THE ORNL BASEMAPPING AND IMAGERY PROJECT: DATA COLLECTION, PROCESSING AND DISSEMINATION

Peter Pace
Mark Tuttle
Richard Durfee
Computational Physics and Engineering Division
Oak Ridge National Laboratory*
P.O. Box 2008
Oak Ridge, Tennessee 37831-6274


*Managed by Lockheed Martin Energy Research Corp. for the U.S. Department of Energy under contract DE-AC05-96OR22464.
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
The ORNL Basemapping and Imagery Project: Data Collection, Processing and Dissemination

Peter J. Pace
Mark A. Tuttle
Richard C. Durfee
GIS and Computer Modeling Group
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6274

ABSTRACT

Over the past three years, the GIS and Computer Modeling (GCM) Group at Oak Ridge National Laboratory (ORNL), has been engaged in creating a very comprehensive geospatial data base for Department of Energy (DOE) installations managed by the DOE Oak Ridge Operations Office (DOE-ORO). This effort encompasses topographic, planimetric, land use/land cover, flood plain, digital elevation, and digital imagery data for the Oak Ridge Reservation (ORR) and surrounding areas. The ORR covers approximately 34,800 acres and includes ORNL, the K-25 Site and the Y-12 Plant. The geographic extent of the Base Mapping and Imagery Project covers the ORR and surrounding area and two other DOE plants (Portsmouth, Ohio and Paducah, Kentucky) for a total of 166,000 acres. The resulting data represent a major improvement in the spatial accuracy and currency of data which are used as a foundation for environmental restoration, facility studies, and other GIS data applications. A GIS data server was also created in order to store and disseminate the new basemapping data.

INTRODUCTION

The DOE-ORO office has responsibility for the ORR, located in Oak Ridge, Tennessee, which covers approximately 34,800 acres. The ORR contains three main plants: ORNL, the K-25 Site, and the Y-12 Plant, with approximately 15,000 employees. DOE ORO also oversees plants in Paducah, Kentucky, and Portsmouth, Ohio. The ORR facilities were created during the World War II and were responsible for research and development in support of the U.S. nuclear weapons program. Today ORNL's mission has expanded into many multidisciplinary programs that address hundreds of complex problems involving energy and environmental issues, national security, and basic scientific research.

One of the major DOE activities in Oak Ridge is the Environmental Restoration (ER) program which is focused on cleaning up the hazardous waste problems resulting from fifty years of nuclear and chemical activities on the ORR. Geospatial data are an important resource used in all aspects of these cleanup efforts. The Basemapping and Imagery Project was established to aid ER and other environmental management projects by supplying high-resolution geospatial information in the form of orthoimages and GIS vector files.
The Basemapping and Imagery Project replaces the Oak Ridge S-16A data base. The S-16A data base is a series of vector coverages which were extracted from a paper map whose title was "S-16A." This map covers the ORR and was compiled from aerial photographs which are now over ten years old. The S-16A map was printed at a 1:24,000 scale resolution. The process used to create the vector coverages and the scale of the original map resulted in data which are unable to support today's more spatially accurate data collection and analyses requirements. In addition, there have been many changes to the ORR since the S-16A map was compiled. The project is a major step towards creation of an ORR-wide map data base that is ten times more accurate than the S-16A (i.e., 1:2,400 scale data) and is more current and consistent.

Planning for the larger Oak Ridge Reservation initiative of creating a new map data base from aerial photography and ground information has brought many people and organizations together to undertake this large initiative. The following is a list of organizations involved in the creation of the project:

1) DOE - Oak Ridge Field Office - responsible for funding this project in support of ORR operations.
2) Lockheed Martin Energy Systems, Inc. (of which GCM is a part) - responsible for specifications and QA/QC of the data, processing the results, managing the project, mapping the data, providing data access, applying the data to ER problems, etc.
3) Tennessee Valley Authority - responsible for creating the data to be delivered to the GCM group. TVA hired PhotoScience, Inc. and Piedmont Aerial Surveys as subcontractors.

