SDE Enterprise Geodatabase Design and Development in Loudoun County, Virginia

Authors: Rupali Kale, David Torraca, Marsil Zook

Abstract

For the past several years, Loudoun County, Virginia has been ranked among the fastest growing counties in the nation. Loudoun's GIS is a mature, enterprise system, having its beginnings in 1986. Loudoun is developing the SDE Enterprise Geodatabase model to standardize GIS operations and business processes throughout the organization. The county uses DB2 for its relational database in conjunction with ArcSDE, which presents its own unique challenges.

This paper explores the challenge of implementing ArcSDE in conjunction with supporting existing procedures and developing new business practices. Loudoun's existing system architecture and database construct will be detailed, with the focus on database design, data integration and publication, and database maintenance procedures. In addition, this paper also describes some out-of-the box tools used to manage the database and some in-house tools developed for monitoring and reporting SDE database activity for data maintainers and users.

Loudoun County in Brief

Loudoun County, Virginia is located approximately 25 miles WNW of Washington D.C. and is part of the Washington D.C. Metropolitan Region. According to the 2005 Loudoun County Department of Economic Development Annual Growth Summary, "Over the past five years, the county led the Northern Virginia region with a population growth of 52 percent, and was recognized by the U.S. Census Bureau as the fastest growing county in the nation with a population over 100,000. Loudoun County has been one of the fastest growing counties in the U.S. since the late 1990's. Projections show that Loudoun will be one of two counties in the region to continue to experience double-digit growth for each decade, 2010 through 2030." The 2007 Loudoun County population estimate is 271,987. Additionally median household income in 2005 was $98,483, which according to the U.S. Census Bureau was the highest household income for a county in that year.

Geographically, Loudoun County encompasses 517 square miles with dense suburban development in the east and rolling farmland and vineyards to the west. Loudoun is bounded on the north by the Potomac River and on the west by the Blue Ridge Mountains. The County has a rich history and is celebrating its 250th Anniversary in 2007.

1. LOGIS (Loudoun County GIS)

Loudoun was one of the first jurisdictions in the nation to acquire GIS, having installed ESRI software in 1986. First used to store environmental, basemap, and land records data for mapping and analysis, the system has grown to become a key component in the county's enterprise system.

1.1 Hardware and Software

The current system components include: 1) IBM AIX Database Server hosting ArcSDE in Unix with IBM's DB2 version 8 as the relational database, 2) Windows 2000 File Server, 3) Windows 2000 Citrix Farm hosting ArcGIS Desktop and Workstation 9.1, and 4) end users with Citrix Clients (thin) and a small number of local ArcGIS 9.1 installs (fat clients). All clients access the server components through the county network. (See Figure 1: Loudoun ArcGIS System Layout).

In addition, GIS data is served by a Visual Basic (VB) and MapObjects (MO) application through ArcIMS on dedicated Windows Internet and Intranet servers.

