

Maps and Legends: Plotting a Course for Geographic Information Systems

*Philip Herold, Public Services Librarian
Albert R. Mann Library, Cornell University*

ABSTRACT

Many academic institutions are experiencing explosive growth in the use of geospatial data and Geographic Information Systems (GIS). As GIS gain widespread acceptance as a research tool, the scholarly community looks to the library as the institution with both the skills to provide intellectual access to data and the technological expertise to guide them in their use. Librarians must be proactive in developing support and service mechanisms for geospatial data and GIS. This paper defines and discusses academic library support issues surrounding GIS and geospatial data. It proposes a model for levels of GIS support and decision-making.

Geographic Information Systems (GIS) merge the classic principles of cartography and geography with modern information technology to create sophisticated, flexible, and powerful tools for spatial analysis, planning, and research. GIS are ubiquitous in fields such as natural resources, environmental studies, and urban and rural planning, but have expanded well beyond these domains. In recent years, GIS have become important research tools in the biological sciences, social sciences, and to a lesser degree in the humanities. At Cornell University, GIS are used to display and analyze geographically referenced data at the Laboratory of Ornithology, the Theory Center (supercomputing), City and Regional Planning, Graduate School of Management, Veterinary School, and in the Human Ecology, Rural Sociology, Soils, Crops and Atmospheric Sciences, Agricultural Economics, and Sociology departments. There is almost no area of study in which a GIS cannot be used to further research. Though they offer limitless new possibilities for the faculty, staff, and students of the university, many academic libraries do not offer support for GIS. Still more lack strong organization of their GIS support. Geographic Information Systems have not yet attained their proper place among the mainstream support services of many libraries.

Libraries charged with the support of academic programs, including instruction, research, and extension, where GIS technology can be used to further research and study, need to address GIS support services. Where pockets of GIS activity exist on a college campus, librarians should glean expertise in the design of services. Where other colleges and universities are setting the pace for GIS, by introducing innovative and practical solutions to GIS service issues, the library should follow. Where there are opportunities for collaboration with other institutions or agencies, librarians should aid in the advancement of spatial data standards, data sharing, and GIS services.

This paper sets forth a model for the integration of GIS services into an academic library. It concentrates on the in-house support of spatial and attribute data. It presumes that a few key trends, evident in the recent maturation of GIS software, hardware, and data, will continue; that the explosive increase in GIS use will persist; and, that libraries will continue to play a vital role in providing access to spatial data and GIS services.

What is GIS?

A geographic information system combines hardware, software, data, and people to enable the capture, storage, retrieval, manipulation, and analysis of geographically-referenced data. This definition may sound complex, but consider that to use a GIS requires little more than does the World Wide Web. The Web requires hardware (PC, Macintosh, UNIX), software (Web browser), data (locally stored or on a remote server), and someone to use it, to make it dynamic. A GIS can be run on a PC, Macintosh, or a UNIX workstation. Software varies from high-end, process-intensive GIS that run on a workstation to small and relatively simplistic packages that perform the basic functions associated with viewing and manipulating data.

Like Web technology, GIS technology is still young and immature. A lack of industry standards for data, and metadata, analogous to the lack of organizational standards for Web-based resources, exists to the detriment of the end-user. Unlike the Web, GIS has passed the initial, awkward phase in which its software was prohibitively complex; its components difficult to grasp and use effectively. Much of the complexity and confusion surrounding GIS has disappeared. Efforts by the Federal Geographic Data Committee (FGDC) and the Open GIS Consortium (OGC) to establish and promote standards for spatial data and metadata are slowly building consensus among data producers and users. These standards will help provide uniformity of data structure and its consequent documentation.

Understanding Spatial Data

Spatial data contains the geographic locations, shapes, and other attributes (e.g., size, name) of features. Features are roads, boundaries, elevation, topography, hydrology, and any other entities that can be represented by either vector data or raster data.

