetings (which were originally monthly and are now are every other month) are the most important time investment. Initially, we discussed data and projects that we hoped to attain for each of our agencies. Then, as the data sets grew, we brought maps or data to show or exchange. The most important local data sources are our county, towns, university, and utility. To varying degrees, we also obtain data from the private sector consisting of developers, contractors, and planners.

The scale of the data is one of the reasons that local data are so useful to our group. We use large-scale data, which simply means that the ratio of map scale to real world units is a larger fraction. For instance, a contractor may use a scale of 1:240 (1” = 20’). Aerial photography planimetrics may use a scale of 1:1200 (1” = 100’). These are large-scale maps. On a large-scale map, a road would generally consist of two lines delineating the edge of pavement and space between them showing the roadway (Figure 2).

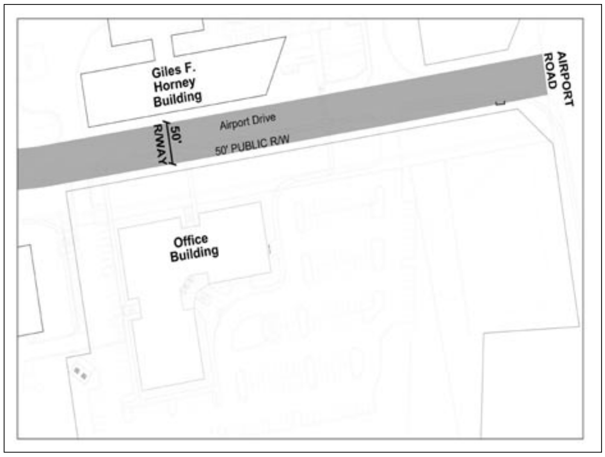


Figure 2. On a large-scale map, a road generally consists of two lines delineating the edge of pavement; the space between them shows the roadway.

The United State Geological Survey (USGS) produces quarterquad maps that are very useful. These maps use a scale of 1:24,000 (1” = 2000’). On a small-scale map, a road would generally be a line (Figure 3). Also, if you had the contractor’s data brought into the scale of the quarterquad maps, it would be a very small 1/100 of the size.

Why is this important? Think about bringing a line into GIS showing a coastline. At a small scale, the coastline would appear relatively smooth. At a large scale, the coastline appears jagged since the pattern

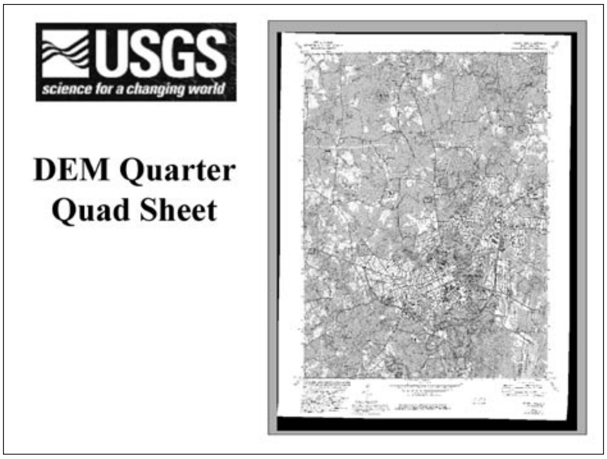


Figure 3. On a small-scale map, a road is generally a line.

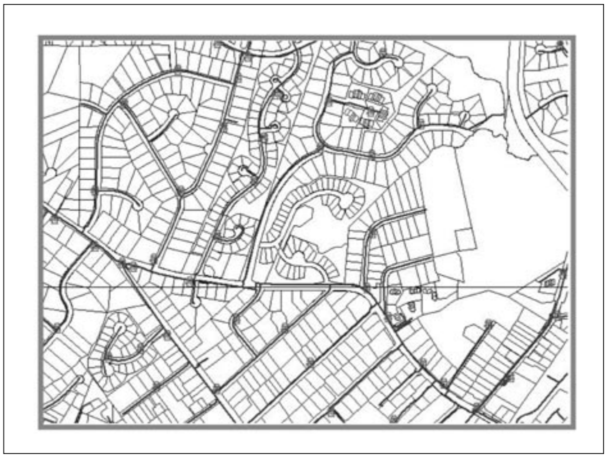


Figure 4. The data have been updated; when the data sets are shown together, the hydrant is on the right side of the street.

of the coastline goes in and out with the jetties and inlets. Large scale picks up details.

Similarly, the accuracy of large-scale data is different from small-scale accuracy. If your hand is close to your eyes, you can see thousands of tiny lines (large scale); if your hand is at arms length, only the pronounced lines are seen (small scale). If we were to draw or map the lines of your hand, large scale would give you one picture and small scale would give you another picture, but both would be a recognizable picture of your hand. The "accuracy" would depend on the way that you chose to view your hand.

Let's continue the analogy of your hand to understand cost in relation to scale and accuracy. Think of your hand up close to your face as large scale, local data. Locally, the tiny lines are important. However, accuracy in recording all of the lines costs money. Also, collecting the location of each tiny line has a time cost as well as a financial cost. Then think about updating such detail as the hand ages or scars. Each change has an additional cost and may not even be possible!

Going back to scale, if we download a small-scale river and compare its location to large-scale property lines bounded by the river, the results are revealing. The river will be a single line and will undoubtedly meander back and forth showing quite different property lines!

So scale, accuracy, and cost are intertwined. Data are collected at many different scales. Accuracy is linked to the scale at which the data are collected. It is true: The greater the scale and the greater the accuracy, the greater the cost.

What is the best scale? What is the best accuracy? In our GIS Community, data are collected at the scale determined by the needs of each agency. The University tries to obtain survey quality data for its buildings, roads, and walks. However, the University traditionally uses the county cadastral data to show its property lines. In our experience, the data move continuously toward improvement.

