

# Geographical Information System Uses for Synthetic Environment Database Generation

Mark B. Field, Pedro Ramos  
Field Consulting & Services, Inc.  
425 South Chickasaw Trail #211, Orlando, Florida, 32825  
Phone: 407-275-9351  
mark\_mail@field-csi.com

## Abstract

*The current trend of simulation database generation system (DBGS) tools is toward the use of more commercial off the shelf (COTS) technology as the foundation of the complete toolset. DBGS tools that support larger database requirements are leveraging COTS geographical information systems (GIS) that provide a solid technology foundation for DBGS development and utilization.*

*The use of GIS tools as the core of a DBGS is a natural progression towards reducing development costs and maintenance costs and increasing productivity. If one looks at the classical DBGS tools that were developed in the 1980s and 1990s, the similarities with today's DBGS tools become apparent.*

*Earlier tools, just like GIS systems, had to deal with real world geographical data and support varying levels of data attribution. This is the fundamental basis of a GIS system: providing a way to store geographic data with attribution describing that data.*

*The most significant historical differences between the data that the typical GIS tool traditionally supports and the data required by the simulation are the storage of three-dimensional data components and the richness of attribution data required. Cutting-edge GIS tools have outgrown the 2D heritage and have become viable 3D data storage and manipulation environments that, with the aid of faster processing power computers allow massive amounts of simulation data to be automatically generated including simulation specific data sets such a texture mapping and the like.*

*This paper will go over the basic technology that is part of a GIS, which is or can be leveraged by DBGS tools to create accurate and robust databases.*

## GIS Technology

Most GIS tools have several key technological features that are clearly useful to DBGS vendors and users. They include spatial data storage, highly configurable interface and data definition, raster support, dynamic reprojections, topology (feature connectivity), robust 2D viewing and editing capabilities, the ability to translate data between different formats to support data fusion, and native support of creating plots of geographic data. DBGS vendors leverage all these technologies, which allow them to concentrate their efforts on solving the simulation specific manipulations.

Since GIS tools are designed as a base technology for many geographic uses, there are several key technologies that may not be available out of the box on some of the GIS tools. These typically require development by DBGS vendors, and include areas such as 3D topology, actual Triangle Irregular Network (TIN) geometry, texture mapping information and true man-made object 3D geometry such as buildings with vertical faces or faces with normals point downward. All of these can be augmented in various ways, but the benefits of using COTS GIS tools typically outweigh the development of processes and support plug-ins or tools to convert GIS data to usable simulation data.

## Topology

Topology is a major technology that is included in most GIS tools that, to a large degree, was rarely available in earlier generation DBGS tools. This is one of the key reasons for DBGS tool vendors to capitalize on core GIS tool technologies.

Topology is defined as a set of defined relationships between links, nodes, and centroids, and describes how primitive geometry constructs (such as lines and polygons) connect and relate to each other. These explicit geometric relationships and associations form the basis for advanced GIS functions such as network tracing and spatial analysis.<sup>1</sup>

Topology allows developers to request basic contextual feature information that is critical to the production of high-quality geometry data models. Some examples of contextual information available are: what other roads is this road connected to, what is to the “right” and “left” of this road, and can I travel from this location to another location on the road network? Are there constraining features whose interactions to other features influence the final geometry?

Applications such as online or in-vehicle navigation tools are modern examples of how useful and widespread the use of topology has become in our everyday lives. When asking for directions from point A to point B, these tools efficiently query the roads that are connected to point A and point B by looking at the topological connections of the roads. Topology is also used in queries, such as requesting points of interest on a given road, such as hotels, restaurants and the like.

Some examples of topology useful to simulation data development include:

- Tracing the road network to verify all the roads in the database are connected and can be used in SAF or applications where road networking information is needed.
- Bridges and overpasses can connect to the road network in the database without causing interactions between themselves and the underlying road or river.
- The ability to determine what kind of intersections should be placed at road connections.

- For directionality attribution of roads, such as one way, exits, and ramps.
- The ability to determine flow of waterways across the entire network.
- Easily create levels of detail by generalizing the shared link between features to ensure there are no gaps or overlaps between the connected features.

Topologically connected data also provide more user-friendly and efficient data editing. This is accomplished by moving topologically connected edges or vertices simultaneously, that is, moving the shared geometry of all the connected features at once. For example, a road intersection can be moved which also moves the vertices of all the connected roads at the intersections at the same time. This reduces the number of edits required and ensures the integrity and quality of the road network and therefore, the derived data that can be generated from it.