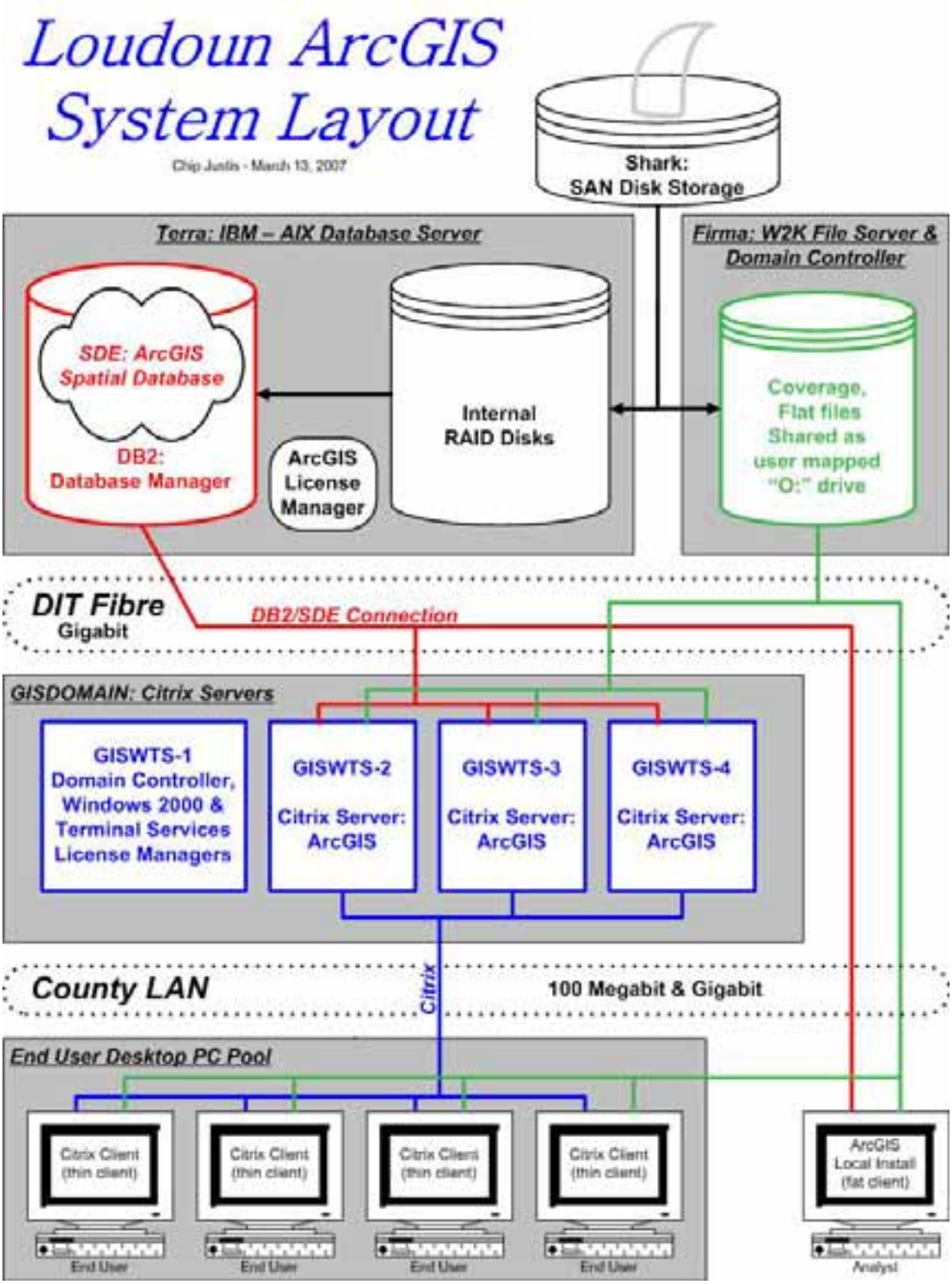

Figure 1: Loudoun ArcGIS System Layout, Diagram by Chip Justis

1.2 People

The County's GIS user profile includes two types of users, those who access ArcGIS Desktop with appropriate access authority, and those who access data through a web browser. Users with access authority include managers, administrators, programmers, analysts, data maintainers, technicians, and map makers. Those users without access authority are classified as browsers and use GIS data through the intranet or internet (www.loudoun.gov/gis/weblogis) and number in the thousands.

The following chart (Figure 2: User Base) details the number of user ids assigned to individuals per department, agency, and jurisdiction. Facilities are located throughout the county, but the majority of county facilities are currently located geographically within the jurisdiction of the Town of Leesburg.

Department	No. of User ID's
Building and Development	54
Economic Development	4
Environmental Health	10
Extension Office	1
Fire and Rescue	4
General Services	11
Information Technology	10
Juvenile Court Services	2
Management and Financial Services	6
Office of Mapping and GIS	24
Parks and Recreation	2
Planning	22
Public Schools Planning and Legislative Services	2
Office of Solid Waste Management	1
Sheriff's Office	2
Town of Leesburg	2
Town of Purcellville	1
Transportation Services	8
Treasurer	1
Total Number of User ID's	**167**

Figure 2: User Base

1.3 Data

The GIS hosts 87 corporate core data layers categorized as follows:

Basemap (13)
Environmental (17)
Land Records (16)
Planning (29)
Public Safety (4)
School boundaries and sites (4)
Utilities (4).

Most of Loudoun's corporate data began as coverages on previous computer platforms. These coverages were migrated to a previous system configuration using ArcGIS 8.3 in order to test SDE. Users were given access to the coverages while the development and testing of data maintenance procedures in SDE commenced. Coverages continued to be updated as data were maintained until recently. Data maintenance is currently supported in SDE, personal geodatabase (pgdb), and coverages. Programmatic data management insures data updates are distributed to the database and to other data formats on the GIS as well as to other enterprise systems.

Land records data are maintained as a versioned geodatabase in SDE. Data are updated on an hourly basis as documents are recorded at the Clerk of the Court's office. These documents, in the form of plats and deeds, are the source material for the maintenance of parcels, street centerline, and addresses, key data used to manage many of the county's services. Parcel boundaries are captured from these developer plats and recorded surveys and Parcel Identification Numbers, (PIN's) are generated using ArcGIS. These identifiers are transferred to the County's Land Management Information System (LMIS) along with other GIS spatial classifications related to environmental characteristics or planning designations. LMIS serves this data, as well as permitting and assessment information, among others, to users throughout the organization in tabular format.

Centerline data are transferred to the County's Computer Aided Dispatch (CAD) system to identify and map caller locations for dispatchers. Addresses are created and stored in the GIS and distributed to other systems such as CAD, schools for planning and student enrollments, and LMIS for permitting new structures.

The county contracted for aerial photography beginning in 1979, and 1:12000 scale hardcopy prints have been produced nearly every year since. A countywide orthophoto was produced from the last three missions and has been made available to users through SDE. The county's base map is maintained by contract from the aerial photography using photogrammetric methods. The base map data includes; edge of pavement, buildings, forest cover, topography, spot elevations, fences, and other data layers. These data are used to assist GIS users in determining precise locations of other mapped features and are used in analyses such as impervious surface and viewshed modeling.

The Office of Mapping and GIS also has a public information counter that uses an ArcGIS application to create 1:2400 and 1:4800 scale map products for the general public and for staff. These maps use standardized symbology and formatting for a consistent product tailored to the end users' needs.

Weblogis, the county's web mapping application, is used by the public information counter to respond to customer queries, and the application is accessible to the public on the internet. The mapping tool interfaces with a portion of the assessment database, and users can toggle from mapping to assessments and back. Both the intranet and internet weblogis applications are updated nightly.

2. SDE Integration

2.1 In the Beginning

In the year 2000, ESRI was consulted to analyze the county's existing GIS and provide a system implementation strategy for both new hardware and software. Loudoun's system at that time consisted of an HP Unix Server using ESRI's ArcInfo software version 5. Once the new configuration was determined and servers installed, Loudoun began migrating its legacy system and business practices to the new platform as they existed in the previous system. This involved moving coverages and programs written in AML to new servers. Users continued to work on both the old and new systems in ArcInfo Workstation version 8 as analysts explored working with the geodatabase and SDE.

There were many things to learn about the new system including; working with Citrix, file management outside ArcInfo Librarian, new methods of printing, Visual Basic programming, DB2 and SDE management, windows security, and the many new features of ArcGIS. Software and hardware anomalies plagued progress from the start but gradually staff, with assistance from ESRI, began to improve system performance and stability.