For instance, a fire hydrant may be close to its actual location. When we look at the fire hydrant in conjunction with the street centerline, both sets of data are misaligned enough so that the hydrant is shown on the wrong side of the street. This hydrant location is managed by the utility, and the street centerline data are managed by the county. Both agencies are continuously improving their data. The street centerline data improve because of new aerial photography, and the hydrant location is improved when its locations are captured by global positioning (GPS). The data are updated and now, when the data sets are shown together, the hydrant is on the correct side of the street (Figure 4). The hydrant location is closer to reality because both data sets have improved.

So far, we have mentioned cadastral or lot line information (used for taxation) from the county. We have mentioned hydrants from the utilities. In GIS, data are generally organized by themes or maps showing similar items. Maps composed of themes or like items are called thematic maps (Figure 5).

Some of the themes collected by our GIS Community could be broken down into the following:

- County – lot lines, soils, geology, wells, road centerlines, rivers, and creeks
- Towns – jurisdictions, zoning, public lands, park and ride lots, land use, land cover, and walks
- University – buildings, parking lots, walks, storm water, and utilities
- Utility – waterlines, sewer lines, meters, hydrants, and reservoirs

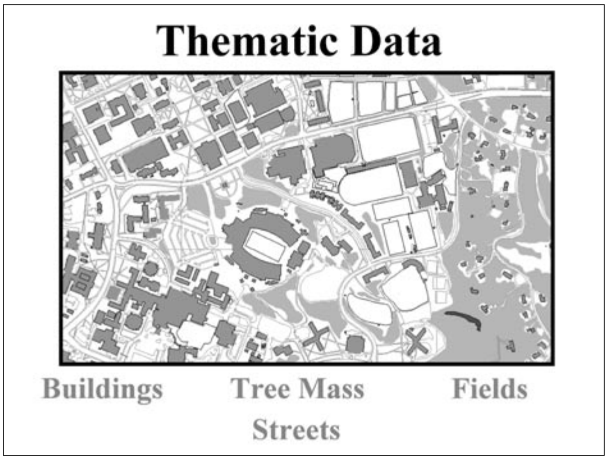


Figure 5. Thematic maps composed of themes or like items.

Data are our most important resource and our biggest investment. The organization of data becomes our GIS. What is the best way to organize a GIS? That depends on the data collected and the organizational processes used to manage that data. The way that the county organizes its data is very different from the way a utility organizes its data. Even when we use the same themes on maps, they look very different because of the boundaries that are important to our organizations and the scales that we use for our maps.

How do we find new data or new applications that might be of use? Our regular GIS Community meetings provide information about existing and new local data sets. We have learned to find, acquire, translate, and use whatever data we can get from wherever we can get it. Why is this important? We have found that data collection is expensive and time consuming.

Even though the Web has changed the way in which we all perform, view our jobs, and present data to the world, the first source of data sharing for our GIS Community is our local organizations. We each bring different skills and perspectives to the community meetings. It is easy to understand, when data are being shared, that we are striving for the "larger picture."

We try to remember that data to answer questions may come from sources we have not yet tapped. Meeting and discussion of projects gives insight into data that may solve problems. We maximize the data investment by sharing our data and ideas with others.

One final note about sharing data: we want to emphasize that we are not in the position of being able to share software licensing and some proprietary purchased data, but we can always share our results. By results we mean digital data that is collected, separate files generated by geoprocessing of data, and database files and printed maps.

CHAPTER 3

Using Standards That Allow Data to be More Universally Shared — Location, Location, Location

Location, location, location? Location, or georeferencing, is the key to Geographic Information Systems. All maps show features in relation to other features. In GIS, the features are located geographically by using a chosen coordinate system. By using the same coordinate system, any other data from any other source will align within the given geographic location. We think of GIS data as having a compass, since the data automatically knows where to go.

Because our local data are georeferenced, it is easier and quicker to share with others. This means that our GIS Community uses data registered to the same coordinate system. (We use North Carolina state plane coordinates, NAD 83, feet). More universally, people speak in terms of latitude/longitude or other coordinate systems. Again, there is tremendous value added to any data that are georeferenced, and georeferenced data can be shifted to other coordinate systems using GIS software.

Georeferencing is a standard intrinsic to GIS. The location of the features in the world is probably the single most important piece of data. Neighborhood, adjacency, and contiguity are based on geographic location and are intrinsic to GIS (discussed in

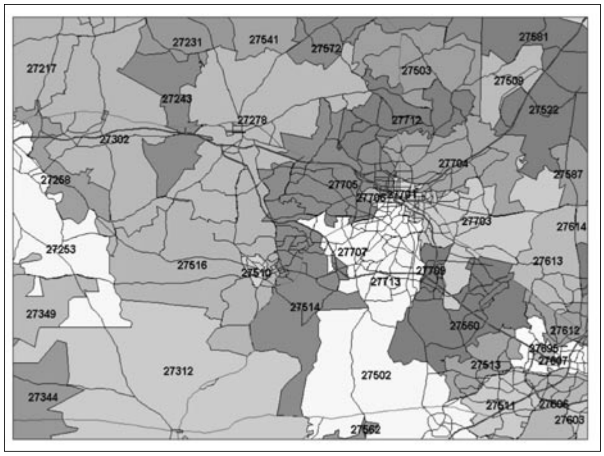


Figure 6. Geocoded data purchased by postal Zip Code zones.

Chapter 1). One result that might emerge by combining geographic data is the ability to see a pattern. Every weather report on television uses Geographic Information Systems to demonstrate complicated weather patterns. We have all watched as hurricanes gain and lose strength. We have all anticipated deep snowstorms in the winter and breaks in summer heat. Just like the information in a weather report, all GIS data have a time and a spatial element. GIS requires looking for patterns in the data and seeing

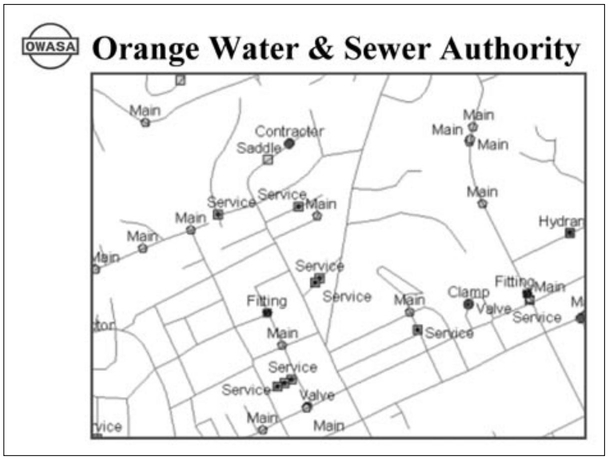


Figure 7. Areas with repeated problems are easily seen by the concentration of points showing pattern in the data.

the manner in which different elements affect the whole system.