## Dynamic Reprojection

Most current GIS applications support dynamic data reprojection. In a nutshell, this implies that the data can be stored in one coordinate system while the users manipulate it in another. For example, data can be stored in WGS84 geodetic, but in order to do manipulations of the data in flat earth ( which is normally most convenient for 2D data displays ), those edits can be done in a local flat earth coordinate system, and the results will be saved in the original WGS84 coordinate system.

More advanced GIS applications can also support various sources in different coordinate systems. The user can select the viewing coordinate system of their choice. This is a key part to data fusion from different and previously underutilized data sources.

Historically, the coordinate set issues that arise when performing data fusion operations have been a significant source of data miscorrelation of imported data, mostly due to the complexity and variety of these data projection models. GIS systems provide a ready-made working environment where those historical difficulties have been largely eliminated.

## Database Storage Technology

GIS applications support a wide variety of formats. Depending on the GIS technology selected, the underlying storage technology can be anything from a simple set of flat files to hierarchical to advanced relational or object database technology. It is generally more efficient for DBGS vendors to focus on the development of advanced algorithms and data manipulation techniques rather than spending critical manpower on data format storage. So many DBGS vendors are utilizing existing GIS storage technology, letting the GIS industry develop the data storage formats and efficient data retrieval algorithms.

Utilizing the data storage capabilities of COTS GIS also allows DBGS vendors to support storage and viewing of multiple versions of the feature data. Users are able to store the original source data, any modified data sets that have been generalized, corrected or otherwise manipulated, as well as different representations of the data to be used in simulations at the same time. The benefits of supporting all these variations can be summarized as follows:

- The ability to do metrics on what was modified from the original source and by how much.
- The ability to store and use multiple representations of the data for supporting different simulation requirements and levels of detail. For example, a forest feature can be represented by an area feature, or by many points representing scattered trees or different resolutions of the feature can be stored for different target systems.
- The original high resolution data can be retrieved and reused on future projects that can support higher fidelity data.

The more advanced GIS database technologies also support restricted user data access and modification, replication, data versioning, redundancy, 24/7 reliability, remote/distant users, and multi-editing users. GIS technology is being used for mission critical applications in the electrical, telecommunications and urban planning industries. The requirements from those industries have expanded the database technologies offered by the advanced GIS providers. Here are a few examples of these technologies:

- Restricted access allows different types of users to edit different types of data. Data collection personnel may create new data layers and/or networks, but cannot edit or view existing restricted data, while those with higher-level data access may view and manipulate the new data.
- This technology could be used to support multiple levels of security classifications.
- Data replication is used for external data development or a rudimentary backup or snapshot of the data set. This can be spatial or data specific. An example of this would be outsourcing data development. A replica of the data in a certain area can be sent offsite to have the data enhanced and then the final product can be rolled back into the onsite production database.
- Only one GIS vendor supports data versioning technology. This technology has the ability to create 'checkpoints' of data at any point in during data development. The data can be roll-backed to a previous version or merged to a new checkpoint. This allows multiple users to edit data within the same geographic area. As long as two people do not modify the same feature, the data can be merged easily, and with tools to support conflicts, the data edited by multiple people can be resolved. This allows large-scale data production to occur without the need to segment and later edge match the data into parcels, which saves valuable manpower and reduces the possibility of data errors.
- Redundancy is available in most advanced GIS. Redundancy allows the database to be running on multiple servers concurrently. When one server fails during database production, the other serves the database without the users being affected.
- Some advanced GIS also support the ability to share a database across multiple sites. This gives the end users the ability to spread the data development across multiple data development sites. This is a very desirable feature for very large database production efforts or in cases when individual training facilities would be interested in updating their own specific database training areas.

GIS technology also allows for the creation of user or application specific data models (also known as schemas). Although this data model is built upon some underlying GIS data model core, it is highly configurable by the DBGS vendor.

The underlying GIS data model requires data types that are not only basic primitive types such as integer, float, and string, but also spatial types such as coordinate, line, and area. These spatial data is where DBGS vendors are creative in using GIS as their core technology. Coordinates, lines, and areas in most GIS are 2D, some GIS use Z as an attribute of the coordinate (2.5D), and a very few actually support an implementation that could support 3D topology. DBGS vendors have to augment these underlying data model restrictions to provide data such as texture mapping or vertex normal data at these coordinates.

