From Coverage to Geodatabase: Moving Towards an Enterprise GIS

Authors
Scott Williams – Irvine Ranch Water District
Haritha Vendra – DCSE, Inc

Abstract
The Irvine Ranch Water District's GIS began in 1994 using existing ArcInfo 7.x technology. Since then, the District's GIS has grown along with the District, both in the amount of data stored due to the rapid develop occurring within the District, and in the data's multiple uses within the District. To continue this evolution, the District recently undertook a project to migrate it's existing GIS data from the old ArcInfo coverage format (stored in ArcInfo Librarian) to the new Geodatabase format. This paper will cover the reasons for migrating, the approach taken, challenges encountered, and finally, lessons learned.

Introduction/Background

Irvine Ranch Water District (IRWD) is a multi-service agency that provides potable and nonpotable water supply and wastewater collection, treatment, and disposal services to a population of approximately 300,000 and covers an area of 179 square miles. The District was formed in 1961 as a California Water District and overlies much of the old Irvine Ranch property in south central Orange County. The District, which includes all of the City of Irvine, is bounded by the Cities of Tustin, Santa Ana, Costa Mesa, and Newport Beach to the west; the Pacific Ocean and Laguna Beach to the south; the Santa Ana Mountains to the north; and the City of Lake Forest to the east. The District extends from the Pacific Ocean to the foothills of the Santa Ana Mountains, covering elevations ranging from sea level to 1,700 feet. The District includes approximately 1450 miles of potable pipelines, 420 miles of nonpotable pipelines, and 880 miles of sewer pipelines. There is rapid commercial and residential real estate development occurring within the District, requiring constant updates to the GIS as facilities are constructed.

GIS at IRWD

The GIS at IRWD was started in 1994 and consisted of Unix ArcInfo 7.0 storing our core (pipelines/basemap) GIS data, supported with Unix ArcView 3.0 for mapping needs. District coverages were stored using the Librarian module of ArcInfo and the District was broken up into 70 Librarian tiles to facilitate handling of the large data set.

The District started by purchasing parcel data from the County of Orange for basemapping, and then having a consultant convert the pipelines and related facilities of the three hand drawn field atlases (potable, nonpotable, sewer). The District also developed several non-network supporting data layers for each pipeline system. All of this data was stored in coverage format within Unix ArcInfo Librarian. This is what we refer to as the "core" GIS data of the District and would be constantly updated as new development within the District took place.

For on-going conversion of new facilities into the GIS, the District utilized a third party CAD based conversion tool. Development data submitted to the District by consultants or developers in CAD format was massaged by the conversion tool to fit our standards/database design and then exported to dxf/txt files, which were then converted to coverages by AML routines and inserted into the existing core data using Librarian protocols. Editing was done originally by another third party software, which was eventually dropped in favor of in-house developed tools running off AML scripts.

Over the years, an extensive collection of shapefiles was generated as data was received from other agencies or developed in-house on an as-needed basis. This miscellaneous shapefile data was stored outside of Librarian, separate from the core system data, as it did not adhere to any strict database designs and was infrequently updated.

Other than several upgrades of Unix ArcInfo from ArcInfo 7.0 to ArcInfo 7.2, the handling of the core system GIS data in Unix ArcInfo Librarian remained in this same environment up until 2006. While the environment for handling our core data remained stable for 12 years, the District was slowly moving forward on the PC/Internet side. From ArcView 3.0 to 3.2, then ArcGIS 8.0 to 8.2 the District has steadily improved its PC side software. In addition the District has implemented an Intranet mapping application that uses .NET and SDE technology. To use our core GIS data stored in Unix ArcInfo Librarian in the PC/Intranet applications, we would run AML export routines to get the tiled Librarian coverage data into a single coverage, convert that to shapefiles, then FTP the shapefiles over to the PC network. For the Intranet application we would then convert these shapefiles into SDE layers. Eventually, the time came when we knew we had to migrate our coverages to geodatabase for the Enterprise GIS to continue moving forward.

IRWD's Project Goals

The goals of the Conversion/Migration project were as follows:

- The top priority was to get all of our Unix ArcInfo Librarian stored coverage data into the modern geodatabase format to facilitate an Enterprise GIS.
- Along with the data migration to geodatabase, the District also needed to develop new conversion, editing, and printing tools to work with the new geodatabase environment
- Upgrade any hardware necessary to support the migration
- Also, in the past few years IRWD had consolidated with the adjacent Los Alisos Water District (LAWD) and taken over sewer operations in an area known as the Irvine Business Complex (IBC). No GIS data was available from the previous agencies responsible for these areas and we received only paper maps. The facility information

for LAWD and IBC needed to be added to our GIS from the hard copy source documents (as-builts).

