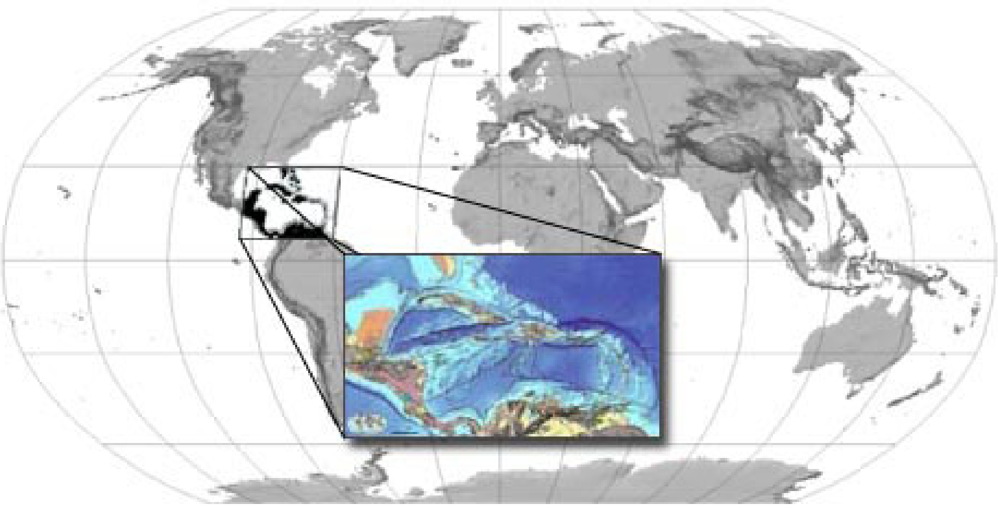
Creating a Geologic Dataset of the Caribbean using ArcGIS

by Christopher D. French
U.S. Geological Survey, Denver, Colorado



Abstract

A geologic dataset of the Caribbean region was recently created for use in the U.S. Geological Survey's World Energy Project. ArcGIS 8.3 and the geodatabase format were found to be effective tools in creating a high-quality and topologically correct dataset. This paper is geared toward novice to intermediate ArcGIS users who would like to know more about the process of creating polygon feature classes in the geodatabase. Issues involved in the process will be discussed in some detail, including digitizing of the source maps, establishing topology rules, attributing with coded domains, maintaining quality control, and addressing challenges throughout the process.

Introduction

A digital geologic map of the Caribbean region was created as part of a series of geologic maps for the U.S. Geological Survey's (USGS) World Energy Project, available on CD-ROM and online. The goal of the World Energy Project is to assess the undiscovered, technically recoverable oil and gas resources of the world. Mapping of the surface geology, geologic provinces, oil and gas fields, and miscellaneous cultural data provides a reference map for the project's research geologists to use throughout a subsurface investigation.

The geology of the Caribbean region is as rich and diverse as the inhabitants living on its islands, isthmuses, and land masses. For the purposes of this project, the Caribbean region extends from the southern tip of Florida to the mainland of South America, and from Barbados to the Yucatan Peninsula of Mexico. In Cartesian coordinates this area ranges from approximately 7 degrees to 28 degrees north latitude and 58 degrees to 93 degrees west longitude. Active volcanism is taking place at the Caribbean plate margins throughout Central America and the Greater and Lesser Antilles, providing a living laboratory for volcanologists. Research involving geomagnetism, seismicity, deep-sea drilling, radar bathymetry, geochronology, and petrochemistry among others, are improving our understanding of the complex geologic evolution of the Caribbean region (Mann, 1999). Some potential for undiscovered oil and gas resources is present in the region and continuing geologic studies will help improve future assessments. Every geologic study in the region plays a part in understanding the geologic story as a whole.

The goal of this mapping project was primarily to convert a hard-copy source map containing surface geology and tectonics into a digital form for use in the USGS World Energy Project and for public dissemination. The original geologic map was created in 1980 by J.E. Case and T.L. Holcombe of the USGS, prepared in cooperation with the U.S. Naval Oceanographic Office and U.S. Naval Ocean Research and Development Activity (Case and Holcombe, 1980). At a scale of 1:2,500,000, it is currently the largest scale geologic map of the entire region. Although additional data were digitized or otherwise compiled for the project, they will not be discussed in this paper, as the focus will remain on creating the polygon feature class depicting surface geology. The resulting map compilation and associated data will be published by the USGS as Open-File Report 97-470-K on a CD-ROM and as an Internet Map Service (IMS) through the USGS Central Energy Resources Team website, available at: http://energy.cr.usgs.gov (fig. 1).

This paper describes the methods and tools used in creating the surface geology dataset. Several steps are explained for data preparation, including scanning, georeferencing, and digitizing into a shapefile. The process of creating the geodatabase and using some of its features for editing and attributing of the geodatabase are also explained. Following is a discussion of challenges encountered and lessons learned from the process of creating the dataset.

