DRAFT -- To be published in DMT'08 Proceedings
(see http://ngmdb.usgs.gov/Info/dmt/)

Rescuing Legacy Digital Data: Maps stored in Adobe Illustrator™ format

By Andrew L. Wunderlich and Robert D. Hatcher, Jr.

Tectonics & Structural Geology Research Group
Department of Earth and Planetary Sciences
and Science Alliance Center of Excellence
306 Earth & Planetary Sciences Building
1412 Circle Drive
University of Tennessee
Knoxville, TN 37996-1410
Telephone: (865) 974-6448
email: gibbon@utk.edu

ABSTRACT

As GIS databases become a standard for spatial data storage, many organizations may be struggling to integrate "legacy" digital data into modern geospatial databases. Map information that is stored in older digital data formats without spatial reference or attribution are in danger of being lost for future mapping and analysis purposes if the data are not converted to a newer digital geospatial database. One popular digital file type for mapping is the Adobe Illustrator™ (AI) format. In our research group at the University of Tennessee, we have a large collection of legacy digital and paper geologic maps covering over 100 7.5-minute quadrangles, which are products of over 40 years of detailed geologic mapping (1:24,000 or larger scale) mostly in the southern Appalachians. Our goal is to transform these data into geospatial databases to enhance their long-term survival and to make them more useful to the geologic community. In response, a six-step method has been developed to convert these maps, virtually intact, to the ESRI geodatabase format: 1) the original files are organized into layers and cleaned up in AI and exported as AutoCAD™ drawings; 2) AutoCAD files are converted to shapefiles and spatially adjusted in ArcMap™ and appended into a geodatabase; 3) geologic point data attributes that were not retained directly are either calculated from the feature (strike/rotation) or added to the features semi-automatically with the help of ArcMap utilities developed in-house (dip/plunge); 4) the entire database is then checked for topology errors, and if the map being processed is adjacent to existing map databases, the adjoining edges are reconciled; 5) a round of quality control (QC) measures are taken, including correcting mistaken attribution; and 6) the finished database is then symbolized, labeled in ArcMap, and exported as a graphic for placement in a final map layout for editing and publication. This is a working method, and as such we are interested in suggestions for improvements.

INTRODUCTION

Many geoscience organizations have large collections of legacy maps in their possession. These maps may be in paper, stable base (Mylar), or other analog forms, or they may be stored in one of various digital formats. Most of these maps were produced for publication and usually have no digital geospatial component, unless they have been digitized. Maps may also be stored in such legacy digital graphics formats as Adobe Illustrator, Macromedia Freehand (no longer in production), and CorelDraw. These formats are common, because these programs have been the best computer graphics solutions for preparing maps to be published. While printed maps are still the standard for many organizations, new technology being utilized in the capture and storage of geospatial data, such as GPS data logging, has created the need to develop geospatial databases in which to store this content. Now, many people are struggling with this duality–new data are stored in databases, old data are stored on paper or in digital graphics–and the two are not easily integrated. Geologic maps stored in legacy digital formats have the potential to be converted to newer geospatial databases. The time, effort, and cost to do research, collect data, and create and edit these maps certainly justifies the comparatively small cost to convert these data to modern formats that are accessible through GIS software and are thus interoperable with new datasets.

Conversion of maps in legacy formats has been the topic of several past DMT presentations. Hatcher (2005) wrote of the importance of converting old maps to digital formats. He stated that old maps are still a vital source of "primary geometric, spatial, and resources data useful for crustal and surficial geologic research, and mining, petroleum, engineering, and environmental applications" (p. 11). He describes the process of scanning and re-compiling old maps in Adobe Illustrator, using the third-party extension MaPublisher by Avenza to georegister the content. The use of MapPublisher to manage the map content has the advantage of being able to export shapefiles of the features to be used in a GIS. However, this method relies on an expensive third-party application to export map data for use in GIS, and the management of geospatial data in Illustrator is not ideal, because that is not its intended use. GIS software packages, such as ArcGIS, are designed for creating, managing, and analyzing geospatial data, and, with recent improvements to the cartographic capabilities, are better equipped to produce maps that are of publication quality, at a price comparable to that of a single-user license for MaPublisher. As a result, more and more organizations are turning to GIS software for creating new geologic datasets and maps, as well as integrating old data and maps through digitization. But the issue of how to integrate legacy digital formats persists. At the Digital Mapping Techniques '06 meeting, Jennifer Mauldin (2006) outlined a basic process necessary to convert maps stored in AI files to ArcGIS-compatible geodatabases. She described the major steps in the process developed at the Nevada Bureau of Mines and Geology to convert their maps, but detailed information was not given. In this paper, we hope to shed more light on the conversion process and give a more detailed account of the process and its nuances.

