GeoBase Communications Data Conversion:

Moving legacy Arc/INFO Coverages to a Geodatabase.

By
Jack Worthy, USAF

Richard Naus

Table of Content Page

1. Abstract:

GeoBase Communications Data Conversion: Moving legacy Arc/INFO Coverages to a Geodatabase
Authors: Jack Worthy, United States Air Force and Richard Naus, Titan Corporation

The 45th Space Wing has converted Cape Canaveral Air Force Station data from legacy Geographic Information System (GIS) data stored in Arc/INFO 7.0 Coverages to a multi-user Geodatabase built on SDE and stored in Oracle 9i. This presentation shall discuss the steps used to migrate the data to the Geodatabase using Arc Catalog and SDE, and the lessons learned during the process, covering some of the highlights and the pitfalls encountered. We shall also discuss some of the benefits that the Air Force has received from the conversion of the legacy data to a multi-user Geodatabase built on SDE, such as: higher data security, multi-user editing, and better query and analysis capabilities.

2. Introduction

The United States Air Force (USAF) 45th Space Wing (SW), 45th Space Communications Squadron (SCS) undertook the task to convert the Outside Cable Plant (OCP) database from legacy Arc/INFO coverages to a multi-user Geodatabase to be stored in Oracle 9i. Titan Corporation was previously contracted to provide GPS data, input the survey's data to create the original base infrastructure database information, so the 45th SCS looked to Titan to migrate the communications (COMM) data to a new data standard developed and provided by the 38th Engineering Installation Group (EIG). The new standards have been named the Geobase Communications Model Data Standards (GCMDS) and are to be implemented USAF wide on approvals from HQ Air Force Intelligence Agency (HQAIA). The conversion at Cape Canaveral Air Force Station (CCAFS) is the first implementation of these new standards.

This paper outlines the process used to complete the conversion of the data from Coverages to Geodatabase, and the migration of the data to the GCMDS format. It also includes some details about the 45th SCS, and some history on the development of the OCP database and the Geographic Information Management System (GIMS).

2.1 45th Space Communications Squadron

The USAF 45th SCS is stationed at Patrick Air Force Base (PAFB), located on the east central coast of Florida. The 45th SCS area of responsibility also includes CCAFS. The Station is also located in the same area of Florida. The 45th SCS mission is to "Assure and protect information access through life-cycle management of information systems and services." Their vision is "World class professionals providing the Space Coast community with premier information systems and services."

The 45th SCS duties are:

- Operation and maintenance of the communications-computer systems, airfield facilities, enterprise network, and range wideband communications to ensure space launch readiness
- Management of the 45th SW communications plans, resources, systems integration and architecture
- Provision of spectrum management direction and support
- Support of Department of Defense, civilian and commercial customers
- Serve as single communications integrator for PAFB, CCAFS, and the Eastern Range.

As part of the 45th SCS's activities and to help the Squadron to complete their duties, GIMS was developed to allow COMM data to be geo-referenced to its true position on the earth. The first phase was to collect the OCP data into a database using custom software developed by Titan called GIMS.

2.2 History of GIMS

The USAF, through 45th SCS, funded the development of the GIMS program and the program commenced in 1997. The program's objective at that time was to develop and deliver a turnkey Geographic Information System (GIS) for the 45th SW under the program guidance of the 45th SCS. The critical element of this program is the OCP application for CCAFS.

The survey and audit for all CCAFS COMM manholes was completed in 1998, with all the cable and manhole data finalized and turned over for production in 1999. The collection of data into GIMS was completed in 2000 and the 45th SW GIMS is accessible for general use via the 45th SW Local Area Network (LAN) with differential Geographic Positioning System (GPS) survey accuracy. During the collection of the data for the CCAFS, more that 1100 manholes and over 1200 miles of cables were populated into the OCP database.

The GIMS applications consist of a GIMS editor and a GIMS viewer. Only a handful of users have access to the Editor with the bulk of the users having access to a read-only Viewer. As of September 2001, Titan had trained more than 100 users on use of the GIMS viewer, with additional users trained every year.