Methodology

Data Preparation

The geologic-tectonic map published in 1980 by the USGS is the most detailed geologic map of the entire Caribbean region. The map was published as two plates in the discontinued USGS Miscellaneous Investigations Map Series as map I-1100. Previous attempts were made unsuccessfully to accurately digitize the map. The integration of powerful graphics programs (Adobe Photoshop) with GIS for heads-up (on-screen) digitizing and merging of the scanned maps allowed for more accurate georeferencing of a single image than merged vector files. Heads-up digitizing is defined here as a process where a registered, scanned image of a map is used as a guide for drafting on a computer screen.

Georeferencing the image was performed in ArcMap 8.3 using a 3rd order polynomial transformation and approximately 56 control points on the map. Political boundaries were used as a guide to match coastline features to the scanned map. A visual comparison of previously created vector files to the newly scanned and georeferenced map revealed errors possibly caused by building and cleaning the coverages with high tolerances, or generalizing, among other steps in the process. A fresh slate was necessary to obtain a more accurate digital geologic map of the Caribbean region.

Digitizing into Shapefile

The "on-screen" drafting was initially saved in the shapefile format because of the ease of using this simplified format. The shapefile was later exported to a personal geodatabase for topology and attribute editing. To begin the process, a polygon was created using the "Create New Feature" edit task from the editing toolbar. Areas adjacent to the new polygon were then added using the "Auto-Complete Polygon" task found under the topology tasks subsection of

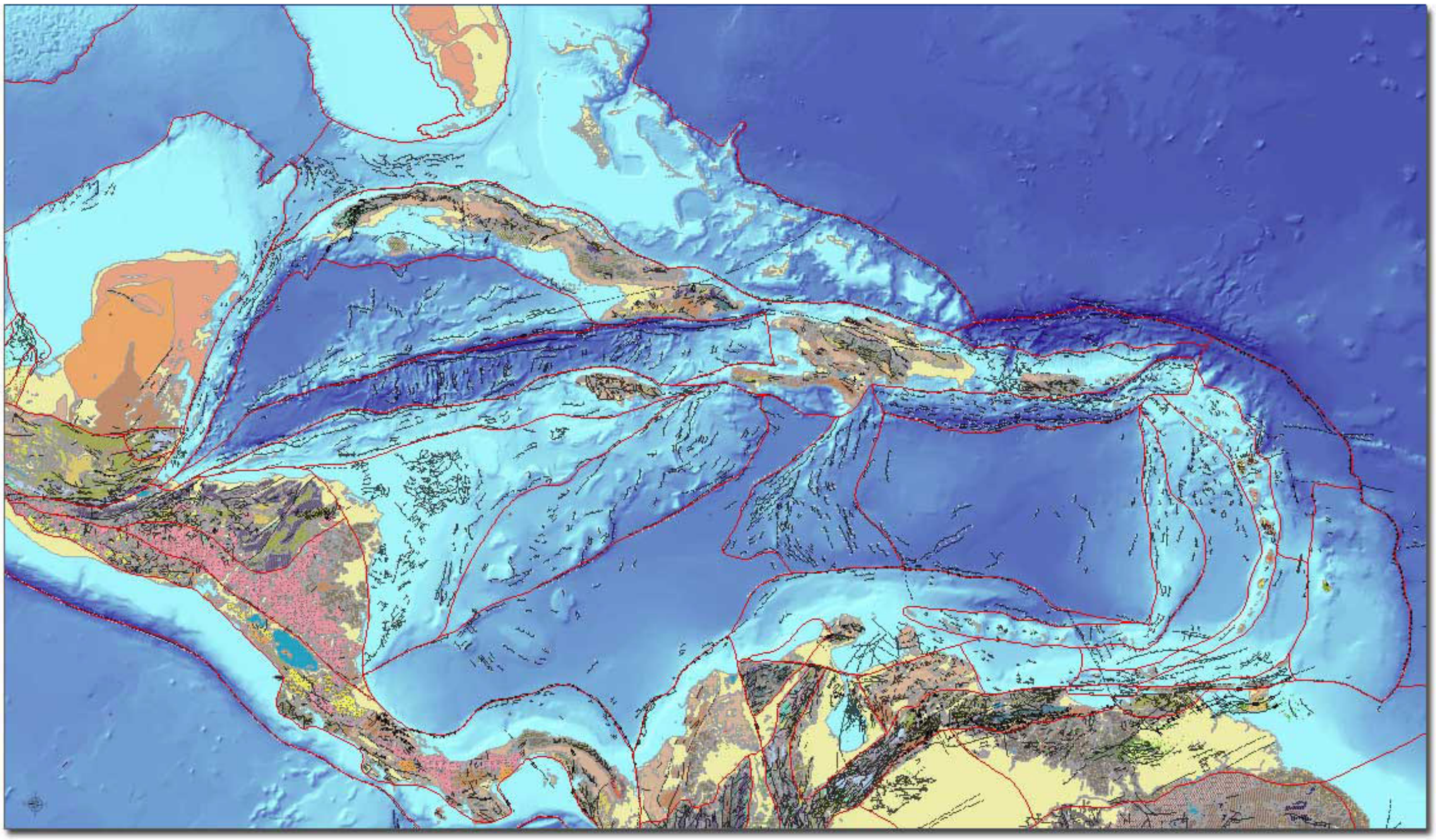


Figure 1. Screen capture of the Map Showing Geology, Oil and Gas Fields, and Geologic Provinces of the Caribbean Region.

