Getting More out of GIS and Database Management Systems through Data Integration and Data Mining

Prasad Chittaluru, Carol L. Hunter, and Jeffrey F. Thompson

Geographic Information Systems (GIS) and databases are now widely used in the day-to-day business operations of government agencies. These tools usually require a significant amount of financial and personnel investment for installation, operation, and maintenance; however, they are typically not utilized to their full potential in the government agency setting because of various issues involving traditional operational procedures and communication across various departments within an enterprise.

Two Central Florida utilities have taken steps to minimize the hurdles and develop procedures toward putting their information technology resources to best use. This article focuses on water and wastewater utilities, but the findings and recommendations are relevant for any government entity with a significant investment in information technology resources.

Information Resources in a Public Enterprise

In many public utility enterprises, a wealth of information is collected by various departments. Here is a representative list of the information that is likely to be available for most utilities and related government agencies:

- *Utility Customer Service* – Water, wastewater, and reclaimed water billing, customer information.
- *Utility GIS* – All spatial information related to the assets owned, leased, and maintained by the utility.
- *Utility Engineering* – Capital improvement projects.
- *Utility Operations* – Water and wastewater production data, SCADA information, water and wastewater treatment parameters information.
- *Utility Maintenance* – Information pertinent to preventive as well as reactive maintenance, such as pump station repair, forcemain or other pipeline repairs, and treatment plant maintenance.
- *Property Appraiser* – Information related to the land parcels with the county or city served by the utility.
- *Planning and Zoning* – Information on current and future land use and zoning and planned developments.
- *MPO* – Population projections, economic/business statistics.
- *Water Management Districts* – Wetland coverages, contamination plumes. and other regulatory permits and coverages.

Information Systems Dilemma

Most government agencies are currently facing the information systems dilemma. The agencies have invested great amounts of financial and human resources in developing information systems across the enterprise. Due to the nature of change in information systems technology, significant additional resources are needed annually for upkeep and upgrading to prevent the systems and data from becoming obsolete.

Each department creates an "island" of data for its use, leading to multiple copies/versions of the same data. This leads to problems with data integrity and updating. Because the different departments in the agency can not seem to agree on an "enterprise system," the existing information resources are seldom put to their best use and full potential.

At best, leaving multiple resources unlinked denies the ability to use the system to its fullest potential. At worst, loss of data integrity can lead to poor project coordination and inaccurate master planning. For example, a maintenance department may repave a road section, only to find that the utilities department subsequently has to dig up the road for a water line upgrade.

By contrast, good data management empowers end users and improves efficiency and decision making by providing data on demand, standardizing data entry, and minimizing duplication. The near future is focused on a consolidated system in which all applications within an agency are tied seamlessly together. Employees, with varying levels of access and authorization, will be able to view agency data.

Web-based database applications integrate data from sources across the organization and provide information and applications that are targeted to users' needs. The information is accessible over intranets, which centralize maintenance of applications and data.

Prasad Chittaluru, Ph.D., P.E., is a program manager for information solutions in the Orlando office of the environmental engineering firm PBS&J. Carol L. Hunter, P.E., was a senior engineer with Seminole County Environmental Services when this article was written. She is currently a senior engineer with Orange County Utilities. Jeffrey F. Thompson, P.E., is a senior engineer with Seminole County Environmental Services.

Web-based GIS applications have all the advantages of database applications, plus they bring the full power of GIS to end users without the need for expensive software and training. GIS is an excellent data-mining and data-integration tool because it correlates information from different sources to reveal patterns. Its spatial component enables users to visualize impacts of data points beyond tables and lists.

Seminole County Environmental Services (SCES) and another Central Florida water and wastewater utility (CFWU) have applied the power of data integration and data mining to improve business process efficiency and customer service. The engineering consulting firm PBS&J used data mining and data integration technology plus the understanding of utility operations to develop a number of applications for the utilities.

