Miami-Dade County Editing and Geodatabase Migration

Ivelisse Torres-Alejo
Julian M. Albo
Elio Elias

Abstract

The purpose of this paper is to describe the various phases involved in the migration from the world of coverages and AML to the world of ArcGIS desktop and the geodatabase. The paper will describe the unique approach taken at Miami-Dade in creating a multi-user versioned environment for the co-existence of various County departments accessing the same centralized geodatabase with the use of ArcObjects editing applications’ deployed via the Citrix environment. The first phase included a single Oracle instance and geodatabase in ArcSDE 9.1, and the second phase included different project geodatabases in the same Oracle instance in ArcSDE 9.2 and ArcSDE 9.3. The paper includes the configuration of the versioning schema, Oracle instances, and the ArcObjects editing extension and supporting Oracle tables, the quality checking tasks enforced within the editing process for data integrity purposes and finally, the replication process from the master editing geodatabases to the publishing geodatabases.

Introduction

Miami-Dade County has an enterprise GIS that includes among its various components a centralized data environment for the production, maintenance and distribution of geographic data. During 2004, the GIS Technical Support team of the County’s Enterprise Technology Services Department (ETSD) started the process of migrating the editing applications and database infrastructure from AMLs and Librarian coverages to ArcGIS/ArcObjects and the geodatabase. When this project, named ArcGIS Geodatabase Migration, was started, Miami-Dade County had an ArcSDE infrastructure used by all County Web applications. The ArcSDE instances were weekly refreshed by converting the appended and single tile coverages into ArcSDE standalone feature classes, but the data editing was still performed in old formats with old tools.

From the start of the migration, it was decided that County staff would be responsible for the management and development of the project. The main objective was to create an integral solution using the new ESRI technologies. The County also reserved the option of using ESRI consulting services when necessary.

The Arc/GIS Geodatabase Migration Project consisted of the following components:

- Complete replacement of hardware platforms, operating systems, database releases and application development tools.
- Architecture design of the ArcSDE/Oracle instances for a single geodatabase and multiple geodatabases.
- Design of new geographic data models planned to coexist in the same enterprise editing environment.
- Conversion of existing applications to Web and client/server desktop solutions integrated to new industry standards.
- Conversion of all batch jobs to new hardware infrastructure, operating systems, programming languages, command line scripts and scheduling agents.
- Retraining of the users’ community in the usage of the new tools and applications.

During the design phase, it was decided that all functions for quality checking/quality assurance (QC/QA), mapping and versioning be provided by a common ArcObjects extension that would coexist with the different custom editing applications. This approach resulted in the Enterprise
Geodatabase Editing Environment (EGEE) which is now the common framework for GIS data editing in Miami-Dade County. The ArcObjects extension also enforced the inclusion of a manager's and editor's role for preserving the integrity of the data. Using this framework, GIS developers avoided repetition of source code and used the same standards for all editing applications. The EGEE also includes an administrative application where modifications to existing editing applications and inclusion of new ones are easily and quickly performed.

The first group of editing applications was migrated when the County ArcSDE infrastructure was under ArcSDE 9.1 release. At the time, ArcSDE architecture supported one Oracle instance and one geodatabase only, but when the second group of applications was migrated, the ArcSDE 9.2 architecture supporting multiple geodatabases in one Oracle instance was introduced. The transition to multiple geodatabases architecture was recognized as a logical decision, due to the characteristics of the County GIS editing with various datasets, editor groups and business requirements. The precedent set by previous experiences demonstrated that recovery and management tasks were considerably more difficult when a single geodatabase model was used.

The objective of this paper is to describe the architecture of the ArcSDE instances and the ArcObjects/ArcMap extension that provides a common framework to all custom editing applications and to compare the differences and similarities in the usage of single geodatabase architecture with the usage of multiple geodatabase architecture. At the end of the paper we will briefly describe the conversion of the Parcel Model to the geodatabase.