2.3 Maintaining the Outside Cable Plant

The OCP data has been maintained using the GIMS editor application. The GIMS editor is a collection of Arc/INFO 7.0 productivity tools implemented through an Arc Macro Language (AML)-based Graphical User Interface (GUI). The GIMS editor application provides a user-friendly approach to developing and maintaining a communications infrastructure, using the GIS software, Arc/INFO. The GIMS editor application is currently designed to facilitate the production of manholes, and to include manhole openings, ducts, and cables. Endpoints of cables are also used to identify equipment and structures linked by cables (e.g., terminals, splices, repeaters, etc.). The GIMS editor application provides a quality control system, a suite of data conversion tools, and a fully automated map generation package. The user can add, modify, and delete graphics and/or information pertaining to manholes, ducts, and cables.

The OCP data is updated using two methods, Manhole Reviews and New Projects. Manhole Reviews are created when a technician physically inspects and produces a drawing of the manhole with its infrastructure. Titan uses the drawing to update the data using the GIMS editor. New projects are delivered to the production staff, where they input the data using GIMS editor. The new projects are usually delivered with GPS coordinates. When they do not provide GPS coordinates, the new infrastructure is field verified and GPS coordinates are collected if need. The data collected from Manhole Review and New Projects are the only new data sources received to update the OCP database.

2.4 The Outside Cable Plant data uses

The OCP data is most readily available at both CCAFS and PAFB for the 100 plus users via the GIMS viewer. The GIMS viewer is an application developed by Titan using ArcView 3.0a and a verity of specialized Avenue scripts. The OCP data via the GIMS viewer is used for determining: the availability of cables required for Launch Missions; maintaining and updating end of life of the OCP infrastructure in support of all 45th SCS duties; and identifying communication routes when issuing dig permits for Station (Base) and Facilities upgrades.

3.0 Conversion of the data

This section of the paper covers the purpose of the OCP data conversion and the step-by-step procedures that were used to complete the tasks.

3.1 Why convert the data

The first question is "why" convert the data from the Arc/INFO coverages to a multi user Geodatabase. There are multiple reasons for moving the OCP data from Arc/INFO 7.0 coverages to a Geodatabase. One reason, which started the idea of conversion, was a need to upgrade the software used for GIMS.

During the past seven years since the beginning of the GIMS project many advancements have taken place in both GIS and database software. The idea of utilizing some of the new software was proposed. The following list highlights the most important reasons for the 45th SCS to convert the data.

- Ongoing projects at CCAFS added new OCP baseline configuration to the CCAFS (e.g. adding many new miles of cables).
- Implementation of a new data standard for COMM developed by the 38th Engineering Installation Group (EIG).
- Once the data is in the GCMDS format, the Arc/INFO application would no longer be able to access the data without major modifications.
- Using a Geodatabase adds the ability to have multiple editors at the same time
- Data can be stored in a Relational Data Base Management System (RDBMS) (i.e. Oracle 9i)
- The support for Arc/INFO 7.0 has been discontinued

All these factors helped convince the 45th SCS it was in their best interest to move the data from legacy Arc/INFO 7.0 coverages to a Geodatabase.

3.2 Step by Step

The next question is "how" Titan was to perform the conversion. The following ten steps provide an easy to follow method of converting Arc/INFO coverages into a multi-user Geodatabase that uses Oracle 9i for storage. The methods Titan used to complete this conversion may not have been the most direct path to our goal, however they were the most convenient. After many different methods of performing the conversion were tested, the following a ten-step protocol was deemed most effective for our circumstance.

(**Note:** The following items were already in place and utilized from the start of the conversion project: Arc/INFO 7.0 coverages in North America Datum 1983 (NAD83), ArcGIS 8.3, ArcSDE 8.3, and Oracle 9i, a standard ArcSDE setup connecting to an Oracle 9i database.)