Typical Uses of GIS in Government

In government agencies, GIS systems are typically used as repositories for asset information and as mechanisms to create *Pretty Maps* for commission meetings and public hearings. It is common to find that similar information is collected and maintained by more than one department in the enterprise. This information, though it might be useful for many departments, tends to become an "island" within the primary department collecting it. Access to the information is restricted due to some real and mostly "perceived" data security issues. The typical explanations provided for the creation of these islands include:

- Reluctance to let go of the control of data critical for their operations
- Perception that data ownership is a meas-

ure of authority
- Lack of trust in other department's data sources
- Different department managers having different opinions and support for technology
- Different hardware and software resources and platforms across departments
- Different amounts of financial resources available across departments
- Data is not maintained/updated regularly

The tendency to create "islands of information" across the enterprise is not specific to water and wastewater utilities alone. It is fairly common in many large government agencies, such as departments of transportation and county public works departments.

It is seldom easy, if not impossible, to access information from the island if you are outside the island. This issue has gained the attention of utility managers who are now faced with decreasing budgets and increased expectations of service quality. The question that comes up commonly is "How do we get more value from our GIS and database management systems to improve our operations efficiency?"

Solution to the Problem

Over the past few years, the paradigm in the information technology world has shifted away from consolidating disparate systems across the enterprise into one centralized system, to one that tries to "bridge the information islands" across the enterprise. Data integration and data mining are now increasingly seen as answers to bringing the information resources of the enterprise together and making them available to the employees and end users.

Data integration is the process of blending data from various data sources to create a larger, more comprehensive information source. Data integration facilitates reporting, query and analysis, and development of analytic applications across various enterprise data resources. It ensures that end users are always working with credible, meaningful, and current information. As a result, the enterprise sees increases in IT productivity, a reduction in total cost of ownership, and improvement in return on investment (ROI) from enterprise applications.

Data mining is the process of finding new, potentially useful knowledge from data that is already available. As a result of the advances in computer hardware and data storage capacity and the development of new technologies such as scanners, organizations are accumulating a large amount of data.

Data mining searches for and presents patterns in data that are not obvious to the end user. It can help organizations extract additional value from piles of accumulated data. Data collected for one purpose can be used for other purposes that may not have been envisioned when it was originally collected.

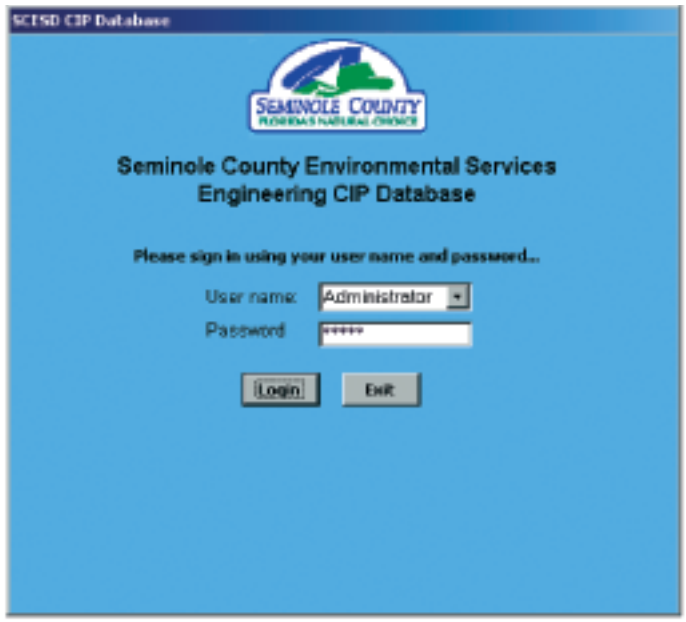

Figure 2: Seminole County Environmental Services ECIP Database sign-in screen.
Prior to this application's development, project information was stored in spreadsheets and text documents, so it was not always easy to retrieve consistent and reliable project information from one location. The ECIP minimizes project tracking difficulties that can accrue when information is stored in multiple documents and locations.

GIS as a Data Integration and Data-Mining Tool

GIS is quickly becoming the data mining and data integration tool of choice for water and wastewater utilities. When set up and used properly, GIS applications can reveal patterns in data that are very useful.

Spatial patterns are easily visible when data is presented using GIS. It is also an excellent mechanism to spatially "correlate" information from different sources. GIS data integration is leading the way for improving the business process efficiency of government agencies such as utilities by minimizing duplication and oversights in project implementation.