Here is a simple example of finding patterns in data from our GIS Community. The water and sewer utility had a collection of 3"x 5" cards of maintenance problems spanning many years. Any data can be considered spatial data if it has an address, and each 3"x 5" card did have an address. The geocoded data were purchased by postal Zip Code zones (Figure 6). By matching the addresses taken from the cards to the geocoded data, each street centerline with an address range indicated where to place the points to match the location of each repair. Consequently, areas that had repeated problems could be easily seen because of the concentration of points in that location. This is an example of pattern in the data (Figure 7). The existing information was more valuable once it was shown geographically.

In addition to the GIS standard of using coordinate systems to organize data, other standards dictate the display of data. In mapping conventions, certain symbols are universally accepted or standard. However, our group found that some symbols were not stable when transferred between certain platforms/programs. We use point, lines, and polygons rather than symbols in data that we share.

Other standards deal with the transfer of data. Over the past few years, mapping data has mushroomed, transfer methods have matured, and import/export features have significantly improved. For instance, the first time the University received lot lines from the county, the data filled 11 5-1/4" floppy disks. Now, because standards are in place, we use CDs and ftp (file transfer protocol) sites. Access is easy and practically transparent to GIS users.

Some of our GIS Community have clear standards for consultants about data that is acceptable. Some prepare a template of information (such as a layered AutoCAD drawing that is ready for the information to be added). Putting such standards in place will make adding the data to GIS projects easier. One of the many challenges in GIS is that hardware and software are constantly evolving so keeping templates up to date may be another challenge.

Our GIS Community User Group meetings have helped provide structure and form to agree about data standards we should apply to use and transfer data. Adhering to a few ground rules about standards has provided a way to use data from many different sources. Some of these standards are intrinsic to GIS (coordinate systems) and some are based on experience (points, lines and polygons rather than symbols). Standards can dictate easy data transfer and are transparent to the GIS user. Other standards are set to help in the exchange of data between our agencies and consultants. In other words, standards help us know what we are getting and help us get what we want.

CHAPTER 4

Obtaining Data Through Short/Long-Range Studies — Plan, Plan, Plan

We mentioned participation of a private consulting firm and a private university in our local GIS Community (Chapter 1). It is true that the role of the private sector in our Community meetings is minimal. However, private consultants play a vital role in the acquisition of GIS data. Consultants for the University have taken the local information and generated GIS data to show existing conditions for storm water runoff. These consultants use GIS models to show proposed conditions, and the GIS information becomes part of the University GIS.

Private consulting firms fly and produce aerial photography and planimetrics for our GIS Community. Many of the GIS buildings, roads, and parking lot plans used by the towns, county, and university to update data come from local surveying and engineering firms. A consulting firm processed the Orange County cadastral data and made the first parcel coverage. Also, the geocoded data discussed in Chapter 3 were obtained from a private firm.

Conversely, the private sector comes to our GIS Community for data to produce their mapping products. Updates to centerline data, jurisdictional boundaries, and addresses are used to update their existing data as well as to correct their data. Since they use the local data, their production is faster and their data are more accurate. It is the interaction of the private and public sector that allows the data sharing efforts and results to expand.

In 1992, the University of North Carolina at Chapel Hill had a tremendous amount of information available, but mostly in hard-copy maps and almost none of it georeferenced. However, during the past 10 years, the University has had two long-range planning studies, one a flood study and one an environmental and storm water study. These studies have greatly benefited from data available through the GIS Community. Access to local data was one key to success. Specific data sets, such as the town of Chapel Hill's land use map, helped the University predict build-out for traffic analysis (Figure 8). This land use map was a resource created and prepared by the town but was readily available for the University to use at no additional expense!

In addition to local data, it is sometimes necessary to purchase data from either private consultants or other governmental agencies. If hard-copy data are available, the digital data may exist for a low cost.

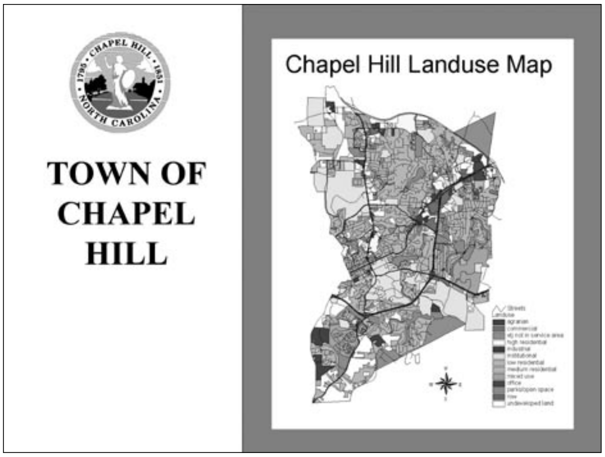


Figure 8. Specific data sets, such as the town of Chapel Hill's land use map, helped the University predict build-out for traffic analysis.

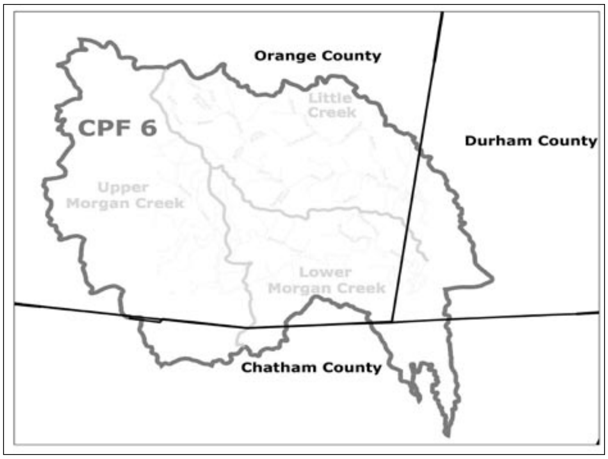


Figure 9. Watershed data have been useful for delineating the perimeters for a proposed local storm water utility. It made sense to look at the storm water within a watershed boundary rather than within a jurisdictional boundary.

