



Leveraging ArcSDE Functionality to Manage Terabytes of Raster Data

Jarrold Skulavik

David Bautts

*Intermap Technologies Inc.
400 Inverness Parkway, Suite 330
Englewood, CO 80112
303-708-0955*

www.intermaptechnologies.com

Paper Abstract

Intermap Technologies is faced with the challenge of storing and retrieving terabytes of data to meet its daily business needs. The majority of this data comprises raster information that stores Digital Elevation Models (DEMs) and radar images for various locations around the world. The sheer volume of data that must be managed provides a unique challenge not commonly addressed in the GIS community. In order to more efficiently manage this invaluable company asset, the GIS management team at Intermap has built a series of tools and workflow procedures based on ArcObjects, ArcGIS, and ArcSDE.

This essay will examine the various tools and procedures created within the ESRI framework and offer insight into customizing the raster data storage and retrieval models provided within that framework.

Introduction

Intermap Technologies' primary business is the creation and maintenance of highly accurate yet affordable Digital Elevation Models (DEMs) and Orthorectified Radar Images (ORIs). Inherently, these products are realized as raster images in several industry-common formats. Intermap's product database spans various locations around the world and, most notably, the entire country of the United Kingdom. Such large data holdings equate to large storage requirements and unique challenges with regard to managing and provisioning such extensive raster data stores. This essay addresses the experience, process, and techniques involved with managing several terabytes of raster data using the ESRI framework of GIS tools, most notably ArcSDE.

Data Descriptions and Specifications

In order to support a broad client base, Intermap generates its DEM products in generic format, namely, Band Interleaved by Line (BIL). This generic format allows for the core data products to be easily ingested into ESRI GIS and most image processing packages such as ERDAS Imagine. For the other core product, the ORI, the standard Intermap format is GeoTIFF. GeoTIFF is a widely accepted format within the industry and allows for rapid use and analysis.

On a more specific note, the data specifications are as follows:

- The DEM product is a 32-bit IEEE floating-point data file. The grid is a 5 meter ground sample distance with meter to sub-meter vertical accuracy and several meter horizontal accuracy
- The ORI product is an 8 bit unsigned file. The grid is a 1.25 meter ground sample distance with sub-meter horizontal accuracy.

The average file size for the DEM data is 23 megabytes, and the average size for the ORI is 90 megabytes. With thousands of individual grids comprising one project, one can quickly grasp the demand for disk space with accompanying data storage and retrieval systems. At the time this essay was written, there were approximately 1.4 terabytes of data residing within Intermap's data store.

Tiled Data Storage

Although raster data models often aim to represent a continuous surface, they are typically subdivided into smaller areas in order to decrease file size and storage requirements. Intermap follows such logic and stores its raster products as adjacent tiles corresponding to a predetermined grid index such

as the USGS topographic quadrangle index. Such an organizational approach generates collections of a tiles that frequently must be combined as a mosaic and then clipped to a given area of interest. In recent years the evolution of the GeoDatabase and the ability to store GIS data in a commercial relational database management system have provided new opportunities for data storage and retrieval. Rather than simply storing files on a file server and accessing them directly via a client GIS software, data can now be consolidated into one common location and rapidly distributed throughout an organization. In addition, GIS data consumers can now leverage the powerful data management capabilities of a large-scale database management system in the same fashion that non-GIS users have for years. It is such benefit types that Intermap hoped to obtain in pursuing the data management and retrieval framework provided by ESRI GIS.