Lessons learned from this migration included the need for proper staffing and training, and better management of both the system migration and the paradigm shift from established business practices. Both are critical for developing a long term support system for GIS operations. Loudoun County's GIS staff, having worked through critical issues in the past, was relatively well positioned to make the transition to SDE. Additionally, having a dedicated SDE Administrator for database analysis and support is a critical component to a successful strategy.

2. 2 Database Structure

The database consists of four instances: Production, Browse, Raster and Test. The Production instance contains versioned data and is the source of updated data for other enterprise systems. The Browse instance contains the latest data updates and, as such, is updated daily from the Production instance. The Raster instance contains all raster data

including orthophotography and Digital Elevation Models (DEM's). The Test instance is where all the development work takes place before it gets implemented in Production.

All changes to layers or to the database are currently done in Test and then implemented in Production and then to Browse for general user access.

2.3 Security

Five levels of security have been implemented based on the roles of the users: Database Administrator, System Administrator, SDE Administrator, Data Managers, and Data Editors. The Database Administrator administers database security policy and manages database creation on DB2. The System Administrator administers security policy for the GIS servers, including Windows Terminals Servers, the Database Server, and File Servers. The SDE Administrator administers security on the DB2 instances as related to GIS applications. The Data Managers manage maintenance processes on QC versions including the processes of reconciliation and posting. The Data Editors edit records.

2.4 Users/Group assignments

SDE access is granted to users in three steps. First, users are added to the Unix server, establishing user access to the server. Second, users are added to the DB2 database through IBM's DB2 Control Center. Finally, users are granted permissions to SDE layers through ArcCatalog.

In April, 2007, group assignment in the Browse instance was implemented. It should also be implemented in the Production and the Raster instances before migrating to ArcGIS Server 9.2.

2.5 Managing the database

The database is managed through ArcCatalog, Hummingbird's Exceed software, and database administration tools such as IBM's DB2 Control Center and Toad by Quest, Inc. GIS tools are used to manage data layers and raster images in functions such as loading, deleting, and editing. Database administration tools are also used to manage DB2 and SDE permissions and database objects. Exceed is used to access and run SDE commands at the command line and to monitor database logfiles.

Log files and the PROCESS_INFO table are monitored on a regular basis. There are approximately 30 session log files which are based on individual connections or giosvrs. This configuration is preferred over user log file tables in which multiple users compete for log file space. However, several months after the creation of these log files a serious problem was encountered when the log files were found to be growing at exponentially increasing rates. After some investigation, it was deduced that the problems were related to an unanticipated buildup in the number of delta table records for certain layers. After instituting nightly and monthly compression tasks, these problems have not reappeared. The PROCESS_INFO table is cleaned up once a week to remove any hung processes which occur

when the application is not shut down properly as a result of computer crash or a network error.

2.7 Versioning and Reconcile/Post

Data maintenance and editing are done using ArcSDE in a versioned environment. Parcels, addresses, centerline, town and county boundaries, and topology rules are maintained in a feature dataset. Topology rules apply to all feature classes within the feature dataset. Parcels, addresses, centerlines, and associated tables are updated to the Production environment on a daily basis.

There are three version levels below SDE Default. A Production version is owned by a Production entity; three QC versions of Production are owned by the parcel, address and centerline coordinators; and several EDIT versions are under the QC versions. Parcel edits are made in personal geodatabase and transferred via "cut and paste" into the EDIT version. The edits are reviewed and then reconciled and posted to the QC version. Once a day the QC versions are reconciled and posted to Production and then to SDE default. Address and centerline edits are primarily done within the EDIT version. After review, the EDIT version is reconciled and posted as described above in the parcel editing process. There are no personal geodatabases involved in the address and centerline processes. See Figure 3: Versioning.