Enterprise Geodatabase Editing Environment
(ArcObjects/ArcMap editing extension and supporting Oracle tables)

As a result of the ArcGIS Geodatabase Migration project, all GIS editing applications are now integrated to Miami-Dade County’s custom ArcObjects/ArcMap extension. For the remainder of this paper, the extension and infrastructure will be referred to as the Enterprise Geodatabase Editing Environment (EGEE). The essential part of the extension was to provide a common framework that would eliminate or reduce (as much as possible) conflicts during editing, create versions and standardize maps, as well as provide tools for editors and managers to validate and quality check the data.

The EGEE is comprised of seventeen (17) interrelated Oracle tables that control the management of the supported editing applications. These tables store the information corresponding to the specific applications, the data that every application has access to (for reading or writing), the users and type of access, the encrypted user passwords, the user roles as managers or editors, as well as other management functions. These tables are stored in a separate Oracle schema aside from the geodatabase, preserving the security and independence of the GIS data.

The EGEE comprises two types of applications: the Administrative that is interrelated with the 17 supporting Oracle tables, and the Interactive Tools that are part of the ArcObjects extension. The ArcObjects extension coexists with the rest of the editing applications in a common ArcMap editing framework.
Administrative Application

Figure 1. Administrative application for the EGEE.

The Administrative Application is a supporting tool for the EGEE. It allows database administrators (GIS DBAs), and editing project managers to add new applications, disable or enable applications, add new users to the system, and assign users to a particular application. The application is very user-friendly, and provides a lot of functionality that would be difficult or time-consuming to perform otherwise. It also serves as a reporting tool for determining the amount of work completed by the different editing applications and users involved.

Interactive Tools (The .NET ArcObjects editing extension)

Figure 2. Editing extension toolbar.
Figure 3. Selection dialog box for editing users.

The interactive tools provide the mechanism for creating user versions (pieces of work), for performing the quality control checking of the data, for the creation of maps for the editors and managers, among other tools. The tools also contain reporting capabilities for the managers, easily allowing them to have a snapshot of the actual work being done at a specific time, by a particular user for a specific application. The interactive tools for managers allow the reassignment of work from one user to the other, and the release of a piece of work (version) by one manager to the other managers in order to continue the quality control checking process. The users and managers received four advanced editing training classes provided by contracted vendors and two in-house training classes for the specific use of the EGEE tools. To support this enterprise environment, a series of batch processes were written by the County GIS staff and are scheduled on a daily or weekly basis.

Every day, all changes approved by the different departmental managers are promoted to the SDE.DEFAULT or project.default in the case of multiple geodatabases. This way, other editors have access to the changes made by other editors the very next day. This task is performed by a program written in .NET Visual Basic, using ArcObjects and then compiled into a command line executable.

The migration to project geodatabases imposed a redesign in the structure of the versioning tables, the executable for the batch reconcile and post and the coding of the ArcObjects/ArcMap extension. The changes were not drastic; they just needed to conform to the new architecture. Most of the source code was reused and as of March 2009, five project geodatabases with thirteen (13) editing applications are coexisting in the same Oracle instance in ArcSDE 9.3. By mid June 2009, other seven (7) applications will be deployed for the Water/Wastewater model for the Water and Sewer department.

**Single Geodatabase Architecture: One Oracle Instance and One Geodatabase**

The first phase of the Migration Project included four (4) data models that excluded the Parcel model. The four models to be migrated were the Administrative data model, the Address model, the Street/Transportation data model, and the Land Use model. The migration of these models included the design of the database and the analysis, design and coding of the editing applications. The legacy editing applications were composed of applications for “editors” and
applications for “managers” and the same conceptual design was followed when migrating to ArcObjects and the Geodatabase.

The typical editing workflow in Miami-Dade County is as follows:

a. A user called the “editor” creates a child version of the surrogate default (see diagram below) using the in-house custom editing extension. In this case, the name of the version follows a naming convention and the version is automatically created. (Note: The extension will be described in another section).
b. Once the “editor” finishes with her/his assigned work with that version and after performing a series of quality checking tasks, the editor “submits” his/her work to the “manager” for approval.
c. The “manager” selects the version submitted for approval and performs in turn a series of quality checking tasks and approves or rejects the work done in the version.
d. If the user version is approved by the manager, during the approval process, an interactive reconcile and post is done against the surrogate default, and the user version is deleted.
e. Steps a. thru d. are repeated in an interactive way.