- Points, lines, and polygons are known as vector data. Vector data are arranged in a series of ordered (x,y) coordinates. Points are often used to represent individual, non-intersecting features like buildings, light posts, or bus stops in large-scale coverages, or airports or cities in small-scale coverages. Lines are frequently used to represent roads, hydrology, and railroads while polygons most often represent land boundaries such as states, counties, school districts, or tax zones. Data attributes, or information describing a feature like the name, length, classification code, and address ranges of a street, are usually associated with the feature that represents them. In the case of a street, this is a line, or "arc" segment.
- Groups of cells arranged to represent a feature are called raster data. Raster data have a cellular structure, in which rows and columns of cells are referenced in an (x,y) system, and each cell, or group of cells, represents the value(s) of that feature. Geographically-referenced digital photography (orthophoto) and remote sensing data (airplane and satellite-gathered) are usually represented by raster data. Raster data are also used in terrain and elevation modelling. This type of data is used heavily in the natural and environmental sciences.

The physical data behind raster or vector data are usually tabular ASCII data which can be translated by a GIS program. Because many diverse geographic features can be represented by geospatial data, and a multiplicity of methods exist for capturing, collecting, and storing information about these features, an alarming number of data formats, each carrying its own set of requirements for display and analysis, now exist. A lack of format standards has not helped to lessen the complexity of this problem. For example,

Digital Elevation Model (DEM) data, which have x and y coordinate (e.g., latitude and longitude) locations, also carry z locations (elevation) and are best represented using the raster format. In contrast, USGS 1:100,00 Digital Line Graph (DLG) data are stored in vector format. The two types of data differ fundamentally with regard to data structure, display, and use. They require different software programs and human expertise to accomplish these tasks.

The Federal government has issued spatial data in many nonproprietary formats, including TIGER/Line, Digital Line Graph (DLG), Digital Elevation Model (DEM), Digital Orthophoto Quadrangle (DOQ), and Digital Raster Graphic (DRG). Hunt and Joselyn describe the key characteristics of several nonproprietary data formats and discuss their use in a library setting.⁽¹⁾ These formats differ from one another in the types of features each represents, in the method each uses to represent features, and in data structure.

Attribute, or descriptive data describe the characteristics of a feature. Statistical data gathered from the decennial census is a good example of attribute data. While these census data do not contain the locations or shapes of features, they contain demographic and other statistics that can be linked to geographic areas, such as states, counties, or census block groups. These geographic areas can be represented in a GIS as polygons. One of the more common uses of GIS in libraries is joining census data aggregated at the county or sub-county level with a spatial data polygon coverage. Once this "join" is achieved, statistics can be represented by a graduated color, or other scheme, in visual, map form.

Establishing the role of the Library

As with other academic library services, librarians must gather information about the library's user community before embarking on a course of action in support of GIS use. It is necessary to know where and by whom GIS is being used within the university. If there are pockets of GIS activity on campus, what is the nature of the activity? Are there instruction programs for GIS? Is the activity largely research based? If so, what types of research are utilizing GIS? Who are the GIS experts among the faculty and staff? What support for GIS already exists? Inquiries into the nature of GIS activities should yield answers about the types of data that are sought after, utilized, and produced on campus. Is there a software type favored by the majority of local GIS users? It is also important to identify the roles of other campus departments.

Depending on the level of campus activity, it may be useful to survey local, state, and national-level GIS programs. GIS are used heavily at all levels of government, primarily for planning, assessment, and management purposes. Local city and county governments often have an active GIS unit with whom collaboration will prove mutually beneficial. Often, local GIS units maintain collections of vitally important geospatial data describing features of the local area. Librarians will find there is high demand for locally-referenced geospatial data. These units will also own the collective expertise of full-time GIS professionals who are immersed in the subject and often eager to share their wisdom.