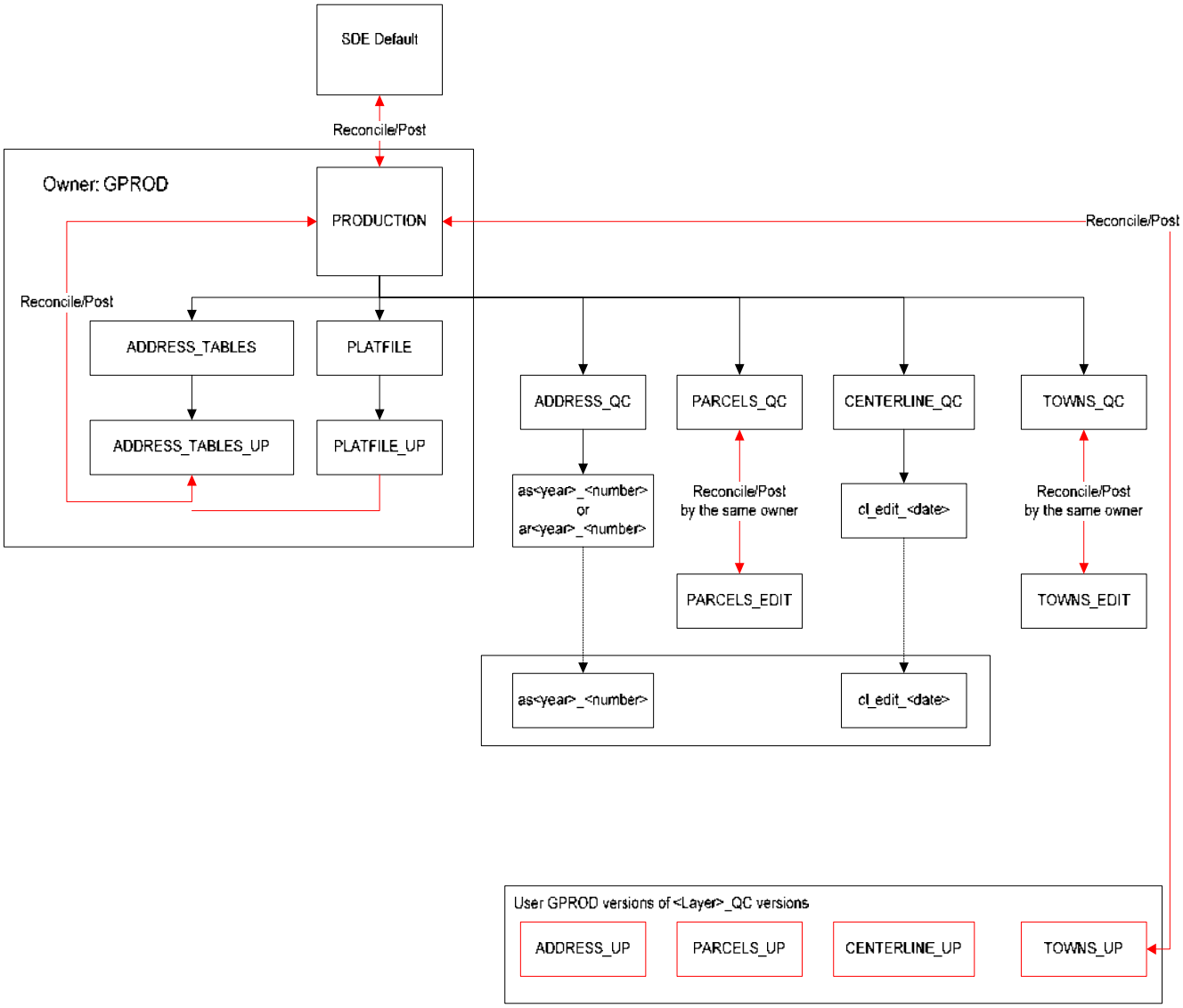

Figure 3: Versioning, Diagram by Mike Fauss

2.8 Compress

Each evening the Production instance is compressed to reduce the number of unreferenced states. Immediately before and after compression, an Analyze is performed to update database statistics. The Analyze and Compression operations are run as a scheduled Unix task. Since the nightly compression is conducted without the deletion of any established versions, the procedure is called a "shallow" compression. On the first Tuesday of each month, a "deep" compression is conducted in which all versions up to DEFAULT are deleted. The deep compression typically reduces the number of referenced states to one. After each monthly compression is complete, the systems analyst and data maintainers restore their versions.

2.9 Conflict management

A number of conflicts were encountered when the reconcile/post process was first initiated, but there has been a gradual reduction as the systems analysts began to understand the various workflow elements. Conflicts are resolved either by the systems analysts or the data maintainers after the program has completed. One type of conflict that still persists, for both spatial and non-spatial attributes, has taken on the name "ghost conflict." In these cases, a feature that is updated to a new state in the Conflict Edit and Pre-Edit versions remains in the original state in the Conflict version. Although it is not known for certain why this is happening, it appears that a failure to delete the Tier II and/or Tier III versions after each posting results in a persistent residual state that is in conflict with the post-edited states.

2.10 SDE Reporting and Management Tools

An in-house built .Net database application is used as a data viewing, reporting, and data repository tool for each database instance. This tool uses views based on SDE system tables, custom tables, and relationships between them as defined in the database.

The Loudoun defined information had historically been stored in a Microsoft Access Database. Known as the Corporate Data Key or CDK, this table was recreated in DB2. It includes information such as staff members responsible for the data, data distribution, and other pertinent remarks. See Figure 4: Corporate Data Key Concept Model in DB2.