Prior to one of the early planning studies, Mylar topological data had been purchased by the University. Since consultants needed digital data, the University contacted the consultant to see whether the digital data were available. This was successful, and data were purchased as a part of the planning study. The University was able to target the areas needed. The cost of the data was low, and delivery was within a week. Once the data were purchased, the University shared it with the county, town, and utility. The data have also been used for subsequent studies as well as by consultants for early site planning.

The University also bought data during the flood study. The flood study was in conjunction with the U.S. Army Corp of Engineers, the State of North Carolina Department of Natural Resources, Orange County, the town of Chapel Hill, the town of Carrboro, and the University of North Carolina. The University obtained watersheds from a state agency (North Carolina Center for Geographic Information and Analysis) and topological data from a federal agency (United States Geological Survey). This flood study provided an opportunity to purchase data with money that would not have otherwise been available.

The watershed data have been useful for delineating the perimeters for a proposed local storm water utility. It made sense to look at the storm water within a watershed boundary rather than within a jurisdictional boundary (Figure 9).

During the environmental study, the University campus was divided into sub-watersheds. The consultant provided the data during the study. The data were useful to analyze spill locations or erosion control violations in conjunction with the university development. In addition, the University now uses watersheds to analyze storm water runoff. When a new building is planned, the entire watershed in which that building is located is analyzed for impact. The University has committed to a zero increase in storm water runoff as part of the University of North Carolina Development Plan.

With a goal of zero runoff, the existing and proposed impervious surfaces had to be analyzed. A map was produced and analyzed by the environmental and storm water consultants. The aerial photography, planimetrics, and survey data were combined to achieve a complete map showing impervious surfaces. The 2003 aerial photography will have a finished product

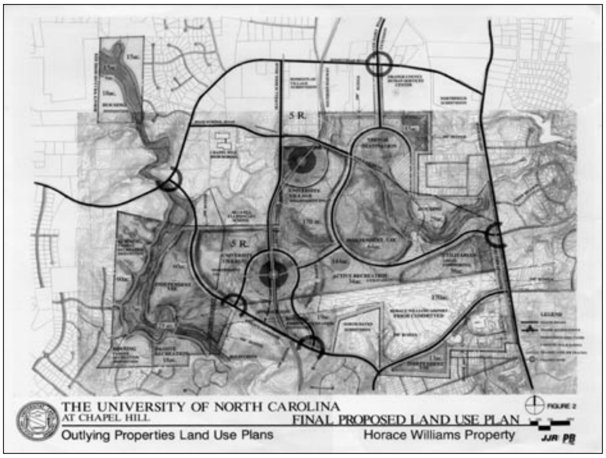


Figure 10. Drawing of the planned development scanned into a raster file.

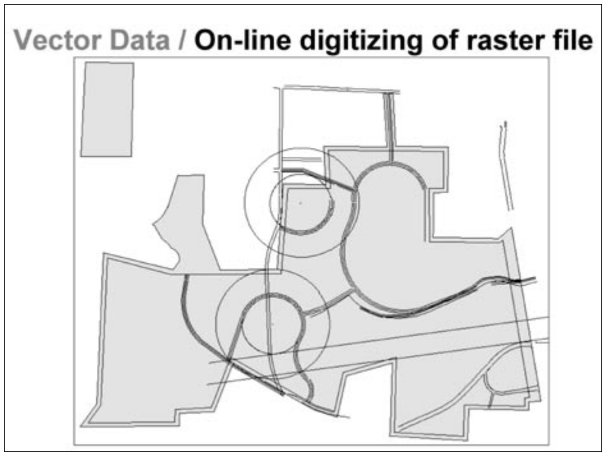


Figure 11. The proposed roads and transit stops were traced from the raster image using heads-up digitizing.

of impervious surfaces for the entire town of Chapel Hill. This information should be available within the next few months.

Our GIS Community is located in an area that is required to hold a Phase II National Pollutant Discharge Elimination System (NPDES) permit. Initially, meetings were held to discuss obtaining a joint permit. In the end, each jurisdiction (as well as the University) decided to obtain its own permit but to combine things such as educational outreach when possible. Watersheds, streams, topography, and the university storm water inventory have all been useful in creating maps and meeting objectives for the University Phase II NPDES permit.

Generally, at the end of the study, an analysis of the area or a long-range plan of the area goes back to increase the pool of GIS or digital information. At the end of the flood study, we hope to have new 100-year flood boundaries. (This would be very useful due to our hurricane-prone location and the amount of upstream development potentially contributing to flood conditions.) From the environmental study, the university has already obtained sub-watersheds and impervious surface.

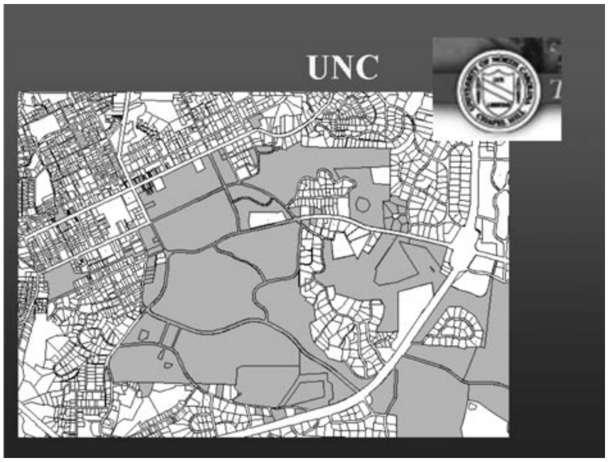


Figure 12. At the University, lot lines from the cadastral data were combined to show larger property holdings.