Both state and federal government GIS activity involve creating digital datasets, establishing geospatial data clearinghouses, and standardizing spatial data transfer formats and metadata. The United States Geological Service (USGS) is the primary producer of geospatial data for this country. Librarian staff should become familiar with USGS data products and standards. Among the USGS products are a number of "framework" datasets. Framework data, as defined by Federal Geographic Data Committee (FGDC) Framework Committee, are those data deemed to be among a core set of geospatial data which serve as a basemap, a set of basic geographic features, to which attribute data can be joined⁽²⁾. Examples are transportation, hydrography, elevation, boundary, and cadastral data. The widespread availability of framework data will provide a strong foundation for the use of GIS in decision-making. The FGDC, charged with the task of facilitating development of the National Spatial Data Infrastructure (NSDI), has taken the lead in developing

content standards for geospatial metadata.(3) Federal agencies which produce geospatial data are required to disseminate their data products via the NSDI Clearinghouse, and to meet FGDC metadata standards. FGDC metadata standards are fast becoming the norm for data access in libraries and other local repositories.

Assessing Library Community Needs

Having accounted for surrounding GIS activities, the role of the library with regard to campus GIS use, and the role of GIS and spatial data in the library, should be defined. Once this has been accomplished, the library should choose a model of data collection and public services to support the use of spatial data that will meet its needs. Or, recognizing the demand for a special set of services which meet the unique needs of the library, a new model should be developed. A library's GIS service role should evolve as users become accustomed to its GIS services and come to expect improvements and additions that will shape the technology. Several more questions should be answered before selecting an initial service model:

- Who will be using spatial data?
- How will users be using Geographic Information Systems?
- What spatial data will users want/need?
- What attribute data will users want/need?
- What spatial and attribute data does the library already own?
- What resources can the library reasonably support?
- What GIS expertise exists within the library?
- Where is data and GIS software/hardware already available?

Invariably, the library serving an undergraduate institution will have a different set of requirements for data and software than a library supporting a research institution. It is unlikely that government-produced DLG, DEM, and DOQ datasets, which require resource-intensive conversion and processing prior to their display and use, will be practical solutions for everyday questions in the undergraduate library.(4) Undergraduate students are rarely afforded opportunities to complete in-depth studies utilizing GIS. Studies involving an expert-level GIS require users to ascend a steep learning curve and to understand geographic and cartographic principles as well as data format issues.

But the undergraduate library should not dismiss the idea of supporting GIS. There are a number of data packages that, when combined with desktop GIS applications, might be described as 'plug-and-play'. Most of these are commercial, not government, products, containing basic sets of geographic features such as transportation, boundary, and hydrology data for United States counties. There are spatial data products to fit the needs of a diverse array of library settings.

GIS Service Model

At their most basic level, GIS services must include both collections of spatial data and access to GIS software and hardware. Most often, the selection of suitable hardware and software will be dictated by the nature of the data being provided. At one end, complex and raw data, requiring time-consuming conversion processing, necessitates the use of a high-end GIS. At the other end, preprocessed, user-friendly data may only require a desktop GIS or data viewer.

The selection of GIS software will usually determine hardware and platform requirements. However, when cost, networking and other issues make the hardware platform of paramount importance, the number of possible GIS applications will be limited to the desired platform (PC, Macintosh, UNIX). The selection of

data, software, and hardware requires looking at all three simultaneously and considering essential support activities: training, staffing, and instruction.

Abbott and Argentati (1995) divide library GIS use into three levels, as defined at North Carolina State University.⁽⁵⁾ Using this approach, the following model integrates the selection of spatial data, GIS software, and hardware, and the necessary internal support roles that will accompany each selection, into three tiers:

Tier I: Data Viewing Support

Data viewing will address the potential mapping needs of many library users. Though most essential components of Geographic Information Systems are absent from this level, a user can quickly and easily create customized, professional looking maps. Many data viewers will provide text and marker labeling, as well as annotation capabilities and some query-based functions. Data viewers are functionally one-dimensional, but will satisfy users with one-dimensional needs.

In most cases, a data viewer will lack the ability to capture or store new data. It will not allow for complex query or statistical and analytical processing of data inherent in the relational database management systems contained within intermediate or high-level systems.