Every night, all the updates posted to the surrogate default version are reconciled and posted against the SDE.DEFAULT, and this is done by a command line executable. The executable performs a second level of backward reconciles traversing the versioning tree and finally compresses the geodatabase. Once the executable is finished, an Oracle analyze job (create statistics) is launched and statistics are recreated in all tables owned by SDE, and the data owner account MDC. The analyze task was a requirement in Oracle 9i, but not in Oracle10g. We will explain this in more detail later.

Versioning schema. Surrogate default. Automatic reconcile and post.

After careful study and analysis of the existing editing workflow, Miami-Dade County’s Enterprise GIS decided to use the Surrogate Default Versioning schema. This is the schema that has a “working” or “management” version that is a direct child of the SDE.DEFAULT version. Consequently, the user versions are all children of the “surrogate default”, and all the interactive reconcile and post tasks among the different user versions and the surrogate default version are performed by the “manager” application users.

Figure 4. Single geodatabase versioning schema.
This architecture can be used in any release of ArcSDE depending on your needs.

Software: Solaris 8, Oracle 9.1 (9.2.0.4) and ArcSDE 9.1 SP5.
Hardware: Sun 6800, dual processors, attached SAN disks.

**General use roles**

- SDE_EDITOR with the following system privilege: CREATE SESSION
- SDE_OWNER, with the following system privileges: CREATE SEQUENCE, CREATE SESSION, CREATE TABLE, CREATE TRIGGER, CREATE PROCEDURE
- SDERESOURCE with the RESOURCE role granted
- SDE_VIEWER with the system privilege CREATE SESSION
- SDE_CONNECT with the role CONNECT

All the editor accounts were given the following default roles and privileges:

SDE_CONNECT, SDE_RESOURCE, SDE_EDITOR.

The data owner account, MDC, is separate from the SDE account in order to avoid the SDE system tables to be in the same tablespaces and schema of the data owner account. The MDC data owner account had the following system privileges:

ADMINISTER DATABASE TRIGGER
ALTER ANY INDEX
ALTER ANY TABLE
ANALYZE ANY
CREATE ANY DIRECTORY
CREATE ANY INDEX
CREATE ANY PROCEDURE
CREATE ANY SEQUENCE
CREATE ANY TABLE
CREATE ANY TRIGGER
CREATE ANY VIEW
CREATE INDEXTYPE
CREATE LIBRARY
CREATE OPERATOR
CREATE PROCEDURE
CREATE PUBLIC SYNONYM
CREATE ROLE
CREATE SEQUENCE
CREATE SESSION
CREATE TABLE
CREATE TRIGGER
And the following database roles:

- SDE_OWNER
- SDE_CONNECT
- SDE_RESOURCE

**Custom Editor and Viewer roles**

The following custom roles were created based on the data that different groups of editors were required to edit, as well as type of access. These roles are derived from the roles recommended by ESRI for editors and viewers. It was necessary to create more roles than the two standard roles, since not all editors edit the same datasets.

- SDE_ADMIN_EDITOR
- SDE_ADMIN_VIEWER
- SDE_ALL_VIEWER
- SDE_LUC_EDITOR
- SDE_LUC_VIEWER
- SDE_PARCEL_EDITOR
- SDE_PARCEL_VIEWER
- SDE_STREET_EDITOR
- SDE_STREET_VIEWER

**Example:** All database users that are editors of the Land Use application were granted the role SDE_LUC_EDITOR. The data associated with the Land Use Model received the privileges directly on the role, not to the individual user accounts. This method simplifies the management of object privileges at the geodatabase level. For the end users that need to edit (read/write) the data, the role called SDE_LUC_EDITOR was granted with the select, insert, update and delete privileges over the Land Use datasets. All users with access to “view” the Land Use data were granted the role called SDE_LUC_VIEWER with select privileges. The GIS Data Manager controls these roles and keeps track of the users using simple SQL statements. Example:

```sql
SELECT grantee, granted_role FROM dba_role_privs WHERE granted_role LIKE 'SDE_%' ORDER BY granted_role;
```
Single geodatabase updates method to publication geodatabases

The updates method is with Oracle export and import tasks. On a weekly basis, the data is extracted and imported into the publication geodatabases, where the entire GIS community has access to the data updated the previous week. These batch processes run in two separate platforms, Solaris and Wintel, both of them controlled by a scheduler system that is hosted by the County mainframe. A group of on-call programmers rotate during the year in order to support these batch processes after regular working hours.