the edit task selection box.
Auto-complete polygon is an editing tool that allows a user to complete a polygon using existing polygon geometry by snapping to the edges of the existing polygon or starting and finishing the sketch just inside the boundary of the existing polygon edge. It can save time by not having to re-digitize a shared boundary. Two precautions should be noted when using this tool. The auto-complete polygon tool needs to begin by crossing the boundary of a polygon and end by crossing the boundary of the same polygon or another adjacent polygon (fig. 2). If the mouse is double-clicked accidentally in an open area without completing the polygon, the tool will fail and leave no trace of the attempt to digitize a polygon. This becomes a problem when digitizing large polygons that might take 15 to 30 or more minutes to complete; appropriate mouse settings and periodic stretching of fingers is recommended. A second precaution is when the auto-complete polygon tool is used among many adjacent polygons, the software cannot detect adjacency and fails to complete the polygon. This took place with more frequency toward the end of the project after several thousand polygons were created.

The snapping tool is another useful feature in ArcMap for digitizing polygons and shared boundaries. This allows a user to define snapping tolerance, which can be set to pixels or map units and can be found under ArcMap's editing options. The actual snapping environment is a dockable window listing all of the layers in a project and can be set to snap to vertices, edges, or ends of any layer in the list. When tools like auto-complete polygon do not do the trick, work around the existing polygons, snapping to each node to ensure the features are properly aligned. Some experimenting with different settings might be needed to get one that works well for the situation. Snapping is also effective for cleaning up the dataset.

Geodatabase Editing

A geodatabase was created after completing the entire surface geology feature class as a shapefile. Waiting to create the geodatabase was decided in part because a shapefile can easily be imported to a geodatabase with no modification, and is essentially handled in the same way, despite lacking topology. Another reason for postponing the geodatabase creation was because additional data were being gathered throughout the mapping process, and the final extent was still being determined.

Geodatabases, as defined by ESRI, are relational databases containing geographic information that store feature classes (points, lines, polygons, and attributes), relationships, annotation, and dimensions. Groups of related feature classes and annotation can be stored in a feature dataset as long as they share the same coordinate system. Once a geodatabase is created, the extent or precision cannot be modified. The extent can be set initially in ArcCatalog with the X,Y domain when creating a feature dataset or by default when importing the first feature class to the geodatabase. Some forethought is required to set the extent appropriately as future layers might or might not share the same extent, and larger extents will generate an error. In order to prevent this problem in the Caribbean project, a simple rectangle was