Most GIS can support data development in the hundreds of square kilometers, and there are only a few GIS database technologies that can support a large database in the thousands to millions of square kilometers. These higher-capacity GIS systems typically use some type of database technology other than flat or hierarchal file formats.

## Data Fusion

There is an increasing availability of diverse "GIS" data in today's market. Clearly some are free, some are not. More local governments are using GIS data to manage many aspects of their infrastructure. These rich data sets can be leveraged to create simulation and training environments, especially for urban close-contact scenarios. This is especially true for projects designed around homeland security requiring training and rehearsal around US cities.

One important and often critical factor in this kind of data exploitation is the understanding of what type and quality of data you are getting. GIS data is only as good as the collection requirements that were in place when the data was collected. This is especially true for attribution. For example, municipalities will not be collecting data with SEDRIS or FACC attribution, so there will still be some form of data conversion and/or inference that will need to be done.

There are many types of "GIS" data formats in the field. Probably the most commonly used one in the simulation industry is the shapefile format from ESRI®. This is a very simple data format that is mostly used for data exchange, although it is somewhat limited because it does not

support the rich functionality of GIS applications. Most GIS vendors are supporting Oracle® Spatial, which stores many types of GIS data as well as the more advanced data types, such as topology.

There are other more advanced GIS vendors that utilize proprietary data formats to support their specialized functionality. For example, General Electric's Smallworld Core Spatial Technology™ GIS supports object-based database versioning, which requires a specific format. An interesting development is the prevalence of GIS data in the industry has allowed niche companies, such as Safe Software, which created a translation application that converts data from one format to another called the Feature Manipulation Engine (FME®) to solve data fusion requirements in the GIS industry. This application supports reading and writing over a hundred different formats. The translation is done with a simple to use translation file that can do basic functionality such as attribute renaming, to calculations and clipping and reprojection. The use of these kinds of ready-made translation engines further expands the amount of data available for any one given application.

## Raster Support

All current GIS support some type of raster display. Some require importing the files into their native file format. Others can natively read several of the readily available formats, GeoTIFF being amongst the most popular.

Part of the raster support includes support of digital elevation models (DEM). DEMs are rasters that describe surfaces derived by one of several means, such as LIDAR (LIght Detection And Ranging) data that are normally delivered in some DEM format. Most GIS do not natively support DEMs and the data must be imported or converted into a format that can be read by the GIS. 3D data, including DEMs are relatively new to the GIS industry. Even though the industry has supported DEMs in a simplistic way, until recently it had been mainly used as background information.

Once raster data is available to the GIS, the GIS can display and reproject the data on the fly just like other data. So here too, the DBGS vendors can leverage existing technology, and use the vector and raster data together to support heads

up digitizing, quality verification, data correlation, or develop specialized algorithms to do things like road cut and fill analysis and river analysis.

## TIN Support

TINs are rather new to the GIS industry. Only a few GIS technologies support true TIN geometry, where the triangles share edges and vertices with other triangles. Other less advanced GIS just use 3D areas to represent triangles. Some GIS have the ability to store data, which were created by DBGS vendors' customized TIN algorithms.

Most GIS can display the TIN in a variety of ways like contours, color elevation, hill shading, etc. This information can be utilized to do additional database analysis to support automated processes like cut and fill analysis, river elevation adjustments, etc.

## Highly Configurable

Most GIS can be highly customized to provide the best usability to the DBGS user. Vendors can add specialized functionalities, modify data models, display multiple data sets, etc. More advanced GIS can support features such as roles and permissions. Giving a better checks/balance for the data being created or modified.

GIS also support several popular software interface languages with C/C++, VB, and Java being the most common. Some still use specialized languages, which could be both an advantage and a disadvantage, depending on the desired application. Some GIS support interpreted languages, which is a benefit to the rapid development of algorithms because the system does not require restarting and reloading of the data. For software development using large data sets this can be a significant savings.

## Large Existing Toolset

Between the GIS vendors and their users community, there is a rich resource of algorithms and knowledge that can be leveraged. There are thousands of existing functions delivered with the GIS that can be pieced together to provide simulation specific tools that perform complex operations that, without support from a shareable GIS core, would require re-development of these readily available

algorithms. There are many user groups available on the Internet that can be used to search for existing solutions. This too saves valuable development manpower for both DBGS vendors and their users.