From planning studies, the University of North Carolina acquired raster (or picture data) and vector (or point/line) data. A planning study at the University combined the raster and vector data so that they could be used for future GIS projects. First, the drawing of the planned development was scanned into a raster file (Figure 10). The raster file was geo-referenced using ArcInfo. Then, the proposed roads and transit stops were traced from the raster image using heads-up digitizing (Figure 11).

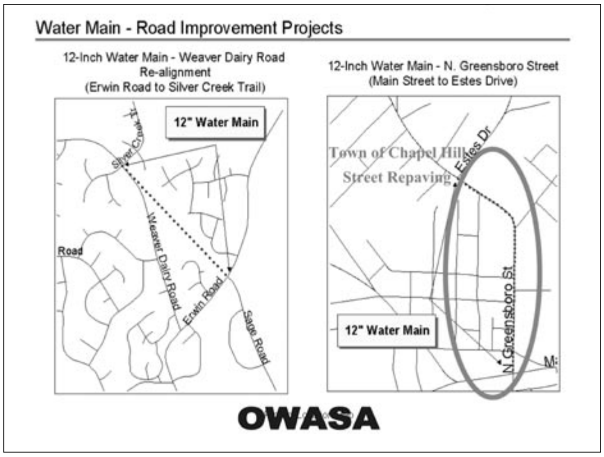


Figure 13. The town of Chapel Hill schedules road maintenance in conjunction with sewer replacement so that citizen disruption is kept to a minimum.

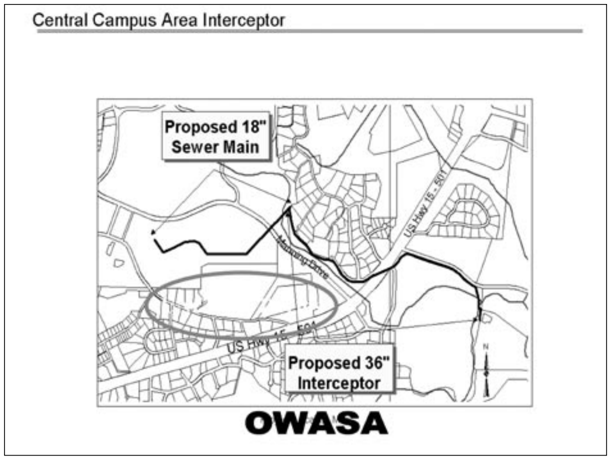


Figure 14. Construction of several new residence halls at the University will need possible sewer line upgrades.

Once there are data, more data can be created from it. At the University, lot lines from the cadastral data were combined to show larger property holdings (Figure 12). This method is used by the University to show all of its property lines, since a boundary survey is unavailable.

A typical shorter term planning study such as Capital Improvement Project by the University, towns, or the Water and Sewer Authority have been an invaluable resource for scheduling specific projects. This is another example of how sharing important data sets benefits each organization. For instance, combining data shared by the individual organizations, will benefit the town of Chapel Hill by scheduling road maintenance in conjunction with sewer replacement so that citizen disruption is kept to a minimum (Figure 13).

As mentioned previously, the University has a development plan that shows construction projects over the next 8 years. In one area, the construction of several new residence halls at the University will need possible sewer line upgrades (Figure 14). This is another productive use of combined data sets.

Some readers may have long-term studies underway, such as a 10-year plan or Capital Improvement Projects. The specific data sets that support such studies are a great way to increase the information base for the community. When data sets are acquired in this way, hopefully, the money to fund the data acquisition or projects will come out of someone else's budget. Once the study is completed, our GIS Community tries to think of ways to bring that data back into the GIS. As we have shown, proposed data can be as useful as existing data. Also, data acquisition is an excellent opportunity to include the private sector into our GIS program.

CHAPTER 5

Updating and Storing Shared Data — Improve, Improve, Improve

Updating and storing shared data are at best a difficult task due to the number of ways it can be done. But not everyone will understand the how and why of "your" system, so be sure to set it up in such a way that is easy to remember and able to be understood by others.

In our GIS Community, any errors found are reported and updated by the creator of the data. By updating and correcting the original data set, everyone benefits when the data are reissued. If managers see one set of data that says one thing and data from another agency says something else, they lose confidence in the data and the data-sharing process (Figure 15).

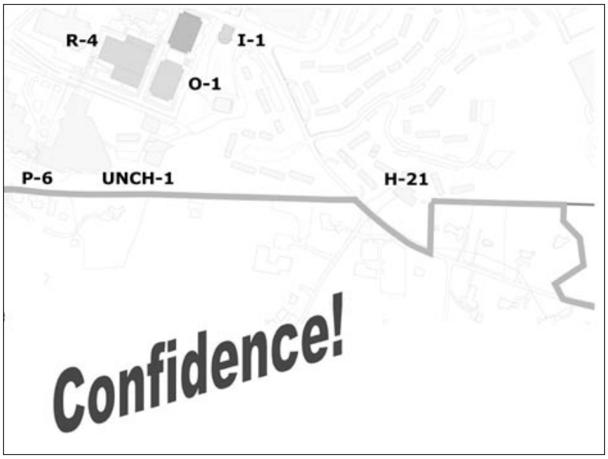


Figure 15. If managers see a set of data that says one thing and another set that says something else, they lose confidence in the data and the data-sharing process.

The University set up a source data folder separate from other data sets. Changes are made on copies of data instead of the original. If the original data need to be corrected, the source of the data is notified. Survey quality data are kept in a separate folder and labeled as such, including the date it was completed. This is a helpful method if metadata is written for this folder at a later time.

Layer-naming schemes can be as unique as they need to be to communicate the type of data that they contain (Figure 16). A naming scheme used by the university is to divide the incoming data into logical groups. CN layers stand for "construction," and NS layers for "not surveyed." Once it is surveyed, the information is entered into its final group and the prefixes are dropped. Another layer has recently been added. FL stands for "flyover" (that data was generated from aerial photography).