Multiple geodatabases’ architecture: One Oracle instance and five project geodatabases. (ArcSDE 9.2 and ArcSDE 9.3)

Hardware: 2 6900 Sun Servers with 4 Sparc IV 1350 MHz CPUs with 40GB Memory.
Software: Solaris 10 5/08, Oracle 10g (version 10.2.0.4) and ArcSDE 9.3, SP1.

The second phase main’s objective was to be able to separate the departmental geodatabases into individual project geodatabases. This would reduce the impact that some data owners could have in any other user data in terms of performance and recovery tasks. This alternative included an extended approach of the same versioning mechanism that Miami-Dade has been using with the single geodatabase, this time extrapolating the concept to project geodatabases. Most of the source code for this approach was part of the initial phase. In separating the departments into project geodatabases, the versioning tree is specific to each geodatabase, making the state lineages and As and Ds tables smaller and easier to handle. For accomplishing this objective, five project geodatabases were created:

<table>
<thead>
<tr>
<th>Geodatabase Name</th>
<th>Data Owner Account</th>
<th>Data Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAR</td>
<td>PARCEL</td>
<td>Parcel</td>
</tr>
<tr>
<td>ADM</td>
<td>ADMD</td>
<td>Administrative</td>
</tr>
<tr>
<td>LUC</td>
<td>LUCD</td>
<td>Land Use and Future Land Use</td>
</tr>
<tr>
<td>STA</td>
<td>STAD</td>
<td>Address and Street/Transportation</td>
</tr>
<tr>
<td>WAS</td>
<td>WASD</td>
<td>Water/Wastewater</td>
</tr>
</tbody>
</table>

Table 1. Project geodatabases.

Figure 5. Multiple geodatabases versioning schema and tablespaces.

With the ESRI introduction of the project geodatabases in ArcSDE 9.2, the Miami-Dade ArcSDE architecture was totally redesigned. The SDE user account continues with its role of the master geodatabase account and the project (or user) geodatabases are self-contained geodatabases in this structure. Every project geodatabase contains its own DEFAULT version, its own DBTUNE table, and its own LAYERS and VERSIONS tables among others. The names of these tables are all prefixed by the project geodatabase names. Example: LUC.DBTUNE, PAR.DBTUNE among others. These tables are created during the process of installing ArcSDE for all geodatabases with the sdesetup command. Example:

```
sdesetup --o install --d ORACLE10G --i 5181:par --u par --p password --l license.txt
```

Besides the creation of the project geodatabases and its corresponding Oracle schemas; each project geodatabase was associated with a data owner account. This approach was similar to the approach used in the single geodatabase architecture, where the SDE account is the geodatabase account, and the data owner account is a separate account in our case called MDC. The division into project geodatabases proved to be a more flexible approach, in the sense that the user geodatabase architecture is more specific to the nature of the datasets that each project geodatabase is containing. In March 2009, the ArcSDE infrastructure was upgraded to ArcSDE 9.3 and Oracle 10g. The Oracle instance and project geodatabases’ architecture remain the
same in design as in ArcSDE 9.2. The biggest change for ArcSDE 9.3 was the upgrade of the hardware platforms to Sun 6900 servers, the operating system to Solaris 10 5/08, and the Oracle release to 10g (version 10.2.0.4). Once Oracle 10g was installed, this release of Oracle has an automatic feature for the creation of statistics. During weekdays, all tables that were changed during regular working hours are analyzed between 10:00 p.m. to 6:00 a.m., creating new statistics. During the weekend, the automatic collection of statistics happens all day. This has eliminated the need to force the creation of statistics and the execution of a batch job.