Migrating to ESRI formats

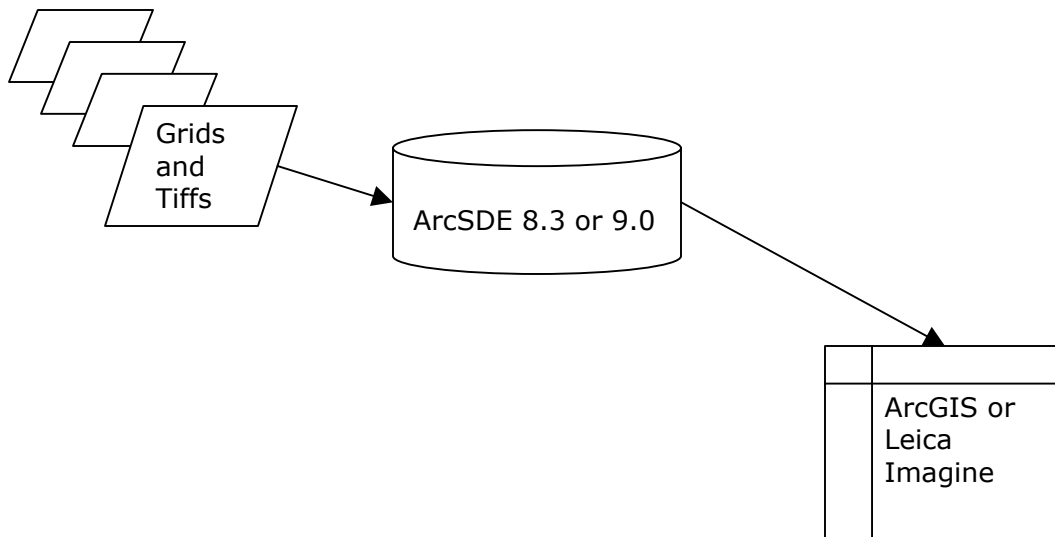
The fact that Intermap stores and distributes its elevation data product in BIL format allows for the data to be imported into most major raster software packages. However, in order to manage and store the data using the ESRI framework, an ESRI-supported format must be adopted and used. BIL files are supported in ESRI GIS, but only with headers that describe the BIL itself. In order to use the data in ArcSDE, or ArcGIS, the file must be converted. For the Intermap data store, all data is stored in ArcGRID and GeoTIFF format. These formats are easily used in all levels of the ESRI GIS framework.

Converting all of the Intermap elevation data to ArcGRID posed a challenge in the beginning stages of database development. The sheer volume of the data made it difficult to efficiently convert the thousands of BIL files residing in the database. With the proper header files built and the location of the floating point grid known on a network, the processing of importing the BIL files is rather easy. By using the *floatgrid* command within ARC/INFO workstation Arc Macro Language (AML), the process of converting BILs to ArcGRIDs was streamlined. However, in order to use the modern framework of ArcGIS and leverage the Windows VBA/VB object model(s), ArcObjects tools were created to customize ArcCatalog. The tools created served two main purposes. First, the Arc header had to be created from metafile information using the Windows framework. Second, ArcObjects, specifically Raster Data Objects (RDO), was deployed to perform the actual file conversion from BIL to ArcGRID. After its completion, it was found that the ArcObjects raster conversion worked far more efficiently than the construction of AML in ARC/INFO Workstation.

SDE Reference Catalogs and true ArcSDE Rasters

The ESRI framework is constantly growing and improving. At the beginning of this data storage project, the latest version of ArcGIS and ArcSDE was 8.3. Currently, the latest version is 9.0. The process of upgrading from version 8.3 to 9.0 created numerous challenges and issues for the Intermap data services.

Within ArcSDE version 8.3, there are three options for storing raster data. Data managers can create embedded raster catalogs, reference raster catalogs, and of course, mosaics. The Intermap data management team examined each of these alternatives and had initially hoped to store all of the data in large ArcSDE binary mosaics. However, the team quickly realized that, in order to store the tremendous amounts of data within the database, far more spinning disk space would be required than the technology budget at Intermap could allow. Therefore, in order to avoid significant hardware expenditure, the Intermap data manager determined the only viable option was to utilize the reference raster catalog. This decision brought with it major implications. By constructing the ArcSDE rasters as catalogs, the data would have to be "fetched" each time it was used. That is, in order to use the data in ArcGIS, Leica ERDAS Imagine, or some other GIS or imaging software, the end user would have to obtain the actual data from the reference (path) in the catalog and not directly from the database server itself. The ArcSDE client (ArcGIS in this case) inherently performed this task. When the reference raster catalog was loaded in ArcGIS, the actual data were shown in the display and viewed by the user. However, the data could not be used until they were moved directly into ArcSDE as binary data or directly accessed using the pointer (path) outside of the reference catalog itself. To accomplish this task, the raster data objects (RDO) of ArcObjects were once again deployed. Within the RDO libraries, there are classes and interfaces that can be used to work with reference raster catalogs. Using these classes and interfaces in combination with other core ArcObjects, the Intermap data manager was able to create an ArcMap application that allowed the user to draw an area of interest where he or she wanted to "extract" data from the reference raster catalogs. The diagram below displays the general data flow process as it was initially constructed within the Intermap data store.