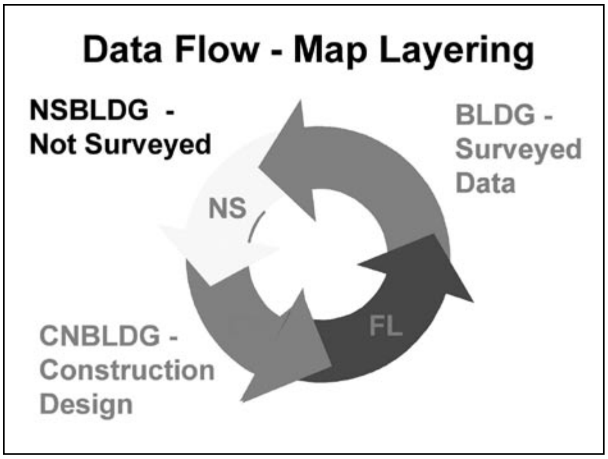


Figure 16. Layer-naming schemes can be as unique as they need to be in order to communicate the type of data that they contain.

Some data sets can be quite large because the information can cover a very large area or have many attributes attached to the graphics. Keeping a separate compact disk (CD) of these data sets, one for back up and one for archival, can be most helpful should something happen to the original file.

Routine updates are handled and reissued by the agency that created the data. For instance, Orange County Land Records continues to update the parcel data set that is used by everyone in the county and disperses a CD to all parties in our GIS Community every month. If we have any questions about a particular area (parcel), we know who to contact for the information.

Another good example of routine updates is provided by our Water And Sewer Utility. New features or attributes are added or changed within the water and sewer data set. Updates are dispersed, via CD, to the parties that need the data within our GIS Community every 6 months.

Within the last 6 months, the University police have switched to software that is used by many law enforcement agencies in North Carolina. This software has a feature that enables it to audit the County centerline data for errors in direction, range, character length, and bad characters. These errors were sent to Orange County Land Records, and they have corrected their data in order to make it work with the law enforcement software. This took days of work, but they did it to make their data better and so that improvements to the data set would be beneficial to all.

For police accreditation, the University has used GIS to locate student housing (including sororities and fraternities) and log crime events within 2 miles. Of course, since some of the crimes were not committed on University property, the Chapel Hill police worked together with the University of North Carolina police to compile the necessary records.

The Orange County Cadastral data are used by the University to track property sales within a 200-foot buffer of campus. As a part of the current University development plan, the University tracts the data to see if housing values are impacted positively or negatively adjacent to the central campus.

At one point, a town-gown committee was formed to study the monetary value of campus property (since there is not a tax-assessed value). GIS was used to evaluate the tax-assessed value per square foot by zoning type. University property was then separated according to zoning type and shown with similar zoning with the tax-assessed value per square foot. This information was given to the town-gown committee for assignment of an approximate value based on the similar property.

These examples show how our community GIS has used data to illustrate the impact to the community. Social impact can be tracked and displayed using GIS. Census data will provide opportunities to locate facilities that are important to at-risk populations and to plan for community needs. Where should our resources be distributed? Playgrounds, ball fields, green space, hospitals, fire stations, and infrastructure improvements can be located equitably. GIS can provide the tool to help community decision makers take the best action.

In conclusion, we always try to get back to the originator of the data to make changes. Changes are sent to the originator of the data so that the source data is corrected, and new copies are distributed to users. This is true for graphical as well as tabular data. Also, we think that it is important to keep source data separate. As a user, we can always go back to the original source and verify that the data are correct. We use whatever naming conventions, layering systems, or data management method help keep the data understandable. Finally, we try to set up interagency verbal agreements for routine updates.

CHAPTER 6

Buying Data across Different Jurisdictions and Different Scales — Buy, Buy, Buy

Buying data from sources outside your GIS Community group may be necessary, but take a good look around at what is available, check out different vendors, and finally ask a lot of questions about the data set before making the final purchase. This is another way that private companies provide a valuable service to our group.

We all acknowledge that sometimes vendor data is such a bargain, you just have to buy it! Here are some shopping ideas that we have used:

- State agencies
- United States Geographic Society
- World Wide Web
- Private companies

Another approach is to get creative and combine them.

Some data are vector data. Vector data consists of points and lines that can be moved or altered as individual entities. Some data are raster data. A raster is simply a picture (Chapter 4).

The USGS supplies a wide variety and scale of data sets. They have data sets available at a 1:100,000 scale that can be downloaded at no cost directly from the World Wide Web. For a recent University project, digital line graphs of five-quarterquad sheets surrounding the University property were purchased from the USGS Website and were delivered on CD (Figure 3 shows a raster image of a USGS quarterquad map). The USGS does not prohibit the sharing of data with others. Sometimes, the USGS digital data might already be in one of your university departments and you could get a copy of it for no charge!

When processing data purchased or received from a source other than your project team, it is imperative to keep good notes. The method used to process it may become important later on, and it is tough to remember all the steps, especially 6 months to 1 year later.

Remember the issues of scale, accuracy, and money that we mentioned in Chapter 2? When we use or process data, right from the start, we try to be aware of its quality. Metadata (data about data) is the supporting document that will tell the details needed for quality assurance. The topic of metadata will be mentioned in more detail in Chapter 8.

Many states have data clearinghouses. In North Carolina, the Center for Geographic Information

Figure 17. BasinPro shows statewide coverage highlighting the watersheds of North Carolina

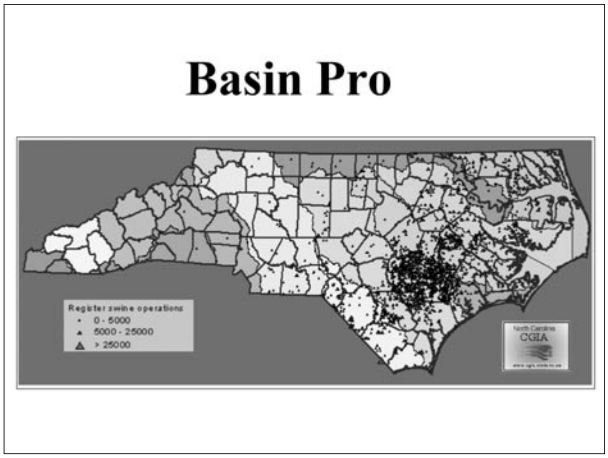


Figure 18. North Carolina planners are concerned for two of the watersheds that are likely sites for hurricanes and have a high density of swine farms.

and Analysis (NCCGIA) is a good data clearinghouse agency. The NCCGIA has a site where data are available quickly and at very reasonable rates. This agency prepares data by clipping the area that is needed or moving the data into a different coordinate system. Therefore, the data are ready to use. Be sure and ask the agency if their data can be shared with other GIS professionals before you pass it along. The NCCGIA requested that we not share copies of their data – but we can share the results derived from the data. It is important to be aware of all agreements and restrictions of data obtained from others.