Data: Most data viewing software arrives coupled with basic sets of geospatial data. The same kind of data that might be found in an atlas--in some cases a statistical atlas--will accompany most data viewers. Examples of this data are streets, railroads, hydrology, and city, county, state, and country boundaries. In some cases U.S. Census-designated (tract, block group) boundaries, or zip code boundaries, will be included. This spatial data will usually come prepared with its basic attribute data. For example, street segments will contain names, road type, and address ranges. Country polygons may contain basic demographic statistics, population figures, etc.

A major limitation of data viewers is that they will generally only read data of a single format, usually a viewer-specific format that is neither compatible with other viewers nor upwardly compatible.

Data viewers will not allow for the conversion of data. As a result, most government-issued spatial data will not be readable in their native format. There are exceptions, such as LandView, a Census Bureau product designed to view TIGER/Line data and the USGS has produced a freeware product to view Digital Line Data. Most data viewers will read only vector-based data or raster-based data, but not both.

Software: Consideration in selecting a data viewer(s) should be given to:

- **Data compatibility:** Will it read more than one type of data?
- **Accompanying datasets:** What geospatial data features are included? Is there an array of additional data than can be purchased? What compatible statistical attribute data are available?
- **Upward compatibility:** Is the data viewer related to a more sophisticated system? Does accompanying data come in an upwardly compatible format? If so, the library could move to the next level of GIS support with greater ease.
- **Mapping features:** Will it allow customized annotation, labeling, coloring, etc.? Is address matching a feature? Address matching will enable users to automatically place features (e.g., homes, businesses, farms) on a street map by typing in their addresses.
- **Ease-of-use:** Is system use seemingly intuitive? Is map creation a progression of logical steps? Is there adequate on-line help and/or print documentation?

- Resulting support requirements: Will self-guided users be successful? Will library staff require extensive training to support the data viewer's use?
- Hardware/platform requirements: Does it run on a PC or Macintosh? Are there special hardware requirements? What are the implications for printing?
- Cost: How does it compare with other, similar products?
- Examples: Data viewers with basic mapping capabilities include BusinessMap, MapExpert, Maps 'n' Facts, LandView.

Tier II: Desktop GIS Support

Desktop GIS are becoming increasingly popular and gaining widespread acceptance as geographic mapping and analysis tools. The maturity of geographic information technology is most evident in desktop GIS applications. Most desktop systems contain at least some of the data query and analysis functionality of full-blown GIS. Most have a level of control over map creation on par with full-blown GIS. Desktop GIS, so-called because they run on PC (Windows, Windows 95, or Windows NT) and/or Macintosh microcomputers (some also have versions for UNIX), are GUI-based, and are much easier to learn and use. Desktop GIS bring the power and functionality of GIS to the novice user.

Support Requirements: Desktop GIS is the right level of support for most academic libraries. A wide range of geographic analysis can be performed with a desktop system of software, hardware, data, and expertise. The support of desktop GIS will require a sizable commitment on the part of the library. Staff training is the most critical element in providing GIS services. Public services staff will need to invest time in becoming familiar with desktop GIS applications, understanding the implications of data formats, and absorbing geographic and cartographic principles, and geospatial analysis techniques. Once a core of GIS expertise exists in the library, other library staff may be trained to provide a wider base of support.

Support of desktop mapping should also contain a user training component. Hands-on computer classroom instruction in the use of a desktop GIS is an effective means for educating large numbers of users. Hand-outs or Web-based tutorials will provide users with a resource for self-guided instruction.

Many libraries currently provide appointment or consultation-based services in support of GIS use. Though a resource-intensive service, one-to-one consulting between a librarian and patron is often the best means of support when the patron has a specific set of questions. In consultation, as in the reference interview, a librarian will identify and relay back to the patron the key components of their request and help to match the library's resources, as well as resources outside the library, with the patron's needs.

Data: Most desktop GIS applications will read only raster or vector data; very few will handle both data types. It is therefore not uncommon for a library to support both raster- and vector-based software application programs.

Many desktop systems will accept a limited number of different data formats, but are designed to work with data in their own, platform-specific formats. For example, MapInfo works with data in native MapInfo format, while ArcView 3 works primarily with Arc/Info coverages or shapefiles. Some systems will support conversion from a foreign format to their native formats.

