GIS DEVELOPMENT GUIDE

Volume II

Table of Contents

SURVEY OF AVAILABLE DATA

EVALUATING GIS HARDWARE AND SOFTWARE

DATABASE PLANNING AND DESIGN

Figures

Table of Contents cont'd

DATABASE CONSTRUCTION

PILOT STUDIES AND BENCHMARK TESTS

GIS DEVELOPMENT GUIDE: SURVEY OF AVAILABLE DATA

1 INTRODUCTION

One of the most important elements of developing a GIS is finding and utilizing the appropriate data. The form of the data is critical to the overall database design and the success of the analyses performed with the system. The quality of the results produced from GIS analyses and applications ultimately resides in the quality of the data used. GIS data can be obtained in various formats from many different sources. Application requirements based upon quality, scale and level of completeness will depend upon the needs of the application. Once data requirements are developed, there are usually a plethora of data options which the potential user can choose from. Some of these choices will include whether to utilize government- or privately-developed data, cost in this case will be a major difference. Other choices may involve data currency, scale, accuracy, and depending upon the application, the data structure, platform specifications or even media format.

This guideline will discuss various information surrounding available GIS data including evaluating data requirements, various types and sources of available GIS data, potential datasets. This guideline will also discuss potential opportunities for data sharing.

2 DATA REQUIRED

Master Data List (from Needs Assessment)

One of the products available from a Needs Assessment is a Master Data List. Based upon descriptions of the tasks future GIS users will want to perform, a listing of the various required data is developed.

From the Needs Assessment you will have identified:

- the data entities
- the attributes associated with the entities

The Master Data list is used to prepare a database plan which includes:

- a logical/physical design of the GIS database
- procedures for building the GIS database
- procedures for managing and maintaining the database

In this guide, the procedures for identifying and documenting existing data will be described.

3 POTENTIAL SOURCES OF DATA

Types of Data

There are many different types of data which can be utilized by a GIS system. Each data type has its own unique properties and potential for contributing to the overall quality and functionality of the GIS database. These various data types are mapped data, tabular data listings, remotely sensed imagery, and scanned images. The following sections describe these data types.

Mapped Data/Map Series

Mapped data may refer to published maps found in an existing map series or collection. These maps should be logically classified based upon their data content (e.g., topographical, hydrological data). Maps which meet National Map Accuracy Standards are usually produced by federal or state government agencies. Paper maps, if not already in digital format, can be utilized in developing the database through vector tablet digitizing or scanning.

Mapped data can also be identified as geographic data which has been digitized into the vector data structure. Vector map data may be found with or without real-world coordinate information and may or may not have topological relationships. Many organizations which digitized their map data in the past, did so utilizing CAD (computer aided drafting), and thus were not able to establish topological relationships between their spatial elements. Today, there exists software which allows CAD data to be quickly converted into topologically correct geographic data which can then be assigned coordinate data within a GIS. Many alternative sources of digital spatial data thus exist, in addition to the volumes of topologically correct geographic data available from local, state and federal governments.

Attribute Tables or Lists

A readily available form of GIS input, data tables and listings are available from many different organizations and government agencies. Various data tables can be obtained as GIS input to provide additional attributes which will be associated with spatial data elements. These elements are easily linked using primary relationship keys. Database, spreadsheet or ASCII-delimited text tables include some of the various import formats available in many GIS systems. Any organization that maintains a database, or uses spreadsheets to organize their records is able to create digital listings. Tables and lists are available from almost any government organization as long as the data does not involve privacy issues which would impede accessing such data.

Image Data (Remotely Sensed Images, Aerial Photos)

Image data is an excellent source of GIS input data. It mainly consists of remotely sensed images which includes both aerial photographs (in analog or digital format) and satellite images. Aerial photos are normally captured with analog cameras. These cameras produce photographs whose data can be very important in a GIS system. Photographs, though not digital, can be digitized by using a vector digitizing tablet, or they can be scanned, and then input into the GIS as an image. In either case, the digital version will normally require rectification and re-scaling in order to correct camera distortions common with most aerial photography.

Until they are converted into a raster GIS format, basic raster images such as satellite imagery or scanned aerial photographs do not offer any topological connectivity or potential for GIS analysis. Satellite imagery is captured in raster digital format. With the advent of an open display architecture, many GIS packages are able to integrate both raster and vector data into the same display. Remotely-sensed image data is useful within an editing environment for display as a backdrop for both heads-up digitizing and updating of vector layers, for verification, or for conversion into raster GIS layers and then subsequently into vector data layers.

Most remote sensing cameras allow for the capture of infrared images, separating different light waves into varying band-widths which together and/or alone may show much more information than a normal camera reading only in the visible spectrum. Most GIS will allow for the display of these images and will allow for the assigning of different colors to the various bands for the effective display of the data. GIS packages today also allow for the processing of these images in order to rectify, warp, and geo-reference the imagery as necessary so that they will be useful as scaled images. After such procedures, geo-referenced images can be overlaid with similarly geo-referenced vector imagery for effective display.

Scanned Images (Pictures, Diagrams)

Scanned raster images are able to be displayed in a GIS the same way that satellite images are displayed. Any raster image, whether it be a scanned map, photograph or diagram, can be easily input into a GIS for display purposes. Integrating scanned images into a GIS display, or converting raster data into raster GIS format are fairly routine capabilities for most high-end GIS packages. As discussed earlier, a GIS allows for the assignment of coordinates to raster image data.

Scanned maps (as opposed to digitized vector representations) can be effective backgrounds upon which other GIS vector layers can be displayed. Scanned maps usually contain much valuable annotation which would be very time-consuming to duplicate in a vector environment. Including raster images allows for the enhancing of an application by providing the user with visual display data which can enhance the user's understanding of the data. Scanned photographs are especially effective. In many GIS packages, links can be established between an image viewer, which displays scanned images, and vector geographic features so that when an event sequence is initiated (e.g. selecting a vector feature), the raster image viewer appears with the specified scanned image.

Formats

There are three major formats in which GIS-usable data can be obtained. They include hardcopy/eye-readable format, analog image format, and in fully digital format. Unique types of information can be accessed from each of these data formats.

Hardcopy (Paper, Linen Or Film)/Eye-Readable

Hardcopy maps are easily accessed from a wide variety of organizations. Hardcopy maps, as a form of GIS source data, can be digitized on a digitizing tablet into vector GIS format, or scanned and then converted into raster GIS format. Although there are potential accuracy problems which are associated with paper and linen maps (related to distortions due to shrinkage/expansion of the media) in capturing geographic features, there is still much unique geographic data which can only be found on these maps. An example of unique data from paper or linen maps is seen when seeking geographic data for a certain time period. Much of the digital data which is readily available may only be the most current, updated data for a region. For example, in order to find geographic data from before 1970, the only choice may be to access a paper or linen map. Use a film copy of the source document where available as this will be the most stable media.

Accessing dated tabular information for the development of an attribute database may be a similar endeavor requiring the use of paper documents. Organizations which have been in existence since before the dawn of digital filing systems all had to keep their data in paper "hard-copy" format at one time. Some of these older records may have been converted into digital form at one point. In other cases, there may be hard-copy documents which are the only versions of dated material. In order to conserve space and the integrity of most documents, many might possibly have been copied onto microfiche.

Image (Picture)

Aerial photography is found to be an abundant geographic data form. Photogrammetry (aerial mapping) is a common way of creating an accurate and up-to-date land base. Aerial photos provide the raw data which is necessary for various planimetric and topographic mapping applications. Photographic images are a very rich data source in that many geographic features can be seen clearly on a photograph but may not be seen in a paper map or a vector digital file (e.g., a large clearing within a wooded area would not be differentiated on most paper maps, but it is clearly visible on the aerial photo).

Aerial photography is available from many sources (i.e.: USGS, DOT, County agencies, etc.) The federal government has recently developed the National Aerial Photography Program (NAPP) in which states that desire to have their counties flown may split the cost with the Federal government. Many useful products are derived from the NAPP including 1:12,000 hard or soft-copy orthophotographs. An orthophoto is a scanned aerial photograph which has been digitally rectified using control points and a digital elevation model. The digital versions are especially useful for GIS applications. If the type of digital aerial photography needed is not available, organizations can create a request for proposal to solicit bids for aerial mapping, although this can be very expensive.

Digital

Within the digital format genre, there are many different varieties of data available. These various options are becoming as numerous as what is currently available in paper maps. In terms of map graphics, there are again two different data structures which are quickly integrated into today's GIS systems: these are raster and vector data formats. Tabular data can be found in digital data format most frequently. Various forms of digital spatial data which are currently available in raster format may include some of the following:

- Scanned maps and aerial photography
- Satellite Imagery
- Digital Orthophotography
- Digital Elevation Models

Some of the various forms of digital spatial data which are currently available in vector format may include some of the following:

- Topological vector linework
- Non-topological vector linework
- Annotation layers

Some of the various forms of digital attribute data which can be input into a GIS includes file types associated with various software components: spreadsheet, database and word-processing. Some of the file formats which can be utilized include: dBase, Excel, and ASCII delimited text.

Government Sources

Government is the largest single source of geographic data. Data for most any GIS application can be obtained through federal, state, or local governments. Various data formats, whether paper, image or digital, can all be obtained through government resources. The following subsections give basic descriptions of the datasets which are available through some federal, state and regional/local government agencies.

Federal Data Sources

The federal government is an excellent source of geographic data. Two of the largest spatial databases which are national in coverage include the US Geological Survey's DLG (Digital Line Graph) database, and the US Census Bureau's TIGER (Topologically Integrated Geographic Encoding and Referencing) database. Both systems contain vector data with point, line and area cartographic map features, and also have attribute data associated with these features. The TIGER database is particularly useful in that its attribute data also contains census demographic data which is associated with block groups and census tracts. This data is readily used today in a variety of analysis applications. Many companies have refined various government datasets, including TIGER, and these datasets offer enhancements in their attribute characteristics, which increases the utility of the data. Unfortunately, problems associated with the positional accuracy of these datasets usually remain as these are much more difficult to resolve. Satellite and digital orthophoto imagery, raster GIS datasets, and tabular datasets are also available from various data producing companies and government agencies.

The following information on federal agencies was taken from the Manual of Federal Geographic Data Products developed by the Federal Geographic Data Committee (FGDC). To contact the FGDC:

Federal Geographic Data Committee Secretariat
US Geologic Survey
590 National Center
Reston, VA 22092
Phone 703-648-4533

The departments all have different agencies and bureaus within them which offer various listings on the types of data which are available (e.g. concerning data structure, scale, software export format, source data, currency, what applications the data can be used for), and from which agencies they can be acquired. The reader is encouraged to consult this manual for further information regarding the geographic data products related to these organizations.

DEPARTMENT OF AGRICULTURE
The Agriculture Stabilization & Conservation Service: **R**
Forest Service: **B, H, L, Sur, T**
Soil Conservation Service: **H, Sub, Sur**

DEPARTMENT OF COMMERCE
Bureau of the Census: **B, S, H, Sur**
Bureau of Economic Analysis: **B, S**
National Environmental Satellite Data & Info. Service: **A, Ged, Gep, H, R, Sub, Sur, T**
National Ocean Service: **Ged, H, R, Sub, Sur, T**
National Weather Service: **A, R, T**

DEPARTMENT OF DEFENSE
Defense Mapping Agency: **B, H, Sur, T**

DEPARTMENT OF HEALTH & HUMAN SERVICES
Centers for Disease Control: B, S

DEPARTMENT OF THE INTERIOR
Bureau of Land Management: **B, H, L, R**
Bureau of Mines: **Sub**
Bureau of Reclamation: **H, Sur**
Minerals Management Service: **B, H, L**
National Park Service: **B, H, Sur, T**
US Fish & Wildlife Service: **H, Sur**
US Geological Survey: **A, B, S, Ged, Gep, H, L, R, Sub, Sur, T**

DEPARTMENT OF TRANSPORTATION
Federal Highway Administration: **Sur**

INDEPENDENT AGENCIES
Federal Emergency Management Agency: **H**
National Aeronautics & Space Administration: **H, L, R, Sub, Sur**
Tennessee Valley Authority: **B, S, Ged, H, L, R, Sub, Sur, T**

Federal Agency Data Product Code:

A = Atmospheric	**H** = Hydrologic	**Sub** = Subsurface
B = Boundaries	**L** = Land Ownership	**Sur** = Surface and Manmade Features
Ged = Geodetic	**R** = Remotely Sensed	**T** = Topography
Gep = Geophysics	**S** = Socioeconomic	

National Spatial Data Infrastructure (NSDI)

There is a wealth of geographic data which can be accessed from federal and state agencies over the internet. Most federal agencies which deal with geographic data have File Transfer Protocol (FTP) servers storing various geographic datasets. These servers allow organizations to download digital data over the internet. One of the most populated servers is the US Geological Survey FTP server, which holds all of the USGS Digital Line Graph files (the USGS server FTP address can be found by calling the USGS at 1-800-USA-MAPS). The Census Bureau also has an FTP server which allows organizations to access portions of its TIGER/Line file database. Government FTP servers can be searched for on the internet using ARCHIE.

Many federal and state agencies and corporations which deal with geographic data have internet home pages which can be accessed on the World-Wide-Web. The US Geological Survey (USGS) home page (URL address: http://www.usgs.gov), like the USGS FTP server, contains a wealth of information about USGS geographic data and how it can be used. From the USGS home page it is possible to search for, view, and download USGS data. One can also obtain USGS Fact Sheets, general information on the USGS, educational resources, publications, research papers, and informational resources on other internet sites. Most federal agencies have their own home page and are structured similarly to the USGS home page. Most major GIS software vendors also have internet home pages. Environmental Systems Research Institute (ESRI), Inc. has an excellent home page (URL address: http://www.esri.com) which contains a wide assortment of useful information.

State Government Agencies

There are many New York State agencies which are good sources of GIS data. Three of these organizations include the Department of Transportation, the Department of Environmental Conservation, and the Office of Real Property Services.

The New York State Department of Transportation (NYSDOT) offers data in paper and digital file formats. Paper topographic maps can be obtained at various scales. Most applicable to GIS needs, the NYSDOT has developed digital spatial files which are part of the New York State County Base Map Series. The Base Map files, though created with a CADD (Computer Aided Design and Drafting), have been designed for use in a GIS. The Department has developed a file structure which will allow for their conversion into a topological GIS format. There are various data layers available within this database including: Roads, Boundaries, Hydrography, Miscellaneous Transportation, and Names (NYSDOT, 1994). For further information, see Digital Files from the County Base Map Series from the NYSDOT.

The New York State Department of Environmental Conservation (NYSDEC) is another state organization which offers GIS data in varying formats. In 1990, the NYSDEC compiled an in-house inventory of its geographic data sources called the "Geographic Data Source Directory. The directory contains information on all of the DEC's geographic data sources with potential GIS applications. The DEC divided its data into the following categories: Air Resources, Construction Management, Fish and Wildlife, Hazardous Substances Regulation, Hazardous Waste Remediation, Lands and Forests, Law Enforcement, Management Planning and Information Systems Development, Marine Resources, Mineral Resources, Operations, Regulatory Affairs, Solid Waste, and Water (Warnecke et al, 1992). A copy of the directory is available from NYSDEC. Call your local office or the main office in Albany.

The New York State Office of Real Property Services (ORPS) has developed a database known as RPIS (Real Property Information System) which contains information on all tax parcels in the state. Each parcel contains a coordinate representing the center point of the parcel and attribute information which includes: unique land-based parcel identification numbers and descriptive information, such as land use, locations, sales information, exemptions, and other parcel attributes. RPIS data is available to local assessors, real property assessment offices , corporations and the general public for a nominal fee.

The New York State Department of Health (DOH) uses GIS in its work in analyzing and mapping environmental health risk areas and hazardous waste sites. The DOH has a database containing Census Bureau TIGER files and parcel maps. These GIS files can be acquired by the public.

Some other agencies which have GIS databases and which may have data usable in a GIS include: the Adirondack Park Agency (APA); the Hudson River Valley Greenway; New York Metropolitan Transportation Council; the Office of Parks, Recreation and Historic Preservation; Department of Public Service; State Emergency Management Office; New York City Department of Environmental Protection (Hilla, 1995); State Data Center Affiliates (various NYS Counties). Please note these are all examples and not intended to be an exhaustive list.

Regional And Local Governments

Many regional and local government agencies and organizations maintain GIS databases. These agencies may have data sharing arrangements with local companies and other municipalities. Information identifying which government agencies and companies have available GIS data layers may be found in regional or local GIS data directories. One such regional data directory developed within New York State is the Regional Directory of Geographic Data Sources for Genesee/Finger Lakes Counties. The directory contains information on participating government agencies and companies which have GIS data layers, then lists information regarding these layers, and provides the name, address and phone number of the person within the organization who can be contacted for further details or data sharing arrangements (GIS/SIG, 1995).

Private Data Firms

There are companies that will develop data for a local government. These companies will develop programs based on contract data conversion or public/private partnerships. Contract data conversion firms are available for those organizations that wish to have custom geographic datasets developed. Usually, the development of these datasets involves the client organization providing existing source data (e.g., paper maps) to the data development firm, which then converts the data into digital format.

In public/private partnerships, the company will work out an agreement with the local government that will provide data conversion but also retain the ability to market, sell and/or use the digital data that was created. Public/private agreements are just emerging as a method for creating GIS databases cost effectively. When considering a public/private partnership, issues such as ownership, access, freedom of information requirements and long-term data maintenance must be addressed as well as the cost sharing of building the database.

4 DESCRIBING AND EVALUATING POTENTIAL DATA

The next step is to actually survey the various departments within the local governments and other external sources to determine what data is available for use in the GIS and what “condition” the data is in.

Metadata Documentation

The first step will be to document the data by developing metadata files for each database available. The metadata file is used for two roles. 1) develop information that will be used to evaluate the data for use in a GIS and 2) fulfill the metadata requirements for data once it is used in a GIS.

For each potential data source for the GIS database, the map series, photos, tabular files, etc. just be identified, reviewed, and ***evaluated for suitability*** to use in the GIS. Maps, photos, and remotely sensed data are the most likely sources and should be evaluated for:

- appropriate scale
- projection and coordinate system
- availability of geodetic control points
- aerial coverage
- completeness and consistency across entire area
- symbolization of entities (especially positional accuracy of symbol due either to size of symbol or off-set placement on map)
- quality of linework and symbols
- general readability and legibility for digitizing (labels)
- quality and stability of source material (paper/mylar)
- amount of manual editing needed prior to conversion
- edge match between map sheets
- existence and type of unique identifies for each entity (often entities shown on in map series used so-called "intelligent" keys or identifiers where an identifier for an object contains the map sheet number and/or other imbedded locational codes - in database design, it is much better to avoid "intelligent" keys of this type, particularly locational codes).
- positional and attribute accuracy

All of the above information needs to be documented for each potential data source. If a particular data source is then used to build part of the GIS database, some of this information will become part of the permanent metadata.