To aid in environmental cleanup by ER programs, facility and environmental databases have been developed to improve understanding of relationships among contaminant sources, surface and subsurface migration pathways, and receptors of environmental contaminants. Three-dimensional modeling, contaminant analysis, and data management have been enhanced through integration of computer tools and geo-spatial data. All these resources are becoming an integral part of the remediation planning and cleanup process supported through communication networks linking scientists, engineers, and decision makers with analytical software and databases. It is expected that the implementation of this project will save time and money for all concerned, avoid costly mistakes in using insufficient and inaccurate data, avoid duplication of effort, improve compliance with regulatory requirements, avoid delays, significantly improve QA involving spatial data, and provide a data resource needed by all activities on the ORR.

DATA PRODUCTS

On April 23 and 24, 1993 aerial photography for the ORR was collected. Photography was collected at an altitude of 7200 feet Above Ground Level (AGL) resulting in a nominal scale of 1:14,400. This allowed for digital mapping that would meet National Map Accuracy Standards (NMAS) for 1:2,400 scale mapping. This aerial photography collection was unique in that photography was captured using on-board Global Positioning Systems (GPS) to control both exposure and flight path. This technology resulted in source photography that far exceeded NMAS standards for this scale of mapping. These photographs were then used in a stereoplotter from which digital information was extracted. The resulting Digital Elevation Model (DEM) was combined with the digital photographs to produce ortho images. These 24-bit raster images were delivered to the GCM in BIP format. A customized color balance algorithm was created and used on the supercomputer facilities at ORNL to minimize the color variations between images. The variations resulted from the unique exposure parameters of each original photograph. The ortho images as well as all the other data sets conformed to the tile structure for the respective study area. The digital imagery for the ORR have 0.5 meter square...
pixels and are 1750 meters on a side.

In addition to the ortho images, planimetric and topographic data were created for each tile in the basemapping and imagery project. Project specifications required that an extensive list of surface features be collected. These features were then sorted into seventeen thematic groups for each tile: building outlines, 5' contour lines, control points, hydrographic lines, hydrographic polygons, hydrographic structures, marginalia, man-made outlines, man-made structures, natural outlines, roads, railroads, street centerlines, spot heights, transportation structures, utilities, and unidentified objects. The various types of data within each coverage were distinguished by a feature code stored in the vector file attribute table. Cartographic data, such as street names, were stored in an annotation layer.

The aircraft used was equipped with dual camera ports from which both natural color photographs and color infrared photographs were obtained. Other data types were created in the project from data other than the natural color aerial photographs. The color infrared photographs were used in combination with the planimetric data to produce land use/land cover vector files. The planimetric and topographic vector data were used with floodplain documentation to create floodplain delineations. In 1994 a low-altitude overflight was conducted for all three facilities in order to acquire more accurate elevation data. This photography supported the creation of 2' contours for flood analyses. Supporting documentation including a data dictionary, progress status maps and metadata also were created.

PROJECT AND PRODUCTION MANAGEMENT CONSIDERATIONS

Many factors influence the success or failure of data conversion projects such as the Basemaps and Imagery Project. Primary among them is the definition of roles and responsibilities of all parties involved in the project. The Tennessee Valley Authority (TVA) was responsible for the production of all products submitted to the GCM. TVA divided their duties between five administratively independent groups within TVA and two subcontractors. The coordination of work flow and duties between these groups presented several potential problems. So that each party was able to work effectively, specific responsibilities for each group and subcontractor had to be defined and monitored. Once the team was assembled, several meetings were required to decide upon a division of duties and a management structure to oversee the project.

The two subcontractors that TVA selected for the project. Photo Science, Inc. (PSI) of Gaithersburg, Maryland and Piedmont Aerial Surveys (PAS) of Greensboro, North Carolina, were responsible for the acquisition of aerial photography and initial data compilation that represents the foundation upon which most of the remaining products were built. Photo Science’s primary responsibility was the conversion of original aerial photographic products into digital ortho images and associated processing. Piedmont Aerial Surveys was responsible for the stereo compilation of both planimetric data and topographic data. Included in the topographic data created by Piedmont was the Digital Elevation Model (DEM) used by Photo Science in the rectification process for the creation of the digital ortho image.