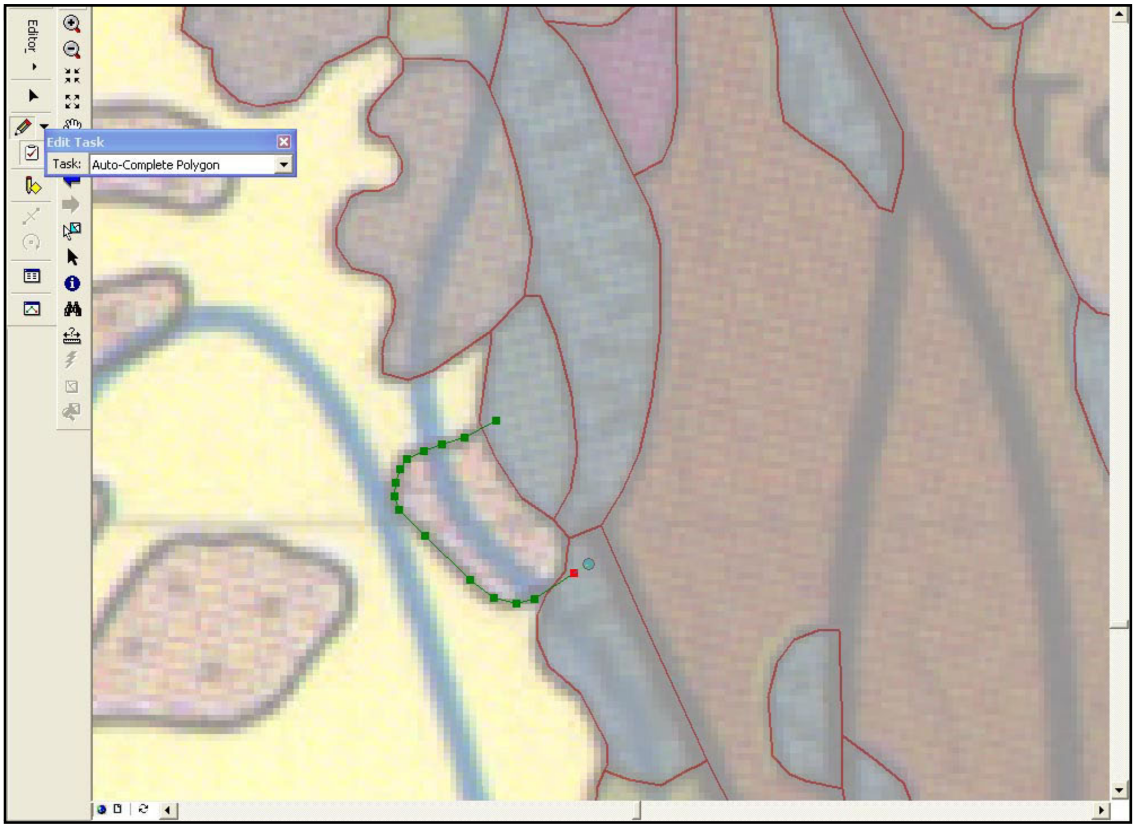


Figure 2. Screen capture of editing session where auto-complete polygon tool is beginning and ending in pre-existing polygons.

created for clipping all datasets before importing to the geodatabase.

Put simply, geodatabases are containers for storing data, where topology, relationships, annotation, attributes, and features all relate and interact with one another based on established rules defined by the user. Geodatabases offer a new way to create datasets with more integrity while more accurately modeling the "real" world. Project and data complexity might dictate how elaborate a geodatabase needs to be. While this dataset contained many polygons, it was considered to be rather straightforward and did not employ advanced geodatabase features.

Two forms of the geodatabase are available—one is considered a personal geodatabase and the other an enterprise geodatabase. At ArcGIS version 8.3, personal geodatabases had size limitations (2GB), no ability to store raster data, and no multi-user editing, while running on Windows platforms in an Access database. Enterprise geodatabases alleviate each of these limitations because they can store any amount of data (depending on server space), store raster catalogs, and allow multi-user editing through versioning. While both enterprise and personal geodatabases provide topology rules and efficient attributing with domains, a personal geodatabase was chosen for portability, considering a CD-ROM would be distributed with ArcMap and ArcReader map documents accessing the geodatabase file. Eventually, the personal geodatabase will be transferred to the Central Energy Resources Team's enterprise geodatabase, powered by Oracle 9i, for serving the Internet Map Service (IMS).

Editing the geodatabase began with establishing topology rules. A total of 26 geodatabase topology rules were available at the ArcGIS 8.3 release for creating complex modeling of topological relationships among point, line, and polygon feature classes. Any combination or number of rules can be applied within a feature dataset as long as they match the feature types they are targeting. Only two rules were necessary for the surface geology polygon feature class; "Must not overlap" and "Must not have gaps". Although the polygon rules were simple for the surface geology layer, they helped achieve a "clean" dataset, where no overlaps or gaps existed. Any sliver gap or overlap could lead to an incorrect analysis or interpretation and add unnecessary polygons, possibly degrading the dataset.

Validation of the geodatabase is the process of the software checking the applied topology rules and relationships against the dataset. Before a validation is run, all areas not yet validated are considered "dirty areas". When a validation is run, the dataset can be modified by snapping nodes to each other depending on the cluster tolerance of the topology. Cluster tolerance of the topology is entirely dependent on the scale, complexity, and accuracy of the data and is considered by the software to be the distance (in map units) where two elements being evaluated are treated as identical. Some thought should be given to setting an appropriate cluster tolerance; testing is necessary to check movement of features after a validation. If modifications to the data are made after they are validated they once again become dirty. Validation can be run on the entire dataset, or on the current extent, saving time for each validation process. Once an area is validated, the error inspector can be used to find and modify errors or mark them as exceptions. Validation of a feature class can be performed in ArcCatalog or in an editing session of ArcMap, although it is highly recommended to validate a topology within ArcMap. When ArcCatalog is used, there is no ability to undo the validation. Undesirable results after a validation could cause a major setback. Keeping many versions of backups are also crucial for working with geodatabases.

After validation, polygons that do not meet the topology criteria are flagged as errors and can be viewed and edited with the error inspector. The error inspector was more beneficial when used in conjunction with the topology, snapping, and editing toolbars, while the dataset was being edited (fig. 3). In the error inspector, a drop-down selection box is used to select the rule type for checking topological relationships. Users can specify whether to check for errors, exceptions or both in the visible extent or the entire dataset. Errors can be investigated and fixed in three ways: by right-clicking and allowing ArcMap to offer a fix, manually adjusting the dataset, or marking them as exceptions. Only a few hundred errors were found in the dataset of 6343 polygons, which frequently consisted of gaps or overlaps where one extra node was inadvertently placed in the digitizing process. The majority of corrections were made by selecting the sliver polygon or gap error, zooming to the selected feature, finding the culprit node and deleting the extra vertex (fig. 4). After removing the sliver errors and marking a few exceptions, the surface geology dataset was topologically sound and ready for attributing.