- The preparation of the Sewer Collection System Master Plan became a top priority for the District and a complete GIS data set was needed to perform the modeling. It therefore became the driving force behind completing the above projects.

IRWD's Project Challenges

A variety of challenges were encountered along the way, including, but not limited to:

Geodatabase Design:

The geodatabase design was the task that took the most time for IRWD staff. Not only were we trying to translate our existing coverage model database design into a geodatabase model design, but we were also trying to add as many features or attributes that we could think about for future uses. First, it took awhile for staff to understand the concepts of the geodatabase design structure and the pros and cons of certain design choices. We started with the ESRI water/wastewater/stormwater geodatabase template and began removing extraneous items. This in itself took quite awhile. Then we needed to add uniquely IRWD features into the geodatabase design and decide where they should go. Then we tried to add unique identifiers or features into the geodatabase that would help with future integrations with other databases. Then, while converting the new LAWD and IBC areas, we found new feature types that needed to be added as we went along. The geodatabase design turned out to be an evolving work in progress as we learned more about how it functioned, learned what would be advantageous strategies, and as we discovered new items to add throughout the project.

Data accuracy/quirks:

Nobody knows the data better than the people who created it, update it, and use it…and sometimes even they get surprised. The migration unearthed an interesting array of errors, both in geometry and attributes. We had hoped to do some data clean up prior to the migration project, but the urgency of the Sewer Collection System Master Plan and an already overcommitted staff precluded that from happening. This led to lots of questions from the consultant during the migration as lots of non-conforming data was found during preparation for migration.

Much of the attribute data entry in the past had been from keyboard input rather than established pick lists so all kinds of typos were found in the data, along with long forgotten placeholder flag values and missing values. As we formalized new valid type code lists and valid range domains, these errors quickly came to the fore. In the geometry we discovered disconnected data, missing data, flow direction errors, and even inconsistent conventions for similar features. Much staff time was spent investigating, confirming and/or correcting anomalies in the GIS data.

Technical Aspects of Project – Consultant's Experience (DCSE)
The conversion project had the specs written out to cover all aspects of the migration starting from geodatabase design to the tools that will make editing easier for the staff. The following section discusses some of the aspects.

Geodatabase Design:
We initially started with the ESRI sample design for Water and wastewater and worked to customize it per District requirements. This was a good beginning point for the design and later the design was guided by the District's schema coming from their existing coverage data. For the landbase part of the GDB design, it was made to closely resemble the City of Irvine schema since the District will be sharing data with them continually. The database was fine-tuned to include any future attributes that the District will be interested in maintaining. The following are the highlights of the design:

- External links to databases were provided wherever required instead of bringing in the data into the geodatabase. This made the design very slim. In the bigger perspective, these external databases could be part of the District's SQL Server database or independent of it.

- The design was made as close as possible to the existing coverage schema so that there is minimal learning curve for the users. This is especially true for the attributes that will be displayed on maps as annotations.

- The design accommodated applications like hydraulic modeling and valve shut down. Attributes required for such applications were added.

- The question of whether or not to use subtypes and what is the criteria for subtyping played a major role in the initial discussions on geodatabase design. Also, in some cases, it was a question of whether to create a separate feature class instead of a subtype. The ruling factors were whether the features shared common attributes and if they had different default values depending on the type.

- All domain values were gleaned from existing coverage data using frequency summaries on fields.

- Separate feature data sets were designed for network and non-network modules of each utility system. The land-base is also a separate feature dataset.

Migrating coverage data:
The coverage data consisted of one coverage per utility system with all the nodes, lines, and annotations in the same coverage. The District had a well defined schema for all the coverages that made facilitated the migration. The following are the highlights of conversion from coverages:

- The utility network coverages had all the features in a single coverage and hence shared common attributes, Most of the times, these attributes were not applicable to all nodes. For example, in domestic water system we have account number as an attribute. This is exclusive to meters only and hence for all other type of nodes this was not applicable. With the new geodatabase format, every feature class had its own specific set of attributes.