At the Tectonics and Structural Geology Research Group at The University of Tennessee, we have a large collection of detailed geologic maps covering over 100 USGS 1:24,000-scale 7.5-minute quadrangles. We developed an efficient process to convert to geospatial databases the maps that had been prepared in a legacy digital format in order to facilitate research, analysis, and regional map compilations, and to preserve the products of over 40 years of detailed mapping mostly in the southern Appalachians. The process had to be compatible with our workflow for digitizing paper and stable-base Mylar maps, i.e., both processes need to utilize the same database schema and quality control (QC) process and produce the same end products. Over the last year, we have developed the process outlined here. Throughout the process description, software commands and interfaces are referenced in the text and may also be shown in the figures.

The software referenced in this paper is Adobe Illustrator CS3 and ArcGIS 9.2 with an ArcInfo license level. While there are not significant differences between recent versions of either software package as they pertain to the process outlined here, some differences may exist. Most of the commands and interfaces should be the same or very similar for Illustrator 10 and newer, and for any ArcGIS version 9.x.

The process consists of six major steps. First, the AI file is exported to one or more CAD drawings. The CAD drawings are inspected and shapefiles of the pertinent classes from each drawing are exported and spatially adjusted in ArcMap. The referenced features are checked and then attributed using custom ArcMap tools and reference imagery exported from Illustrator. The database is checked for topology errors and edge-matched with adjoining map databases if necessary. The finished database is QCed, and any necessary revisions are incorporated. Finally, the feature classes are symbolized, annotated, placed in a layout in ArcMap, and exported for final editing. This is an evolving process that is constantly being revised, and as such is open to comments and suggestions from the geologic mapping/GIS community.

ILLUSTRATOR FILE PREPARATION AND EXPORTING

One of the most powerful features of the Adobe Illustrator software package is the ability to store a virtually unlimited amount of vector artwork and text in layers. This permits individual features to be organized by type and the layers to be turned on and off, greatly improving map organization and editorial workflow. When converting an AI file to a GIS-compatible format, object layering in the Illustrator file becomes critical. Using AutoCAD drawing files (DWG) as an intermediary, the attributes of object outline color, line weight, and containing layer are maintained in the attribute table of the CAD feature class. This enables feature attribute querying in GIS software based on those fields and the unique values within them. Accessing the features and their attributes through a GIS, however, is dependent on properly setting up the AI file for export to CAD and the subsequent processing steps during the conversion to a vector GIS format.

The steps to prepare the AI file for export to an AutoCAD drawing (DWG) file are as follows: 1) separate the artwork into layers named by the geologic feature type (for example, inclined bedding, antiform overturned, contact certain, etc.); 2) densify and straighten linework that has been drawn with Bezier curves; 3) export the cleaned-up AI file to a DWG file and examine it for errors in content; and 4) export images (TIFF or JPEG) of the complete map, and/or any other important components, to use as a reference later in the feature attribution process. The following discussion of these steps will detail the most critical actions that lead to successful conversion. Individual results will vary based on the quality and complexity of the input artwork, so testing these steps on a small AI file that is easy to manipulate will be the best way to develop a method that is appropriate for your project.

Layering

Separating the artwork into layers in Illustrator is the first and most critical step in the conversion. At this point, think about how the artwork in the AI file will be stored as features in a geodatabase. If your organization has an existing database schema where feature types, attributes, and feature representations are already defined, the process of organizing the AI file is simplified considerably, because these parameters can be applied to a layer-naming convention in Illustrator rather easily. If this is not the case, careful planning has to be done to ensure that there is consistency (and a level of universality) in the arrangement of the artwork in layers, since the layer name becomes the most critical attribute in the DWG file. In this case, you may need to review the content of several AI files to evaluate the type and scope of features contained in your maps.

Setting up a simple geodatabase to house the results of the conversion is also recommended. It should be structured to hold a minimum set of attributes so features can be identified, symbolized, and manipulated easily and consistently. There are many resources to aid in geodatabase design, namely those available from the ESRI Support website (http://support.esri.com; search "geodatabase design"). Once the feature types or classifications have been established, sorting the features into layers can commence. Illustrator has the capability to select objects based on certain aspects of their appearance, such as stroke and/or fill color, stroke weight, or graphic style. Using these tools, all polygons representing the same rock unit can be selected simultaneously based on their fill color, or all lines representing concealed contacts can be identified by their stroke color and weight, then placed in the appropriate layer, for example. Figures 1, 2, and 3 are examples from a geologic map that was separated into layers in this way. In each figure, the name of the layer where the object resides (e.g., map unit "Omb" in Figure 1) becomes the content of the "Layer" attribute of that object in the exported DWG file. Other attributes from the artwork that are recorded in the attribute table of the exported DWG file are stroke color (indexed value from 1 to 255) and stroke width (in hundredths of millimeters). Unfortunately, fill colors are not recorded in the attribute table of the Polygon class in the DWG file.