The metadata software accompanying this guideline provides three tables for recording the basic metadata about a potential data source. The content of these tables is listed below. The first table contains information on the source document (or file); the second table can describe each entity contained on a source document; and the third table can describe each attribute of an entity. Once again, only the most basic entries have been included in the supporting software in order to keep the software simple an straightforward. A particular user may wish to expand the tables provided to meet his/her specific needs.

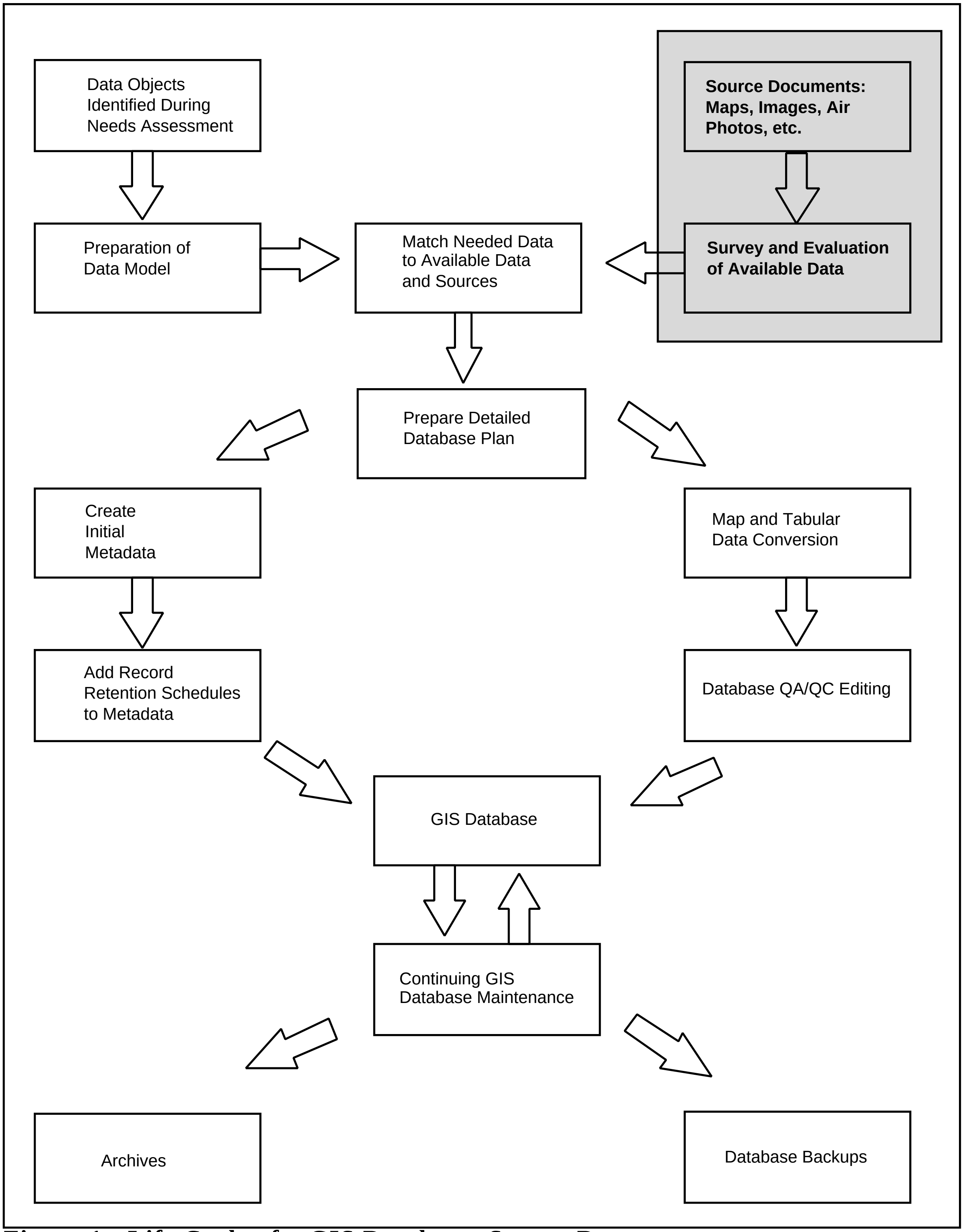


Figure 1 - Life Cycle of a GIS Database: Source Documents

The following lists the fields of the three tables that contain source data information:

Source Documents

Source Document Name:	Parcel Map
Source ID #:	1
Source Organization:	Town of Amherst
Type of Document:	Map
Number of Sheets (map, photo, etc):	200
Source Material:	Mylar
Projection Name:	UTM
Coordinate System:	State Plane
Date Created:	5-Oct-91
Last Updated:	8-Nov-95
Control Accuracy Map:	National Map Accuracy Standard
Scale:	Variable; 1" = 50 ft To 1" = 200 ft
Availability:	Current
Reviewed By:	Lee Stockholm
Review Date:	19-Dec-95
Spatial Extent:	Town of Amherst
File Format:	N/A
Comments:	

Entities Contained In Source

Source ID #: 1

Entity Name: Parcel

Spatial Entity: Polygon

Estimate Volume Spatial Entity: 126 per map sheet

Symbol: None

Accuracy Description Spatial Entity: National Map Accuracy Standard

Reviewed By: Lee Stockholm

Review Date: 02-Jan-94

Scrub Needed: Yes

Comments:

Attributes By Entity

Source ID #: 1

Entity Name: Parcel

Attribute Name: SBL Number

Attribute Description: Section, Block, and Lot Number

Code Set Name: N/A

Accuracy Description Attribute: N/A

Reviewed By: John Henry

Review Date: 08-Feb-93

Comments:

Additional Criteria For Evaluating Potential Data Sources

As the survey is being conducted, it is important to consider the following issues about the data:

- Is the data current and what is it's continuing availability?
- Is the data suitable for intended applications?
- Is the quality of the data appropriate for the type of applications needed? This should include both locational and attribute accuracy.
- Is the data cost effective?

FOR FURTHER INFORMATION:

The Manual of Federal Geographic Data Products, developed by the Federal Geographic Data Committee, is an excellent source for information on geographic datasets produced by agencies within the federal government. Listed by federal agencies and bureaus within each federal department, there are listings on the types of data which are available (e.g. concerning data structure, scale, software export format, source data, currency, what applications the data can be used for), and from which agencies they can be acquired.

To order contact:
Federal Geographic Data Committee Secretariat
US Geologic Survey
590 National Center
Reston, VA 22092
Phone 703-648-4533

New York State Department of Transportation data listing: Digital Files from the County Base Map Series.

Map Information Section
Mapping and Geographic Information Systems Bureau
New York State Department of Transportation
State Office Campus
Building 4, Room 105
Albany, New York 12232
Phone: (518) 457-3555

Example of a Regional Level GIS Data Directory:

1995 Regional Directory of Geographic Data Sources, developed by the GIS/SIG (Geographic Information Sharing/Special Interest Group) for New York State's Genesee/Finger Lake Region Counties. The directory is a listing of the various data sources which are available from local companies, and local government agencies in the Genesee/Finger Lakes Region.

The International GIS Source book, published by GIS World, Inc. is an annual publication which contains an excellent "Data Source Listings" chapter. It provides a wealth of information on companies which produce GIS datasets and also provides descriptions of the data they produce. The chapter also lists the different types of spatial data produced by public agencies, and lists data availability and contacts.

REFERENCE

Hilla, Christine M. "The Revolution of Geographic Information Systems in Land Use and Environmental Planning in New York State," Environmental Law in New York, Vol 6, no. 3., March, 1995.

Montgomery, Glenn E. and Harold C. Schuch, 1993. GIS Data Conversion Handbook. Fort Collins, CO: GIS World, Inc., pp. 89-91.

NYSDOT (New York State Department of Transportation), Digital Files from the County Base Map Series, mapping and Geographic Information Systems Bureau (1994).

Warnecke, L., J. Johnson, K. Marshall and R. Brown, State Geographic Information Activities Compendium, 294 Council of State Government (1991).

GIS DEVELOPMENT GUIDE: EVALUATING GIS HARDWARE AND SOFTWARE

1 INTRODUCTION

Purpose of Guide

A GIS is more than just hardware and software. It is a complex system with multiple components: Hardware, Software, People, Procedures and Data. The purpose of this guide is to focus on the hardware and software components of the system and how to acquire information on what is available.

Deciding what hardware and software to use for your GIS is a difficult yet important task. It will make up the foundation on which you will build your system. There is no clear-cut formula to use to make the selection process easier. In this guideline we will give you suggestions that you can use to evaluate various systems and sources for additional information.

2 SOURCES OF INFORMATION ABOUT GIS

To develop an understanding of GIS, you will need to get information about GIS systems. Here is a sampling of references to start with. This is not a comprehensive listing. Use it as a starting point and spread out from there.

GIS Source Book

The GIS source book is a good reference book that will give you a great deal of information about software vendors, trade associations, product specifications and more. This book is published by:

GIS World, Inc.
155 E. Boardwalk Drive, Suite 250
Fort Collins, CO 80525
Phone: 303-223-4848
Fax: 303-223-5700
Internet: info@gisworld.com

Other Publications

Conference Proceedings
Each major GIS conference publishes the proceedings from their event. Contact the association listed in Attachment A for information on how to obtain these documents.

Scholarly Journals

There are a number of scholarly journals that deal with GIS. These are published on an on-going basis.

Cartographica - Contact: Canadian Cartographic Association
Cartography and Geographic Information Systems - Contact: American Cartographic Association
International Journal of Geographical Information Systems - Contact: Keith Clark at CUNY Hunter College, New York City
URISA Journal - Contact: Urban and Regional Information Systems Association

Trade Magazines

There are a number of trade magazines that are focused on GIS. They are:

GIS World
GIS World Inc.
155 E. Boardwalk Drive
Suite 250, Fort Collins, CO 80525
Phone: 303-223-4848
Fax: 303-223-5700
Internet: info@gisworld.com

Business Geographics
GIS World, Inc.
155 E. Boardwalk Drive, Suite 250
Fort Collins, CO 80525
Phone: 303-223-4848
Fax: 303-223-5700
Internet: info@gisworld.com

Geo Info Systems
Advanstar Communications
859 Williamette St.
Eugene, OR., 97401-6806
Phone: 541-343-1200
Fax: 541-344-3514
Internet:geoinfomag@aol.com
WWW site:http://www.advanstar.com/geo/gis

GPS World
Advanstar Communications
859 Williamette St.
Eugene, OR., 97401-6806
Phone: 541-343-1200
Fax: 541-344-3514
Internet:geoinfomag@aol.com
WWW site:http://www.advanstar.com/geo/gis

Association Newsletters

Many associations have newsletters that cover GIS topics and can be a good source of information. Contact the organizations listed in attachment A for more information

Books with vendor specific information

There is a number of books published about GIS and related topics. Here are some of the publishers:

Onword Press
2530 Camino Entrada
Sante Fe, NM, 87505-4835
Phone: 505-474-5132
Fax: 505-474-5030

John Wiley & Sons, Inc.
605 Third Avenue
New York, NY, 10158-0012

ESRI, Inc.
80 New York Street
Redlands, CA 92373-8100
Phone: 909-793-2853
Fax: 909-793-4801

GIS World, Inc.
155 E. Boardwalk Drive, Suite 250
Fort Collins, CO 80525
Phone: 303-223-4848
Fax: 303-223-5700
Internet: info@gisworld.com

Vendor Booths at Trade Shows

A wealth of information is available at trade shows from vendor booths. These can range from the general product literature to white papers and technical journals. This is also a good time to gather a large amount of information on different companies in a short period of time.

User Groups

User Groups are another source of valuable information and support. There are a number of user groups that have formed to provide support and professional networking. GIS user groups are formed around a geographic region or by users of specific software products. New users are always welcome to these groups. A listing of users groups is contained in Attachment A

Current Users

The best way to gauge a vendor is by talking to their installed sites. The information that you get from talking to these users will be valuable insight into the type of company you will be working with. Ask the vendors you want to explore for a list of all of their users in the area or that are similar to your organization. Ask for contact names and phone numbers/e-mail addresses.

3 SELECTION PROCESS

Initially you will need to evaluate the software independently of hardware. The software will be selected based on the functionality it offers. Your hardware selection will be based on the GIS software you select and the operating system strategy your organization uses. You will need to test the hardware and software together making sure it works as advertised.

The nature of hardware and software technology is that it changes. In recent years it has been changing very quickly. Don't let this stop your efforts. It is easy to get intimidated. The important thing to remember is to get a product that has been proven in the marketplace and continues to have a clear development path. Avoid technology that is outdated or is on the bleeding edge and has not been proven.

Software

Software is evaluated on functionality and performance. In the Needs Assessment guide the need to identify the functionality was discussed. Here is where you will begin to use this information.

Functionality

What is important here is the ability of the software to do the things you need it to do in a straightforward manner. As an example, if the intended users are relatively new to using computers, the software has to have an easy to use graphical user interface (GUI). If the organization needs to develop specific applications, the software should have a programming language that allows the software to be modified or customized.

In the Needs Assessment Guide, the final report contains tables and references to the functionality you will need. Use this in developing the overall functionality required for the system.

Standards

Standards are a way of making sure that there is a common denominator that all systems can use. This can be in the form of data formats that can import and export data into the system, guidelines used for developing software, supporting industry developed standards that allow different applications to share data. Standards are generally developed by a neutral trade organization or in some cases are defined by the market.
There is a group that has formed for the GIS industry called Open GIS. This organization is developing standards for developers to use as they engineer software. Open GIS is made up of representatives from the software developer companies.

Performance

The performance of the software is dependent on two factors, 1) how it is engineered and 2) the speed of the hardware it is running on. GIS software is complex and will use a large amount of the system resources (memory, disk, etc.). The more complex the software, the more resources it will need.

Performance will be impacted if you have a minimally configured computer. Look for the developer’s software specifications to see what configuration is needed to run the software. This will give you the minimum requirements. Follow this up by getting the recommended specifications from the developer or a user group. These recommendations will give you a more accurate idea of the type of configuration you will need.

Expandability

The software needs you have today will change over time. More than likely your system will need to expand. Is the software being evaluated able to provide networking capabilities? Will it share data with other applications? Will it grow as the organization's GIS grows? Evaluate software based on the ability to grow with you. This may mean that there are complimentary products that can be used in conjunction with the package you are evaluating today or the developer has clearly defined plans for added functionality. Talk with other users to see if the developer has a good track record for providing these enhancements.

Licensing

GIS software is not purchased, it is licensed. There is normally a one-time license fee with an on-going maintenance fee that provides you with the most current versions of the software as they are released. In large systems this will be spelled out in a licensing agreement with a corresponding maintenance agreement. For desktop software a shrink wrap license is used with subsequent releases being offered to existing users through a discounted upgrade. The maintenance fees and upgrade costs generally run between 15% to 30% of the initial license fee.

The terms in most software packages spell out how the software can and cannot be used. Have the terms of the license reviewed by an attorney before signing up. This can save hassles later as you are developing and using your system.

Hardware

When discussing hardware, there are terms/concepts that you need to understand. The following is a discussion of these. However, GIS software selection drives the hardware requirements. Therefore before launching a full scale evaluation of hardware, make your selection for the GIS software you will be using.
Hardware can be broken down into the following basic components:

Operating System
Processor
Disk
Memory
Communications

Operating Systems

An operating system is the software that runs the computer hardware. It is this program that tells the computer what to do and how to do it. You may already be familiar with some of the operating systems that are on the market such as Microsoft's Windows product or various brands of the UNIX operating system.

It is important to have an Operating System plan within your organization. The plan should take into account the departments that will be using the computer system, the type of network being used (or being planned), what operating systems are currently being used, how large the database is and what kind of technical support skills you have access to (in-house or contractor).

The GIS will need to fit into your operating system plan. This will be important as you add other departments onto the system.

Processor

The processor or CPU (central processor unit) is the part of the computer that actually does the calculations or "processes" the instructions being sent to it. The most common term that describe the processor's capabilities is the clock speed. This is stated in terms of MHz (MegaHertz). The clock speed simply describes how many cycles per second the processor works. The higher the clock speed the faster the processor.

Another description of the processor's capability is how many bits it can access at one time. Many of the new processors are 32-bit processors. This means that the CPU can access or "grab" 32 bits of information during each cycle. Older computers such as a "386" machine where 16-bit machines. There are some machines on the market that manufacture a 64-bit machine (such as Digital Equipment Corporation). These are very fast CPU's but are hampered by the lack of a 64-bit operating system that can take advantage of it's speed. It is the direction the hardware industry seems to be heading.

Disk

The disk or hard drive is the device used to store the operating and application software. It is also used to store data and images. In working with a GIS you will quickly find out that GIS uses a large amount of disk space. It is not uncommon to have multiple gigabytes of hard drive on a single end-user machine and 10 - 20 gigabytes on a central data server. Luckily the prices of hard drives have been coming down and will continue to be affordable.

Memory

Memory or random access memory (RAM) is used as a temporary storage space by the operating system and by the application software which is running on the computer. Most applications run better as the amount of memory increases. This is true up to a point. At some point, the performance increases will begin to taper off as additional memory is added. Most software developers can give you configuration data that indicates where this point is.

Communications

The trend in most systems today is to link up users throughout the organization on a network. This is an area in the computer industry that is advancing very rapidly. It is recommended that you retain a competent consultant who works with networks to give you detailed and current information.

In simple terms, a network is a connection between computers that allows information to be passed around from computer to computer. In a typical organization, this is a local area network (LAN). In order to connect a computer to the network it will need a network card for the wiring to plug into and network software to allow the computer to transmit and receive signals over the wiring. Of course the physical network (wiring) is also needed.

A small network within a department is inexpensive and can allow the users to share network resources such as printers and database servers. The network can provide services like e-mail and disk sharing. It can also be the entryway into larger networks that go outside the building or campus your organization is located on. This is called a Wide Area Network (WAN). A WAN requires a more structured network architecture. It does give users access to more resources.

Another important point to consider is developing access to the Internet. This specialized network is growing rapidly and provides an incredible amount of resources for a user. The Internet is an area to share ideas in a GIS forum, download data for use in the system, get technical support for a problem, get the latest information on a product from a vendors home page or develop one of your own. The amount of information is overwhelming and too diverse to list in this guide. The point is that you should seriously be considering getting a connection to the Internet. When considering your network, factor this into the equation.

Benchmarking a System

Benchmarking a GIS can be a very involved process. The level of effort needed for the benchmark should be proportional to the size and complexity of the overall system being developed. A benchmark is the process of testing various combination of hardware and software and evaluating their functionality and performance. The benchmark is usually part of an RFP process and is only done with a limited number of selected vendors (i.e.: those that have been shortlisted). Each combination is tested under similar conditions using a predefined data set that is indicative of your database. This data set should be used with all of the hardware / software configurations selected for evaluation. When completed, an organization will have results that can used to objectively evaluate the systems.