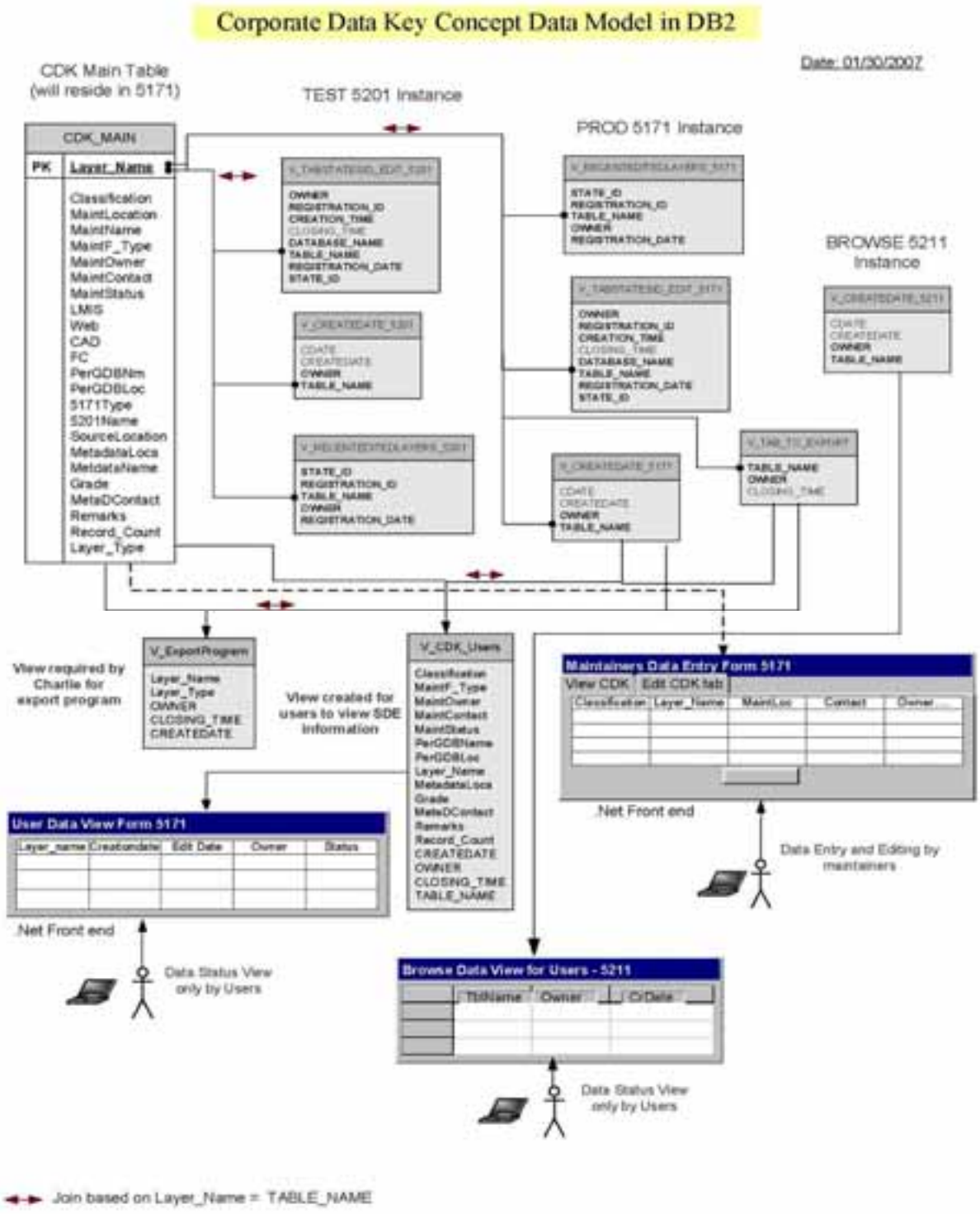

Figure 4: Corporate Data Key Concept Model in DB2, Diagram by Rupali Kale

A .Net database reporting application has been developed to report SDE specific data information and additional Loudoun County specific metadata information for all the SDE data instances. Reports contain data status information, live from SDE tables and views.

Pertinent information for management and users include creation date and modification date. Creation date is the date when layer was created in SDE. Last modified date is the date when the last edit state was closed for that data layer.

This .Net application will eventually have three different front end interfaces: One for users, who will be able to see the data status for each SDE data layer in all instances; a second front end for data maintainers who will edit and maintain the main CDK table; and a third interface for general GIS users to see only Browse instance reports. Currently the SDE status reporting application, which reports the creation date, modified date, feature class names and their owners, has been implemented. See Figure 5: SDE Reporting.

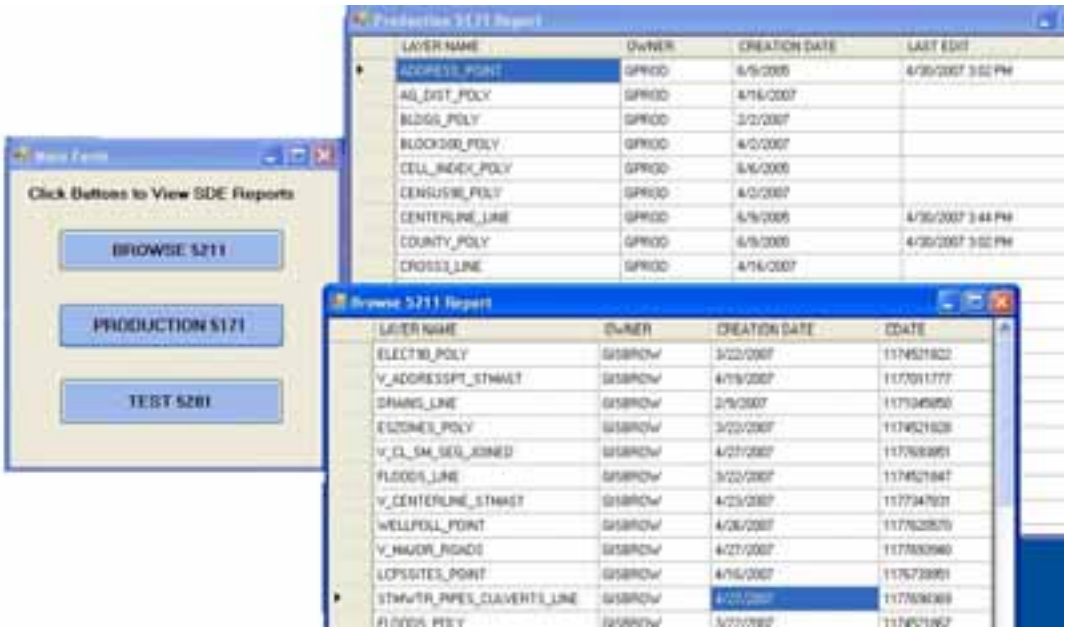

Figure 5: SDE Reporting

3. Data Maintenance

3.1 Data Updates from outside SDE

Data updates to coverages and SDE instances are accomplished through VB, VBA and ArcObjects custom programs and some out of the box tools. Loudoun County's current versioning scenario, as explained earlier, includes QC and Edit versions for data sets like parcels, addresses, and centerlines which are edited, reconciled, and posted to the Default version on a daily basis. Once the edits are posted to Default, another in-house built program, SDE2COV, runs as a scheduled task. This program converts data from SDE to Personal GeoDatabase and subsequently to coverages.

Database updates are performed by data owners within their respective departments. In the existing system, the data owners edit either shapefiles, coverages, or feature classes. The developers run programs to validate and update the data to SDE Production and Default versions, and then back to coverage data.