Geodatabase Attributing

Attributing a digitized dataset can sometimes be a long and tedious chore that the geodatabase can help simplify by improving the attribute accuracy and reducing errors. Users can define allowable attributes for a given field and select the attributes during an editing session by using a drop-down selection box. This is referred to as creating “coded domains” in the geodatabase properties. For the Caribbean geodatabase, coded domains were created that define the geologic codes and descriptions of the surface geology. These domains were then used to attribute the Caribbean geodatabase. Domains

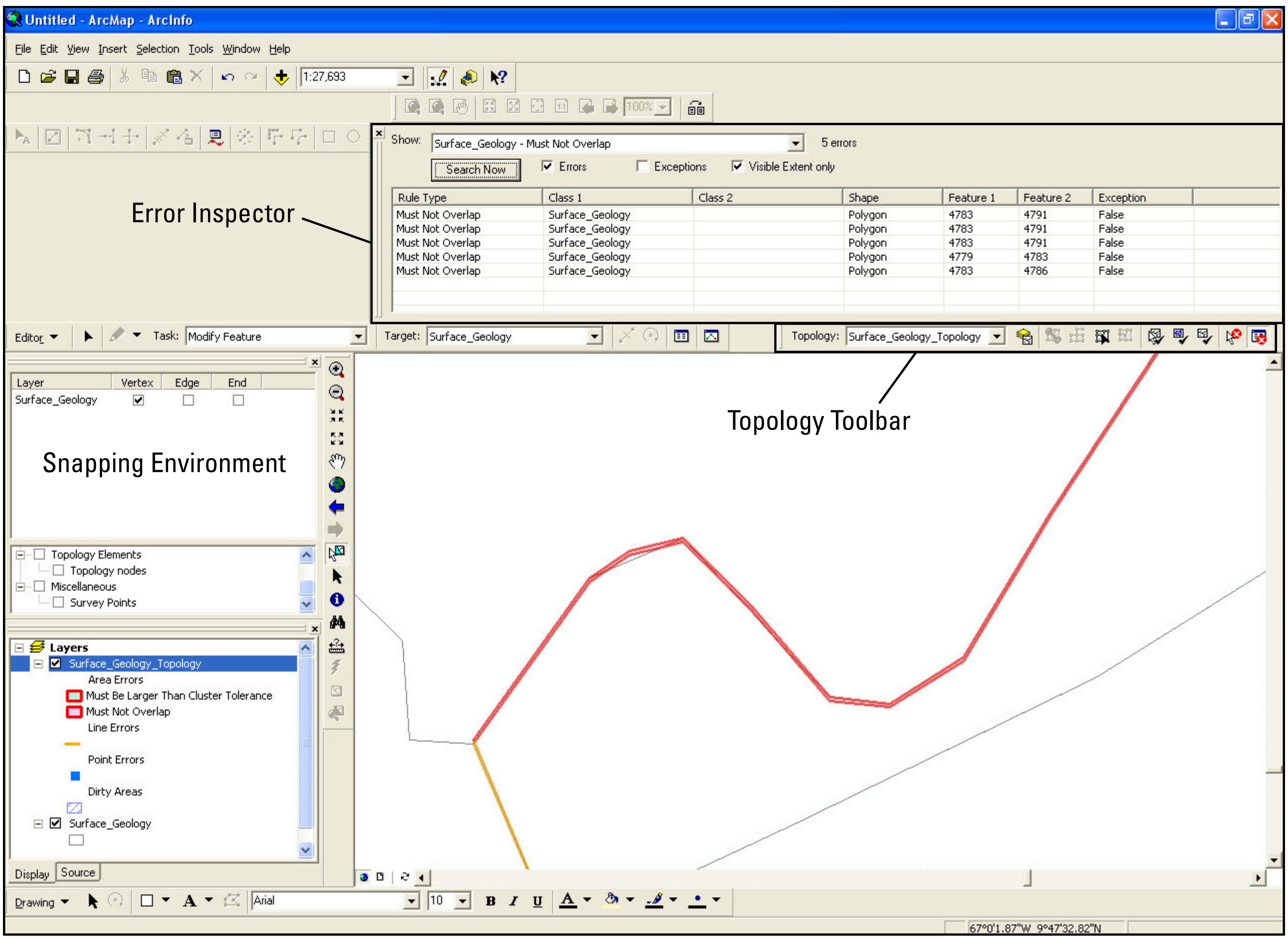


Figure 3. Screen capture of ArcMap 8.3 editing environment showing combination of error inspector, topology toolbar, and snapping environment.