Setting It Up

When putting a benchmark together there is strength in numbers. Get a committee together. A committee will take the burden off of one person and give the process more objectivity. Have representation from all the interested departments and agencies within the organization. A working group of about 8-10 committee members is reasonable.

The committee will develop the criteria that will be used to evaluate the systems. Use the Needs Assessment documentation as a reference for this. These criteria will form the basis of the benchmark. Develop a series of tasks that each vendor will need to complete during the benchmark. The tasks should be measurable (i.e.: time, ease of use, can the function be done). Also prepare a form that each of the committee members will use to rate the tasks performed in the benchmark.

In your benchmark you will not only be to rating various aspects of the system, you are also going to be rating the vendor. Be sure to include some measurement for teamwork, communication, and technical skills of the vendor. It might be useful to work with a consultant that has experience setting up benchmarks or to get advice (and examples of documentation) from another local government who has recently completed a benchmark.

Well in advance of the scheduled benchmarks, send out information that outlines the tasks the vendor will need to perform and any rules they will need to follow (how much time for set up, time given to perform various tasks, how many people can be present for the benchmark, etc.).

Vendor Support

The vendor you select will become an extended team member for your GIS. There needs to be a good “fit”. The vendor will be a good source of support and information. All vendors provide some type of technical support. Ask current users how it has worked for them. If there have been problems in the past, do existing users see improvement? The GIS industry has been growing very fast over the last few years, there are bound to be some growing pains. What you should be looking for is a vendor who listens to what you need and makes improvements based on user input.

Attachment A - User Groups

New York State

Western New York ARC/INFO Users Group (WNYARC)
Buffalo area:
Contact: Graham Hayes
GIS Resource Group, Inc.
716-655-5540

GIS/SIG
Rochester Area:
Contact: Scott Sherwood

Multi-County GIS Cooperative
Statewide:

Tri-County GIS Users Group
Southern Tier:
Contact: Jennifer Fais

GISMO
New York City:
Contact: Jack Eichenbaum

Capital Region ARC/INFO User Group (CAPARC)
Albany Area:

URISA New York State Chapter
Contact - Lee Harrington, Professor
SUNY College of Environmental Science and Forestry
Syracuse
Phone: 315-470-6670
Fax: 315-470-6535

Long Island GIS (LIGIS)
Contact: Joseph P. Jones

National

American Congress on Surveying and Mapping (ACSM)
5410 Grosvenor Lane
Bethesda, MD, 20814
Phone: 301-493-0200
Fax: 301-493-8245

American Society for Photogrammetry and Remote Sensing (ASPRS) & (GIS/LIS)
5410 Grosvenor Lane
Bethesda, MD, 20814
Phone: 301-493-0290
Fax: 301-493-0208

Association for American Geographers (AAG)
1710 Sixteenth St. N.W.
Washington D.C., 20009-3198
Phone: 202-234-1450
Fax: 202-234-2744

Automated Mapping/Facility Management International (AM/FM International)
14456 East Evans Ave.
Aurora, CO, 80014
Phone: 303-337-0513
Fax: 303-337-1001

Canadian Association of Geographers (CAG)
Burnside Hall, McGill University
Rue Sherbrooke St. W
Montreal, Quebec H3A 2K6
Phone: 514-398-4946
Fax: 514-398-7437

Canadian Institute of Geomatics (CIG)
206-1750 rue Courtwood Crescent
Ottawa, Ontario, K2C 2B5
Phone: 613-224-9851
Fax: 613-224-9577

Urban And Regional Information Systems Association (URISA)
900 Second St. N.E., Suite 304
Washington, D.C. 20002
Phone: 202-289-1685
Fax: 202-842-1850

GIS DEVELOPMENT GUIDE: DATABASE PLANNING AND DESIGN

1 INTRODUCTION

The primary purpose of this phase of the GIS development process is to specify "how" the GIS will perform the required applications. Database planning and design involves defining how graphics will be symbolized (i.e., color, weight, size, symbols, etc.), how graphics files will be structured, how nongraphic attribute files will be structured, how file directories will be organized, how files will be named, how the project area will be subdivided geographically, how GIS products will be presented (e.g., map sheet layouts, report formats, etc.)., and what management and security restrictions will be imposed on file access. This is done by completing the following activities:

- Select a source (document, map, digital file, etc) for each entity and attribute included in the E-R diagram
- Set-up the actual database design (logical/physical design)
- Define the procedures for converting data from source media to the database
- Define procedures for managing and maintaining the database

The database planning and design activity is conducted concurrently with the pilot study and/or benchmark activities. Clearly, actual procedures and the physical database design cannot be completed before specific GIS hardware and software has been selected while at the same time GIS hardware and software selection cannot be finalized until the selected GIS can be shown to adequately perform the required functions on the data. Thus, these two activities (design and testing) need to be conducted concurrently and iteratively.

In many cases, neither database design matters nor hardware and software selection are unconstrained activities.. First, the overall environment within which the GIS will exist must be evaluated. If there exist "legacy" systems (either data, hardware or software) with which the new GIS must be compatible, then design choices may be limited. Both GIS hardware and software configurations and database organizations that are not compatible with the existing conditions should be eliminated from further consideration. Secondly, other constraints from an organizational perspective must be evaluated. It may, for example, be preferable to select a specific GIS or database structure because other agencies with whom data will be shared have adopted a particular systems. Finally, assuming that the intended GIS (whether it will be large or small) will be part of a corporate or shared database, the respective roles of each participant need to be evaluated. Clearly, greater flexibility of choice will exist for major players in a shared database (e.g., county, city, or regional unit of government) than for smaller players (town, village, or special purpose GIS applications). This does not mean that the latter must always go with the majority, but simply that the shared GIS environment must be realistically evaluated. In fact, one way for the smaller participants in a shared GIS to ensure their needs are considered, is to fully document their needs and resources using procedures recommended in these guidelines.

Finally, with the completion of both the database planning and design and the pilot study/benchmark activities, sufficient detailed data volume estimates and GIS performance information will be known to calculate reliable cost estimates and prepare production schedules. This becomes the final feasibility check before major resources are committed to data conversion and GIS acquisition.

What is already known about the GIS requirement

Prior phases of the GIS development process should have produced the following information which is needed at this time:

- A complete list of data, properly defined and checked for validity and consistency (from the master data list, E-R data model and metadata entries).
- A list of potential data sources (maps, aerial photos, tabular files, digital files, etc.) cataloged and evaluated for accuracy and completeness (from the available data survey). This inventory would also include all legacy data files, either within the agency or elsewhere, which must be maintained as part of the overall shared database.
- The list of functional capabilities required of the GIS (from needs assessment).

2 SELECTING SOURCES FOR THE GIS DATABASE

This activity involves matching each entity and its attributes to a source (map, document, photo, digital file). The information available for this task is as follows:

- List of ***entities and attributes*** from the conceptual design phase

Master Data List

Entity	**Attributes**	**Spatial Object**
Street_segment	name, address_range	Line
Street_intersection	street_names	Line
Parcel	section_block_lot#, owner_name, owner_address, sites_address, area, depth, front_footage, assessed_value, last_sale_date, last_sale_price, size (owner_name, owner_address, assessed_value as of previous January 1st))	Polygon
Building	building_ID, date_built, building_material, building_assessed_value	Footprint
Occupancy	occupant_name, occupant_address, occupancy_type_code	None
Street_segment	name, type, width, length, pavement_type	Polygon
Street_intersection	length, width traffic_flow_conditions, intersecting_streets	Polygon
Water_main	type, size, material, installation_date	Line
Valve	type, installation_date	Node
Hydrant	type, installation_date, pressure, last_pressure_test_date	Node
Service	name, address, type, invalid_indicator	None
Soil	soil_code, area	Polygon
Wetland	wetland_code, area	Polygon
Floodplain	flood_code, area	Polygon
Traffic_zone	zone_ID#, area	Polygon
Census_tract	tract#, population	Polygon

Water_District	name, ID_number	Polygon
Zoning	zoning_code, area	Polygon

- The list of ***surveyed data sources from the Available Data Survey*** and their recorded characteristics in the metadata tables ***Source Documents, Entities Contained in Source, and Attributes by Entity.***

Source Documents

Source Document Name: Parcel Map

Source ID #: 1

Source Organization: Town of Amherst

Type of Document: Map

Number of Sheets (map, photo, etc): 200

Source Material: Mylar

Projection Name: UTM

Coordinate System: State Plane

Date Created: 5-Oct-91

Last Updated: 8-Nov-95

Control Accuracy Map: National Map Accuracy Standard

Scale: Variable; 1" = 50 ft To 1" = 200 ft

Availability: Current

Reviewed By: Lee Stockholm

Review Date: 19-Dec-95

Spatial Extent: Town of Amherst

File Format: N/A

Comments:

Entities Contained In Source

Source ID #: 1

Entity Name: Parcel

Spatial Entity: Polygon

Estimate Volume Spatial Entity: 126 per map sheet

Symbol: None

Accuracy Description Spatial Entity: National Map Accuracy Standard

Reviewed By: Lee Stockholm

Review Date: 02-Jan-94

Scrub Needed: Yes

Comments:

Attributes By Entity

Source ID #: 1

Entity Name: Parcel

Attribute Name: SBL Number

Attribute Description: Section, Block, and Lot Number

Code Set Name: N/A

Accuracy Description Attribute: N/A

Reviewed By: John Henry

Review Date: 08-Feb-93

Comments:

If there is a choice between sources, that is, two or more sources are available for a particular entity attribute, then criteria for deciding between them will be needed. In general, these criteria will be:

- Accuracy of resulting data
- Cost of conversion from source to database
- Availability of the source for conversion
- Availability of a continuing flow of data for database maintenance.

Occasionally, alternative sources may result in different representations in the database, such as a vector representation versus a scanned image. In this situation, the ability of each representation to satisfy the requirements of the GIS applications will need to be evaluated.

Once a source has been selected, the metadata tables that record source data information need to be completed as appropriate. These are:

- Data Object Information
- Attribute Information
- Spatial Object Information
- Source Document Information

To complete the accuracy information, the accuracy expected from the conversion process will need to be determined. This accuracy target will also be used later in the database construction phase by the quality control procedures. The metadata tables that need to be completed at this time are shown below:

Data Object Information

Data Object Name	Parcel
Type:	Simple
Data Object Description:	Land ownership parcel
Spatial Object Type:	Polygon
Comments:	

Attribute Information

Data Object Name:	Parcel
Data Attribute Name:	SBL Number
Attribute Description:	Section, Block, and Lot Number
Attribute Filename:	Parcel.PAT
Codeset Name/Description:	N/A
Measurement Units:	N/A
Accuracy Description:	N/A
Comments:	

Spatial Object Information

Data Object Name:	Parcel
Spatial Object Type:	Polygon
Place Name:	Amherst
Projection Name/Description:	UTM
HCS Name:	State Plane Coordinate System
HCS Datum:	NAD83
HCS X-offset:	1000000
HCS Y-offset:	800000
HCS Xmin:	25
HCS Xmax:	83
HCS Ymin:	42
HCS Ymax:	98
HCS Units:	Feet
HCS Accuracy Description:	National Map Accuracy Standard
VCS Name:	
VCS Datum:	
VCS Zmin:	0
VCS Zmax:	0
VCS Units:	
VCS Accuracy Description:	
Comments:	

Source Document Information

Data Object Name:	Parcel
Spatial Object Type:	Polygon
Source Document Name:	Parcel Map
Type:	Map
Scale:	Variable: 1" = 50 feet To 1" = 200 feet
Date Document Created:	17-Nov-89
Date Last Updated:	05-Oct-94
Date Digitized/Scanned:	24-Apr-95
Digitizing/Scanning Method Description:	Manual digitized with Wild B8
Accuracy Description:	90% of all tested points within 2 feet
Comments:	

For some of the above tables, information will be available for only some of the entries. The remaining entries will be completed later as the database is implemented. The examples shown are from the metadata portion of the GIS Design software package that accompanies these guidelines. This package is a Microsoft Access™ program that runs "stand-alone" (you do not need a copy of Microsoft Access™) on a regular PC. Where the same information is needed for multiple tables, this information is only entered once. The information is then automatically transferred to the other tables where it is needed.

3 THE LOGICAL/PHYSICAL DESIGN OF THE GIS DATABASE

This activity involves converting the conceptual design to the logical/physical design of the GIS database (hereafter referred to as the physical design). The GIS software to be used dictates most of the physical database design. The structure and format of the data in a GIS, like ARC/INFO™, Intergraph™, MapInfo™, System 9™, etc. have already been determined by each vendor respectively. If one separates the conceptual entity and its attributes from the corresponding spatial entity and its geometric representation, it can be seen that the physical database design for the spatial entity has been completely defined by the vendor and the GIS designer does not need to do anything more for this part of the data. The attributes of the entities may, however, be held in a relational database management system linked to the GIS. If this is the case, the GIS analyst needs to design the relational tables for the attribute information. Figure 1 illustrates the split between the entity's attributes and the spatial information. This example is based on the ARC/INFO™ GIS and a relational database system.

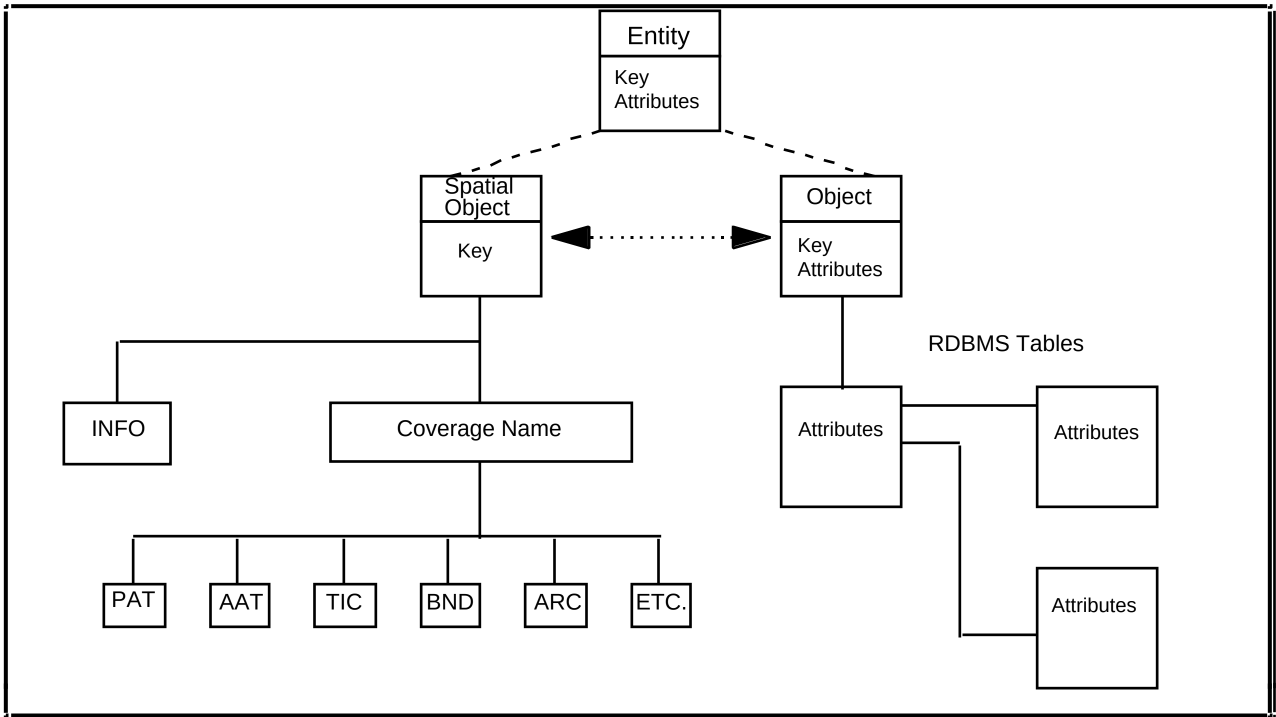


Figure 1 - GIS Representation of Object and Associated Spatial Object

The translation from the entity representation in the E-R diagram to the physical design of the database for a single entity is shown in Figure 2:

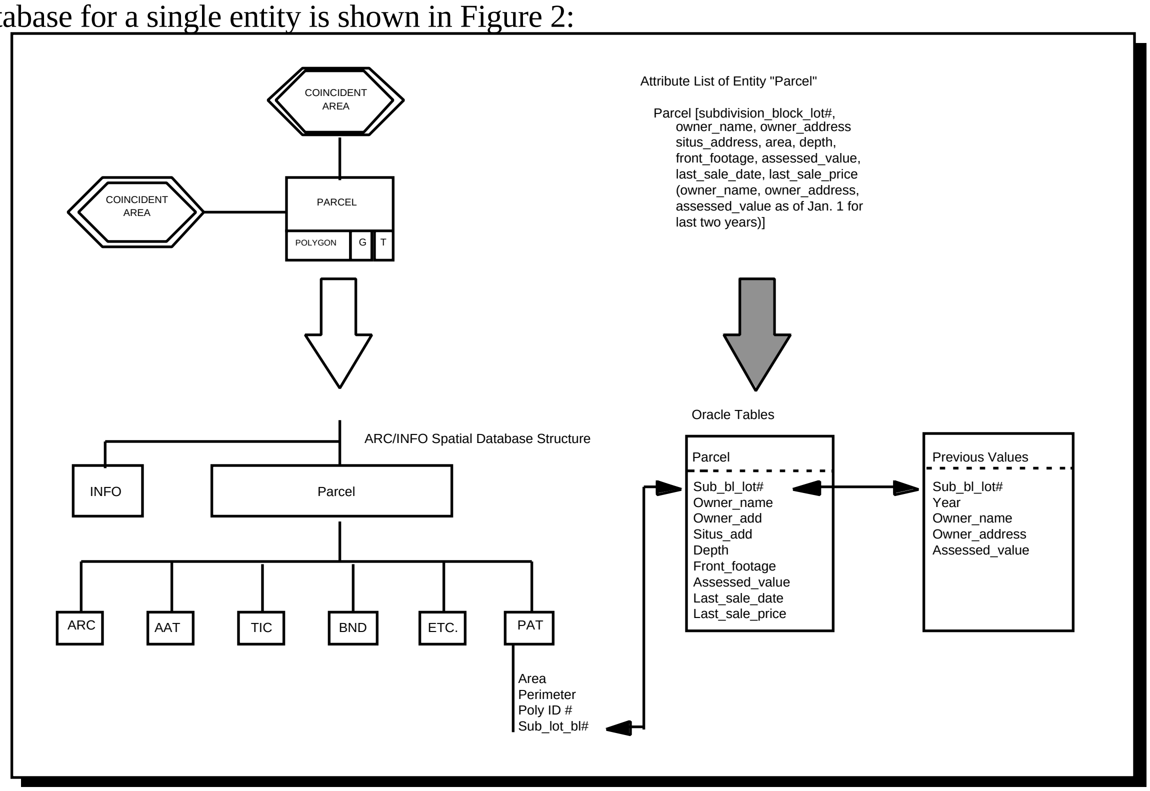


Figure 2 - Example of Mapping of E-R Entity and Attribute List into ARC/INFO™ & ORACLE™ Logical Database