Over the past few years, SCES and another CFWU have made tremendous progress in harnessing the power of information resources available across the enterprise by creating data integration and data mining applications as "bridges across the islands of information." This development has made the information readily accessible by all authorized staff across the intranet. As a result, the departments are able to minimize duplication in data collection and storage, and there is increased confidence in the information available as a result of the significant reduction in data redundancy.

Here is a summary of some of the applications that *represent* Getting More out of your GIS and Database Management Systems through Data Integration and Data Mining:

Putting Technology to Work

Both utilities are developing an integrated suite of business applications that share data across divisions and processes, making information available on demand to county staff via the county intranet. These streamlined applications are used for plan review, customer service, transmission systems assessment, and GIS.

Before these applications were developed, data was entered over spreadsheets, text documents, and client/server applications. The new suite of applications is primarily

Continued on page 50

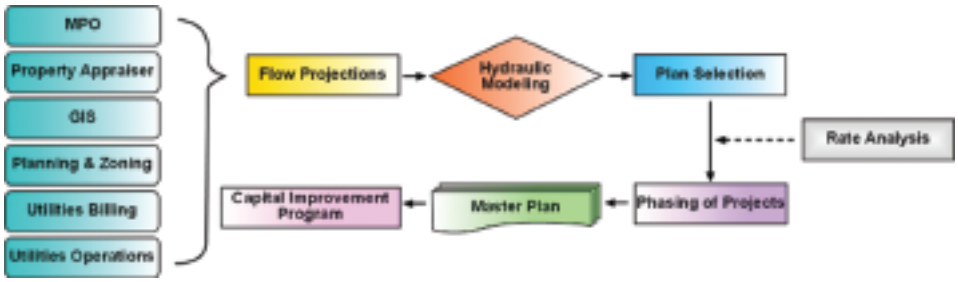

Figure 1: Integrated Master Planning Process

Figure 3: DMIS Application Screen Capture

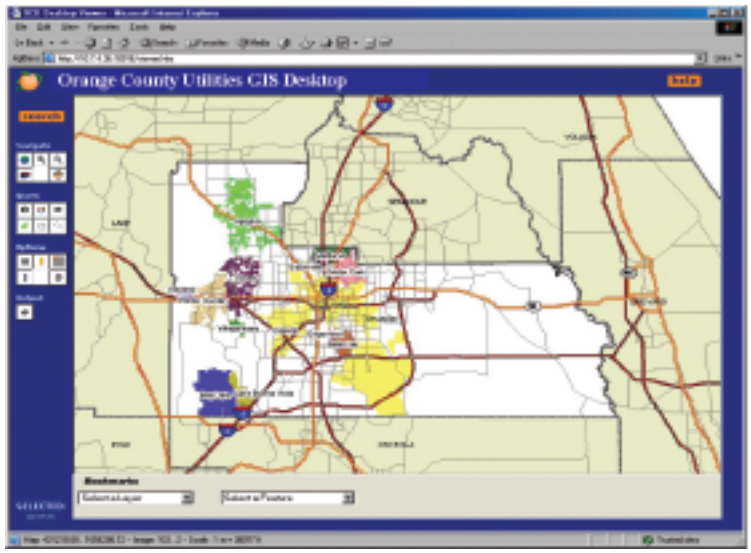

Figure 4: Utilities GIS Desktop

Continued from page 49

web-based GIS and database applications integrated through a common attribute called a sequence number. This attribute is now assigned to all new projects, as well as to GIS data, to facilitate cross-referencing data across the multitude of applications. Identification of such common attributes is key for successful integration of applications across the enterprise.

The utilities staff ultimately plans to integrate all major applications used across the enterprise and make them accessible to end users through a common, seamless interface with a shared user name and password. The new applications, described below, represent a significant step in this direction, and the utilities are poised to put their information resources to their highest, best use.

Utilities Master Planning

The two utilities have applied data integration and data mining to their master planning activities. An innovative master plan methodology was developed that integrated property appraisals, planning and zoning information, utilities billing, and data gathered from operations, SCADA, engineering, GIS, and metropolitan planning organizations. This approach provided a representative spatial distribution of current and future water and wastewater demands and facilitated development of more reliable hydraulic modeling to simulate system operation.