Desktop GIS packages most often come with a suite of geospatial and attribute data. A core of streets, hydrology, and some combination of country, state, county, local, and zip code boundary data coverages is a standard package for vector-based data. Sets of statistical, demographic, and other attribute data associated with the geospatial features (e.g., U.S. Census data) are also commonly included.

A desktop GIS' suite of data products will not be comprehensive enough to adequately answer many of the geographic questions that are asked by patrons in a diverse academic community. Libraries will want to develop geospatial and attribute data collections to match the capabilities of the desktop system they have chosen to support and the demands of their user populations. Librarians should look to both governmental and commercial data providers in creating a well-rounded geospatial and attribute data collection. Large volumes of geospatial data are available at little or no original cost from Federal agencies like the USGS. A solid understanding of data formats and the library's desktop GIS will enable librarians to select appropriate data products. Likewise, attribute data is abundantly available from state and federal agencies, many of which are mandated to freely disseminate their data products. The Census of Population and Housing, Economic Census, Census of Agriculture, are examples of large sets of attribute data that can be used in a desktop GIS.

Software: Consideration in selecting a desktop GIS should be given to:

- Data compatibility: Does it support raster data, vector data, or both? Does it support only data in its own software-specific format? Will it convert data from other proprietary and nonproprietary formats into its native format? If so, specifically which formats?
- Database compatibility: What database or spreadsheet formatted data will the system read and use? Does the desktop system allow interfacing with a relational database management system (RDBMS)?
- Data capture: Does it support map digitizing? Does it allow for point, line, or polygon layer (coverage) creation? Does it permit raster coverage creation?
- Data storage: Will it be able to manage large volumes of data? Does it have a data library or other system for data management?
- Data manipulation: What editing capabilities exist? Does it allow editing tabular data? On-screen feature editing? Is it capable of performing topological cleaning and building?
- Analysis capability: Is there advanced query-based selection by feature attributes? Does it handle calculations of spatial and statistical data? Will the application analyze spatial relationships and perform buffering, networking or routing?
- Data display: Does it allow thematic mapping (e.g., graduated color, dot density)? Can one perform map overlay, interactive feature modification, and produce charts and reports? Will it allow geocoding?
- Programming: Does the desktop GIS support a programming/scripting language for customizing use?
- Ease-of-use: How user-friendly is the software? Is data display a straightforward set of steps? Are complex tasks performed without great difficulty? Is there a comprehensive system of on-line help? Is there adequate technical documentation? Is installation quickly achieved?
- Hardware/Platform: With which operating systems is the software compatible? What are the recommended memory, storage, and processing speed requirements? Is it possible to network the application? What peripherals (e.g., printers, plotters, digitizing units) are supported?
- Cost: What is the actual cost? What are the hidden costs, including upgrades?
- Examples: MapInfo, ArcView 3, Maptitude, Atlas GIS, IDRISI, ERDAS, GRASS

Tier III: Full Function GIS Support

By definition, full-function GIS contains all of the current technologies comprising GIS. Full-function GIS will enable a person to capture, store, manipulate, analyze and display geographic information. They will support data conversion, data management, map overlay, and spatial analysis; provide tools for interactive display and query, on-screen feature and tabular editing, and geocoding; and include modules for relational database management system integration and custom programming.(6)

Supporting full-function GIS in an academic library requires a high level of commitment to staff training. The nature of GIS is technically complex. It cannot be expected that high-level expertise can be spread throughout a library's staff unless GIS support is that library's primary mission. Instead, a small core of experts will carry the greatest burden in supporting high-level GIS activity. This core group of experts will benefit from formal training by experienced GIS professionals. They will need to be able to set aside blocks of time in which to practice and develop their skills. They will need to install and test systems; and develop internal GIS policies. Hiring staff who can bring high-level experience can help to jump-start the library's support program.

User support is time-consuming. Most projects involving a full-function GIS will require consultation between project participants and a GIS expert. Developing means by which to "bring the technology to the user", when the technology is highly complex, alleviates some of the time requirements over the long term. But in the short term, creating systems to simplify GIS tasks, including instruction modules, user guides, on-line tutorials, or macro programs to simplify complex GIS tasks for the user, means a substantial commitment in development time.