Again, this example is based on ARC/INFO™ and the Oracle relational database system and shows how one entity from the E-R diagram would be represented in a single layer (coverage in ARC/INFO™ terms) and two Oracle tables. It will not always be the case where one entity from the E-R diagram translates into a single layer. More complex representations will be needed. Generally this will involve two or more entities forming a single layer with, possibly, several relational database tables. For example, Figure 3 from the conceptual design guideline shows, in part, the following entities:

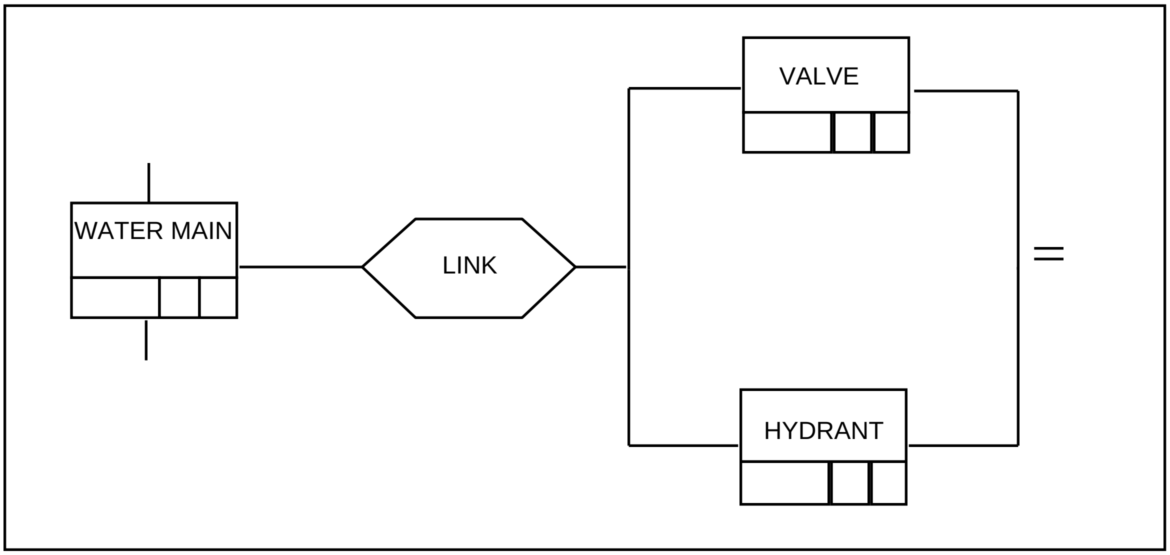


Figure 3 - E-R Representation of Elements of a Water Distribution System

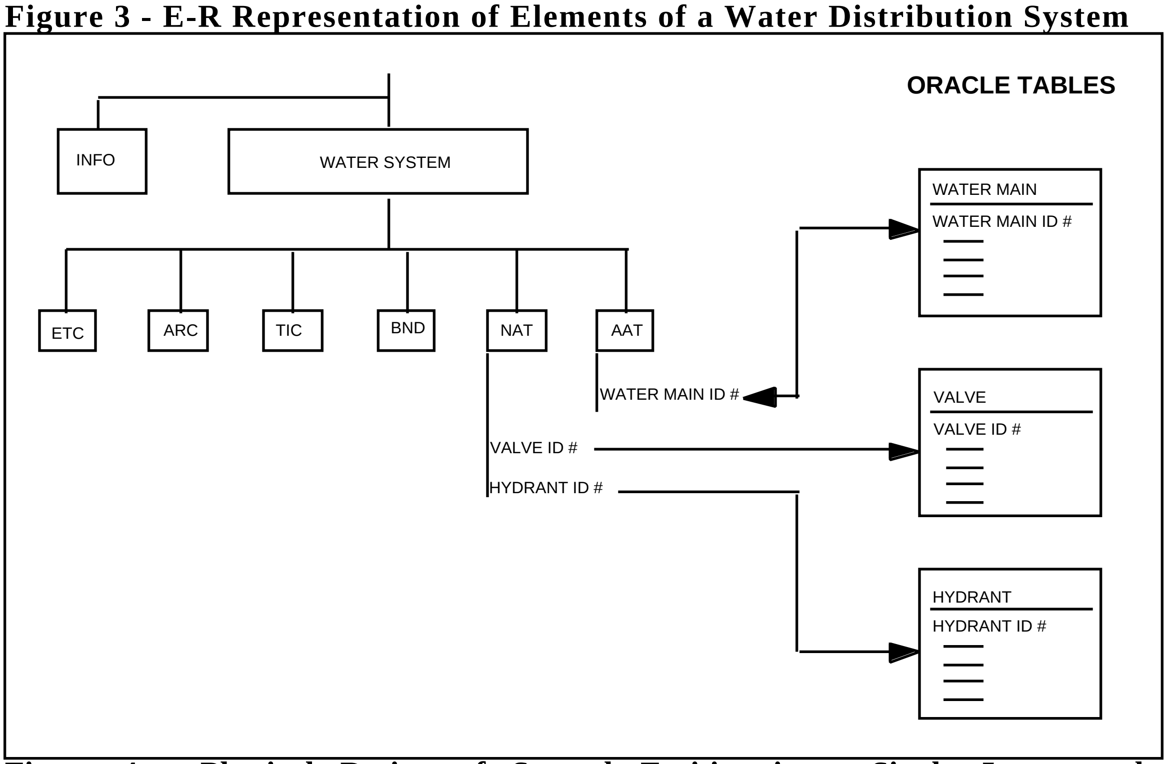


Figure 4 - Physical Design of Several Entities in a Single Layer and Three Relational Tables

In figure 4, the water main segments, the valves and the fire hydrants have been placed together in one layer as line segments, and two sets of nodes. However, each entity has its own relational table to record its respective attributes (see Table 1, page 2). The relationship is maintained by unique keys for each instance of each entity.

Every entity shown on the E-R diagram must be translated to either a GIS layer, a relational table(s), or both, as indicated by the information to be included. In addition, every relationship of the type "relationship represented in database" (single line hexagon on the E-R diagram) must be implemented through the primary and secondary keys in the tables for the entities represented.

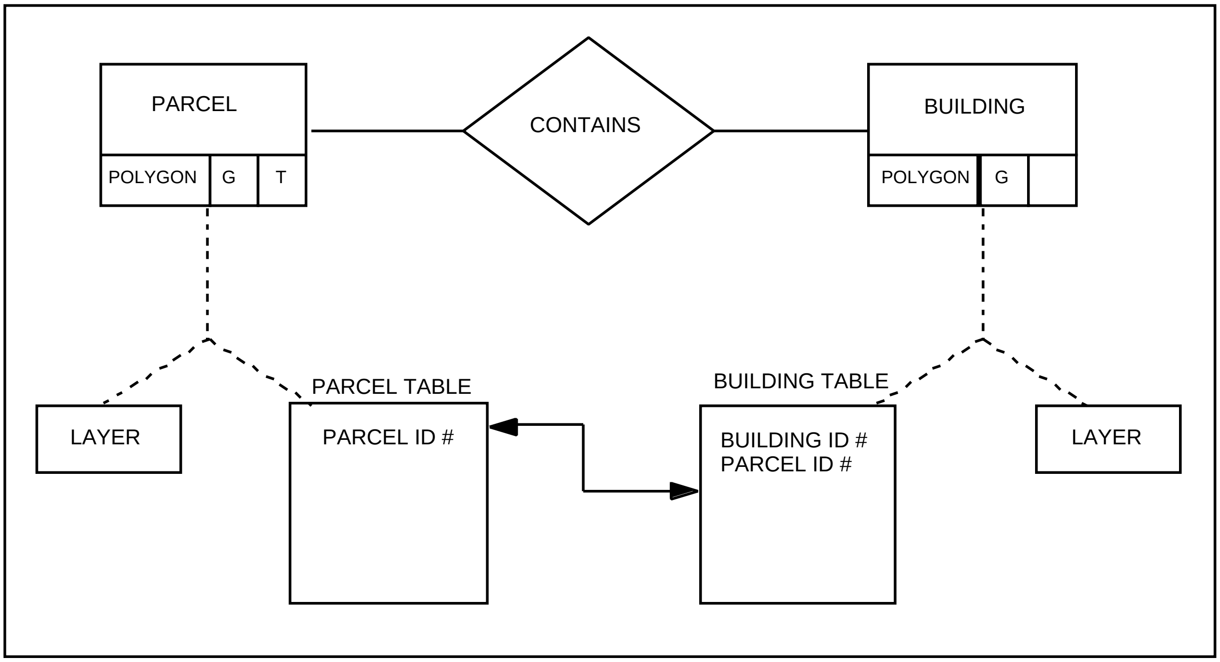


Figure 5 - Standard Database Relationship with Primary and Secondary Keys

As shown in Figure 5, the entity "**parcel**" may "**contain**" the entity "**building**." The table for each entity would have its own primary key (ID#), however, the table for **building** must also have a secondary key (parcel ID#) to maintain the relationship in the database.

The completed physical database design must account for all entities and their attributes, the spatial object with topology and coordinates as needed, and all relationships to be contained in the database. The remaining items on the E-R diagram, the two types of spatial relationships, must be accounted for in the list of functional capabilities, that is, the implied spatial operations must be possible in the chosen GIS software

4 PROCEDURES FOR BUILDING THE GIS DATABASE

Developing a GIS database is frequently thought of as simply replicating a map in a computer. As can be inferred by the nature and detail of the activities recommended up to this point in these guidelines, building a GIS database involves much more than "replicating a map." While substantial portions of the GIS database will come from map source documents, many other sources may also be used, such as aerial photos, tabular files, other digital data, etc. Also, the "map" representation is only part of the GIS database. In addition to the map representation and relational tables, a GIS can hold scanned images (drawings, plans, photos), references to other objects, names and places, and derived views from the data. The collection of data from diverse sources and its organization into a useful database requires development of procedures to cover the following major activities:

- ***Getting the Data*** which may include acquiring existing data from both internal and external sources, evaluating and checking the source materials for completeness and quality, and/or creating new data by planning and conducting aerial or field surveys. Contemporary GIS projects attempt to rely on existing, rather than new, data due to the high cost of original data collection. However, existing data (maps and other forms) were usually created for some other purpose and thus have constraints for use in a GIS. This places much greater importance on evaluating and checking the suitability of source data for use in a GIS.

- ***Fixing any problems in the data source***, often focused only on map source documents, this activity has been called "map scrubbing." Depending on the technology to be used to convert the map graphic image into its digital form, the source documents will have to meet certain standards. Some conversion processes require the map to be almost perfect which other processes attempt to automate all needed "fixes" to the map. What is required here is for the GIS analyst to specify, in detail, a procedure capable of converting the map documents into an acceptable digital file ***while accounting for all known problems in the map documents***. This procedure should be tested in the pilot project and modified as needed.

- ***Converting to digital data***, the physical process of digitizing or scanning to produce digital files in the required format. The major decision here is whether or not to use an outside data conversion contractor or to do the conversion within the organization. In either case, specifications describing the nature of the digital files should be prepared. In addition to including the physical database design, specifications should describe the following:
  - Accuracy requirements (completeness required, positional accuracy for spatial objects, allowable classification error rates for attributes).
  - Quality control procedures that will be conducted to measure accuracy.
  - Partitioning of the area covered by the GIS into working units (map sheets) and how these will be organized in the resulting database (including edge matching requirements).
  - Document and digital file flow control, including logging procedures, naming conventions, and version control.

- ***Change control***, most map series are not static but are updated on a periodic basis. Once a portion of the map has been sent to digitizing (or whatever process is used), a procedure must be in place to capture any updates to the map and enter these into the digital files.

- ***Building the GIS Database,*** once digitizing has been completed, the sponsoring organization has a set of digital files, not an organized database (illustrated in Figure 5). The system integration process (a subsequent guideline document) must take all the digital files and set-up the ultimate GIS database in a form that will be efficient for the users. The several considerations required for this process are covered under GIS Data Database Construction, GIS System Integration and GIS maintenance and use.

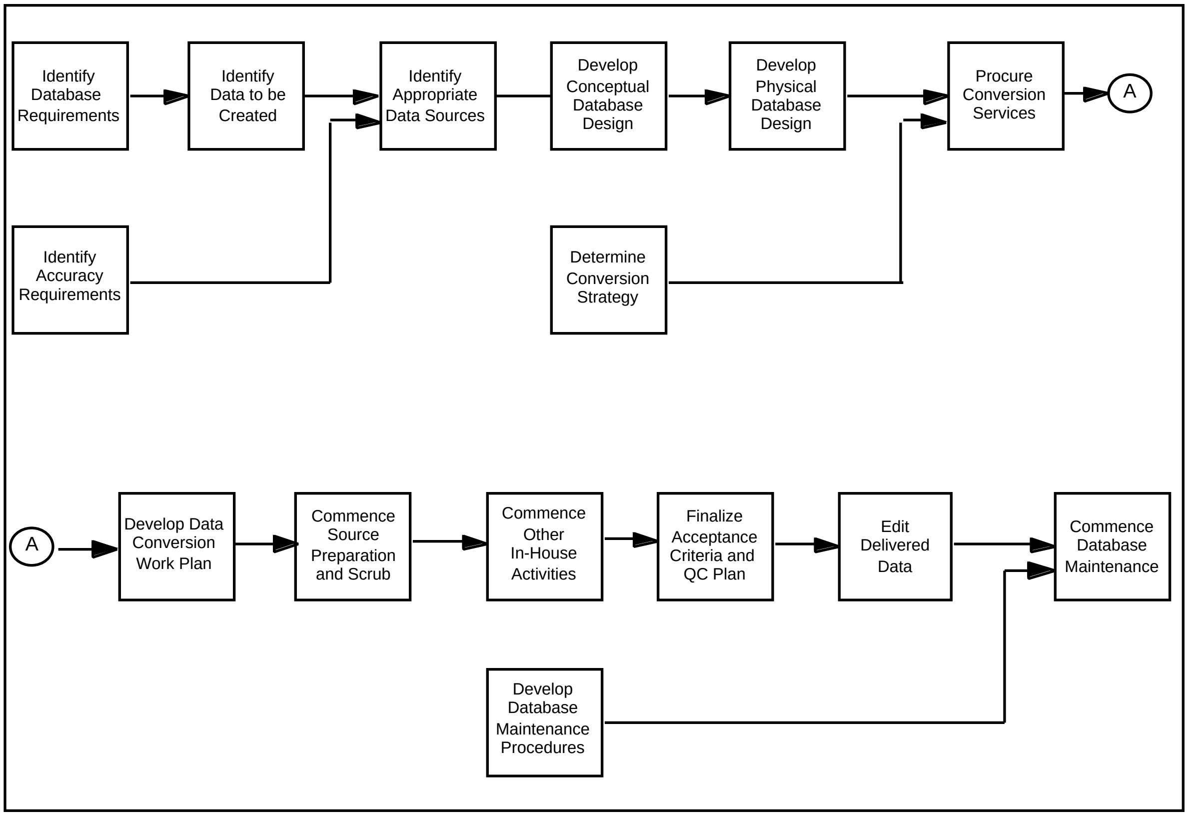


Figure 6 - Guide to Data Conversion/Database Creation - *GIS Data Conversion Handbook*

5 PROCEDURES FOR MANAGING AND MAINTAINING THE DATABASE

Because the physical world is constantly changing, the GIS database must be updated to reflect these changes. Once again, the credibility of the GIS database is at stake if the data is not current.

Usually, the effort required to maintain the database is as much as, or more than that required to create it. This ongoing maintenance work is usually assigned to in-house personnel as opposed to a contractor. The entire process should be planned well in advance. Once again, the equipment and personnel must be ready to take over the maintenance of the database when the data conversion effort and database building processes are complete.

Database maintenance requires two supporting efforts: ongoing user training and user support. Ongoing user training is needed to replace departing users with newly trained personnel. This will enable the data maintenance to be carried out on a continuous and timely basis. It is also important to offer advanced training to existing users to provide them with the opportunity to improve their skills and to make better use of the system.

GIS is a complicated technology, making operating problems inevitable. User support will help users solve these problems quickly. It will also customize the GIS software to enable them to execute processing tasks more quickly and more efficiently. User support is usually provided by in-house or contract programmers. It requires a knowledge of the operating system and macro programming language as well as troubleshooting common command and file problems.

6 GIS DATA SHARING COOPERATIVES

The establishment of data sharing cooperatives within the public sector is a cost-effective means of database development and maintenance which is encouraged. Cooperative-multiparticipant database projects allow for data exchange, and the opportunity to create new means for developing, maintaining, and accessing information. The sharing of data in the public sector, especially between government agencies and offices which are funded by the same financial resources, should be expected. It does not make fiscal sense for public funds to be utilized in the development of two GIS databases of the same geographic area for two different agencies. Benefits of data sharing thus would include: the development of a much larger database, for far less cost; the development of more efficient interaction between public agencies; and through the utilization of a single, seamless database the availability of more accurate information, since all agencies would share the same, up-to-date information. Following pages represents a matrix which indicates in general opportunities for data sharing between municipal operating units/functions.

The goal of a data sharing strategy is to maximize the utility of data while minimizing the cost to the organization. It is important that all parties involved have clear and realistic expectations as well as common objectives to make the data sharing work. Under any circumstance, however, database management and maintenance will require us to redefine our relationships with those we routinely exchange data with, whether they are within an organization or part of a multiparticipant effort including outside agencies. Work flow and information flow must be reviewed and changed if necessary. Procedures and practices for the timely exchange and updating of data must be put in place and data quality standards adhered to, whether it be hard copy data which must be converted for inclusion or digital files which might be available for importing to our system. Systematic collection and integration of new and/or update data must be employed in order to safeguard the initial investment, maintain the integrity of the database and assure, system reliability to meet function needs.

GIS DEVELOPMENT GUIDE: DATABASE CONSTRUCTION

1 INTRODUCTION

Scope Of Database Construction

A database construction process is divided into two major activities

- creation of digital files from maps, air photos, tables and other source documents;
- organization of the digital files into a GIS database.

This guideline document describes the first process, digital conversion, and the subsequent guideline entitled "GIS System Integration" deals with the organization of the digital files into a database.

Figures 1 and 2 are two versions of the digital data conversion process (Burrough, 1986; and Montgomery and Schuch). Only the second half of figure 2 describes the actual digital conversion process, the first half identifies previous planning activities. In both figures, the end product(s) are digital data files which, if passed through quality control, are suitable for inclusion in the GIS database.