- The first step Titan performed was to match the existing data to the new data schema of the GCMDS. This was a significant undertaking. Titan employees had to list out each category for the old and new tables and relate each category from the old to the new, then verify that the new domain tables contained the same categories that the old data contained. In some cases Titan added more categories to the new domain tables and in others the existing data was changed to adhere to the categories of the new domain tables.

- The second step was to create two Geodatabases, one in NAD83 and one in World Geodetic Survey 1984 (WGS84). Using the two Geodatabases was not necessary, however the use of the two Geodatabase did make it easier to convert the data to the new format. (See Lessons Learned, part 1) The old data was in NAD83 and the new standard is in WGS84.

  (For more information on how to create the Geodatabases please refer to ESRI publication "Creating and Managing Geodatabases".)

- The third step moved the data from the coverages to NAD83 Geodatabase using ArcCatalog's "Export Coverage to Geodatabase" function. Once completed for the first coverage, the process was repeated on the other coverages. Titan performed this function on each of the original eight coverages.

  Once all the coverages were exported Titan had created a fully populated NAD83 Geodatabase with the original data in its original format. The data then needed to be converted to the WGS84 coordinate system to comply with the GCMDS format.

- The forth step, converted the data from NAD83 to WGS84. Titan used a process called "Project Wizard (shapefiles, geodatabase)" available in Arc ToolBox to move from one coordinate system to the other. This is a nice little tool ESRI provides to make coordinate system changes on the fly.

- The fifth step created the final Geodatabase, which was be used in production. The final Geodatabase needed to have all the proper table names, column heading, and data types when it was created. The different domain tables also needed to be created and populated. These things **must** be completed before any data is loaded into the final Geodatabase (See Lessons Learned, part 2).

The domain tables were created in ArcCatalog by right clicking on the geodatabase and go to properties. Under the domain tab, all the relevant data must be added (e.g. data type, maximum length).

- The sixth step was use of ArcCatalog to load the data from the WGS84 Geodatabase into the final Geodatabase. During this loading process the domain tables were utilized to place data. If the data that is being loaded does not match up with the entries in the domain table, the load will not populate any data into that column of the Geodatabase.

  Data from the old format was moved to the tables and column that the data will reside in the new format and data type (e.g. such as Long Integer to Double).

  Some of the loading was straight forward, table to table with no changes. In other cases we used PLSQL script to standardize and clean the data that was loaded. For example, Titan used the scripts to remove extra spaces on the end of text and standardize the capitalization in a text string (See Lessons Learned, part 3).

- The seventh step was to export feature classes to Oracle using ArcCatalog. For new feature classes that do not have corresponding data that is being moved from old to the new, some dummy data **must** be moved into the new column (See Lessons Learned, part 4).

- The eighth step was removal of the dummy data from Oracle tables. Titan used a SQL script to remove all the dummy data.

- The ninth step was to create the table relationships. The relationships were created by using a set of processes or "wizards" in ArcCatalog. One of the more difficult tasks of setting up the relationships was working out the Foreign and Primary keys. The 38th EIG provided the Foreign and Primary keys in the GCMDS. (See Lessons Learned, part 5)

  (For more information on how to create the relationships please refer to ESRI publication "Creating and Managing Geodatabases".)

- The tenth and final step was performance of a sanity check and quality control on the data in the new Geodatabase. For the sanity check, Titan used the matching table from the first step to review the data and make sure the data ended up where it was expected to end up. As part of the QC we used statistical comparisons between the old and new data (e.g. number of manholes, cable, and so on).

4. 0 Lessons Learned

Titan learned many lessons while testing the processes that were used in the conversion of the Coverages to a Geodatabase and migration of the data to the new GCMDS format. The following five items were the most significant. The project would have been easier to complete and shorter in duration, if Titan had been aware of these things from the start of the project.

1) Titan discovered during testing that if all the coverages are put into a Geodatabase and the moved to another Geodatabase with the new coordinate system, it made the process run much smoother. This process presented less chance for corruption of the data, as opposed to converting each coverages coordinate system and then individually loading each of the coverages into a Geodatabase with the new coordinate system.