As a result of several factors, GCM’s role in the project went well beyond being the recipient of data. GCM played a primary role in developing the majority of QA/QC procedures as well as the work flow processes that were ultimately implemented by TVA for the project. In addition, GCM also had the more traditional responsibility of QA/QC of all incoming data delivered as well as data tracking and organization upon receipt of deliveries. The definition and assignment of these duties took place over a six month period that included the pilot phase of the project. Typically the pilot phase is intended to implement a proposed production process and work flow. For the basemaps project, not only did a proposed work flow and process need to be tested, but exact data product specifications and media needed to be determined. The pilot project produced a data flow and responsibility matrix that was
modifiable and flexible. As a result of the complexity of the project as well as the diverse nature of the project, the production flow did in fact alter during the course of the project. The final version of the work flow and production process is shown in Figure 1.

WORK FLOW

Data collection for each tile began with the stereo compilation of planimetric and topographic data by PAS. A Digital Terrain Model (DTM) consisting of break lines and mass point data defining the surface were compiled first. These topographic data were converted by PAS into a DEM which was delivered to PSI for use in producing the digital ortho image. Planimetric features were then extracted using the stereo plotter. The digital ortho products, ortho photographic prints and the planimetric/topographic data files were then submitted to the TVA. The digital and print ortho products were forwarded to the GCM. Concurrent with submission of the planimetric and topographic files to the GCM, the topographic files were provided to TVA’s Hydrologic Modeling Group. The topographic data were then used as an integral component in the delineation of the 100- and 500-year flood plain.

The planimetric and topographic files were then converted by TVA from their native format to the appropriate GIS format. Included in this phase is extraction and creation of each of the seventeen layers identified in the GIS data dictionary. Immediately upon completion of the conversion of the initial planimetric/topographic files, the GIS data where then enhanced with a variety of attribute and cartographic data. Where identifiable, TVA assigned proper names to all transportation features (roads and railroads), water features (streams and lakes), and building names. Each name was added as annotation only to facilitate production of cartographic products. Using data provided by the GCM, TVA also assigned a building ID to each building polygon within the reservation that would provide a link to existing tabular database(s) that used the reservation building identifier as a primary key.

Upon receipt of a delivery from TVA, the GCM first logs the submission and then loads the data into the system. Each digital ortho image delivered by TVA first must pass QA/QC tests and then is color balanced before being loaded. All other products are loaded directly to their appropriate system and each product is subjected to comprehensive QA/QC review by GCM staff.

QUALITY CONTROL/QUALITY ASSURANCE

The greatest challenge in implementing a QA/QC plan for a project of this scope is devising a scheme that accounts not only for the diverse products being provided, but also avoids or limits duplication of effort as much as possible. The QA/QC activities for a large scale data conversion effort such as this must be an integral part of the whole production process. The QA/QC plan should not only ensure that the highest possible quality product is produced, delivered, and received, but it must not be an impediment to the overall process as well. A pilot phase was created as a first draft of the original QA/QC scheme. During the course of the pilot phase, it was determined that there were two critical points throughout the production process that required a QA/QC review: 1) each time a data file changed hands, and 2) each time a data file was translated from one tile format to another. These checks were in addition to routine QA/QC reviews performed within each group as well as any quality reviews that each functional group performed as part of their day-to-day activities.

The first major QA/QC review occurred when the topographic and planimetric data files were submitted from PAS to the TVA. Not only was adherence to the technical specifications reviewed, but the complete data capture process also was examined. This single review is perhaps the most important during the entire production process. The output that is generated as a result of stereo compilation is literally the source document for all other products. The topographic data generated is concurrently provided to PSI for use in the ortho rectification process and delineation of flood plain boundaries. In a more direct fashion, the data from stereo compilation is translated to the appropriate GIS formats. The
GIS data is then used to create a source file for the land use/land cover photo interpretation work.