Steps in creating a topologically correct vector polygon database

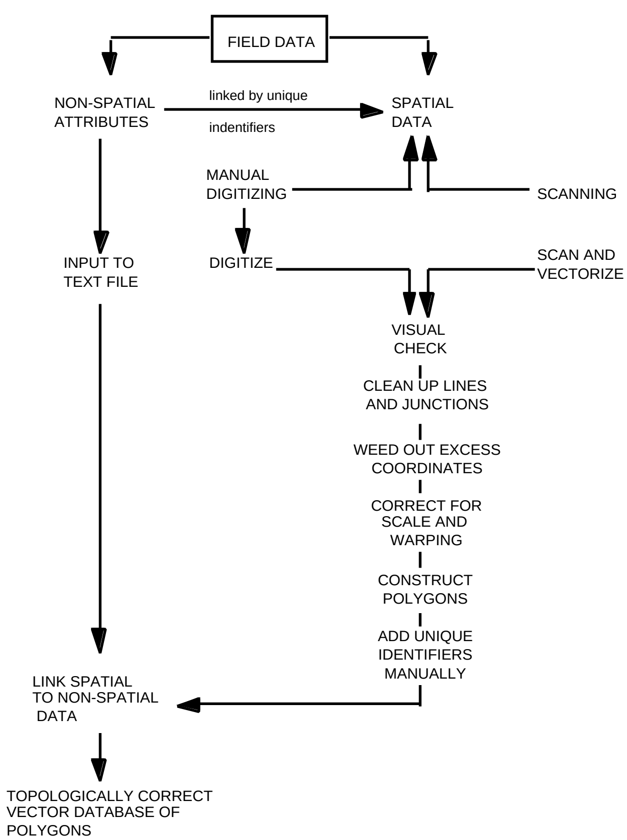


Figure 1 - Source: ***Principles of Geographic Information***

Systems for Land Resources Assessment, **Burrough, P.A., 1986**

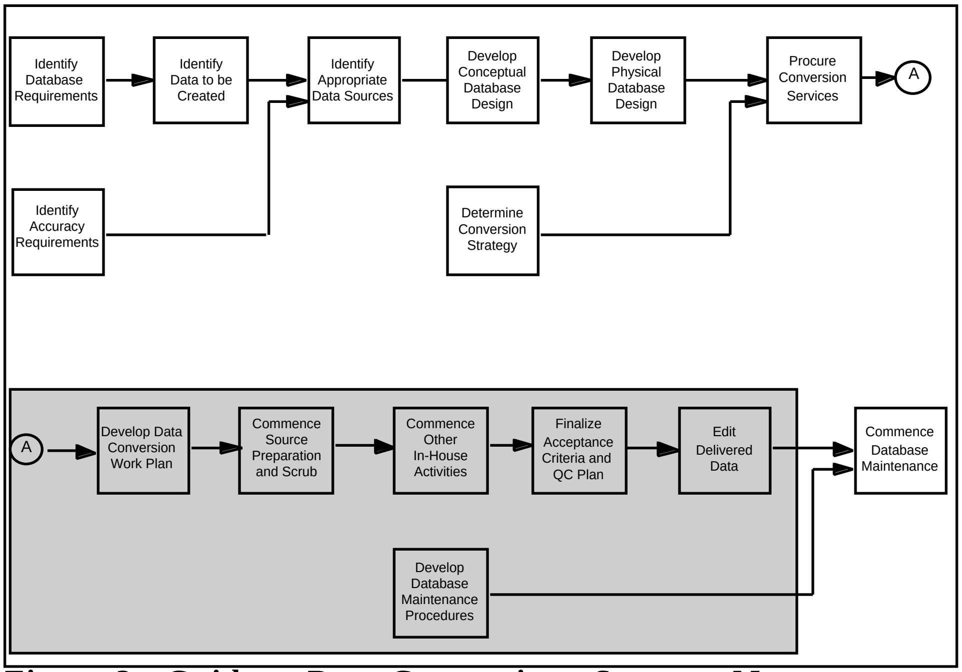


Figure 2 - Guide to Data Conversion. Source: ***Montgomery and Schuch***

2 INFORMATION REQUIRED TO SUPPORT DATA CONVERSION PROCESS

Data Model

GIS technology employs computer software to link tabular databases to map graphics, allowing users to quickly visualize their data. This can be in the form of generating maps, on-line queries, producing reports, or performing spatial analysis.

To briefly summarize the characteristics of GIS software and the data required for operations, we offer the following diagram:

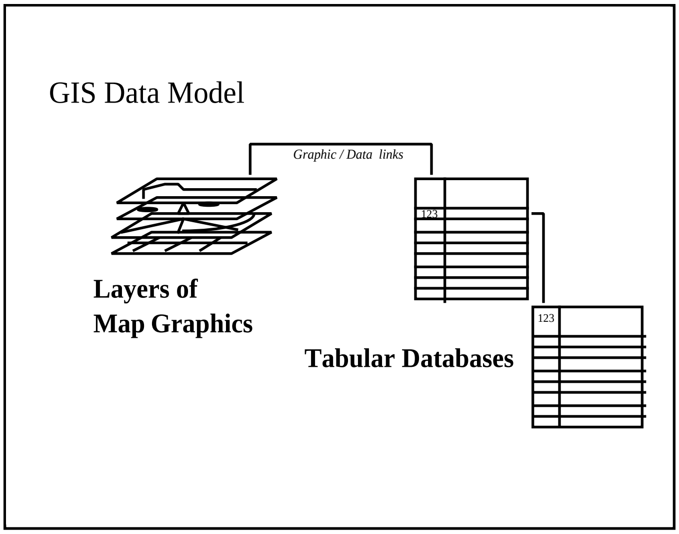


Figure 3 GIS Data Model

GIS (Spatial) Data Formats

In digital form, GIS data is composed of two types: *map graphics (layers)* and *tabular databases.*

- Map graphics represent all of the features (entities) on a map as points, lines, or areas, or pixels.
- Tabular databases contain the attribute information which describe the features (buildings, parcels, poles, transformers, etc.).

GIS data layers are created through the process of *digitizing*. The digitizing process produces the digital graphic features (point, line or area) and their geographical location. Tables can be created from most database files and can be loaded into a GIS from spreadsheet or database software programs like Excel™, Access™, FoxPro™, Oracle™, Sybase™, etc. A common key must be established between the map graphics and the tabular database records to create a link. This link is usually defined during the “scrubbing” phase (data preparation) and created during data capture (digitizing). For parcel data, the parcel-id or SBL number (section, block and lot) is a good example of a common key. The map graphic (point or polygon) is assigned an SBL number as it is digitized. The database records are created with an SBL number and other attributes of the parcel (value, landuse, ownership, etc.).

Raster and Vector Format

GIS allows map or other visual data to be stored in either a *raster* or *vector* data structure:

There are two types of raster or scanned image: 1) remotely sensed data from satellites; and 2) scanned drawings or pictures. Satellite imagery partions the earth's surface into a uniform set of grid cells called pixels. This type of GIS data is termed raster data. Most remote sensing devices record data from several wave-lengths of the electromagnetic spectrum. These values can be interpreted to produce a "classified image" to give each pixel a value that represents conditions on the earth's surface (e.g., land use/land cover, temperature, etc.). The second type of scanned image is a simple raster image where each pixel can be either black or white (on or off) or can have a set of values to represent colors. These scanned images can be displayed on computer screens as needed.

Raster data is produced by scanning a map, drawing or photo. The result is an array of pixels (small, closely packed cells) which are either turned "on" or "off." A simple scanned image, for example, in TIFF (Tagged Image File format) format, does not have the ability to be utilized for GIS analysis, and is used only for its display value. The "cells" of the digital version of the image do not have any actual geographical nature as they represent only the dimensions of the original analog version of the image. Raster data in it’s most basic form is purely graphical and has no “intelligence” or associated database records.

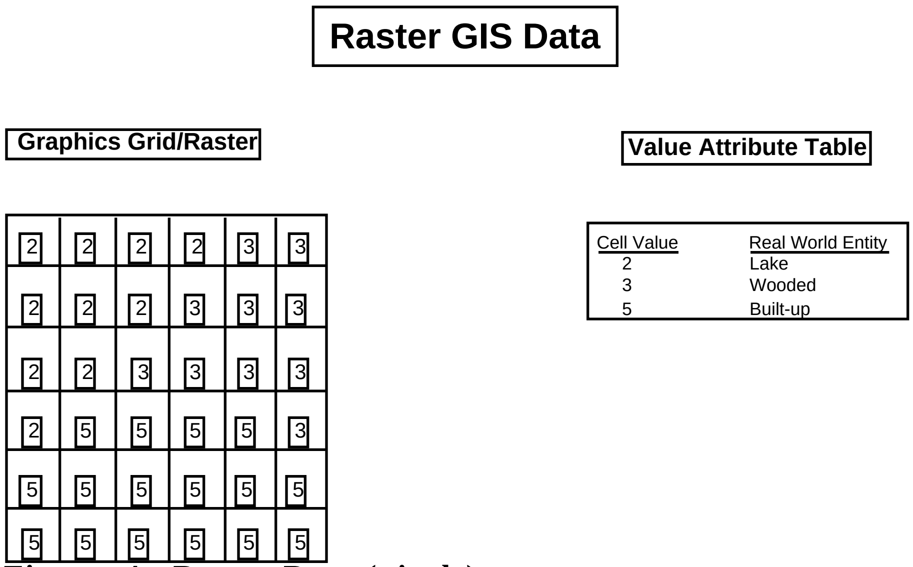


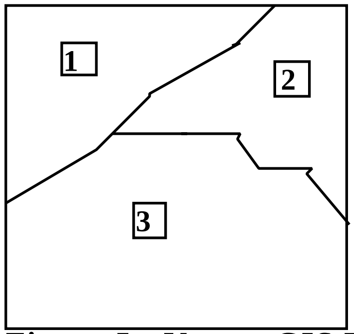
Figure 4 - Raster Data (pixels)

Raster data can be enhanced to provide spatial analysis within a GIS. Pixels or cells represent measurable areas on the earth's surface and are linked to attribute information. These cells are assigned numeric values which correspond to the type of real-world entity which is represented at that location (e.g., cells containing value "2" may represent a lake, cells of value "3" may represent a particular wooded area, etc.).

- *Vector data* represents map features in graphic elements known as points, lines and polygons (areas).

Vector GIS Data

Vector GIS Polygon Layer



Polygon Attribute Table

Polygon Number	Identity Attribute
1	Lake
2	Wooded
3	Built-up

Figure 5 - Vector GIS Data

Vector graphics coordinates are represented as single, or a series of, xy-coordinates. Data is normally collected in this format by tracing map features on the actual source maps or photos with a stylus on a digitizing board. As the stylus passes over the feature, the operator activates the appropriate control for the computer to capture the xy-coordinates. The system stores the xy-coordinates within a file. Vector data can also be collected on-screen (called "heads-up" digitizing), by tracing a scanned image on the computer screen in a similar manner.

3 DATA CONVERSION TECHNOLOGIES AVAILABLE

Manual Digitizing

Manual digitizing involves the use of a digitizing tablet and cursor tool called a “puck,” a plastic device holding a coil with a set of locator cross-hairs to select and digitally encode points on a map. A trained operator securely mounts the source map upon the digitizing tablet and, utilizing the cross hairs on the digitizing “puck,” traces the cross hair axis along each linear feature to be captured in the digital file. The tablet records the movement of the puck and captures the features’ coordinates. The work is time-consuming and labor intensive. Concentration, skill and hand-eye coordination are crucial in order to maintain the positional accuracy and completeness of the map features.

Traditional data conversion efforts are based on producing a vector data file compiled by manually digitizing paper maps. Vector data provides a high degree of GIS functionality by associating attributes with map features, allowing graphic selections, spatial queries and other analytical uses of the data. Vector data also carries with it the highest costs for conversion. The industry average for a complete data conversion project to digitize parcel lines, dimensions and text is between $3.00 - $5.00 per parcel. The price is determined by the complexity and amount of data. To keep costs down, data can be selectively omitted from conversion (i.e. not all text and annotation will be captured). The resulting vector data can reproduce a useful, albeit more visually stark version of the original map. A bare bones data conversion project can be conducted by digitizing only the linework from the tax maps. The minimum industry cost for digitizing parcel line work with a unique ID only is between $1.00 to $1.50 per parcel.

Scanning

Scanning converts lines and text on paper maps into a series of picture elements or “pixels.” The higher the resolution of the scanned image (more dots per square-inch), the smoother and more accurately defined the data will appear. As the dots per inch (DPI) increases, so does the file size. Most tax maps should be captured with a scan resolution of 300-400 DPI. One of the main advantages to scanning is that the user sees a digital image that looks identical to their paper maps -- complete with notes, symbology, text style and coffee stains, etc. Scanning can replicate the visual nature of the original map at a fraction of the cost of digitizing. However this low cost has a "price". The raster image is a dumb graphic -- there is no “intelligence” associated with it, i.e. individual entities cannot be manipulated. Edge-matching and geo-referencing the images (associating the pixels with real world coordinates) improves the utility of the scanned images by providing a seamless view of the raster data in an image catalog. Scanned images require more disk space than an equivalent vector dataset, but the trade-off is that the raster scanning conversion process is faster and costs less than vector conversion.

Raster to Vector Conversion

Scanned data, in raster format, can be "vectorized” (converted into vector data) in many high-end GIS software packages or through a stand-alone data conversion package. Vectorizing simply involves running a scanned image through a conversion program. In the vectorization process, features which are represented as pixels are converted into a series of X,Y points and/or linear features with nodes and vertices. Once converted within a GIS environment, the data is in the same format created using a digitizing tablet and cursor. Many vectorized datasets require significant editing after conversion.

Hybrid Solution

Since both vector and raster datasets have decided advantages and disadvantages, a hybrid solution capitalizes on the best of both worlds. Overlaying vector format data with a geo-referenced backdrop image provides a powerful graphic display tool. The combined display solution could show the vector map features and their attributes (also available for GIS query), and an exact replica of the scanned source material which may be a tax map or aerial photography. If needed, individual parcels, pavement edges, city blocks or entire maps can be vectorized from the geo-referenced scanned images. This process is called incremental conversion. It allows the county to convert scanned raster data to vector formatted data on an as-needed basis. There are a plethora of raster to vector conversion routines on the market, but it is important that the conversion take place in the same map coordinate system and data format as your existing data. The key advantage to the hybrid approach is this: even after full vectorization, the scanned images continue to provide a higher quality graphic image as a visual backdrop behind the vector data.

Entry of Attribute Data

Additional attribute data can be added to the database by joining a table which contains the new attributes to an existing table already in the GIS. To join these tables together a common field must be present. Most GIS software can then use the resulting table to display the new attributes linked to the entities. There are various sources for building an attribute database for a GIS. From CD-ROM telephone and business market listings with addresses, to data which is maintained in various government databases in “dbase” or various other database formats.

Acquisition of External Digital Data

The availability of existing digital data will have an effect upon the design of the database. Integrating existing databases with the primary GIS will require the establishment of common data keys and other unique identifiers. Issues of data location, data format, record match rates, and the overall value of integrating the external data should all be considered before deciding to purchase or acquire existing datasets.

GIS Hardware And Software Used in Digital Data Conversion

Most contemporary GIS software packages are structured to operate on computer workstations to accomplish digitizing and editing tasks.
Four basic types of workstations can be identified:

- A digitizing station, a workstation which is connected to a precision digitizing tablet, which utilizes a high-resolution display terminal, and which also has all of the analysis functions necessary for querying, displaying and editing data
- An editing workstation, which is used for conducting most of the QA/QC functions of the conversion process, having all the functionality of the digitizing station except for the ability to digitize data via a digitizing tablet
- Graphic data review/Tabular data input workstations are used for displaying and reviewing graphic data, and for the entering of tabular attribute data associated with these features
- X Terminals are the fourth type of workstation and these allow for graphic display and input of data utilizing the X Window System communications protocol.

With the increasing power of today's personal computers, many GIS analysis packages are being designed for PC's. As GIS data files are very large, PC-based GIS packages usually require a PC with minimum requirements including a 486 processor and 16 megabytes of RAM. Hard-drive disk space depends upon how large the datasets are which are being used. A safe bottom-line for hard-drive space with a PC is 500 megabytes. For most data conversion projects, much more hard-drive space will be needed in order to store data as they are converted. Tape storage hardware is also necessary in order to efficiently backup the many megabytes of files created in the conversion process. Just to provide an idea of the storage requirements necessary for basic scanning conversion, the file-size of one tax map alone, in (Tagged Image File Format or TIFF) image format, scanned at a 500 dots per inch (dpi) resolution, can range anywhere from 1-3 megabytes alone.

Digitizing hardware requirements vary according to the conversion approach which is applied. For vector conversion, a digitizing tablet will be necessary in usually a manual digitizing process. Another piece of digitizing hardware, a scanner, is used to create raster images. Automatic digitization, through the use of a scanner is a very popular approach for capturing data. Raster data can subsequently be transformed into vector data in most turn-key GIS packages, through the use of raster-to-vector conversion algorithms.

After the conversion of map data into digital form, hardware will be needed for outputting digital data in hardcopy format. When handling a data conversion project, a necessary piece of output hardware is a pen or raster plotter. GIS software allows for the creation of plots at any viewscale. The plotter, with its ability to draw on a variety of materials (including paper, mylar and vellum), allows for the creation of quality map plots. Most plotters usually have a minimum width of three feet. Vector and raster plotters are both available on the market. Vector, pen plotters utilize various pens for the drawing of linear features on drawing media. Pen plotters can handle most plotting jobs, but they do not produce good results in area shading such as in the production of cholorpleth maps. Raster plotters, on the other hand, are excellent in producing shading results. Raster plotters usually cost more than vector plotters, but are substantially more versatile and have better capabilities.

Other output devices for the creation of hardcopies of GIS data include: screen copy devices, used for copying screen contents onto paper without having to produce a plot file; computer FAX (facsimilie) transmissions, often used in communications between conversion contractors and clients, produce small letter-size plots, and the fax transmission files (as raster images) can be saved and viewed later; printers are used to output tabular data which is derived from the GIS, and if configured correctly, can produce small letter-size plots.

Pilot Project/Benchmark Test Results

The pilot project is a very important activity that precedes the data conversion project. The pilot project allows you, the GIS software developer, and the data conversion contractor the ability to test and review the numerous steps involved in creating the database. Defining the pilot study area involves selection of a small geographic area which will allow for a high degree of being successful, that is, that it will be completed in a relatively short period of time and will allow for the testing of all project elements which are necessary (conversion procedures, applications, database design). Test results which are obtained from the pilot project usually include assessments of: database content, conversion procedures, suitability of sources, database design, efficiency of prepared applications on datasets, the accuracy of final data, and cost estimates.

Identified Problems With Source Data

The pilot study involves testing and finding successes and problems in procedures and designs for the GIS. It involves looking for problems that occur due to lack of, or inadequacy in, source data. It is important to identify problems especially at the source data level since it is usually the easiest and cheapest to correct errors prior to data conversion.