Though simple in concept, the operation of the reference raster is somewhat limiting in its generic form. The data that are being referenced must actually be referenced based on the client's location and the file server's location to it. According to ESRI technical support, the reference can be made from the server. However, the Intermap data manager found that this was not the case. When using the reference raster catalogs that stored the paths relative from the database server to the file server, only the footprint was returned. Whereas, when the reference path was relative from the client (ArcGIS) to the file server then data were in fact returned. It should be noted here that when qualifying the data as returned, it is meant that the end user can literally see the rasters. For Intermap's end users, there exists a need to actually query and analyze the data, as much of the company's data holdings are elevation data and not image data. In order to do accomplish the task of allowing the end user to fully obtain the data from the catalog, the tool that was written took the reference and dynamically loaded the raster into ArcSDE as a mosaic. Therefore, the end user could reference all of the rasters in the database using the reference raster catalog, select the tiles of interest, then deploy the Raster Data Objects (via ArcMap tools) to mosaic the selected tiles together as a true ArcSDE raster dataset.

One of the more challenging aspects of the reference raster catalog is that it stores the corner coordinates of each raster in the database using the coordinate reference system assigned to the tile to which it refers. For the UTM coordinate reference system, this poses as a challenge since each zone has the same metric range of coordinates. To address this issue, a separate reference catalog had to be created for each UTM zone within the Intermap database.

Once all of the reference raster catalogs had been built, and all of the issues had been addressed, the system operated quite well.

In May of 2004, ArcGIS and ArcSDE 9.0 were installed and migrated to within Intermap's GIS department. With ArcGIS 9.0 came a new suite of tools and advanced geo-processing tools. Of more relevance to Intermap, improved raster data management was also included with the release of ArcGIS 9.0. Immediately following the installation of the new ArcGIS and ArcSDE platform, it was discovered that the reference raster catalog was no longer supported. Though the reference catalog (referred to as table-based raster catalog in ArcGIS 9.0) is documented and does in fact exist and as object class within ArcGIS 9.0, it failed to operate after extensive troubleshooting and technical assistance from ESRI. The Intermap data manager was forced to quickly find a new solution to this new problem.

After extensively searching the ESRI user forums, developer help, and ArcScripts, the "RasterIndexPolys" tool written by ESRI was found. Ironically, the product of this tool is an output table very similar to the reference raster catalog. The tool searches all of the rasters in a directory, loads their spatial references into memory, then generates a shapefile containing the raster's path, name, and corner coordinates. This tool was found to be very useful and has become the newly preferred means for indexing large numbers of rasters. The tool does have one shortcoming though: The topology of the resulting polygons in the shapefile is always incomplete. That is, the resulting polygons are not closed. In fact, they are missing their final side altogether. To address this, a second tool from ESRI is deployed, namely, the "CleanShapfiles" tool. This tool corrects the topology for each polygon in a shapefile. In the case for the polygons created by the Intermap data team, the tool closes the fourth and final edge of a square polygon building corrected topology in the layer and producing a useful shapefile.

At the time this paper is being written, the approach to retrieving the tiled raster datasets via the reference raster catalog had to be abandoned and replaced due to the lack of support for the reference catalogs in ArcGIS 9.0. To replace said approach, the raster reference catalogs were deleted from ArcSDE and replaced with the feature classes produced by the RasterIndexPolys tool. Fortunately, the overall approach described earlier can still be used—it simply must be modified to reference the "pointer" in a ArcSDE feature class rather than the "pointer" in a ArcSDE raster reference catalog.

As previously stated, the raster data management and tools were improved in the release of ArcGIS 9.0. As it currently exists, the GeoDatabase supports raster catalogs as true ArcGIS data types. The rasters can be "unmanaged" or "managed." What is meant here is that the unmanaged rasters are referenced within the GeoDatabase and accessed with RDO,

though the data exist elsewhere on a network. A "managed" raster catalog creates ERDAS .img files and stores those locally on the client machine. Oddly enough, the "unmanaged" and "managed" raster catalogs were designed to be used within a personal GeoDatabase and are not fully implemented in ArcSDE 9.0. Both types can be created and used in ArcSDE 9.0, but the data are always resident within the database and are thus no different from the previous ArcSDE 8.3 raster catalog, with the exception of several new tools imbedded within ArcGIS 9.0. It is highly anticipated that a true "unmanaged" raster catalog will be available in ArcSDE 9.x in the near future. The Intermap data team feels this will dramatically streamline the management of raster datasets, both within the GIS and remote sensing industries.

Author Information

Jarrold Skulavik

GIS Data Manager
Intermap Technologies, Inc.
400 Inverness Parkway, Suite 330
Englewood, CO 80112
Phone: 303-708-0955
Fax: 303-708-0952
email: jskulavik@intermaptechnologies.com

David Bautts

GIS Data Manager
Intermap Technologies, Inc.
400 Inverness Parkway, Suite 330
Englewood, CO 80112
Phone: 303-708-0955
Fax: 303-708-0952
email: dbautts@intermaptechnologies.com