The typical editing workflow with this new architecture is similar to the single geodatabase approach. It only differs now in that the editors create their new versions under their project surrogate default versions and that the “managers” perform their interactive reconcile and post against the project surrogate default version. Each night, five (5) occurrences of the initial executable are run to perform the automatic reconcile and post against the five (5) project geodatabase DEFAULT versions. These executables perform a second round of backward reconciles and a compress of the project geodatabases at the end of the execution. Every night five (5) SQL scripts are executed for all five (5) geodatabases in order to rebuild indexes. (Example with the PAR geodatabase)

```sql
declare
    CURSOR index_cur IS
        SELECT owner, index_name FROM dba_indexes
        WHERE owner in (select distinct(owner) from PAR.TABLE_REGISTRY) and INDEX_TYPE = 'NORMAL'
        ORDER BY owner, index_name;
    SQL_STMT VARCHAR2(200);
begin
    FOR IndexRec in index_cur LOOP
        SQL_STMT := 'alter index ' || IndexRec.owner || '.' || IndexRec.index_name || ' rebuild';
        EXECUTE IMMEDIATE SQL_STMT;
    END LOOP;
end;
exit;
```

Since the migration to ArcSDE 9.2 (and ArcSDE 9.3) the default storage precision was set to high and the log files were created using session logs. See parameters below:

```
DEFAULTPRECISION    HIGH
ALLOWSESSIONLOGFILE TRUE
LOGFILEPOOLSIZE     200
```

**Multiple geodatabases’ data updates method to publication geodatabases**

With the introduction of ArcSDE 9.3, Miami-Dade County changed the updates method to publication geodatabases. The publication geodatabases are using one Oracle instance and only one geodatabase. There are two Oracle production instances, one used for editing with project geodatabases and one used for batch updates with a single geodatabase architecture that refreshes the contents of the weekly editing performed in the project geodatabases along with the updates to non-versioned data. This is a sequential process that starts with the data extraction from the editing geodatabase to the second production geodatabase, then in this production single geodatabase other batch jobs are executed that refresh the non-versioned data. Finally, this Oracle instance with a single owner account called MDC is exported in Oracle and imported into four (4) more Oracle instances all of them with single geodatabase architecture. The non-
editor users have access to these instances via Citrix, ArcIMS, ArcGIS server and Web services using read-only connections.

![Diagram](image.png)

**Figure 6. Multiple geodatabases update process.**

**Geodatabase Design and Data Models - Parcel Model Migration**

Miami-Dade County’s Enterprise GIS comprises more than 200 layers grouped by more than 20 different categories: educational, health, emergency management and police, fire rescue, planning and zoning, environmental, cadastre, addressing and street maintenance, among others. Historically, Miami-Dade has maintained its street network, addresses, parcels, as well as zoning using in-house customized AMLs accessible by internal County staff. Other layers like municipalities boundaries, commission districts and police grids were maintained with ArcInfo and ArcView using out of the box (OOTB) functionality.

Miami-Dade grouped all the existing GIS layers into five (5) data models: Address, Street/Transportation, Administrative, Land Use, and Parcel. Each model was associated to a customized ArcObjects editing application using a multiuser versioned approach. With the introduction of the geodatabase in the new technology, the usage of subtypes, domains, relationships and topology rules at the database level helped achieve a higher quality of data integrity.

The Parcel Model in Miami-Dade County has the particularity that the subdivisions, lots and easements are created and maintained by the department of Public Works and the parcels are maintained by the department of Property Appraiser. This fact by itself imposed workflow integration for the case of the new plats. The Parcel Model design included the main feature dataset with the feature classes that participated in topology rules and some stand alone feature classes and supporting Oracle tables.

The Parcel model migration comprised the following tasks:

1) Preprocessing and collection of data in existing ArcInfo/Librarian formats.
2) Transfer of all preprocessed datasets from the DEC Unix server to a Windows workstation.
3) Execution of batch commands, Python scripts, SQL scripts and VB code for the loading of the appended coverages and INFO files to feature datasets, feature classes and Oracle tables into the Sun Solaris/ArcSDE development server and instance.

4) Transfer of the quality checked ArcSDE datasets and Oracle tables from the development instance to the Solaris production editing instance.

The Parcel Model for Miami-Dade consisted of the following Librarian coverages:

1) Parcels (the lines, the polygons and the annotations)
2) Lots (the lines, the polygons and the annotations)
3) Subdivisions (the lines and the polygons)
4) Easements (the lines and the annotations)

There were 917 tiles (square mile sections) in the Librarian index.