As a result of this approach, many other uses came out of the master plan. For example, the staff has more defensible consumptive use permit data. Retrofitting projects for areas served by septic tanks and private wells has been prioritized. Portions of the service area most suited for water conservation were identified. Figure 1 shows a schematic of how data is gathered to advance this master plan process.

Engineering CIP Database

An application created for SCES is being used to track and report on the planning, development, and construction projects handled by its staff. The Engineering CIP Database (ECIP) provides a common, intuitive interface allowing quick access to a variety of project information. County staff members use the tool to assist in project management from the planning phase through completion.

The ECIP tracks financial information such as engineers' estimates, budgets, work orders, invoices, and change orders. It also links and retrieves project documents, prioritizes activities, tracks progress, and assists the engineering staff in managing the work of consultants, contractors, and other county staff members. Figure 2 shows the user sign-in screen used to access the system.

The ECIS application was originally developed in Microsoft Access and VBA (Visual Basic for Applications). The backend database was migrated to Microsoft SQL Server to address data access needs of SCES staff and to provide a more robust relational database management system (RDBMS) for data management.

Data Management Information System (DMIS)

The Data Management Information System (DMIS) supports the business activities of the plan review section, utilities planning section, and customer service division. This application facilitates centralized tracking of planned development project review, calculation of user connection fees, hydraulic model demands, and estimation of available system capacity.

Migration to the DMIS application included development of complex data import routines and creation of intricate interface screens and database reports. Data entry, queries, and reporting are handled through the web-based interface screens. User names and passwords allow varying access levels to protect data integrity. Figure 3 shows a screen capture of the DMIS application.

Utilities GIS Desktop

A web-based mapping application, Utilities GIS Desktop, was developed using ArcIMS to provide search, viewing, and printing tools for utilities staff members who are not proficient in the use of GIS. Figure 4 shows a screen capture of the Utilities GIS Desktop application. This application provides GIS data and map products to the utilities staff over the intranet and has increased the use of GIS data across the enterprise. Using the GIS Desktop application, staff members can print standard map products, such as map books and system maps, from their desktop without requiring a license for GIS software.

Call Tracker Application

A web-based GIS application, Call Tracker, integrates GIS and a work-order management system used by a Central Florida utility's solid waste division. Figure 5 shows a screen capture of this application. An ArcIMS map interface was developed to the work management system for creating the work orders and graphically displaying the type and status of existing work orders for call center personnel. The map portion of the application was developed using ArcSDE views and Oracle database links to dynamically show the change in status of work orders on the map interface.

For the web applications discussed above, Active Server Page (ASP) and .NET

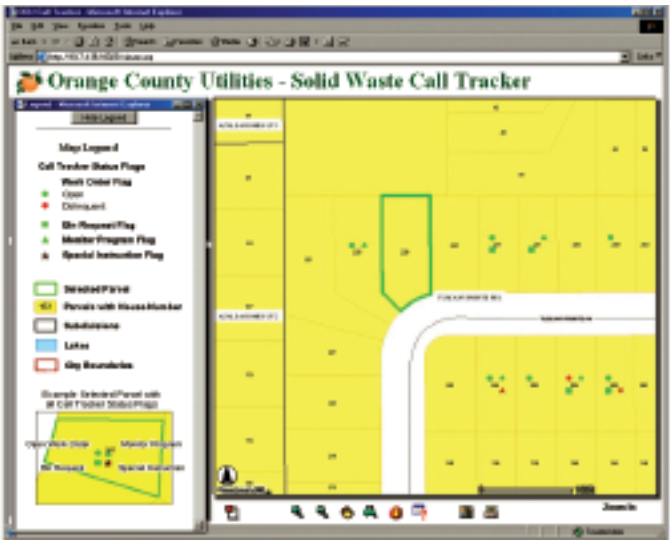

Figure 5: Call Tracker Application

technologies were used for backend functionality development. DHTML, JavaScript and Cascading Style Sheets (CSS) were used to develop the graphical user interfaces (GUI). Oracle 8i was used as the RDBMS to store and retrieve data. The GIS applications were developed using ESRI software including ArcIMS, ArcGIS, ArcObjects and ArcSDE.

Future Applications

In the future, utilities and other public agencies can create an Enterprise Applications Center on their intranet. From this site, the agency's staff can access all the major business applications used across the enterprise.