Upon delivery of the ortho image products to the TVA, the paper print is reviewed. Specific items appearing on each ortho are reviewed, such as correct grid notation, tile identification, and proper scale. The digital ortho image is reviewed both radiometrically and spatially. A set of control points is established from the planimetric data and these points are then verified from the image. A series of statistics are generated which identify errors of omission and commission.

GCM staff begin QC activities on a tile when the original topographic and planimetric file is received. A rigorous review of technical specifications is performed and the data are either returned to TVA for corrections or accepted and loaded onto the system. The DTM submitted by TVA is reviewed against the topographic data submitted as well as against existing vertical elevation data. An extensive review of these files is performed when TVA completes the enhancement process on the GIS data and submits the GIS data files to GCM.

The review of GIS data continues with a check for correspondence between the data and the original planimetric and topographic files. Primarily, an attempt is made to determine if all features contained in the original files have been translated to their correct layers. Attribute information that can be extracted from the original files is reviewed as well. The next step is a review of the topological organization of each layer as well as a review of topological relationships between layers. Adherence to the technical specifications also are reviewed at this time using customized Arc/Info Arc Macro Language (AML) programs. Finally, a review of street names, water body names, and building identifiers is performed.

DATA TRACKING

All basemapping data is entered into a tracking system as it is delivered to the GCM. There are four categories of basemap data which are tracked with this system: documentation, paper maps, photos, and digital data. The tracking system is a FoxPro-based database running on a PC using a customized user interface which uses Windows-like pull down menus. These menus permit the user to add data about the basemapping project, edit data or create a report. The database structure is designed to support the "shipment" nature of basemap data deliveries from the contractor. Each shipment is assigned a unique number. All documents, maps, photos and digital media in the shipment are assigned the same number which becomes the primary key throughout the system.

From the initial FoxPro page which assigns the shipment number, the user inputs whether the shipment includes documents, maps, photos and/or digital data. For each of these categories which have data, the system prompts the user for appropriate information. Stored information regarding documentation include type, source, title, and content. Map information includes details regarding the map, scale, date, context, geographic extents and lineage. Aerial photos provide a unique challenge because some photos are bound in books according to flightpaths and others are stored as single prints. In order to avoid entering a separate record for each bound photo, information representative of all the bound photographs are stored for the volume. Database information for aerial photographs include scale, altitude, date, film type, flight line and print number.

Digital data is the most complex type of data to track. For each shipment there is a one-to-many relationship between the shipment and the digital media. For example the shipment could include five diskettes and two tapes. Then for each digital item (i.e. a diskette) there is a one-to-many relationship between the item and its digital files. For example one diskette could hold several GIS digital files. Therefore this one-to-many relationship has to be modeled within the FoxPro database. This is accomplished with an initial digital data screen which assigns a secondary key for each digital data item.
For each secondary key (diskette or tape) there is another FoxPro screen which assigns a tertiary key which is a unique number for each file on that item. These three keys represent the one-to-many-to-many relationship of digital data. Metadata, or data about data which permit users to understand the characteristics and lineage of the data, is stored for each digital data file.

DATA MAINTENANCE AND ENHANCEMENTS

One aspect of maintaining this database is the correction of errors that went undetected during all phases of the QA/QC plan and its processes. This type of error is completely unavoidable as a set of spatial databases such as these literally contain billions of discreet "pieces" of information. It would prove to be extremely costly to examine and review each of these. It is hoped that the QA/QC plan will identify the greatest majority of these errors, but some should be expected to go undetected. For example, in a 1750 m by 1750 meter tile of contour data, there are as many as 3,000 arcs, each must have connectivity with an adjacent arc; within each tile there may be as many as 14 different feature codes identifying the contours; and each arc must have an elevation attribute that is topologically correct in association with its adjacent arcs. Within the GCM group, a database administrator has been designated to perform these edits to the original data files. Appropriate system and software measures must be taken to ensure that no other staff can accidentally or intentionally modify the data sets. Another factor that must be considered in maintaining these databases are comments from the GISST user community. As users become more accustomed to working with data of this detail and accuracy, it is expected that some changes to the data structure may be required. To date, few comments and suggestions have been received from users although a number have been pleased with the resolution and image quality so far.