Data updates to the Web service run nightly as scheduled processes pulling data from SDE using AML. These data are synthesized, converted to shapefiles, and then pushed to the web servers daily. See Figure 6: Weblogis Process Diagram.

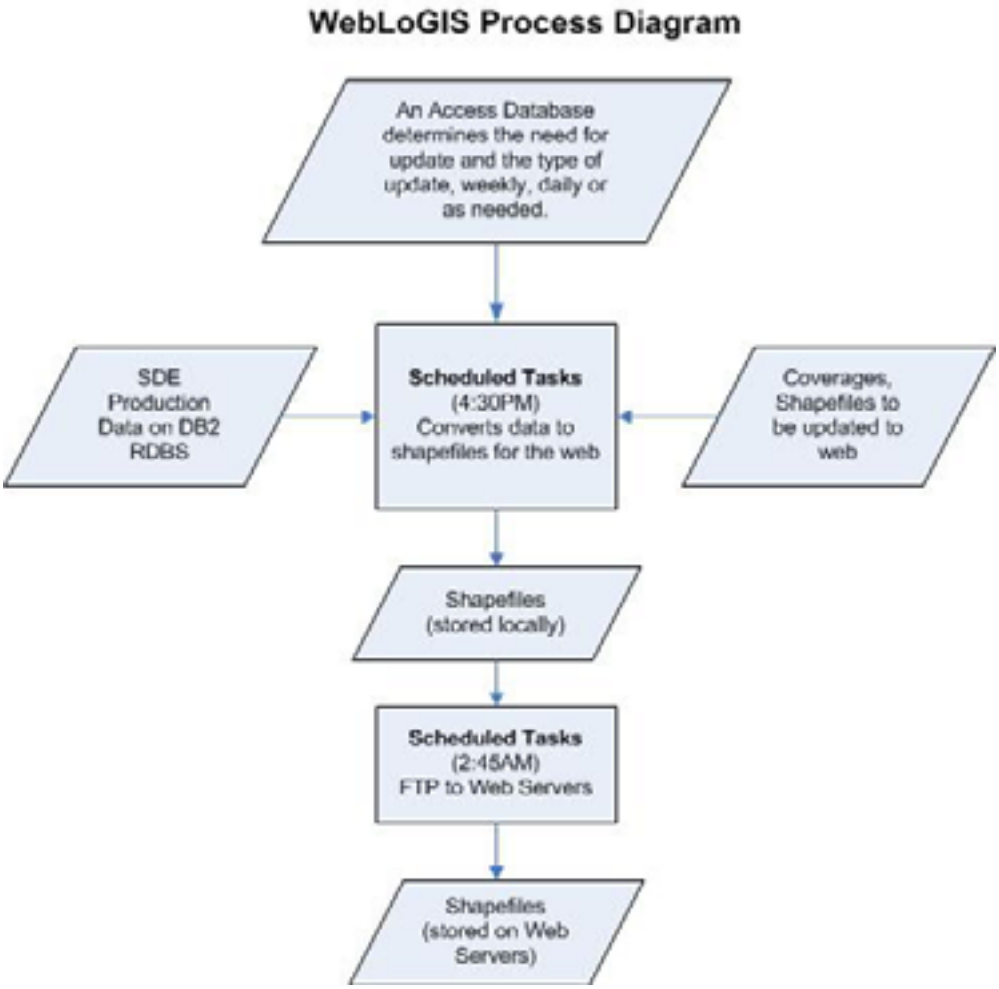

Figure 6: Weblogis Process, Diagram by Sue Carlson

Some in-house built tools are used to customize functionality for editors and to update data from non-SDE sources like shapefiles and personal geodatabases to SDE. For example, Loudoun County's Wells data is a point feature class maintained in all three SDE instances, Production, Test and Browse. These data are edited in personal geodatabase by the data owner in the County's Health Department using custom built tools to easily edit and validate the data. The edit and other tools are built using ArcObjects and VB, and VBA. Once the editor completes the edits for the week, scheduled scripts are run to validate and update the data to SDE Production and Test and Browse instances. See Figure 7: Well Data Processing.