When evaluating the results of a pilot study, problems with digital data accuracy resulting from source data flaws, are bound to arise. Usually, the source data used for a project are not in the proper format required for the best possible data result. For example, problems may arise when the source data for a certain data layer consists of maps which are at various scales. These various scale differences can create error when these digitized layers are joined into a single layer. Other problems arise when there are not adequate control points found upon map sheets in order to accurately register coverages while they are being digitized. At times, even adjacent large-scale source map sheets may have positional discrepancies between them. Such inconsistencies will be reflected in the corresponding digital data. Procedures for dealing with all known source data problems need to be specified prior to the start of data conversion.

4 DATA CONVERSION CONTRACTORS

Firms Available And Services Offered

There are different types of firms which can handle GIS data conversion. There are some firms which specialize in GIS data conversion, and sub-contract out the services of other firms as needed. Some other firms which handle data conversion but do not particularly specialize in data conversion alone include: aerial mapping firms, engineering firms and GIS vendors. Various firms will offer standard data conversion services, but based upon their main type of work, may offer some unique services. For example, a firm specializing in GIS data conversion may have a wide variety of software options which the client company can choose from. Such a firm usually will have numerous digitizing workstations and a large staff, and be able to complete the project in a shorter period of time than other firms which do not particularly specialize in GIS data conversion. If needed, a specialized GIS data conversion company could subcontract services from another company.

Aerial mapping firms can offer many specialized data conversion services associated with photogrammetry, which will not be available directly through a general data conversion contractor. Many aerial mapping firms now have considerable expertise with the creation of digital orthophoto images, rectified and scaled scans of aerial photography, which can be displayed and utilized with vector data. Engineering and surveying firms are well-equipped to deal with most data conversion projects, and will usually have a major civil engineering/surveying unit within the organization. These firms usually will focus upon certain aspects of GIS systems and approach conversion projects with stress upon the extent of construction detail, positional accuracy requirements, COGO input, scale requirements and database accuracy issues. At times, GIS software vendors will handle data conversion projects in order to test their software in benchmark studies and pilot projects.

The main conversion services which are usually offered include: physical GIS database design and implementation, deed research, record compilation, scrubbing, digitizing, surveying, programming and image development and registration.

Approximate Cost of Services

Outsourcing data conversion with data purchase/ownership

CONVERSION METHOD	PER-PARCEL COST
Manually digitized vector data (linework alone)	$1.20 / Parcel
Manually digitized vector data (linework & annotation)	$5.00 / Parcel
Vector data developed from the vectorization of scanned maps (linework & Annotation)	$3.00 / Parcel

Raster image data (registered to a coordinate system)	$50. / map = $0.55 / Parcel

Outsourcing Data Conversion and Licensing Data

CONVERSION METHOD	PER-PARCEL COST
Manually digitized Vector Data (Linework and Annotation)	$1.50 / Parcel

(No cost estimates are available for Raster Data)

(Note: All of the above cost estimates are based upon average prices offered by various data conversion vendors)

Making Arrangements For External Data Conversion

There are a number of ways of obtaining the digital conversion of map data. Arrangements are usually made through the development of a Request for Proposal (RFP), and then evaluating the proposals submitted by various conversion contractors. Some of the criteria which are desired in selecting a conversion contractor include: the company's technical capability, the company's experience with data conversion, the company's range of services, location, personnel experience and the overall technical plan of operation. Balanced with all of these items is usually the organization's budget and the costs associated with the project.

5 DATA CONVERSION PROCESSES

Digital Conversion Of Mapped Data

Digital data conversion of mapped data is a costly and time-consuming effort. The more closely the digital data reflects the source document, and the more attributes are associated with the map features, the higher the map utility but also the higher the cost of conversion. Because of the high cost of digitizing all graphic map features, and text/graphic symbology, conversion efforts may compromise data functionality by limiting the number of features captured in order to keep costs down. The actual processes involved with digital conversion of mapped data are usually the most involved, and most time-consuming of all. These two traits together explain why data conversion is usually the highest cost of implementing the GIS.

Planning The Data Conversion Process

The data conversion process needs to be planned effectively in order to minimize the chance of data conversion problems which can greatly disrupt the normal workflow of the organization. It is necessary to plan all of the physical processes which will be involved in data conversion and to develop time-estimates for all work. These main processes include:

- Specifications
- Source map preparation
- Document flow control
- Supervision plans
- Problem resolution procedures

These procedures allow for the efficient conversion of mapped data. Guidelines for normal data capture procedures such as scanning and table digitizing should be developed to ensure that all data are consistently digitized. Particularly when an organization is conducting conversion in-house, a small amount of time invested in developing error prevention procedures will greatly benefit the organization by saving time in the correction/editing phase of the conversion. It is easier to prevent errors than to go ahead and try to correct them after the actual digitizing has been conducted.

Data Conversion Specifications: Horizontal And Vertical Control, Projection; Coordinate System, Accuracy Requirements

Any discussion about data conversion should start with the topic of accuracy. We've all heard the expression, "Garbage In, Garbage Out." Without the ability to meet the proper accuracy standards established early in a GIS conversion project, the resulting GIS may be useless based upon its lack of accuracy. Even still, in reality, when building a GIS and handling data conversion, we are faced with a variety of source documents which may each carry a different scale, resolution, quality and level of accuracy. Some source map data may be so questionable that it should not be loaded into the GIS. Extracting reliable data later-on from the GIS will depend upon either the converting of data from reliable source documents, or the development of new data "from scratch."

Map projections affect the way that map features are displayed (as they affect the amount of visual distortion of the map), and the way map coordinates are distributed. Before any GIS graphic data layers will be ready for overlay functions, the layers must be referenced to a common geographic coordinate system. GIS software can display data in any number of projection systems, such as UTM (Universe Transverse Mercator), State Plane Coordinate Systems, and more. For scanned maps and aerial photos (which are simple non-GIS raster images), to be displayed effectively with vector data, the images need to be registered and rectified to the same coordinate system.

Establishing specific requirements for map accuracy should be done at the beginning of a project. If a certain level of accuracy is desired, it is this level which will have to be developed in future aspects of the project. Procedures should be standardized in order to ensure the best and most consistent results possible.

Source Map Preparation (Pre-Digitizing Edits)

Preparing the analog data that will be converted is an important first step. This needs to be done whether the data will be scanned or digitized, and whether you are outsourcing the work or completing it in-house. This pre-processing is also referred to as “scrubbing” the data. The process involves coding the source document using unique ID’s and/or using some method to highlight the data that should be captured from these documents. This makes it clear to the person performing the scanning or digitizing what they should be picking up. It will also be important later for performing quality control checks and to make sure that the digital data has a link to the attribute database needed for a GIS.

Document Flow Control

Without a clear system for monitoring and planning the flow of map (and attribute data) documents between the normal storage locations of map documents and those parties handling the actual data conversion, problems will usually arise in tracking the location of maps. When a large number of maps are being converted, it is important to maintain a full understanding between both the conversion contractor or in-house conversion staff, and the normal user group of the source documents about exactly which documents are being handled, and at what time. Source maps are delivered to the conversion group or contractor as a work packet, usually consisting of a manageable number of maps of a certain geographic region, which is pre-determined within the data conversion workplan. A scheme for tracking packets of source documents, as well as the resulting digital files is needed. This scheme should be able to track the digital file through the quality control processes.

In addition to tracking the flow of documents and digital files through the entire data conversion process, a procedure needs to be established for handling updates to the data that occur during the conversion time period. This change control procedure may be quite similar to the final database maintenance plan, however, it must be in place before any of the data conversion processes are started. Also, if this procedure will likely be very different from the previous manual map updating methods used and may involve substantial restructuring of tasks and responsibilities within the organization.

Supervision Plans (Particularly For Contract Conversion)

When planning the data conversion process, it is important that attention be given to the development of detailed plans for supervising the data conversion process. Supervisory plans allow the organization to distribute responsibility for the many different facets of the data conversion project. When data conversion has been contracted out, it is important that communication be maintained between the client company and the contractor. The development of specific variations normal administrative tools used for scheduling and budget control can be very useful (e.g., CPM/PERT scheduling procedures; GANTT charts, etc.)

Problem Resolution Procedures

In order to ensure the efficient progress of all aspects of the data conversion project, it is important to develop formal procedures for problem resolution. Editing procedures and data standards should be developed for such items as: major and minor positional accuracy problems; inaccurate rubber-sheeting, or map-joining/file-matching problems; attribute coding errors, etc. Other procedures for events such as missing source data, handling various scale resolution issues, and even hardware and software system problems should also be created. Establishing such procedures and assigning responsibilities for resolution are extremely important, particularly when outside contractors are involved.

Converting The Data

As stated earlier, it is important to follow consistent pre-established procedures in the actual digitizing of the datasets. Consistently using a tested and approved set of conversion guidelines and procedures will eliminate any chance of ambiguity in methods. Using established procedures will allow for the most consistent product possible.

Reviewing Digital Data

The digital data review process involves three issues:

- data file format and format conversion problems
- data quality questions
- data updating and maintenance

The review process must first be handled before the decision to rely on other digital data sources is made. Additionally, formal data sharing agreements should be made between the two organizations.

Quality Control (Accuracy) Checking Procedures

A quality assurance (QA) program is a crucial aspect of the GIS implementation process. To be successful in developing reliable QA methods, individual tasks must be worked out and documented in detail. Data acceptance criteria is a very important aspect of the conversion program, and can be a complex issue. A full analysis of accuracy and data content needs will facilitate the creation of documentation which may be utilized by the accuracy assessment team.
A combination of automatic and manual data verification procedures is normally found in a complete QA program. The actual process normally involves validation of the data against the source material, evaluation of the data's utility within the database design, and an assessment of the data with regard to the standards established by the organization handling the conversion project. Automated procedures will normally require customized software in order to perform data checks. Most GIS packages today have their own macro programming languages which allow for the creation of customized programs. Some automated QA procedures include: checking that all features are represented according to conversion specifications (e.g., placed in the correct layer); features requiring network connectivity are represented with logical relationships, for example, two different diameters of piping or two different gauges of wire must have a connecting device between them which should be represented by a graphic feature with unique attributes;

relationships of connectivity must be maintained between graphic features (Montgomery and Schuch, 143).

Manual quality control procedures normally involve creating and checking edit plots of vector data against source map data. QA requirements which will have to be met include: absolute/relative accuracy of map features should be met and all features specified on the source map should be included on the edit plot; map annotation should be in required format (e.g., correct symbology, font, color, etc.) and text offsets should be within specified distance and of correct orientation; plots of joined datasets should have adequate edge matching capability (M&S, 145).

Final Correction Responsibilities

Quality control editing of the digitized product is a crucial step in preparing spatial feature data. After initially digitizing a data layer, an edit plot is produced of those digitized features. The edit plot is a hard-copy printing of the digitized features. The edit plot is printed at the same scale as the source data and checked by overlaying the plot with the source map on a light table. This edit check allows for the determination of errors such as misaligned or missing features. Corrections may then be made by adding or deleting and re-digitizing features. When on-screen digitizing, feature placement errors may be corrected by “rubbersheeting” the graphic features to fit the source data. Rubbersheeting is the process of stretching graphic features through the establishment of graphic movement “links” with a from-point (where the feature presently is located), and a to-point (where the feature should be placed). GIS graphic manipulation routines then move graphics according to these specified links.

File Matching Procedures (Edge Match, Logical Relationships Within Data, Etc.)

Files which are going to be spatially joined must first have adequate edge-matching alignment of their graphic features. This entails a number of basic GIS graphic manipulation procedures: (1) coordinate transformation, which projects the data layer into its appropriate real-world coordinates; (2) rubbersheeting of the graphic features in one data file to accurately coincide with the adjacent graphic features in another file; (3) spatial joining, the combining of two or more data files into one seamless file spanning the geographic area of all files.

Coordinate transformation is the process of establishing control points upon the digitized layer and defining real-world coordinates for those points. A GIS coordinate transformation routine is then used to transform the coordinates of all features on the data layer based upon those control point coordinates. Once transformed, spatially adjacent data layers may then be displayed simultaneously within their combined geographic extent. A determination may then be made as to the effectiveness and accuracy of the coordinates assigned to the data layers. If necessary, graphic features found in both data layers may be rubbersheeted to better align features which will need to be connected. For example, if the endpoint of a graphic feature representing a street centerline is not reasonably close to its corresponding starting point on the adjacent data layer, one or both of these graphic lines will have to be moved so that the graphic feature will connect. An alignment problem such as this can signal possible errors in the coordinate transformation and/or the source data. After features are accurately matched, the data files may be combined into a single data file. The combined data file will afterwards require editing and the development of new topological relationships in the new dataset. An example of one post-spatial join editing procedure is the removal of graphic line-connection points called "nodes" which may interfere with various elements of the attribute database.

Final Acceptance Criteria

Standards for appropriate quality assurance, and accuracy verification procedures in general, depend greatly upon the data sources, the schematics of the database for which data is being prepared, and the actual data conversion approaches applied. Acceptance of the joined digital map files depends upon the data's meeting certain criteria. Criteria usually relate to accuracy, such as the determination of whether the product meet National Map Accuracy Standards at the appropriate scale. Other criteria may relate to whether attributes are in order, if they have been added. Most acceptance determinations should be made on whether the feature data is meeting standards of accuracy, completeness, topological consistency, and attribute data content.

Building Main Database

One of the final stages involved in developing a GIS database involves putting all the converted data together. Establishing one uniform database involves entering all attribute and feature data into a common database with an established workable file/directory structure, sometimes known as a "data library." As the database is developed and data is ready for use, it can be released to the various data users for analysis. Once the database is designed, it then becomes important to maintain data accuracy and currency. If changes are made within the confines of the data layers, these changes must be defined and updates made to keep the integrity of the database. Subsequent guideline documents deal with data integration and database maintenance.

6 ATTRIBUTE DATA ENTRY

Source Documents

There are a number of source documents which can be utilized as data for the attribute database. Many organizations are able to utilize their existing electronic database files and import this data directly into their GIS database. In the case of paper files relating to geographic areas, and attribute data existing on paper maps, this data will have to be manually entered into GIS attribute data files in the form of tables. Before this information is entered into a database, it must first be reviewed and edited. It is also important to have a procedural plan designed for the entry of this data in order to coordinate the flow of these source documents.

Pre-Entry Checking And Editing

A review of GIS attribute source documents can oftentimes reveal an unorganized mass of maps, charts, tables, spreadsheets, and various textual documents. The checking and editing of source documents is handled in the scrubbing phase of the project. Without a specific plan designed for the entry of these various data elements, it is highly likely that error will be introduced into the GIS database. It is ***crucial that all source documents are readable and properly formatted*** to allow for the most efficient entry of numerical and textual data. If the database conversion is being outsourced, and the contractor is unable to read the source data, the resulting database will be inaccurate, more costly, or both. It is recommended that a formal scrub manual, designed according to the database and application requirements, be developed to help facilitate the supplementing of source data and its entry into the database. Logical consistency is an important element for both graphic and attribute elements. Records and attributes which are related to graphic elements within a network system must maintain logical relationships.

Document Flow Control

An organization will typically have a multitude of different document formats which it will need in coding all of its GIS attribute data. It is crucial that tracking mechanisms be implemented in preparation for the key entry process. Usually duplication of source documents which will be used in the key entry process will not be feasible. As many source documents to be key entered are used on a regular basis within the organization, it will be important to develop guidelines for tracking these documents if they are needed during the process. Timing and coordination will be factors in planning document usage.

Key Entry Process

As stated earlier, some organizations will be able to enter much tabular data into the database simply by way of importing existing tables or files into the GIS, or relating tables which exist in their external DBMS. Normally, it will be necessary to enter attribute data into the system utilizing a keyboard. Many organizations choose to use lists when entering data from the keyboard. It is much more efficient during conversion to enter a 2 or 3-digit code which has a reference list associated with it. Typing in a full description of the graphic into the text field takes longer, and increases the chance of typographical error.

Digital File Flow Control

Numerous files will result from the key entry process. These files will need to be given proper names and directory locations in order to track and prepare the data logically for use within the GIS.

Quality Control Procedures

Most databases allow the user to specify the type of field for each data element, whether it is numeric, alphanumeric date, etc; whether it has decimal places, and so on. This feature can help prevent mistakes as the system will not allow entries other than those specified in advance.

There are a number of automated and manual procedures which can be performed to check the quality of attribute data. Some customized programs may be required for the testing of some quality control criteria. Some attribute value validity checks which may be performed include: verifying that each record represents a graphic feature in the database, verifying that each feature has a tabular record with attributes associated with it, determining if all attribute records are correct, and determining that all attributes calculated from certain applications must be correct based upon the input values and the corresponding formulas. The translation of obsolete record symbology into a GIS usable format, according to conversion specifications, is one procedure which will have to be conducted manually (Montgomery and Schuch, 145).

The responsibility for checking and maintaining automated quality control procedures can be placed in the hands of the staff responsible for actual data conversion. When outsourcing data conversion, one of the most time-consuming aspects of the project is the evaluation of converted data once it has been received from the vendor. Usually, automated routines are developed which can be utilized in the evaluation of the datasets, and in determining if the data fulfills all of the requirements and standards stated in the contract. This process can be simplified by the client company delivering automated quality control checking routines to the data conversion vendor. The vendor is then able to run these routines, evaluate and edit the data so that it will meet requirements before it is even shipped to the client. Such a procedure saves valuable time and expenses which would otherwise have been spent on quality control evaluation, shipping and business communication.

Change Control

Final editing procedures and data acceptance are based upon whether major revisions in the data will need to be performed. After data verification and quality assurance checks, it may be necessary to again re-evaluate database design, technical specifications of the data, and conversion procedures overall. Ideally, the planning and design of the database will be sufficiently comprehensive and correct such that the logical/physical database design will not have to be modified. However, it is rare that a data conversion project will be able to push through to completion without some changes being necessary. Many conversion projects develop procedures which are used to identify, evaluate and then to approve or disapprove the final products. A form should be developed which is used to list desired changes which have been identified. The listing of desired changes is then evaluated in terms of both the volume of the data which has yet to be edited, and the amount of data which has already been converted. The conversion vendor will usually develop documentation which describes the estimated cost/savings which will be associated with the changes and final edits. Most organizations now accept the fact that changes will be a normal part of data conversion and change requests are usually expected. The challenge then lies in the methods by which change mechanisms are developed and agreed upon between client and vendor.

Final Acceptance Criteria

Acceptance criteria are the measures of data quality which are used to determine if the data conversion work has been performed according to requirements specified. In the case of outsourcing of conversion, these criteria will determine if the data has been prepared according to the contract specifications. If the data does not meet these specifications, the conversion contractor will be required to perform any necessary editing upon the data to reach acceptable standards. Acceptance criteria and standards may vary between organizations.