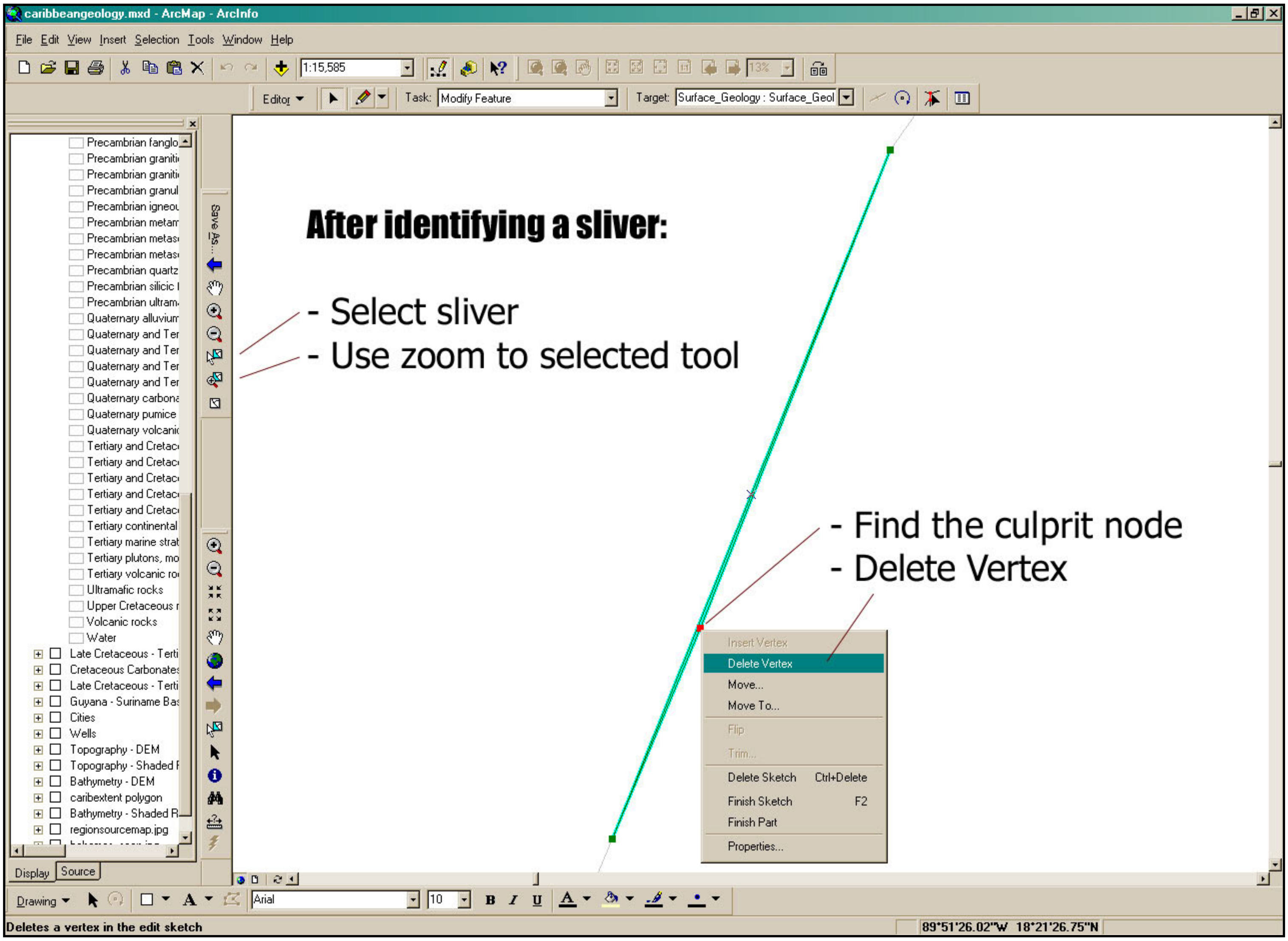


Figure 4. Screen capture demonstrating the process of manually removing extra nodes causing gaps or overlaps. Deleting a single node corrected the majority of topology errors.

consist of three parts when established in ArcCatalog; a name, properties, and coded values (fig. 5). Domains are defined and can be elaborated in the description portion of the dialog for clarification. Domain properties are set for field type and domain type. Coded values are the attribute choices provided in the pull-down selection box during the attributing process (fig. 6). To further speed up the attributing process, a plain text document was created in order to reference the geologic codes to the correct descriptions.

Two other ArcMap features were helpful in the attributing process. One was the ability to select and attribute multiple polygons at one time. The other was the use of transparency to delineate between attributed and non-attributed polygons (fig. 7). The overall attributing and spot checking for this dataset took only days compared to weeks using other methods.

Discussion

Methods chosen for digitizing the map previously might have led to discrepancies and are worth mentioning. Another attempted method was automated raster to vector conversion that involved drafting the polygons on Mylar film, scanning as a binary image, and allowing ArcInfo to decipher the linework. The vector method introduced errors during vector processing including those caused by cleaning multiple generations of coverages. The heads-up approach proved to be more effective in reproducing the surface geology that existed on the original map.