- All the lines were in same coverage for the utility systems. For example, the sewer mains and laterals were in same coverage. In order to make the sewer system data compatible for hydraulic modeling and per the GDB design, the laterals were moved to a separate layer. The coverage data had mains split wherever there was a lateral joined per coverage topology. However since the Geodatabase supports complex edges, the mains were un-split and they now run between manholes or cleanouts.

- The District's coverage data was stored in the ArcINFO Librarian format. It was separated at tile boundaries using null nodes. This resulted in undesirable breaks in pipes crossing tile boundaries. There are also chances that the attributes of pipes on both sides of the boundary are not the same or the snapping was not done properly. The reason for these discrepancies is that the pipes were digitized separately on either side of the tiles in the coverage system. The null nodes were all inspected manually and the pipes were snapped and merged wherever applicable. The District also did a lot of field research for some instances which could not be resolved using as-built drawings.

- Connectivity is very crucial in utility network data for various applications that involve tracing the network. The data was checked for connectivity using various programs like flow split, and the ArcGIS utility network analyst. Disconnects in the network were filled using either as-builts or information from District. Also, since the District has rapid on-going development there are chances that as-builts of some parts of the system may not have been received by the District yet.

- The coverage data also had Jumps / crossovers / hooks. Since these are actually a distortion of the reality, it was decided to straighten pipes with jumps and instead add a graphic symbol at these locations to simulate jumps.

- The annotations in the coverages were converted to geodatabase format. They were arranged into separate annotation feature classes depending on the text content. These were then feature-linked using automated as well as manual procedures.

- Fields with data type decimal had issues in migration while using the personal geodatabase format. This was due to the fact that personal geodatabase takes a default of 32 and 8 for precision and scale. It will be circumvented by calculating the values again using alternative methods.

LAWD and IBC Data Conversion:
The LAWD and IBC data was converted using as-builts and atlas maps depending on availability. This data was then merged with the existing IRWD data. Since the LAWD

data had some missing as-builts, field checks were required to fill in some required attributes for the Sewer system.

QA/QC:
During the migration project, DCSE has been employing various out-of-box ESRI tools like polygon topology, network validation tools and the utility network analyst. The District is testing the Geodata Sentry tool from Laurel Hill GIS for their QC requirements. This can enable the District to frequently check the data using established routines.

Tools:
The District has a unique tools requirement owing to the progress of new developments in their service area. There are both major edits like adding new developments and minor edits like adding a single hydrant that the District needs to do to maintain their GIS. The developers in the area have been providing the District with CAD drawings of their as-builts which the District used to bring over into GIS using automated as well as manual procedures. The drawings are not consistent in the layer naming convention or drawing convention. In order to leverage the CAD drawings for the Geodatabase updates, tools are under development that will convert the cad drawings to geodatabase format. This involves calibrating the tool to a drawing format from a developer, converting the drawing to a cleaned format, transferring it to geodatabase format. The minor edits will be handled using a set of tools that works exclusively in ArcGIS. This involves a tool library that inserts features at specified location.

Missing information:
The District's goal for the GIS data is to attain 100% completeness and accuracy of the data. This cannot necessarily be attained in the duration of this project. However, it was important that the District know where the gaps are in the data in order to fill them later. Hence features with missing attributes were flagged in a separate text field called "Flag" with names of attributes that were missing. The District plans on visiting these eventually and fill in correct information.

Future Goals

While the current conversion/migration project is still not finished, we anticipate that migrating to the geodatabase environment will open up many avenues to create a true Enterprise GIS at IRWD. Some of the future projects include:

Field solutions for viewing and editing GIS data:
The District has already begun developing a field solution for viewing GIS information in the field that will leverage the new geodatabase format. Once that is successful, we would move on to developing field editing capabilities.

Integrations with other District databases:

The District has several databases that would be natural partners to the GIS. We would like to integrate spatial data into our CMMS work order system, our Customer Service Database, and SCADA database.

Internet applications:
The District has already developed an intranet GIS data viewer for employee usage. The next step would be to develop internet applications to serve customer or public needs.

Author Information

Scott Williams
GIS Coordinator
Irvine Ranch Water District
15600 Sand Canyon Avenue
Irvine, California 92618
949-453-5584
Williams@irwd.com

Haritha Vendra
GIS Analyst
DCSE, Inc.
95 Argonaut
Aliso Viejo, California 92656
949-465-3417
hvendra@dcse.com