Another product available through the NCCGIA is a CD with statewide data, called BasinPro. These data sets provide a digital set of information – statewide – that would be near impossible for us to collect at the local level.

BasinPro provides statewide coverage showing the counties of North Carolina, watersheds, major roads, and rivers (Figure 17). There are many views preset to show National Pollutant Discharge Elimination System (NPDES) discharge sites, voter districts, and swine farm locations. You can also add additional data, turn layers off and on, and change colors and labels customized to your project. North Carolina planners are concerned for two of the watersheds that are likely sites for hurricanes and have a high density of swine farms (Figure 18).

We mentioned geocoded data sets in Chapter 3. The geocoded data were purchased to use in conjunction with point data that contained addresses. Any data that have an address can be shown geographically as a point theme to reveal a pattern.

Free data are available on the Web. Generally, the data are of a small scale; therefore, based on discussion in Chapter 2 about combining large- and small-scale data, the user must be careful.

If buying data, ask the vendor questions before you make the purchase. There are definitely bargains out there. Once you buy data, ask if you can share it.

Within the University environment, sharing data is facilitated by the University Intranet. The data are available for all for student GIS projects. Even data that have been purchased can sometimes be shared for student projects based on permission of the vendor.

CHAPTER 7

Storing Data at a Central Location for University/Community Projects — Warehouse, Warehouse, Warehouse

A great example of how our GIS Community User Group has been instrumental in data sharing is to describe the central database for addressing that has been created as a cooperative effort. The central database idea began about 2 years ago. Each jurisdiction needed one place that could house and update their address data as needed. The idea was then slowly developed and tested by all members of our User Group with each phase of the project. Each agency gave valuable suggestions about the final data structure, how access to the data should be done, and specific requirements needed to update the data.

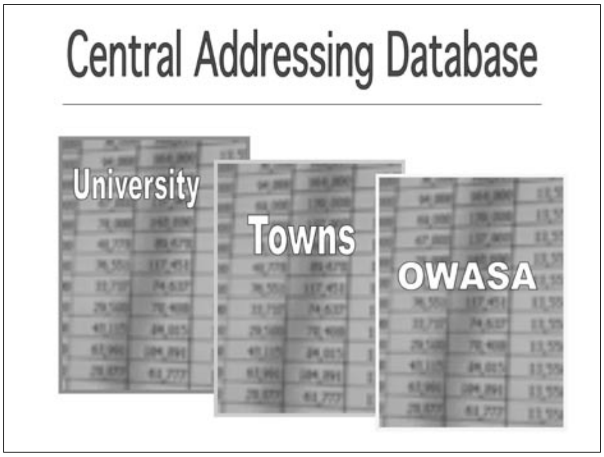


Figure 19. Each user can view and update addresses in specified jurisdictions directly through the secured Internet site.

After a few months of final adjustments, the final version was released to GIS professionals throughout the County. Today, there is a live link via a secured Internet site where each can view and update addresses in our own jurisdictions (Figure 19). This project is successful.

In the near future we are hopeful that, in conjunction with the database, we will have buildings or lots linked to each address. At the present time, only the property lots are linked to the central database. However, we are not stopping with these improvements. The county is about to launch an address-verification program that will ultimately be used to correct any information currently in the address database that is incorrect or missing.

The University is a good example of the importance of having data sets in one common place. The University stores its data so that the data are available for student use. A type of clearinghouse was set up so that only one person is notified when a student needs a specific data set. If a student requests data from the county, towns, or utility, the data are provided to the University contact and posted on the University Intranet. This increases efficiency within our GIS Community. Multiple data sets for student projects have been eliminated. Specifically, the campus computer department has allowed for 8 gigabytes of space to store local data for student projects. Any local data are posted to the site so that the data are available to all students/faculty/staff all the time. Within the past year, a GIS librarian was hired and will now manage this site.

Someday we hope to set up such an Intranet site for the GIS Community. All of the data will be available in one site. The long-range plan is to store all local data in one central location. Data would be online and would be the most up-to-date.

Metadata would define the quality and content. Rather than supplying customized products, the data are available. The accompanying metadata will allow the user to have all the information he or she needs to use the data in the most appropriate manner.

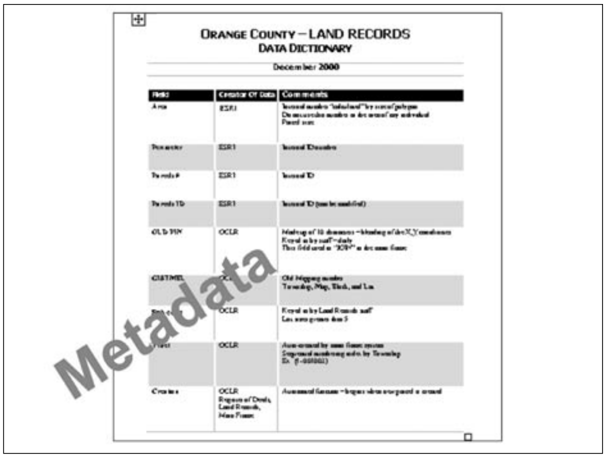


Figure 20. Once metadata was created to accompany county parcel data, the data became much more valuable to everyone.

CHAPTER 8

Producing Metadata so that Ordinary Data Turns into an Extraordinary Resource — Metadata, Metadata, Metadata

You most likely have heard the phrase "The software/hardware that you use may come and go, but your data will go on forever." And it will. Hopefully, your data will last and be used by you, your organization, and by others, and that is just what you had intended when you so painstakingly collected it. However, in order for the data to go on and on, a record needs to be kept of where it originated, who created it, and how it was maintained along the way; that record is called metadata – data about your data.