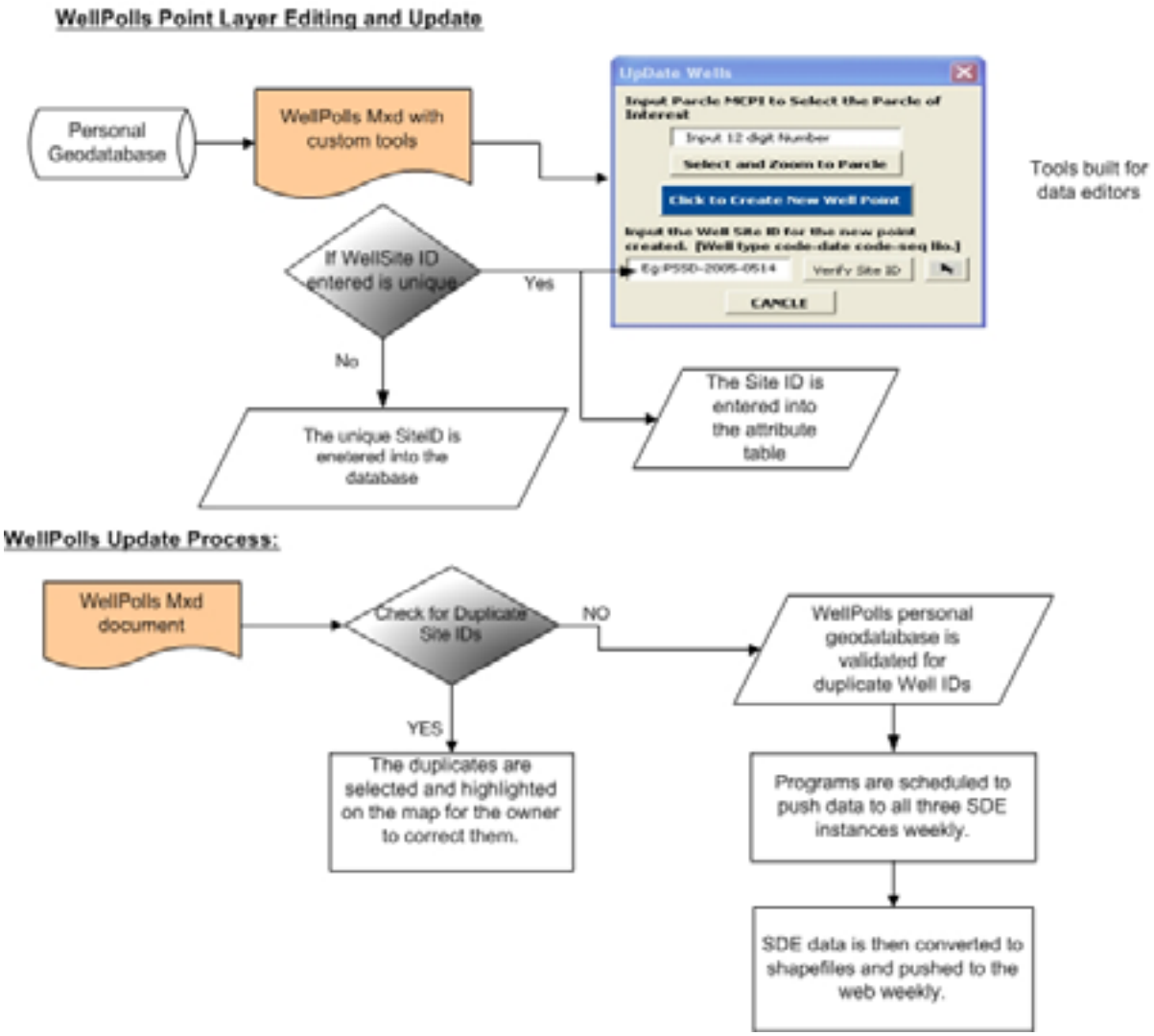

Figure 7: Well Data Processing, Diagram by Rupali Kale

3.2 Data Transfer to LMIS

The Land Management Information System (LMIS) is the County's central enterprise database that consolidates information from several other data systems in the organization. There are several spatial data layers, such as zoning and floodplain, for which acreage information is transferred to LMIS when parcel data changes. Alternatively, when zoning or floodplain information changes, these changes also need to be updated in LMIS. These data updates are transferred from the GIS to LMIS using two different approaches; 1) Pushing only edited data elements within a data layer to LMIS, also known as a Big Crayon Process, and 2) Pushing all the data elements within a layer to LMIS, also known as a Full Conversion.

Custom applications were developed in VB, ArcObjects, and SQL for both these approaches. In the first approach, whenever there are minor updates to a data layer, the previous version

of the data and the new changes are combined with the current parcel data, compared, and then filtered to reflect only the updated areas. Parcel Identification Numbers (PIN's), feature codes, and acreage for each changed polygon are transferred to LMIS. In the second approach, when there are multiple changes to a layer the data are combined with the current parcel data and all records, including the same field values as noted previously, are transferred to LMIS.

3.3 Data Update Process between SDE Instances

Tools were developed in-house to keep Land Records data synchronized between Production and Browse instances. The Browse instance is updated nightly using a scheduled geo-processing model. The update program does not re-create data in Browse, since there are views and metadata that need to be maintained. The process truncates all the records in feature classes and tables in the Browse instance, then updates all the records from Production to Browse. This method assures the persistence of relationships, views, and metadata.

Existing spatial views include commonly joined table relationships and subsets of feature classes. For example, one of the common joins users perform is a join of address points or centerline arcs to the table that contains all the street names. By creating a view of these table relationships and by keeping the base information up to date, users no longer have to perform the join themselves. Another example of a fairly complex view is a major roads view in which the centerlines feature class is joined to two different tables and filtered, based on a related field, to give only major roads along with street names and routing information.

The Browse instance was implemented to allow for load balancing between data editing in the Production instance and data viewing. Since Production data are versioned, single version views will not show any updates unless the instance is compressed to state 0. Spatial views are created and maintained in the Browse instance for ease of maintenance, updates, and to provide custom data to users. Multi-version views can show dynamic updates based on versioned data, but these are attribute only for DB2 and do not contain any spatial information.

3.4 The Future: Database Replication Tools

One of the main advantages of migrating and maintaining an enterprise SDE database is that distributing and synchronizing data becomes very easy. Distributing data between instances in the future will be done using database replication tools that come with ArcGIS 9.2. Distributing data across platforms from SDE to LMIS, CAD, the Public Information Counter, and the Web will involve some custom in-house program changes. Current data distribution to these systems will undergo some changes based on this future implementation. In addition, most if not all data will be edited and maintained in SDE.