While digitizing on-screen, a user can “zoom” in to check the accuracy of line placement while digitizing. Features like snapping and auto-complete polygon in the ArcGIS editing environment added efficiency to the process and created an initially “cleaner” dataset. Striking a balance between the accuracy of linework and timeliness should be determined early in the process in order to complete the project within a reasonable timeframe. Common sense, scale of the source data, and required scale for the GIS feature class are important parts of the equation. Deadlines also play a role in determining the digitizing scale, but one factor remains true regardless—keep it consistent and well documented. A consistent scale of 1:250,000 was

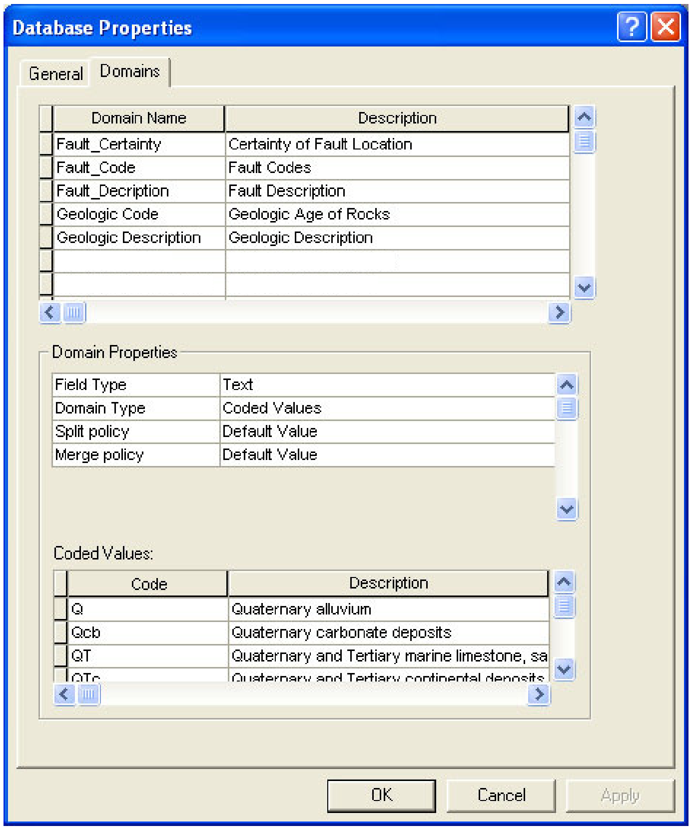


Figure 5. Screen capture of domain dialog window showing names and descriptions of domains, domain properties, and list of coded values.

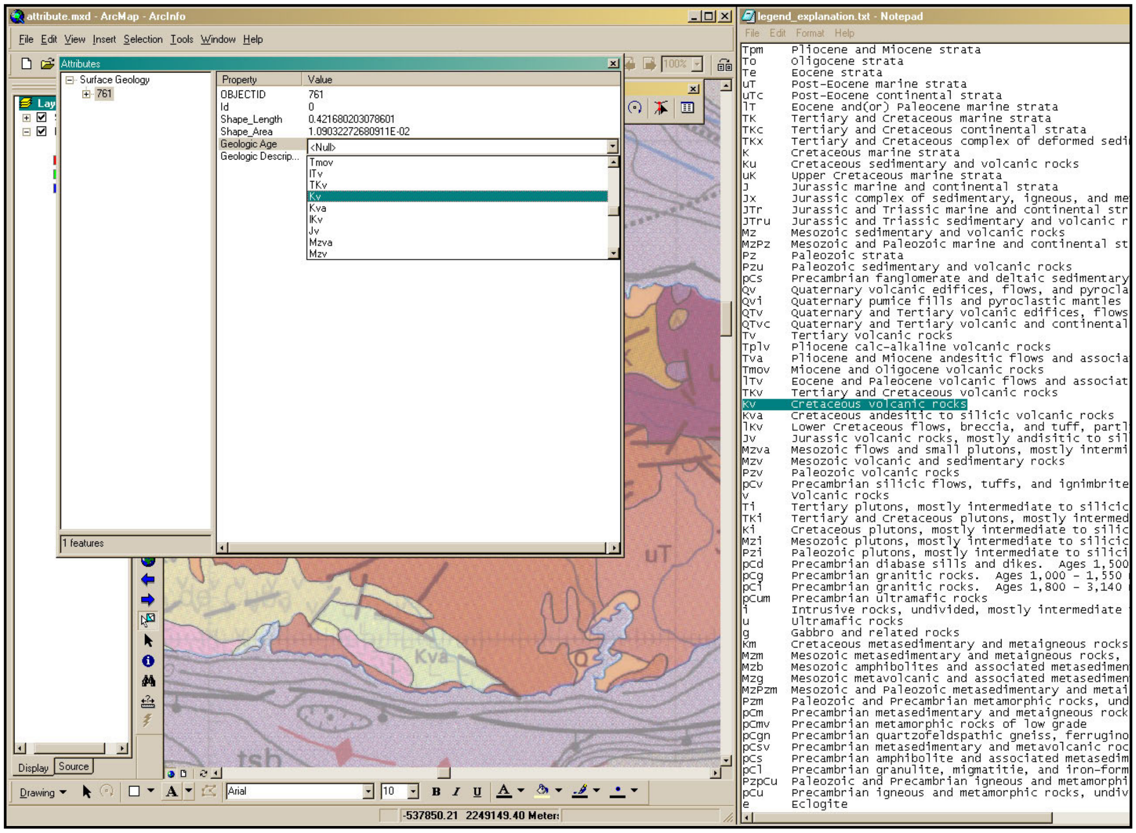


Figure 6. Screen capture of attribute dialog with pull-down menu from coded domain list and text document on right to assist with code-description matching.