User instruction will also be accomplished largely on a one-to-one basis. Full-function GIS are open systems, which contain the power and tools to perform a multitude of tasks. Most systems do not have intuitive, graphical-oriented interfaces. The raw tools exist, but their use cannot be taught in a short time. A workshop series for potential users can serve to introduce the technology, but will not empower users far beyond that point. Academic library support of full-functional GIS will depend on an existing foundation of campus GIS activity including formal classroom training and an existing expertise in various "pockets" where GIS are being used.

Selecting a fully-functional GIS is simplified by the fact that only a few exist. Environmental Systems Research Institute (ESRI), makers of BusinessMap, ArcView, and Atlas GIS, offers its flagship product, Arc/Info, the market leader amongst GIS software in academic settings (75% of universities using GIS) and a full-functional GIS (Foresman 1996). Intergraph's Microstation MGE is another full-function GIS. Although it remains popular among commercial and government users, MGE has captured less than 20% of the academic market.(7) Other software products can be considered full-function. One such example, MapInfo, has most of the functionality of this level. Others are designed to meet highly specific GIS market needs, and have much less widespread use. In selecting a full-function GIS, one should consider the software preferences of current and potential users, existing campus GIS software licenses and preferences, and desktop GIS software the library already supports (e.g., ArcView is closely linked with Arc/Info). Comparisons between the major brands of GIS will help determine the right product for a specific library.

Concluding Comments

Since 1992 the Association of Research Libraries (ARL) has led the ARL GIS Literacy Project. The project has helped educate and equip many librarians with the necessary skills to facilitate access to geospatial data and the GIS tools to use them, and has relied on collaboration, involving both public and private sectors, to successfully integrate GIS services and resources into libraries.(8) Involvement in the ARL GIS Literacy Project is a good way for libraries to introduce GIS into their mainstream services. The project

provides an established support mechanism for libraries interested learning about GIS. It can provide libraries with some of the data and software components, as well as the expertise to use those pieces. For libraries who have existing GIS services, the project is an information resource to help develop services further.

For librarians who have considered integrating GIS technology into the library but have hesitated, the time has arrived. Once prohibitively complex, GIS software applications have migrated to the desktop and become much easier to use. Geospatial data is much more accessible, both intellectually and physically, that it has ever been. Although there are still a few serious obstacles to data access, the collaborative efforts of educational and governmental institutions are helping to remove these obstacles. It is time for librarians to integrate GIS technology into the library; to claim GIS support as a component of its mainstream services.

NOTES

1. Li Hunt and Mark Joselyn, "Maximizing Accessibility to Spatially Referenced Digital Data," *Journal of Academic Librarianship* 21 no. 4 (July 1995): 257-265.
2. Federal Geographic Data Committee, *Development of a National Digital Geospatial Data Framework* (Washington, D.C.: Federal Geographic Data Committee, 1995).
3. Text of the FGDC Content Standards for Metadata are available at URL:
<<http://www.fgdc.gov/Metadata/metahome.html> (<http://www.fgdc.gov/Metadata/metahome.html>)>.
4. Hunt and Joselyn, "Maximizing Accessibility to Spatially Referenced Digital Data," 259.
5. Lisa T. Abbott and Carolyn D. Argentati, "GIS: A New Component of Public Services," *Journal of Academic Librarianship* 21 no. 4 (July 1995): 251-256.
6. Keith C. Clarke, *Getting Started with Geographic Information Systems*, (Upper Saddle River, N.J.: Prentice-Hall, 1997), 235.
7. Timothy W. Foresman, "Academic Research and Education in GIS," in *GeoDirectory 1997: Volume 2 Academic Institutions*, ed. Gayle K. Rodcay (Fort Collins, Colo.: GIS World, 1997), 4.
8. Nancy M. Cline and Prudence S. Adler, "GIS and Research Libraries: One Perspective," *Information Technology and Libraries* 14 no. 2 (June 1995): 111.