Metadata is especially important when you have a central data clearinghouse because most often the person who will be cataloging and dispersing the data does not know any of the critical details about it. Even if you have to create word documents for each drawing or data layer that describes the origin and processing of the information, a word document is better than nothing. The essential information about scale, accuracy, and ownership need to be passed along with your data in order for it to have real value.

Some feel that metadata is a luxury to have and never quite seem to find the time to produce. But data that you receive, not created by you, can be confusing or misleading to use if you do not know who, what, why, when, where, and how it was created. A good example is data received from various consulting firms. Most consultants do not send a detailed metadata sheet showing their layering and symbol libraries, and their drawings are often not georeferenced. How will you know where to file it and what it represents without the metadata? Ultimately, the solution is to work on design and data standards for individual and collaborative projects.

A good example of a collaborative project was when metadata was needed to accompany county parcel data. As is often the case, the county parcel data are the culmination of many different divisions contributing their data to make the final database complete. The county responded quickly to a request, and a GIS volunteer in the community generated metadata for the county parcel data. Once metadata was created to accompany the county parcel data, the data became much more valuable to everyone (Figure 20). Sometimes, if you need the metadata, you will have to do your best to create it, even if the data are not your data. Similarly, metadata was created to go with the town zoning information because it was needed for consultants.

There are several commercial programs to enable the process of creating metadata. Our goal is to create data that is compliant with the Federal Government Data Compliant (FGDC) standards. Actually there are many ways to create metadata. NCCGIA creates a .txt file containing metadata for their data. In Basin-Pro, you select an "M" button on any view and the metadata appears.

Metadata gives your data value – it is definitely "value added." The only hope for data to be used appropriately is by providing information about the source and quality of the data. The format of the metadata produced by some agencies ranges from simple text files or word documents to Federal Geographic Data Committee-compliant documents. Anything is better than nothing.

In a GIS project, the biggest source of error may come from not having time or personnel to verify the sources of data. Metadata is a great asset if it is available. Also, the data itself should also be at least spot checked for accuracy.

GIS has the ability to answer questions that would have taken much longer or would have been impossible manually. The viability of GIS to answer such questions correctly will be dependent on the processes to define, gather, and correct the data in a timely manner.

For data that are dependent on several sources to reach an answer, a team should be formed to include jurisdictions with a stake in the data. Orange County, the town of Chapel Hill, the town of Carrboro, and the University should each have a GIS representative for quality assurance. Each jurisdiction must be involved to the extent that the data they contribute is the most current and the process to extract the data meets standards on which the team agrees. Our GIS Community does not have a process like this now, but maybe this will be a way in which we assure data quality in the future.

Many GIS projects are conducted by one person in isolation without the checks and verification of the other jurisdictions. On large-scale project with results that may affect many people, additional checks are necessary. Just as a team is assembled to acquire data, a team should be used to verify that the data are correct and used appropriately.

CONCLUSION

Our vision is to have a community data warehouse for our local data sets. It is hoped that the Web will hold the data and that the data will be easy to access but very secure. It has been a long climb and we are not there yet, but we are close.

We are working toward using the Federal Data Standards for metadata. Data would be located by using key words (probably in xml) to match the Federal Data Standards.

Buying data can be helpful in providing a larger area context. The USGS and some state agencies have excellent data. Private companies also provide GIS services and use our shared data. We have shown that buying geocoded street centerline data has been very useful. Any data with addresses can be matched to the geocoded data so that a point is positioned at every address. This has been useful to display patterns for maintenance calls by the utility company and for police calls.

If data are created by others, it must be updated by that agency rather than changing the data of others. Source data are kept separately as a record document.

Data beget data. Lot lines from cadastral data can be combined to show larger property holdings if boundary surveys are not available.

When studies are required, planning data can be extremely valuable to the GIS community. More and more, proposed data are quite useful for planning interagency coordination. Studies also provide opportunities to purchase data with other funding.

Standards have been an asset to our GIS Community. Standards promote the easy transfer of data and help consultants provide data in a format such as digital templates that are useful to the end user. Also, layering standards or naming schemes are useful for the smooth transfer of data.

Collaborative projects such as county-wide aerial photography use economy of scale, and taxpayers and customers benefit. Another example of collaborative projects is the master addressing database project. This project has been a great tool for updating addresses for the county, towns, and university.

Remember that data to help with projects may come from sources that have yet to be tapped. Meeting with the local GIS community provides a forum to discuss data that is available, projects in the making, and coordination among agencies.

ACKNOWLEDGEMENTS

We wish to thank a number of people for support and data. First, our GIS colleagues and friends who have been sources that we tapped to solve problems and to ask to share their data readily: Chris Dwinell, GIS Applications Developer, Land Records, Orange County, NC; Deborah Fredricks Squires, GIS Technician, Chapel Hill, NC; and Ruth Heaton, GIS Specialist, Carrboro, NC.

In addition, other groups have been quite helpful: National Collegiate CAD Conference (NCCC) North Carolina Professional Engineers (NCPE) Public Utilities, University of North Carolina Geography and Information Science Departments, University of North Carolina, Facilities Planning and Construction, Engineering Information Services, NCCGIA, City of Durham, North Carolina, and Chatham County.

ABOUT THE AUTHORS

Paula Gee Davis is Mapping Manager at the University of North Carolina at Chapel Hill for about 10 years. She is a Professional Engineer.

Janet Jackson is GIS Coordinator at Orange Water and Sewer Authority. She is a recent graduate of the University of North Carolina at Chapel Hill and the 2003 recipient of the McNally Award for Excellence in Geography.

Ms. Davis and Ms. Jackson have made presentations on "How to Get What You Don't Pay For: Sharing Data" and "Give Them an Inch, and They Will Map a Mile."

REFERENCES

Parr, Daniel M., GIS Glossary of Terms. Park Ridge: Urban and Regional Information Systems Association (URISA), 2000.