The task of updating large and complex spatial databases such as that created by this project presents many challenges, some of which are easily overcome while others are considerably more problematic. With the use of GPS on board the aircraft, cost associated with acquisition of new aerial photography has decreased to a point where new photographs are easily obtained. Unfortunately, costs associated with converting the photographs to a useable digital format remain high. Currently under consideration is the acquisition of new aerial photography during the 1996 flying season. Only selected portions of this photography will be converted to new digital ortho photographs, and if needed, new topographic and planimetric data could be generated. An advantage of this approach would be to minimize the associated cost. This option must be carefully considered as it carries the disadvantage of mixing data from different dates within the database. In addition, any users who have extracted data to their own system and included original basemap data in their research must be advised that portions of the database have been updated. This is particularly critical for any users who have created composite data sets from the original tiled data sets.

A second approach to updating these data sets is to design and implement a system whereby the GCM group is notified of any construction activities within the jurisdiction of the ORR, Portsmouth or Paducah facilities. This approach would require that digital 'as-built' documents be obtained and these would be converted from their native format to the appropriate GIS format. From these files, the original tiled databases could then be updated. As with the selected compilation method above, this would have the same problems with notification to users and repetition of effort for composite databases.

THE GISST SERVER

Getting the new basemap data to the users is very important. As a result, the organization and management of the data and ancillary supporting files is critical. The GIS and Spatial Technologies (GISST) data server, under development by the GCM, is designed to disseminate these data. The design is a large and logically organized set of directories and subdirectories. The GISST seeks to create
and maintain raw and processed GIS databases for use by all ORR spatial data users. The data are accessed and downloaded via the File Transfer Protocol (FTP). The GISST has been developed as a complement to the Oak Ridge Environmental Information System (OREIS), which includes a toolset for accessing, analyzing, visualizing, and reporting tabular environmental data stored in an extensive Oracle database and spatial data using ArcView and ArcInfo. GISST aids in the creation and transfer of composite high-quality geographic data to OREIS in a form that is easy to understand and use.

The GISST directory tree is divided into eight levels of subdirectories, with each level corresponding to a logical subdivision of the data. As users choose each subdirectory, they traverse the directory tree towards the data of interest. The first directory contains two choices: raster and vector. All imagery, Digital Elevation Models (DEMs), and supporting ancillary files are stored within subdirectories branching indirectly from the raster directory. Conversely, all vector files and supporting ancillary files are stored in subdirectories branching from the vector directory. The ordering of these subdirectory types was designed to reflect the way users would access the data. Subdirectories representing broad concepts, which the user would tend not to change during the data access process, were located at the beginning of the tree. Subdirectories that might be changed several times were placed nearer the end of the directory tree. The final subdirectories are the ones that actually hold the datasets. Therefore, the user would traverse the entire tree to get at the data that met the conditions represented by the preceding subdirectories. When finished downloading the datasets, the user would "climb back up" the directory tree until the appropriate branch to the next set of data is found. The design of the directory tree structure minimizes the amount of "tree climbing" required for the "average" user.

Part of the data management problem is to maintain different versions of each dataset in multiple formats and coordinate systems since there are no enforced standards imposed upon the users. The GISST stores the GIS files in one primary format due to harddisk space constraints: ArcInfo uncompressed export files in Tennessee State plane (meters). Users who download the data using FTP therefore must convert the data into other formats and/or coordinate systems themselves.

GISST/WWW

The main disadvantage of the FTP-accessed GISST server is that it is difficult to use: The user must type in the lengthy path names, determine what data are available, download a copy of the data, and move to another directory to repeat the process. Using NetScape server software, the GISST was enhanced to produce the GISST/WWW (GISST/ World Wide Web) screen-based point-and-click interface. The GISST/WWW allows registered users to quickly change directories and access the data of their choice by substituting clicking for typing.