File Matching And Linking

In most GIS packages which utilize relational database technology, the file matching and linking is a fairly simple process. Most GIS packages contain straight-forward procedures for joining and relating attribute files, which normally entails the selection of the unique identifying key between the graphic feature attribute table and any other data attribute tables. Once the identifier-link has been specified, the GIS software automatically establishes the relationship between the tables, and maintains the relationship between them.

7 EXTERNAL DIGITAL DATA

Sources Of Digital Data

Digital spatial and attribute data can be found from a variety of sources. Various companies today produce “canned” digital spatial datasets which are ready for use within a GIS environment. Utilizing an existing database is a good way to supplement data in the conversion process and is one of the best ways to save money on the cost of producing a database. Most federal, state, and local government agencies have data which is available to the public for minimal cost.

Two of the largest spatial databases which are national in coverage include the US. Geological Survey’s DLG (Digital Line Graph) database, and the U.S. Census Bureau’s TIGER (Topologically Integrated Geographic Encoding and Referencing) database. Both systems contain vector data with point, line and area cartographic map features, and also have attribute data associated with these features. The TIGER database is particularly useful in that its attribute data also contains valuable Bureau of the Census demographic data which is associated with block groups and census tracts. This data is used today in a variety of analysis applications. Many companies have refined various government datasets, including TIGER, and these datasets offer various enhancements in their attribute characteristics, which increases the utility of the data. Unfortunately, problems associated with the positional accuracy of these datasets usually remain and are much more difficult to resolve.

Satellite and digital orthophoto imagery, raster GIS datasets, and tabular datasets are also available from various data producing companies and government agencies.

Transfer Specifications

Many government agencies produce spatial data which is in its own unique format. Many full-feature GIS packages have the ability to import government spatial datasets into data layers which are usable within their own environment. Some agencies or companies may produce their data in the most common data formats for government data in the transfer of their data (e.g. TIGER or DLG format). Such policies allow for easy transfer to various systems.

Quality Control Checks

Quality control checks on external datasets will be necessary. Many government datasets, although extensive in their geographic coverage and in the utility of the associated data, do not always have the most accurate or complete data, particularly in terms of positional accuracy. It is always advisable to be skeptical of a dataset's accuracy statement and compliance with standards and to fully test and evaluate the data before purchasing it or incorporating it into the database. Various automated and manual quality control procedures, discussed for both assessing cartographic feature and attribute characteristics should be utilized in a quality assurance evaluation of the external data.

8 ACCURACY AND FINAL ACCEPTANCE CRITERIA

Acceptance criteria determine to what standards data must comply in order to be usable within the system. Graphic acceptance standards for external digital data may be identified in three different cartographic quality types which include: relative accuracy, absolute accuracy and graphic quality. Standards for GIS data will normally depend upon the accuracy required of the dataset. In the GIS environment, accuracy will depend upon the scale at which the data is digitized, and at which scale it is meant to be used.

- Relative accuracy is basically a measure of the normal deviation between two objects on a map and is normally described in terms of + or - the number of measurement units (normally inches or feet) the feature is located apart from its neighboring map features, as compared to their locations in the real-world.
- Absolute accuracy criteria will evaluate the measure of the maximum deviation between the location of the digital map feature and its location in the real-world. Many organizations set their absolute accuracy standards based upon National Map Accuracy Standards.
- Graphic Quality refers to the visual cartographic display quality of the data, and pertains to aspects such as the data's legibility on the display, the logical consistency of map graphic representations, and adherence to common graphic standards. Placement and legibility of annotation, linework, and other common map elements all fall under graphic quality.

Informational quality is another accuracy criteria component which should be given much attention in building a database. Informational quality relates to the level of accuracy for both map graphic features and to their corresponding tabular attribute data. There are four basic categories for assessing these qualities:

- completeness
- correctness
- timeliness
- integrity

Together, these aspects of informational quality comprise the extent to which the dataset will meet the basic requirements for data conversion acceptance.

Completeness is an assessment of the dataset's existing features against what should currently be located within the dataset. Completeness may relate to a number of digital map features: annotation symbols, textual annotation, linework. Completeness will also relate to the attribute data, and whether all of the necessary attributes are accounted for. A typical requirement for the bottom limit of dataset completeness, when outsourcing conversion, is that not more than 1% of the required features and attributes will be missing from the digital dataset. For example, out of 80 roads that are located within a geographic area, if only 72 are included on the map, then only 90% of the data is included, and thus the map is only 90% complete.

Correctness is that quality which relates to the truth and full knowledge of the information contained. If a map shows a number of roads, and the linework is positioned correctly, but is not labeled correctly, there is a problem with correctness. Correctness applies both to map features and to attribute data. If a dataset has the positional accuracy, or the completeness in terms of placing an object, but does not have the correct label for that object, this is a problem with the correctness of the dataset. Evaluating correctness can be done through automated or manual procedures. Validation procedures are those which would be utilized in the testing of the datasets. An example of assessing correctness might include the matching of one dataset source against another to check for data accuracy from the various matching qualities. Every graphic and database feature has the potential for error.

Timeliness is another measure of informational quality, and it is a unique form of correctness. Timeliness is based upon the currency of a dataset, and if it is not up-to-date, or current, then the dataset must be of a specified age. The timeliness of a dataset begins from the date the dataset arrives at the client's door. From that point on, it is the responsibility of the client organization to maintain the data, and its currency.

The integrity of a dataset is a measure of its utility. Graphically, database integrity means that the dataset is maintaining its connectivity and topological consistency. In it, all lines are connected, there are no line overshoots or undershoots, and all feature on the display are representative of real-world features. In order to maintain database integrity, there should not be any missing or duplicate records or features.

GIS DEVELOPMENT GUIDE: PILOT STUDIES AND BENCHMARK TESTS

1 INTRODUCTION

Prior to making a commitment to a new technology like GIS, it is important to consider testing concepts and physical designs for development of such a system within a local government. This can be done by performing a pilot study to determine if GIS can be useful in the daily conduct of business and, if so, further conducting a benchmark test to determine the best hardware and software combination to meet specific needs.

Numerous GIS pilot studies and benchmark tests have been conducted by local governments within the state and across the nation. Decisions on deployment of GIS should not be based solely on other experience. Managers and end users respond best to relevant local data and actual applications, and will learn more readily if they have first hand experience defining and conducting a pilot study on benchmark test in-house.

2 PILOT STUDY: PROVING THE CONCEPT

Planning a Pilot Study

A pilot study provides the opportunity for a local government to evaluate the feasibility of integrating a GIS into the day-to-day functions of its' operating units. Implementing GIS is a major undertaking. A pilot study provides a limited but useful insight into what it will take to implement GIS within the organization. ***Proving the concept, measuring performance, and uncovering problems during a pilot study, which runs concurrent with detailed system planning, database planning, and design, is more beneficial than pressing forward with implementation without this knowledge.***

To maximize the usefulness of the pilot study, it must be planned and designed to match the organizations work flow, functions, and goals as described in the GIS needs assessment. The pilot study will be successful if it has the support and involvement of upper management and staff from the outset. This involvement will provide the opportunity to evaluate management and staff ability to learn and adopt new technology.

Objectives of a Pilot Study

A pilot study is a focused test to prove the utility of GIS within a local government. It is not a full GIS implementation nor is it simply a GIS demonstration; but rather a test of how GIS can be deployed within an organization to improve operations. It is the platform for testing preliminary design assumptions, data conversion strategies, and system applications. A properly planned and executed pilot study should:

- create a sample of the database
- test the quality of source documents
- test applications
- test data management and maintenance procedures
- estimate data volumes
- estimate costs for data conversion
- estimate costs for staff training

The pilot study should be limited to a small number of departments or GIS functions and a small geographic area. The pilot study should be application or function driven. Even though data conversion will take a major portion of the pilot study development time, it is the use of the data that is important. What the GIS can do with the data proves the functionality and feasibility of GIS in local government. The Needs Assessment document has identified applications, data required, sources of data, etc. In addition, a conceptual database design has been previously developed. Following is a list of procedures for carrying out a pilot study:

- select applications from needs assessment
- determine study area
- review conceptual database design
- determine conversion strategy
- develop physical database design
- procure conversion services and develop conversion work plan
- commence source preparation and scrubbing
- develop acceptance criteria and qc plan
- develop data management and maintenance procedures
- test application
- evaluate and quantify results
- prepare cost estimates

Selecting Applications to Include

Care must be taken to select a variety of applications appropriate to test the functional capabilities of GIS and the entire database structure. A review of the Needs Assessment report should provide selective applications to meet these requirements. Make sure to include data administration applications along with end user/operations applications. Data loading, backups, editing and QC routines have little user appeal, but they represent important functions that the organization will rely on daily to update and maintain the GIS database.

Selecting Data

Data to be tested in the pilot study can either be purchased from external sources or converted from in-house maps, photos, drawings, documents and databases. In any event, the data should represent the full mix and range of data expected to be included with the final database. It should include samples of archived or legacy system records and documents if they are planned to be included in the GIS in the future. All potential data types and formats should be considered for the pilot. This is the chance to test the whole process of integrating and managing data, together with the utility of the data in a GIS environment and different conversion and compression methods, before final decisions are made.

Spatial Extent of the Pilot Study

Selection of the study area should address several issues:

- Data density
- Representative sampling
- Seamless vs. sheet-wise conversion or storage

Choose an area (or areas) of interest that represents the range of data density and complexity. Make sure that all data entities to be tested exist in the area of interest. This will provide a representative dataset and allow the extrapolation of data volumes and conversion costs for the range of data over the entire conversion area.

To measure hardware performance the selected area should be chosen to match the file or map sheet size the end user will normally work with. Be aware that even if the data is currently represented as single map sheets at a variety of scales, the GIS will store the data as a "seamless" dataset.

Preliminary Data Conversion Specifications

A set of data conversion specifications need to be defined for each of the required data layers in the test datasets. The conversion specs need to address....

- Accuracy
- Coverage
- Completeness
- Timeliness
- Correctness
- Credibility
- Validity
- Reliability
- Convenience
- Condition
- Readability
- Precedence
- Maintainability
- Metadata

The foundation of the GIS is derived from the conversion process which creates a topologically correct spatial database. The following diagram identifies in detail the steps necessary to create this database.

Steps in creating a topologically correct vector polygon database

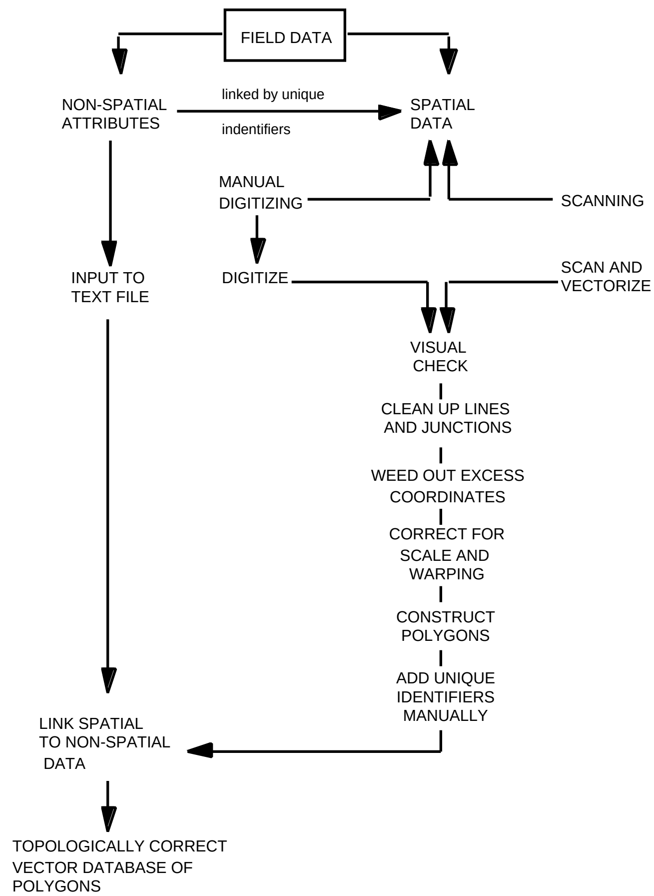


Figure 1 - Source: ***Principles of Geographic Information Systems for Land Resources Assessment,*** **Burrough, P.A., 1986.**

Selecting GIS Hardware and Software

To provide for continuity and to minimize added expense for total system development, select the most likely choice of hardware and software based on the database design specifications, and purchase or borrow that necessary for the pilot study from the hardware and software vendors.

Selecting a Data Conversion Vendor

Even though this is only a pilot study, it also serves as a test of likely suppliers of hardware, software and data conversion services. Therefore, a respectable data conversion vendor should be selected to perform the work, and prior uses of the vendor services should be contacted to confirm their ability to meet expectations. It shouldn't matter what method the conversion vendor uses to convert the data. Be open to suggestions from the potential conversion vendors as to the most cost effective methods to convert the data. As long as you get the data in the correct and usable format to satisfy your database plans, the method for data conversion used should not be an issue.

However, you will get much better results if the vendor has first hand experience with the chosen GIS software and the data conversion takes place in the same GIS software package. There is always a chance of losing attributes or inheriting coordinating precision errors converting from one format to another.

Defining Criteria for Evaluating the Pilot Study

The pilot study performance must be evaluated in measurable terms. By its very name, a pilot study implies an initial investigation. An investigation implies a set of questions to ask and a set of answers to achieve. For clarity, the questions can be addressed to match the major component of GIS plus others as needed.

Database

- Were adequate source documents available and was their quality sufficient?
- How much effort was involved in "scrubbing" the data before conversion?
- How long did the conversion process take?
- Were there any problems or setbacks?
- Was supplemental data purchased, if so, what was the cost?
- Did the data model work for each layer as defined?
- Was the data adequate (i.e. all data elements populated)?
- What errors were found in the data (closure, connectivity, accuracy, completeness, etc.)

Applications

- Were the applications written as specified
- Did the applications fit smoothly in the GIS or was a separate process invoked?
- Are the required functions built into the GIS or will applications need to be developed?
- Is the GIS customizable?
- How responsive and knowledgeable is the software developer's technical support staff?
- Were expectations met?

Management and Maintenance Procedures

- How will the data be updated, managed, and maintained in the future?
- Have all those who will contribute to the updating and maintenance been identified?
- Have data management and administration applications been developed and tested?
- Have data accuracy and security issues been addressed?
- Who will have permission to read, write, and otherwise access data?
- How will using GIS change information flow and work flow in the organization?

Costs

- How large a database will be created?
- What will be the required level of existing staff commitment during the data preparation and GIS construction process?
- What will be the cost for data conversion of in-house documents?
- What will be the cost for obtaining supplemental data from outside sources?
- How will GIS impact or interface with existing hardware and software?
- What new hardware, software and peripheral equipment is required?
- How much training of staff is required?
- Will additional staff with distinct GIS programming and analysis capabilities be required?

3 EXECUTING THE PILOT STUDY

Data Preparation (Scrubbing) and Delivery

Document preparation of source data representing the entire range of data to be included in the database must be completed before the conversion contractor can begin work. Data preparation includes improving the clarity of data for people outside the organization who are unfamiliar with internal practices. This pre-conversion process is referred to a "scrubbing."

Scrubbing is used to identify and highlight features on maps that will be converted to a digital format. The process provides a unique opportunity to review or research the source and quality of the documents and data being used for conversion.

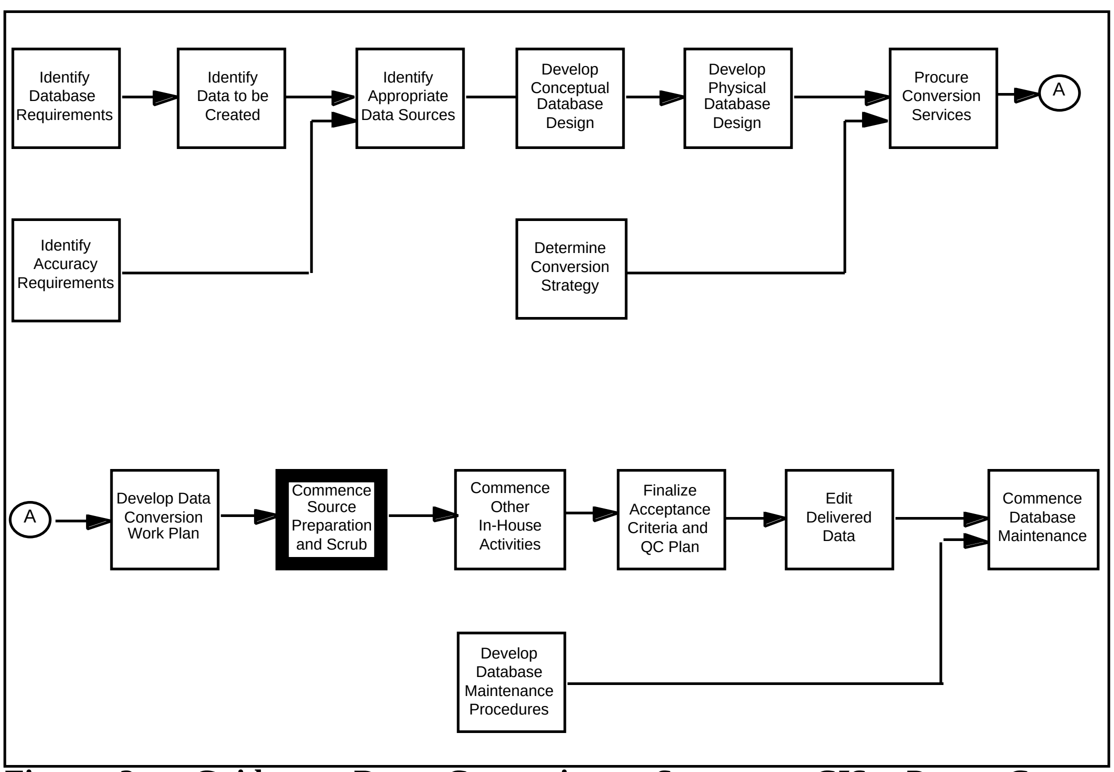


Figure 2 - Guide to Data Conversion Source: ***GIS Data Conversion Handbook***

Scrubbing is generally an internal process, but may also be performed by the conversion vendor. The conversion vendor will need to be trained on how to read your maps or drawings. The first map (or all maps) may need to be marked with highlighter pens and an attached symbol key to define what features need to be collected.

At the same time the maps are marked-up, coding sheets are filled out with the attributes of the features to be captured and a unique id number is assigned to both the feature and the coding sheet to create a relate key. This key is critical to connecting the attribute records to the correct map feature defined in Database Design.

The best key is a dumb, unique, sequential number that has no significance. The key should never be intelligent, that is contain other information. The key should never be a value that has meaning, or has the potential of changing. Don't use address, or map sheet number or XY coordinates or date installed. These values are very important and should each have their own field in the database. Don't use them as the primary key. The reason is very simple. If you use a smart key like SBL number and you have to change the number, you run the risk of losing the connection to all other related tables that key on the SBL number. Make the change and the records no longer match. However, if the key is unique and has no meaning it will never have to be changed. Street names change, numbers get transposed, features are discovered to be on the wrong map sheet or at the wrong XY coordinates. If any corrections need to be made, a large defensive programming effort must be in-place to guarantee the integrity of the intelligent key. Avoid the grief and use a dumb, unique key.