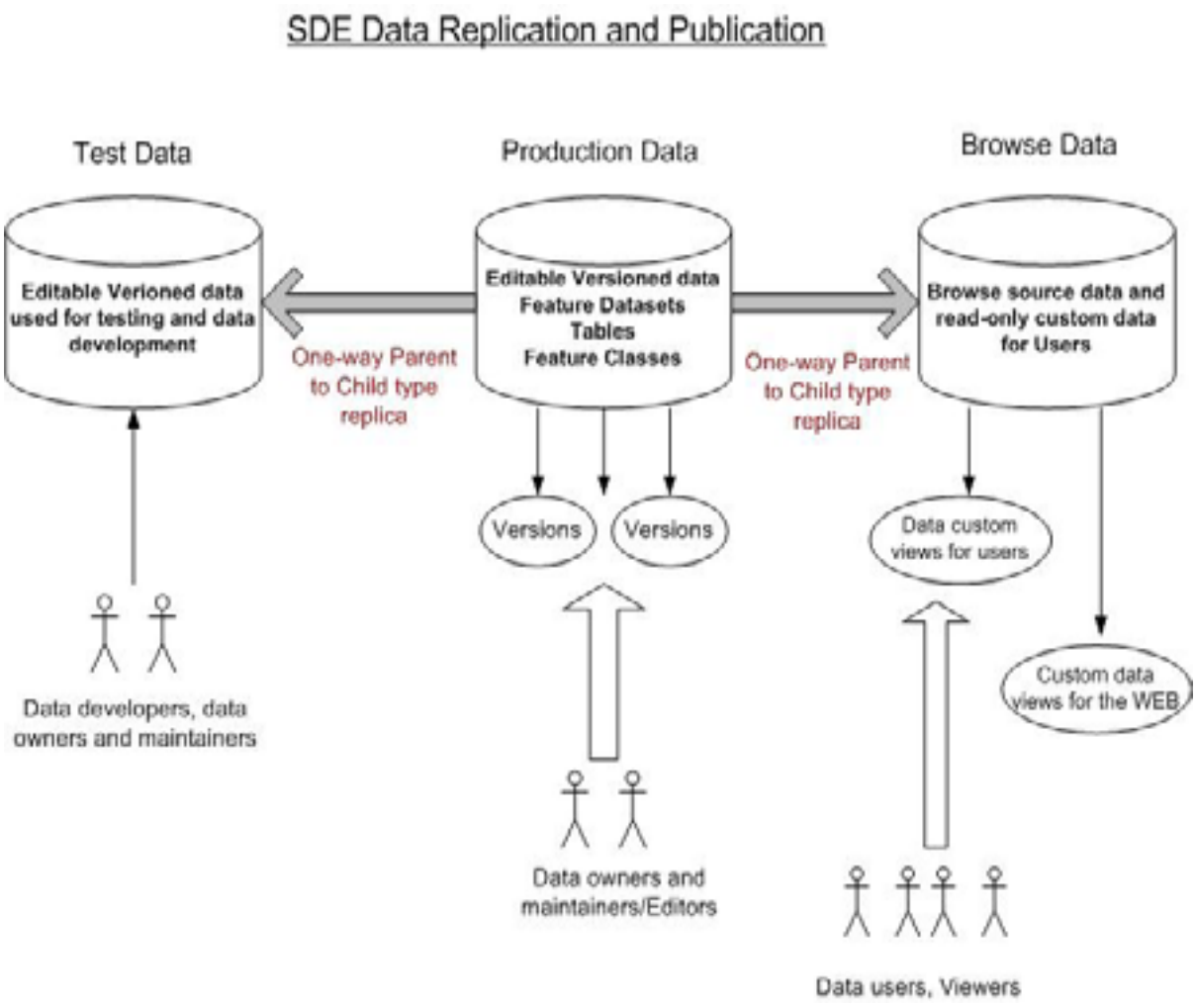

Figure 8: SDE Data Replication and Publication to be Implemented, Diagram by Rupali Kale

Database replication will play a very important role in data publication to users. To be able to update data to both the Browse and Test instances, the strategy is to implement one-way replicas since most of the editing will be performed only in the Production instance. This will help synchronize data one-way from Production to Test and one-way from Production to Browse. This will be a parent-child replica type. The main advantage of using this tool is that the data can be synchronized a number of times a day, but does not require that all the source data be versioned. The option to select a replica by re-using the existing schema will be implemented. By reusing the existing schema, permissions and single version SDE views can be kept intact. The database replication method will be tested on a feature dataset in the Test instance to test parent-child replicas while re-using the existing schema. It will also be tested on the current edit state of data to be replicated and on views created in Browse. See Figure 8: SDE Data Replication and Publication.

Conclusion

Loudoun County will continue to grow in population and is dedicated to serving the public's needs by leveraging advances in technology. This paper discussed the County's GIS and its implementation of SDE as an important part of the strategy for accomplishing this goal. Through SDE, GIS data can be managed more efficiently and effectively to a broad range of users.

The release of ArcGIS 9.2 provides yet another opportunity to enhance existing procedures through utilities such as database replication and improvements in database management. Working with views will assist us in meeting user needs by presenting data as business processes require. Experience with system migration has formed a good foundation to understand the implications of future changes to GIS.

Acknowledgements: The authors would like to thank the following Loudoun County employees for their assistance in making this paper possible: Susan Carlson, GIS Web Analyst; Mike Fauss, GIS Land Records Manager; Chip Justis, GIS System Administrator; Charles Shore, GIS Programmer Analyst; Larry Stipek, Director Office of Mapping and Geographic Information; Shaning Yu, GIS Programmer Analyst.

Authors Contact Information:

Rupali Kale
GIS Programmer Analyst
Office of Mapping and Geographic Information
1 Harrison St., MS#65
Leesburg, VA 20175
703-777-0190
703-771-5075 (fax)
Rupali.Kale@loudoun.gov

David Torraca
GIS Manager
Office of Mapping and Geographic Information
1 Harrison St., MS#65
Leesburg, VA 20175
703-777-0367
703-771-5075 (fax)
David.Torraca@loudoun.gov

Marsil Zook
SDE Administrator
Office of Mapping and Geographic Information
1 Harrison St., MS#65
Leesburg, VA 20175
703-771-5092
703-771-5075 (fax)
Marsil.Zook@loudoun.gov