2) The final Geodatabase needs to have proper table names, column headings, and data types in use when it is created. The different domain tables also need to be created and populated with correct information.

3) Titan discovered during the test conversion process that the data being converted was not standardized very well. Over the years of collection one users method varied from another's (e.g. capitalization and spacing were not consistent). This played havoc with inputting data into tables controlled by domains.

   Titan created scripts to standardize and clean the data to allow for implementation of the domain tables. Using the scripts provides a higher level of consistency and a lower error level than if this operation were performed manually.

4) Move dummy data over into Oracle if there is no data in that table or column. This is done because the table or column will not be created in Oracle if there is no data in them. This is cause by a limitation in ArcCatalog "export" command, not the database.

5) The relationships between tables must be established using ArcCatalog instead of Oracle itself. When the relationships were established using Oracle, ArcSDE could not access the data properly. The table relationships were not maintained. However, if the relationships were established using ArcCatalog, ArcSDE could access the data properly and the relationships were maintained. The relationships were managed in the Geodatabase.

5.0 Benefits

Many benefits have been realized as a result of the OCP data being housed in a multi-user Geodatabase stored in Oracle 9i. One of the biggest benefits is that many of the inherent features of a multi-user Geodatabase can be utilized. For example the name "multi-user Geodatabase" describes one of the advantages. Multiple editors may be editing the data at the same time, while with the old method; using Arc/INFO 7.0 only one editor could work at a time.

Another added benefit of storing the data in a Geodatabase was access to the data was changed to use COTS software. A new GIMS editor was developed using ArcGIS 8.3. This software comes with many of the applications that in the old editor were developed using AML. The move to ArcGIS 8.3 provides supported software and limits the risk associated with using discontinued support for software such as Arc/INFO 7.0.

Using ArcGIS 8.3 as the editor, the geodatabase is accessed through ArcSDE 8.3 and the data is stored in Oracle. Oracle itself provides many inherent advantages. A few of these advantages are higher levels of security, greater access to the data, and accessibility by many users at one time. This functionality was not available in the INFO Coverage format.

Titan has also developed a new GIMS viewer application utilizing the data stored in Oracle, and accessed via ESRI's ArcIMS 4.01. This new application replaces the old desktop ArcView 3.0a driven viewer with an Intranet viewer. ArcView 3.0a is also no longer supported by ESRI. The new viewer limits the risk associated with using discontinued support for software. This "thin client" also provides a great cost savings benefit. It removes the cost of new installations and upgrades deployment to an individuals PC (in the 45th SW case this is more than 100 users) and places the application in a central server with access available via the base's Intranet.

6.0 Future

The completion of the data conversion from the legacy data to the Geodatabase provides the 45th Space Wing and the 45th Space Communications Squadron with the ability to adapt quickly to the changing needs of the 45th Space Wing and their clients. The new format also provides the 45th Space Wing with the ability to server out data to not only to GeoBase requirements, but also to whenever future communication paths will be used in the Common Installation Picture (CIP).

7.0 Acronyms

AML	Arc Macro Language
CCAFS	Cape Canaveral Air Force Station
CIP	Common Installation Picture
COMM	Communications
EIG	Engineering Installation Group
GCMDS	Geobase Communications Model Data Standards
GIMS	Geographic Information Management System
GIS	Geographic Information System
GPS	Geographic Positioning System
GUI	Graphical User Interface
HQAIA	HQ Air Force Intelligence Agency
LAN	Local Area Network
NAD83	North America Datum 1983
OCP	Outside Cable Plant
PAFB	Patrick Air Force Base
RDBMS	Relational Data Base Management System
SCS	Space Communications Squadron
SW	Space Wing
USAF	United States Air Force
WGS84	World Geodetic Survey 1984

8.0 References

GIS Educational Solutions from ESRI "Creating and Managing Geodatabase (for ArcEditor 8 and ArcInfo 8)" 2000, Course version 1.1, Revised June 2001