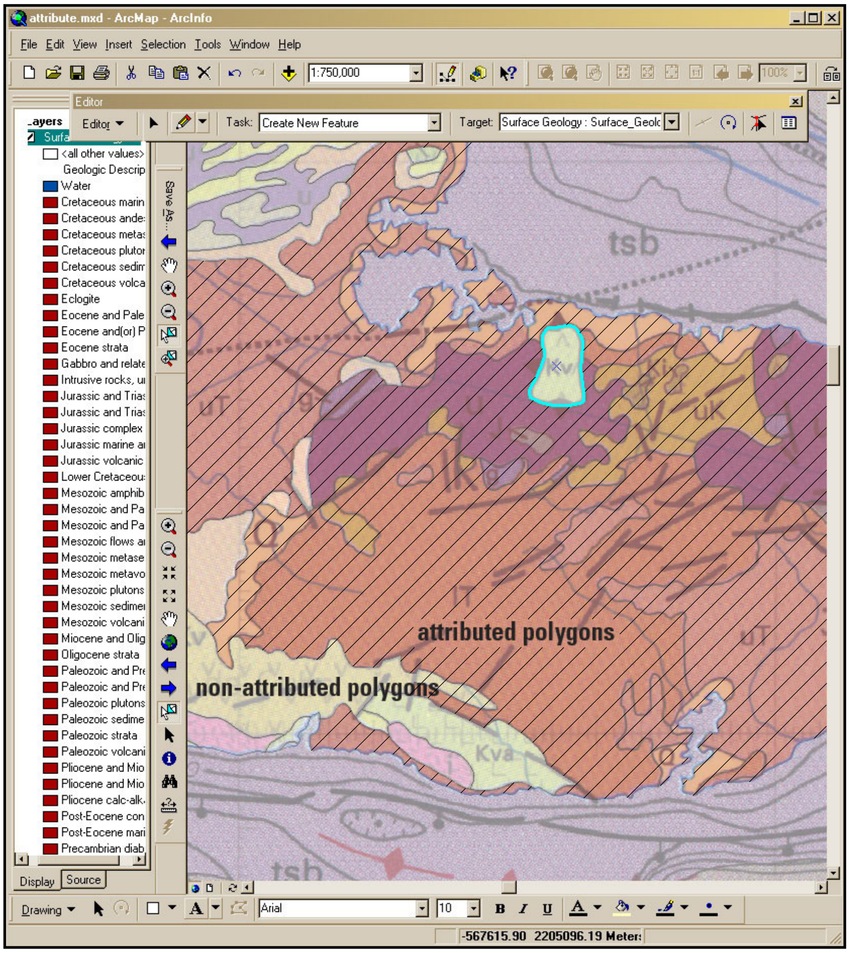


Figure 7. Screen capture of red semi-transparent polygons with hatching to compare attributed and non-attributed polygons.

determined to be appropriate while digitizing to assure accurate and smooth lines.

One aspect of setting up and using the coded domains that made it counter-intuitive at times relates to the "Coded Values" portion of the domain dialog. For this property, the code is the alpha-numeric identifier that is detailed in the associated description. The description is shown when using the identify tool, selecting by attributes, or when labeling. The fact that the description is used for labeling became an issue in the Caribbean dataset because traditionally geologic maps use the geologic code for describing a polygon. This problem had to be alleviated by creating another coded domain with the description matching the code. In addition, exporting the geodatabase feature class as a shapefile or coverage will produce a field with the code as the attribute—not the description as it is shown in the geodatabase. Therefore, a lookup table in DBF format was created to join the description field based on the geologic code.

Summary

ArcGIS and the geodatabase format were found to be effective tools in the creation of the geologic dataset of the Caribbean region. Establishing topological rules and coded domains in the geodatabase for attributing led to a higher-quality data product because of enforcement of the rules and selection of attributes. ArcMap helped in preparing the source data through georeferencing and offered many tools for efficient heads-up digitizing. Heads-up digitizing was superior to traditional methods of digitizing because of its ability to enlarge the original work and digitize at a larger scale than would normally be feasible.

This project scratched the surface of geodatabase capabilities. Many more features are inherent in the geodatabase for editing and modeling than were utilized in this project. A recommendation to GIS users who have not experienced the geodatabase is to sample its functionality and usefulness for their projects.

References Cited

Case, J.E., and Holcombe, T.L., 1980, Geologic-tectonic map of the Caribbean Region: U.S Geological Survey Miscellaneous Investigations Series Map I-1100, scale 1:2,500,000. [Prepared in cooperation with the United States Naval Oceanographic Office and the United States Naval Ocean Research and Development Activity.]

Mann, P., 1999, Caribbean sedimentary basins—Classification and tectonic setting from Jurassic to Recent, *in* Mann, P. ed., Caribbean basins. Sedimentary Basins of the World 4, Elsevier Science B.V., Amsterdam, p. 3-31.