NetScape provides built-in functions that are used extensively throughout the GISST/WWW. A script programming language known as the Practical Extraction and Report Language (PERL) was used to customize hypertext documents for GISST/WWW applications. The GISST/WWW uses a series of screens to replicate the directory structure of the GISST. These screens are created on-the-fly using PERL code based upon user-supplied parameters. For a given directory in the GISST, a screen is displayed which represented choices associated with lower-level subdirectories. The user would point-and-click on one of the icons to move to the corresponding subdirectory. At that point another screen would appear with icons representing the next level of choices. This method was visually appealing and greatly improved the ease of movement between subdirectories. The PERL script was also used to communicate with format and coordinate system conversion programs. The GISST/WWW converts on-the-fly from the Arc/Info export file in Tennessee State Plane (meters) into five other combinations of formats and coordinate systems. The user can download all the data in either Arc/Info export format or MapInfo mid/mif format. In addition, the coordinate system can be converted automatically to either
Tennessee State Plane (feet) or an ORR-developed "Administrative Grid" system. Similarly the TIFF formatted ortho images can be download on-the-fly into BMP format. These built-in conversions permit the GISST/WWW to support a wide range of GIS data users within the ORR community.

Further development of the GISST server is underway to meet a variety of user data needs at the early stages of ER projects. Several enhancements for greater ease-of-use or capability are under consideration. In addition, the field of Internet browser tools is rapidly changing and applications such as the GISST/WWW which are based on these tools must keep pace. The GISST/WWW will migrate to Netscape 2.0 which permits concurrent screens. This will allow more flexibility in how the user communicates with the GISST/WWW. One screen can be reserved for graphics while the other is used for user interactions such as commands and queries. The advent of Java and other Internet tools will permit applications such as the GISST/WWW to download part of their functions to the user's computer. Depending upon the software used, the application could draw user-requested vectors within the GISST/WWW screen. Currently the user must rely on stored GIF raster images of vector screens.

Other enhancements under consideration to the GISST/WWW involve data access. The user currently has to determine which data sets are available by telling the system the required data type, format, area, and coordinate system, through the series of screens. An additional method of data access would be based on the metadata files. Currently these metadata files are ASCII text files containing relevant information about their respective GIS data files. If these metadata files were instead stored within a database they could be queried directly by the user. The result of the query could lead directly to a list of choice(s) from which the user could quickly choose. This would have the added advantage of permitting the user to select data based on geographic areas rather than the hard-coded tile structure by permitting the user to specify the spatial extent of the area of interest. If more than one file is requested for downloading, the user currently has to wait for each download to complete before starting the next. A batch downloading capability is being investigated where the user could submit a "shopping list" and be free to pursue other tasks.

CONCLUSIONS

The Basemapping and Imagery Project's raster and vector GIS datafiles deliver high-resolution QA/QC'd data to the data users at the ORR, Portsmouth and Paducah facilities. These data are a much needed improvement over the former available basemaps in terms of resolution and attribution. When completed, these data will be the most comprehensive GIS and orthoimage coverages available for any DOE reservation. The development of the production system and quality control procedures was often difficult but hopefully future projects of this nature can benefit from this effort. While the delivery of the GIS data files will be completed in 1996, the maintenance of these files must continue. These data will be updated with information incorporated from ancillary datasets and key groups across all three project sites. These data changes must be controlled and disseminated to the user in a structured manner.

ACKNOWLEDGEMENTS

The authors would like to thank the rest of the GIS and Computer Modeling Group for their collective support in the Basemapping and Imagery Project. Special thanks go to Tom Evers, Phil Coleman, Michael Miller, Marsha Henley, Yang Cheng, Richard Goeltz and Michael Sherman. We would also like to thank Cathy Shappert and Jim Hines for their timely and helpful editing.
Figure 1. Work flow and production process of the Basemapping and Imagery Project.
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.