Migrating a Large Dataset to a Personal Geodatabase

Billie Leff and Windy Robertson, National Geographic Society – Maps Division
1145 17th Street NW, Washington, DC 20036
bleff@ngs.org and wroberts@ngs.org

Abstract:

Most people are familiar with a product such as the National Geographic World Atlas, but few are aware about the GIS intricacies behind the printed plates. Join us in a journey from the recently antiquated coverage format to the jazzy geodatabase format for this cartographic product. We will take a look at some of the challenges we were faced with and some of the solutions we have developed. Would we recommend the process - DEFINITELY, but unless you are Indiana Jones, with innate intuition about the safe route, come find out about the boards/rocks that you CAN step on.

Introduction:

A lot has changed since National Geographic launched its Cartographic Division in 1915. The exclusively artistic hand-drawing and hand-lettering process of creating maps was replaced over time with a process that skillfully combines the art and science of cartography. While new technology has allowed us to dramatically increase the number and variety of maps published in a year, it hasn't diminished our strong emphasis on visual presentation.

From the smallest 1-inch locator map to the World Atlas, the National Geographic map creation process is unique and involves the work of designers, researchers, GIS staff, editors, and production cartographers. To ensure consistent quality results, we have created custom tools, including our own fonts, a projection, a plot tool, a QC tool, a tint-band generation and a river-tapering tool. ESRI is our GIS provider and we have worked with them from day one to further develop and improve these tools. Incidentally, though most geospatial data begins its journey on a PC, all cartographic finishing is completed on a Mac.

Recently, both organizations worked overtime to produce the 8th Edition World Atlas. This project has completed yet another transition, first from film to digital, then into coverages and now into GDBs. But before immersing into subtypes and layer symbolization, I would like to review the National Geographic Atlas history. In 1958, the Society published an empty cover of the Atlas Folio. Each year, subscribers received a new plate to hand-glue into the blank folder, and sometimes even enter the plate name in the table of contents. It wasn't until the late 1980s that we transitioned from chronoflex to GIS. Even today, all plates are a compilation of outside and in-house data, with a varied scale, most are derived from hand-drawn maps and all relief is also hand-drawn.

To take advantage of some of the newer tools available to ArcGIS 8 users (such as QC through PLTS), we transitioned all our GIS data for the 8th Edition Atlas to the Geodatabase format. If this process were as simple as it seems, I wouldn't be sharing our experience with you…

Discussion:

The GIS Data Dictionary (our metadata equivalent) is the foundation of the GDB design. Each of the 70 plus Atlas plates resides in its own GDB with two geodatasets (one for geospatial elements, such as political boundaries; and one for page elements, such as map title and bar scale). Within the geospatial elements dataset, there are 27 classes with about

8.5 subtypes each. There are also 8 annotation classes with about 22 attribute codes each. Rather than just migrating the data as it was, we decided to first improve the organization of features within classes. However, regrouping the features caused problems with the creation of the index. For example, FIPS codes with country names now resided in multiple classes, thus complicating the indexing process.

To facilitate the data migration and to meet our deadlines, we divided the 103 personal geodatabases (70 plates and 33 insets) into 10 task orders (TOs). For each TO, we allotted one week for editorial comments, two weeks for editorial changes, one week for QC, and one week for final changes and map production. This method allowed us to balance the amount of work assigned to the contractor (for example, not much has changed in North America, however this is not the case in Eastern Europe). An overlap in the processing of the TOs providing consistent work for the various groups involved.

Managing quality control for such a project was a mammoth undertaking. Having an established protocol, rather than a constantly evolving one, would have sped up the process, ensured consistency, and eliminated the need for secondary quality control. We struggled with automating this process for a product with such a strong visual emphasis. A QC protocol was eventually developed for each phase of TO processing: QC after database design ensured that the correct classes, subtypes and attributes were included and that they were derived from the correct coverage data; QC after database migration ensured feature existence, integrity, and visibility; QC after edit ensured topology validity and edit completeness; QC after map production ensured feature symbolization, map positioning and scale, and text correctness; and QC of the index ensured consistency in abbreviations, indexing of physical features, feature restructuring, and countless other requirements.

The next processing phase for each task order to be reviewed is the data migration phase. This phase was completed by our contractor using the advanced data loading capabilities of the PLTS Data Loader. During this phase, an automatically generated QC report for any non-migrated data facilitates the QC process. We often did not identify such data until later phases, which lead to the continuous editing of the GDB design and the addition of data.

After the GIS data was migrated from the coverage to the new GDB format, our editors provided our subcontractor with Digital Edit Overlays (DEOs). These layered Illustrator files identified all updates that needed to be completed for the 8th Edition Atlas. There were over 15,000 edit calls that ranged from feature additions (high-speed train routes, space ports and new towns) to text edits (changes in the transcription of towns, political unit renaming).