A total of twenty one (21) INFO tables were converted to Oracle tables with its new design. The preprocessing of the data in the ArcInfo librarian layers comprised a series of customized AMLs that included the appending of all tiles in double precision coverages including the arcs (lines) and polygons for polygon coverages. The polygon coverages included a special field called LINE_TYPE in the AAT (Arc Attribute table). This special field was maintained by the two departments that maintain the Parcel Model. The field identified which lines were blocks, lots, subdivisions, easements and parcels. The field was converted into a subtype field in the line feature classes and is now maintained with the usage of the new applications. The coverage fields were preprocessed having in mind the new data design for every feature class in the Parcel Model feature dataset. All the coverage related fields were dropped and new fields were added to conform to the new design. The day of the data extraction, a UNIX script launched all the collected AMLs and scripts, since this task needed to be done as fast as possible. The data extraction took nine (9) hours in the old environment and the data loading took two (2) days in the new environment and this included a combination of Python scripts, SQL scripts and executables for the loading of the preprocessed coverages and INFO tables into the ArcSDE instance. The data was loaded in a development instance and after passing quality control tests, the final datasets were loaded in the production editing instance where the only pending step was the registration as versioned and the assignment of privileges over the custom database roles explained above. A total of 554,275 parcels were loaded to the new system, with 603,415 lots and blocks, 21,986 subdivisions and 311,370 condominiums.

During the migration process, the end users were able to test the applications and quality of the data. The implementation of the model along with the editing applications and supporting batch jobs comprised two weeks and three weekends where a parallel process was in place using the old applications and the new applications. The results of the editing of the new datasets were checked with printed maps and with the updates of the data to publication geodatabases that were in test mode. All internet applications tested the newly created data and all legacy applications tested the backward converted datasets in shapefiles and coverages derived from the newly edited data in the Parcel model. The Parcel Model was initially implemented in August 2008 in ArcSDE 9.2 being the first one to use the multiple geodatabase architecture. In March 2009, it was migrated to ArcSDE 9.3 along with twelve other editing applications in five (5) project geodatabases.
Conclusion

During the different phases of this migration, the County GIS technical staff learned some lessons that helped in understanding the complexity of the new technology. First of all, the decision of having one single geodatabase or multiple geodatabases depends on the size and the type of business processes the data being edited is related to. This means, that if the GIS manager has only one type of business and a small group, using the single geodatabase approach could be the best solution. In the case of multi-departmental applications and therefore, different groups of editors and type of data, the multiple geodatabases is a better approach. The county GIS decided to create a custom versioning solution and editing framework because of the need to introduce quality checking standards enforced by the manager editors that are mainly concerned with the business rules and the quality of the data. The editing with a single geodatabase started on November 2006 and the editing with project geodatabases started on August 2008. The combination of the usage of geodatabase features like subtypes, domains, relationships, topology rules, along with a careful designed ArcSDE/RDBMS architecture and the usage of quality checking tools, security standards and an on-going training of the end-users have proved to be very effective in our case. In fact, the data creation process has been enhanced. This data is published weekly in all our web servers and provides supports to critical systems such as CAD 911, 311, tax collection, hearing notices, and telecommunication systems among many others.

References

- ArcGIS 9, ArcSDE® Configuration and Tuning Guide for Oracle®
- Geodatabase: Administration for Oracle –An Introduction to the Multiuser Geodatabase by Kasia Tuszynska and Jim Gough (ESRI Workshop 2007)
- Versioning Workflows - An ESRI ® Technical Paper • January 2004
- Versioning - An ESRI ® Technical Paper - January 2004
- Job Tracking for ArcGIS®, Workflow Management Solution, 2007

Authors

Ivelisse Torres-Alejo, Senior Operating Systems Programmer, Senior GIS Technical Project Manager, Enterprise Technology Services Department, bitorre@miamidade.gov
Julian M. Albo, Senior Operating Systems Programmer, Senior GIS Technical Project Manager, Enterprise Technology Services Department, jma@miamidade.gov
Elio Elias, Operating Systems Programmer, Applications developer, Systems Analyst, eliase@miamidade.gov