The access level to these applications can be determined automatically from the user ID and password detected by the application, providing one central access point for all the critical business information for the staff and eliminating the need for the staff to remember multiple user names and passwords. At the same time, the agencies can also identify business processes that can be improvised using web-based GIS and database applications.

Working toward Achieving Enterprise Data Integration

An agency's motivation to create a fully integrated, enterprise-wide system sometimes runs into stumbling blocks during development. Departments may have different hardware and software resources, incompatible platforms, even varying budgets—all factors that make integration difficult.

There is also the perception that whichever group is considered the data's "owner" or "manager" has the most power. One group may be reluctant to relinquish control to an integrated system. Some users, formerly charged with maintaining a particular database, may not trust the new system or may not believe that the data they see is correct.

Once acknowledged, such factors can be avoided. The first step is to build an effective team for the project. Identify a "champion" within the organization, someone who understands the value of fully accessible data. Then, identify key players familiar with systems and applications used in each department. Team members should have a strong understanding of the agency's responsibilities and good GIS/IT knowledge.

If an agency does not have a large in-house computer staff, an outside consultant may best handle design of hardware and software used for the integrated system. In either case, the key to a successful transition is starting small. Implementing the changeover in phases allows employees and managers to

Continued on page 54

Continued from page 51
gain familiarity and trust in the system. Buy-in across the enterprise is required.

The Starting Point

Some tasks and processes are natural starting points for creating a GIS-based system. Good candidates include data-intensive tasks, repetitive activities, standard reporting, sharing secure data among multiple users and locations, and projects with computer model simulations. A recommended application design and development process is shown in Figure 6.

To integrate systems across the enterprise, identify the following:

- Major systems and applications across the enterprise.
- Hardware and software types and versions.
- Systems and applications that duplicate data collection and storage.
- All business processes (to eliminate redundancy).

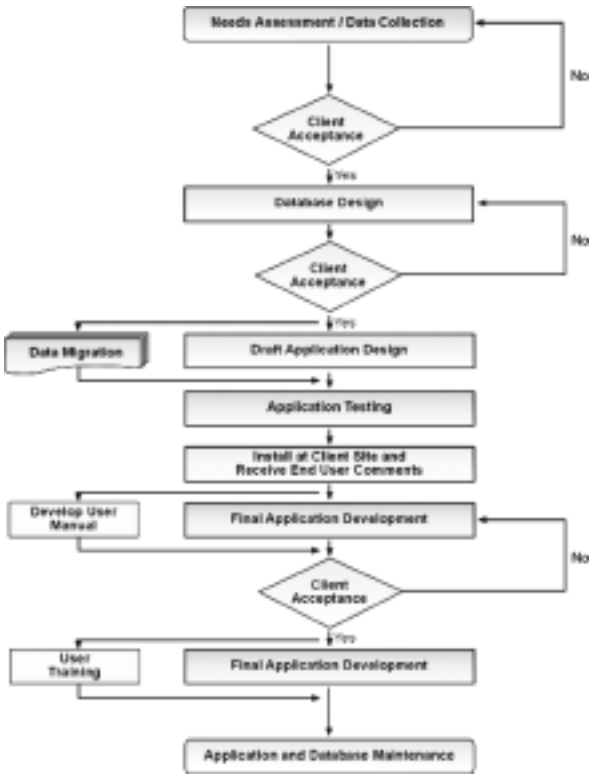

Figure 6: Recommended Application Design and Development Process for Enterprise Systems

If systems cannot be eliminated, the agency can design and implement a "data bridge" to integrate them.

Following a needs and systems assessment, the agency can develop a prototype. Then, the selected application can be developed, tested, and deployed. Training for users and personnel maintaining the system ensures that the application will be used to its fullest potential. The same team selected for the project can watch for new uses and suggest future enhancements so the system grows with the agency.

Conclusion

The recent developments in web-based GIS and database technologies provide excellent data integration and data mining tools for public works and utilities. The applications discussed provide a flavor for the variety of activities that can benefit from application of these technologies. Regardless of the specific data collected or the size of the agency, it is now feasible to get more value out of the investments made in the information resources across the enterprise employing the tools and technologies discussed in this article.