The last GIS processing phase for each plate was plate map production. All plates feature a pre-defined legend, scale, and map angle. We achieved this using ArcMap query definitions (to further break down particular classes), layer symbolization, and having a fixed scale and map angle. We then exported the maps as layered Illustrator files for final cartographic production. At this time we ran into Microsoft rounding errors: map scales and feature placements changed because all coordinates were rounded for the sake of speed. As a result, all map scales were finalized in Illustrator and text adjustments had to be made. Additional finishing touches were also performed in Illustrator (such as plate number and plate grid placement, stacked type adjustments, and hillshade image insertion).

The last phase of the project was generating the Index. This text document features well 100,000 place names and was developed using prototype indexing software. To help you understand the complexity of the process, following are some of our guidelines: feature names cannot be abbreviated in the index (though they can be on the maps), individual pieces of annotation which are part of a piece of stacked-type must be indexed correctly,

rivers bordering two countries must list both country names, features on island possessions must include the "mother" country, features appearing on more than one map must be indexed to the map on which the political unit is shown in its entirety (the "best plate"). As you can imagine, this is all accomplished through data attribution and programming magic.

Conclusion:

Having a committed partner and dedicated staff was critical to the completion of this project. Though several organizations may have completed the migration of such a large database to the new GIS format of GDBs, when we began, there were few precedents. Having started the project ahead of schedule, and by maintaining an open mind to untested solutions, we were able to work through the challenges and print an Atlas that we are all proud of. Following is a summary of our obstacles, solutions, and methodology changes.

Although the data migration to a new format was simple (just a continuously evolving process repeated almost 100 times), we encountered obstacles and made numerous changes:

1. The GDB design was not completed until more than half-way through the project.
2. No provisions were made for the generation of metadata.
3. Most insets, which previously were not projected, had to be projected (including projecting their annotation).
4. In addition to data migration, all custom tools had to be adapted.
5. We encountered font and feature placement differences between PCs and Macs.
6. In the past, custom site symbols were placed in text fonts for ease of symbol rotation, resulting in storing all point features as both points and annotation.
7. We ran into difficulties with the handling of annotation in geodatabases.
8. In the past, "holding lines" were used for annotation placement.
9. Generating the index was a long and painful process and we all underestimated its complexity.

We were however, able to better organize our data and learn more about the capabilities of the software (we were also able to influence the development of ArcGIS 9.0). Because we had a highly professional and knowledgeable partner and because both organizations were highly committed to the end goal, we were able to resolve all problems:

1. Once the DGB design was finalized, all data was migrated a second time (and sometimes a third time).
2. Creating metadata for a constantly evolving design would have been a waste of time. It is never too late to create metadata (unless the GIS team bails ship).
3. Now that all insets have been projected, they can be used as bases for other map products.
4. We are still in the process of developing and improving our custom tools and some of the tools have been replaces by tools inherent from the new software (e.g. map plotting or map production from ArcMap to Illustrator).
5. All text placement is finalized on a Mac. Feature placement and map size problems appeared to result from a Microsoft rounding error, which we compensate for on the Macs.
6. Site symbols are now stored only as points saving time in their maintenance and reducing the file size.
7. We have been assured that annotation handling has been greatly improved in ArcGIS version 9.0 (however, we now have to migrate the annotation portion of the data to that format).
8. Having dealt away with the "holding lines" saved a lot of time in data maintenance and attribution.

9. As long as it works (and we all eventually get compensated for it), the double time we all put in for the generation of the index is well worth it.

Acknowledgements:

We would like to acknowledge the tireless efforts of our subcontractor ESRI and their subcontractor AIS. ESRI went beyond their call of duty (and contract obligations) to "read our minds" in a continuous effort to improve the Atlas product. They were highly responsive to all our needs. We would also like to acknowledge our Chief Cartographer Allen Carroll and our Division Director Kevin Allen, who were able to guide us through the project completion; our Atlas Project Manager, Rick Bullington, who has not had a day off from Thanksgiving to the end of May (he almost had to learn to ride his bike again); our editorial team, headed by Michael Horner, which put in large amounts of overtime and ensured that all edits were as consistent as humanly possible; our designer Sally Summerall who did not let the tiniest lake on the one-inch locator globes slip by with a 25% color difference; our production team, headed by Jehan Aziz, who was able to automate the map production; our research team, headed by Amy Jo Woodruff, who had to put up with unsolicited design comments from all of us; and finally, the GIS team which pulled off double-time literally for the index compilation, crashing at the apartment closest to our office, so as to not waste time commuting, and leaving behind screaming children in the hands of exhausted fathers (they all survived).

Author Information:

Billie Leff
GIS Project Manager
National Geographic Society – Maps Division
1145 17th Street NW, Washington, DC 20036
(202) 828-6664 office
(202) 429-5704 fax
bleff@ngs.org

Windy Arrow Robertson
GIS Analyst
National Geographic Society – Maps Division
1145 17th Street NW, Washington, DC 20036
(202) 857-5808 office
(202) 429-5704 fax
wroberts@ngs.org