## COTS User Interfaces

Another advantage of leveraging COTS technology such as GIS is the configurable user interface. A vendor can leverage the GIS user interface configurability to build a DBGS application. This too is a technology that is better for the GIS companies to deal with. It is more efficient to build upon the existing GIS user interfaces, and more quickly achieve the desired development goals, while preserving a uniformed look-and-feel in the tools.

DBGS vendors can componentize functionality using the capabilities of the GIS. This can be used for packaging functionalities based on customer requirements. Users of a GIS-based DBGS also have the ability to expand the DBGS functionality delivered from the vendor to support their unique requirements that may not be available from the DBGS vendor.

## Paper Map Generation

One obscure and unique requirement in most simulation database development environments is paper map production, especially for ground-based applications. Using GIS technology provides a means of creating good quality maps. All GIS products support the ability to plot data on any size plotting or printing device. Although there is still quite a bit of work that is required to create completely realistic maps, the end product can be extremely close to the real maps. The majority of the work that is required is the labeling of all the data so the labels are informative, but not confusing from overlapping or misplaced labels. Tools can be (and have been) developed to support the creation of these labels. Those algorithms are not yet mature enough to fully automate the process. The main reason for this lack of maturity compared to other processes is because map generation is always last in the data development cycle and not as high priority as other process development that ensure data correctness.

Any data that are part of external models that must also be included in paper maps need to be imported into the GIS to display on the paper

maps. For example, instanced data sets containing road network relevant features are sometimes stored in external models, but must be displayed on the paper maps. This is not out of the box functionality and the GIS data must be created from the external models, which must seamlessly integrate with the land-based vector data.

## Conclusion

The continuing escalation of requirements and demands for higher quality data, along with the overall industry pressures and customer desires for enhanced data longevity, data interchange mechanisms, data correlation, and quality have made the use of GIS-based systems a natural step in the evolution of simulation database generation tools.

Given the degree of uniformity and present GIS data interchange standards and translation capabilities, databases created with GIS-based DBGS tools can be more easily shared. This is especially true when considering the advanced database management technologies that lie at the core of most GIS systems, which include remote and multi-site data sharing capabilities. When considering the large amounts of data that current simulation environments demand and the historical expenses and challenges related to the management of such large data sets, the advantages of more sophisticated database management capabilities start to weigh in significantly in the life cycle and database development costs.

In addition to these data management advantages, the availability of time-tested and robust geometry manipulation algorithms at the core of the GIS tool data sets provides a solid foundation upon which more advanced and stable data manipulation and enhancement algorithms can be built. The availability of this wealth of knowledge not only improves the development life cycle of new extensions (or "plug-ins") to the DBGS/GIS systems, but also drastically improves the quality of the data being produced, largely because of the maturity of the underlying software.

When combined with the innate GIS multi data source importing capabilities and the rich variety of coordinate set translation, correlation and manipulation capabilities, a very appealing

picture emerges. This makes GIS-based DBGS tools not only an attractive, but desirable environment upon which present and future DBGS systems can be effectively developed and, most importantly, sustained.

DBGS vendors will need to look at GIS technology for their land-based applications to reduce the costs of DBGS software development as well as overall database management and production expenses. The GIS community will be continually enhancing their technology to support higher fidelity data development to support the requirements of their large customer base. Just as the gaming community has served as a performance and capabilities increase and cost reduction showcase for 3D visualization techniques and simulation, the GIS community offers great potential for cost reduction and enhanced data life cycles of database production, while increasing the quality and performance of DBGS applications and its derived data sets.

By using a GIS-based DBGS, the end user can attain significant productivity increases with fewer data anomalies and process-based errors. The higher data quality will support the reduction in the cost of database production at the same time increasing data fidelity and realism for the best possible training benefits.

## Author's Biographies

**Mark B. Field** is President and founder of FCSI. He has been working in the simulation industry since 1991 and was the system architect for Lockheed Martin's GIS-based DBGS successfully used for the UKCATT project.

**Pedro Ramos** is a highly qualified independent contractor who has been working in the simulation industry since 1983. He has worked with many simulation leaders including, but not limited to, Lockheed Martin, Evans & Sutherland, Raydon, and Quantum3D.

## References

<sup>i</sup> KAM GIS Dictionary,  
<http://www.kam.to/kam/services/gisdictionary.cfm#T>

Notice: All trademarks are property of their respective owners. Copyright 2005, Field Consulting and Services, Inc.