Coding sheets are only required if the attributes of the features are not readily available from the map document. For example, if all the required attributes for a feature are shown as annotations on the map (e.g. the size, material and slope for a sanitary sewer line), then a coding sheet is unnecessary. If additional research is required to find the installation date, contractor name, flow modeling parameters or video inspection survey, then a coding sheet needs to filled out for each feature. Again it is critical to create and maintain a unique key between the map feature and the attribute data on the coding sheet.

Once the data has been prepared for conversion, make copies of everything being sent out and make an inventory of the maps, coding sheets, photos, etc. that will be sent to the vendor. Ask the vendor to perform an inventory check on the receiving end to verify a complete shipment arrived.

Change management is essential. If the manual maps or data will be continually updated in-house during the conversion process, keep careful records about what maps and or features have changed since the maps have been sent out. This is an important process that needs to be fully in-place if the pilot study leads to a full GIS implementation.

When and Where to Set Up the Pilot Study

Expect the pilot study to have an impact on daily work. Choose participants where the pilot will not have a negative impact on the daily workload. Even if the GIS is to assist a mission critical process like E911, conduct the pilot as a parallel effort, don't expect it to replace an existing system. At the same time try to make the GIS a part of the daily workflow to test the integration potential.

To ensure some level of success of the pilot study, choose willing participants to act as the test bed/ pilot study group. Make sure they understand the impact the pilot will have on the organization and the level of commitment from the staff members. Use educational seminars to inform the employees about GIS technology and the purpose of the pilot study. Communicate very clearly what the objectives of the pilot study will be, what functions and datasets will be tested and which questions will be investigated. Describe the required feedback and the use of questionnaires or checklists that will be used. Above all else, communicate to keep staff informed and to control expectations.

Who Should Participate

A team representing a cross-section including managers, supervision, and operations staff should be assembled for the pilot study. Choose the staff carefully to assure objective and thoughtful system evaluation. If possible, choose the same people that were involved in the needs assessment process. They will be more aware of GIS technology and may be eager to see the project move forward.

Testing and Evaluation Period

Have a pilot team kickoff meeting with the conversion / software / hardware vendors present. Restate the objectives of the pilot study and responsibilities of each party. Review Needs Assessment, database design documents and assess training requirements. Define communication protocol guidelines if necessary to keep key players communicating and resolving problems.

Before the data arrives, install the software and or hardware in the target department. Conduct user training to familiarize employees with the use of the GIS software. If employees are unfamiliar with computers, allow more time for training and familiarization.

Once the data has been converted and delivered, have the conversion vendor or the software vendor load the data on the target machines. Be sure that this step and all preparatory efforts are monitored and treated as a learning process for your staff.

Begin a through investigation of the capabilities and limitations of the hardware and software. Keep user and vendor defined checklists beside the machines at all times. Have each user log their observations and impressions with each session. Make sure to note any change in performance as a function of time of day or workload. Also note if the user's level of comfort has increased with time spent using the software.

Log all calls to the data conversion, software and hardware vendors. Note the knowledge and skill of the call takers, responsiveness and turn-around time from initial call to problem resolution. Some problems may be addressed on the phone, others may take days. If the call cannot be handled immediately, ask the outside technical support person for an estimated time.

Obtaining Feedback From Participants

It is imperative that all individuals involved in the pilot study provide input before during and after the pilot study is complete. The best method to guarantee feedback from the participants is to have them help formulate the objectives of the pilot, the questionnaires and checklists. Sample questions to address were listed earlier in this document. Augment these with questions from your own staff. Some questions can be answered with a yes/no checklist, some answers will be a dollar figure, and some will require a scoring system to rate aspects of the system performance from satisfactory to poor or unacceptable. Other issues that may effect information flow, traditional procedures and work tasks will require participants to write essay questions or draw sketches of changes they would like to see in the user interface or in the map display. All responses should be compiled in such a way that the responses can be measured and rated numerically.

4 EVALUATING THE PILOT STUDY

What Information Should Be Derived From the Pilot Study

The first question to be addressed is whether the pilot study was a success. Success doesn't necessarily mean that the process went without a hitch. A successful pilot study can be fraught with problems and GIS can be rejected as a technology for the organization. The success of the pilot study should be measured by whether the goals and objectives defined for the pilot were achieved. Most issues listed below were covered in earlier portions of the document, but are summarized again.

Data Specific Issues

Many issues to be assessed in the pilot study are data specific and are related to data quality, volumes and conversion efforts.

Source Document Quality

Most first time GIS users are so awe struck by seeing their maps on the computer screen or on colorful hard copy plots that they overlook the importance of reviewing the quality and usefulness of the source documents and the utility of the final product. Many original maps are so old and faded, that they are unusable as a source document to create a GIS dataset. Some municipal agencies have scraped the existing maps and re-surveyed the entire town's street and utility infrastructure. This is not a cheap alternative, but digitizing bad maps is not a good investment.

Quality Control Needs

There is a danger present in any data conversion project (even for a pilot study) that the vendor will perform the conversion and deliver the data to the client without an adequate Quality Control process in place. If the client is new to GIS, they may not be able to determine if all the data is present, if the data is layered correctly or if all attributes are populated.

Because a GIS looks at map features as spatially related, connected or closed features, GIS query and display functions can be used to identify features that are in error. By displaying each map layer one at a time using the attributes of the features, item values that are out of range (blank, zero, or extreme values) will show up graphically on the maps in different colors or symbol patterns. Erroneous values should be reported to the conversion vendor immediately for resolution.

The client may consider using a third party GIS consulting firm to review the quality of the data and verify the map accuracy.

Data Availability

Before an attribute field is added to a coding sheet as a target for data capture, be sure the value is readily available and has importance to the operation of the agency. Many data fields would be nice to have, but may not be cost effective. For example, a sidewalk and driveway inventory for a community would be a useful data layer to capture. However, if there are no existing maps showing sidewalk locations, using aerial photos and photogrammetry is a costly approach to capture sidewalks and driveways. A cheaper alternative may be to create two single digit fields in the street centerline attribute table to hold flags indicating the presence or absence of sidewalks on the left or right side of the street. An operator looking at the GIS screen and air photos can assign the values to the flags without a large amount of effort. Based on these values, different line styles or colors can be used to symbolize the presence of sidewalks in a screen display or hardcopy maps.

Pre-conversion Editing

Be sure to track and review the number of man hours and problems encountered during the pre-conversion scrubbing effort. These steps will undoubtedly be performed again during the full conversion and now is the time to assess the impact on the organization.

Data Volumes

Data volumes and disk space is an important issue to evaluate in the pilot study. The pilot by design covers a small area of interest. Use the same data cost ratios discussed above to extrapolate data volumes for the entire GIS implementation effort. Data volume is not only a disk space issue. There are inherent problems associated with managing large datasets. Large files take more computer resources to manipulate, backup, restore, copy, convert, etc. A tiling scheme (i.e. breaking the data into smaller packets for storage and manipulation) should be investigated in the pilot study as a future solution for full implementation.

Assessing the Adequacy of the Data Conversion Specifications

Data conversion specifications are provided to give the conversion vendor and the client organization a set of guidelines on what layers, features and attributes should be captured, at what precision, level of accuracy and in what format is the data to be delivered. Best intentions and reality need to meet in the pilot study to evaluate the expectations and the level of effort (costs) involved with converting the target dataset.

Ask the conversion vendor for feedback on the clarity of the specifications. Do the specs make sense? Some vendors, holding to the adage "the customer is always right", will not question your specifications and will do whatever you ask no matter how in-efficient the process. Others will openly suggest alternatives approaches and will seek clarifications. Note the kinds of questions they present and be open to changes early in the process.

Evaluation of logical data model and applications

Not only should the quality of the data conversion and the GIS software be reviewed in the pilot, but just as important, the logical data model needs to be reviewed. The logical data model describes how map features are defined (points, lines, polygons, annotations) and the relationships between these map features and related database tables. Running applications against the data model will allow measurement of response time that is a function of data organization.

The bottom line is does the data model make sense for all the applications being addressed in the pilot and will it be useful in the full implementation. Ask the conversion and software vendors to explain the organizational structure of the GIS data model. What are the advantages, disadvantages and tradeoffs for the model used in the pilot and ask if the same structure would work comparably in a full implementation. Look carefully for short cuts or data model changes to make a dataset work in the pilot. It may work very well for a demo on a small dataset, but it may be unwieldy in a large implementation.

GIS hardware and software performance

Test the GIS running under a variety of scenarios ranging from single to multiple users performing simple to complex tasks. Ask your software vendor to write a simple macro to simulate multiple users running a series of large database queries. Test the performance of query and display user applications while data administration functions are running.

Were the users able to learn to use the system and perform useful work?

Refined GIS Cost Estimates

By requiring the conversion vendor to ***keep detailed logs of conversion times for each data layer and feature type by map sheet***, the client organization can project or extrapolate from the pilot data conversion to a cost for full conversion. One approach that has work well in the past is to use parcel density as an indicator of manmade features. For example, if you compute a series of ratios of the number of buildings, light poles, miles of pavement edge, manholes, hydrants, and other features against the number of parcels in the pilot area, you can compute with pretty good certainty the number of manmade features in the remainder of the GIS implementation area. The Office of Real

Property Services has a low cost ($50 / town) parcel centroid database in a GIS format that can be used as a guide for parcel density. Unfortunately physical features like streams, ponds, contours, wooded areas, wetlands, etc., do not have a direct correlation to parcels. In fact there seems to be an inverse relationship between parcel density and number of physical features. The point to be learned is that the pilot study should provide an indication of costs for a full featured/full function GIS implementation effort.

Analyzing User Feedback

Tally the number of positive responses to yes/no questions, compute an average score for user satisfaction, and compile the essay responses for content and tone. Review the complied results with all team members and management. Interview team members to clarify questions with unclear or strong responses to gain more insight. From response scorecards and comments develop an overall score to determine user satisfaction, completion of goals and objectives.

5 BENCHMARK TESTS: COMPETITIVE EVALUATION

The purpose of a benchmark is to evaluate the performance and functionality of different data conversion methods, hardware and software configurations in a controlled environment. Each software package can be compared in the same hardware environment or one software package can be compared across different hardware platforms.

By defining a uniform set of functions to be performed against a standard dataset, key advantages and disadvantages of the different configurations can be compared fairly and objectively.

Planning a Benchmark Test

As with any successful project, a detailed, thought out plan needs to be devised. It should be noted that performing a benchmark takes a large amount of effort by both the local government agency and the vendors taking part. Few firms can afford to devote large amounts of staff time and computing resources competing in benchmark tests for free. Keep that in mind as you design the benchmark to focus the tests on key issues that can be readily compared. If the benchmark will be extensive, associated costs may be incurred.

Objectives for the Test

A benchmark provides an opportunity to evaluate the claims of advanced technology and high performance presented by the marketing/sales force of competing data conversion, hardware and GIS software vendors.

The objectives of the benchmark should be defined clearly and communicated to all parties involved. Suggested objectives for each of the different types of benchmarks include testing:

Conversion Methods

- Cost effective procedures

- Sound methodology
- Quality control measures
- Compliance with conversion specifications

Hardware

- Computing performance
- Conformance to standards
- Network compatibility and interoperability
- Future growth plans and downward compatibility

Software

- Conformance to standards
- Computing speed / performance
- GIS functionality (standard and advanced)
- Can the software run on your existing hardware system
- Ease of use - menu interface, on-line help, map generation, etc.
- Ease of customization for non-standard functions
- Licensing and maintenance costs

This list of objectives is not all inclusive and should only be used as a guideline or a starting point for your organization to design a benchmark study.

Preparing Ground Rules

Based on the defined objectives, all parties involved should be aware of what will be tested, how they will be judged and what criteria will be used as a measure (i.e. low cost, high performance, good service, quality, accuracy, etc.).

- Tests to be performed should be as fair as possible
- The exact same information and datasets should be given to all vendors
- A reasonable time frame should be provided to perform the work
- No vendor should be given preferential treatment over any other and clarifications of intent should be offered to all
- Tests should be quantitatively measurable
- Hardware tests should use comparably equipped or comparably priced machines
- Software tests should be performed on the same hardware and operating system

Create scoring sheets for each aspect of the test. For subjective tests, like ease of use, have each user rate their satisfaction/dissatisfaction with the results of each phase using a numeric rank-order scheme. This won't eliminate bias but will allow impressions and opinions to be compared. For objective tests, like machine performance, record the clock speed, disk space requirements, number of button clicks, error messages, response time, etc. for each test conducted.

Preparing the Test Specifications (Preliminary Request for Proposals or RFP)

The test specifications need to outline the type of test to be conducted (conversion, hardware or software); objectives of the test; detailed description of the test; measures for compliance; and a time frame for completion.

Selecting the Participants and Location

In order to conduct a benchmark, you need knowledgeable participants (both internal and external). The internal participants should be knowledgeable regarding the topic to be tested (data conversion, hardware or software).

Selecting external participants is more involved. Situations range from not knowing any vendors to invite to how to limit the number of vendors. The smaller the number of participants the easier the final selection process will be for the local government agency.

The Request for Qualifications (RFQ) process can be used to filter or pre-qualify potential participants. GIS is a specialized field and not every business involved with computers is qualified.

Several factors should be considered when selecting vendors for a benchmark test

- Are they knowledgeable about local government agency operations
- Are they a well known company
- Are they technically qualified
- Are they experienced and have a successful track record
- Are they financially sound, insured or bonded
- Are they going to be around 5 years down the road
- Are they local or do they have a local representative
- Would their previous clients hire them again

If the RFQ and/or the RFP are written clearly and succinctly, the process will filter the participants and only those companies that specialize in the subject in question will respond.

The benchmark can occur either at the client’s site or the vendor’s offices. Some tests like data conversion are best conducted at the vendor site to minimize relocating staff and equipment for a test. Hardware and software benchmarks are commonly conducted at both the vendor and client site. The initial data loading, customization and testing is performed at the vendor site. Once the operations are stable, the client is invited to view the results at the vendor site, or the system is transported to the client site.

Preparing the Data

For a data conversion benchmark, provide each vendor with a set of marked up (scrubbed) set of maps, documents and coding sheets as described in the pilot study section above. If possible, provide the data conversion vendor with an example dataset from the pilot study which shows the appropriate data layering, tolerances and attributes to be captured. If not a dataset, clear specifications for how the data should appear when complete. Specify what data format (*.dxf, *.e00, *.mif, tar, zip, etc.) and what type and size of media (1/4”, 8mm or 4mm tapes) you want the data delivered in.

For a hardware or software benchmark, provide a sample dataset which contains all possible layers for inclusion in the GIS. The data could be purchased, converted during the pilot study or could be the results from a data conversion benchmark noted above. Provide sufficient documentation with the data to describe the use of the data, the organizational structure and contents.

Scheduling The Benchmark Test

Once the benchmark has been defined and agreed to by the participants, set a time for the testing to occur. Schedule a start date and a duration. Unless you specifically want to use company responsiveness as part of the test (i.e. how fast can they respond to a problem), don't require an immediate start date or extremely short time frame. There is no need to cause undue panic and stress, you want a good test.

Transmitting Application Specifications And Data To Participants

Before transmitting maps, documents or data to any vendor, make an inventory and backup copies of all items. Either specify to the vendors that the data will be provided in a single data format on a specific media, or make arrangements to provide the data in a format they can read. Be sure to test the readability of the tape or disk on a target machine in your office before sending the data out. Once the data has been verified as complete and readable, make two copies of the tapes or diskettes, one to send and one to keep as a recoverable backup for documentation of the delivery. Provide detailed instructions as to the contents of the tapes or disks and how to extract the data. List phone numbers of responsible persons should problems arise with delivery or data extraction. Ask the vendor to perform an inventory at the receiving end to acknowledge receipt of the data or documents.

On-Site Arrangements

If the tests are to be conducted at your site, make sure you have the authorization and backing of management and all personnel to be involved. Provide plenty of advanced notice and time to setup. If you are conducting hardware tests you have to decide if more than one vendor's machines will be present at the same time for comparative testing. With both machines setup in the same room, you can conduct the exact same tests in "real time" and visually compare the results, but this will require more setup space and logistic leeway in the schedule to accommodate multiple vendors. Make sure you have a suitable environment for equipment with adequate power, air conditioning and security. Also make sure you have all required utility software in place to read and write compressed files from tape and virus detection software.

If you are performing software tests, make sure you have two or more machines with the **exact** same hardware and operating system configurations. If you can't have multiple machines, be sure to backup and restore the current operating system files before testing each software package to ensure a fair test of disk space requirements, resource usage and functionality. Always use the same datasets for each test.

Identifying Deficiencies In Specifications

Although the tests were well thought out and carefully followed, you will probably wish you had performed additional tests during the benchmark. If short comings are discovered early on and they do not involve major changes in direction, additional tests could be incorporated. Be sure to notify the local management, staff and vendor participants of the change in objectives.

Defining benchmark criteria

Data Conversion Issues

A standard set of tests need to be performed to evaluate the results of a data conversion benchmark. Overlaying checkplots with the source documents on a light table is a straightforward but time consuming way to compare the conversion results. Suggestions made in the Pilot Study section of this document, outline methods for using GIS query and display functions to determine if all the data is present, layered correctly and attribute values are within range. Displaying map features by attributes will highlight errors or items out of range in different colors or symbol patterns.

GIS Software Performance

Software tests can be classified into 2 groups - capabilities and performance. Capabilities tests if the software can perform a specific task (i.e. convert DXF files, register image data, access external databases, read AutoCAD drawings, etc.) Performance deals with how well or how fast the software performs the selected task. How fast can be measured with a stopwatch, how well is open to interpretation.

The operating system on the machines in question will play a big factor in how GIS software will perform. GIS software written to run on a 32 bit operating system will not perform as well in a 16 bit environment without work arounds. Likewise, a 16 bit application will run faster on a 32 bit machine, but will not run as well as 32 bit software on a 32 bit operating system like UNIX, Windows 95 or Windows NT.

Hardware Performance

The goal is to find the fastest, cheapest hardware to meet your budget. Take advantage of computer magazine reviews of hardware. They conduct standard benchmark tests involving word processing, spreadsheets and graphics packages. The test results won’t be GIS specific, but will show the overall performance of a given computer. Oddly enough, two computers with seemingly identical hardware specifications (clock speed, memory, and disk space) can perform very differently based on internal wiring, graphics acceleration and chip configurations.

Evaluating Benchmark Results

If the questions were formulated clearly, and the results were recorded honestly, evaluating the results of the benchmark should be process of simple addition. Essay responses and comments will have to be followed up with further tests to clarify any problems or differences encountered.