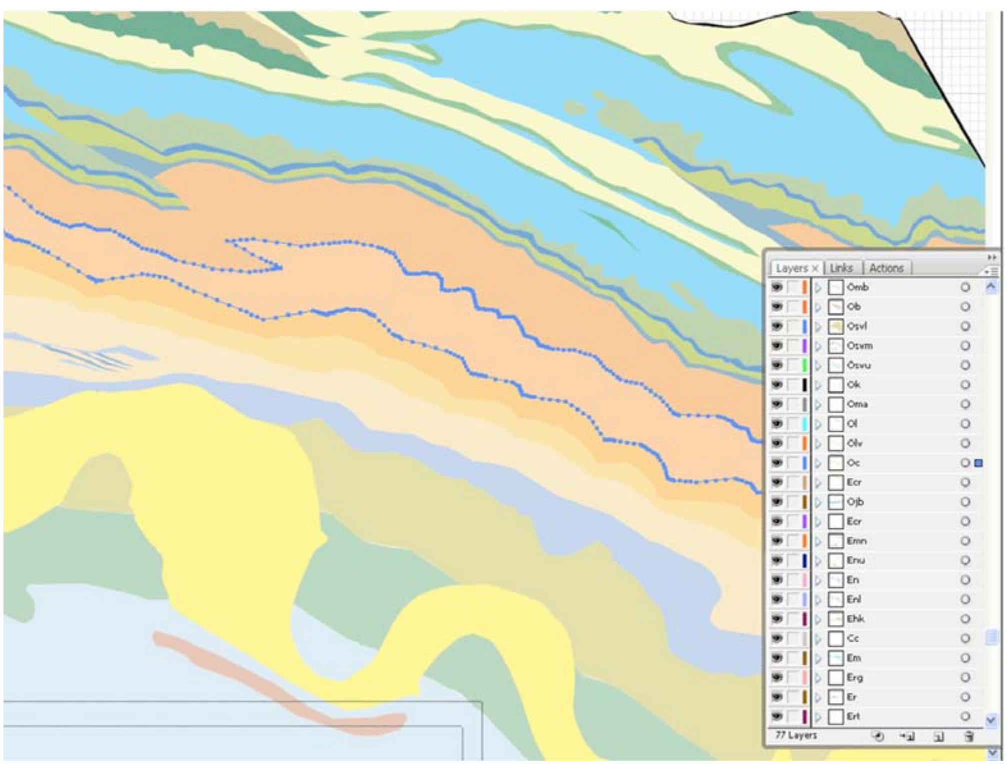

Figure 1. Rock units are separated into layers by unit code in Adobe Illustrator. The "Select > Same > Fill Color" command is very helpful in selecting all of the map areas or polygons for each rock type at one time.

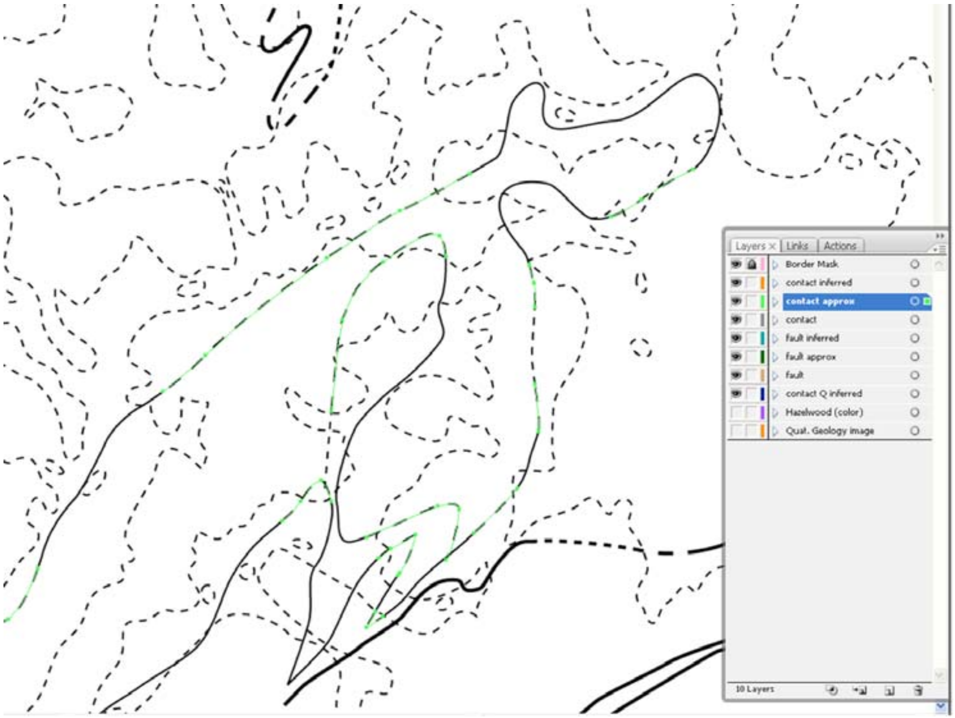

Figure 2. Faults and contacts are separated into layers by type in Adobe Illustrator. The "Select > Same > Fill & Stroke" command is very helpful in selecting lines that share the same dash and weight properties.

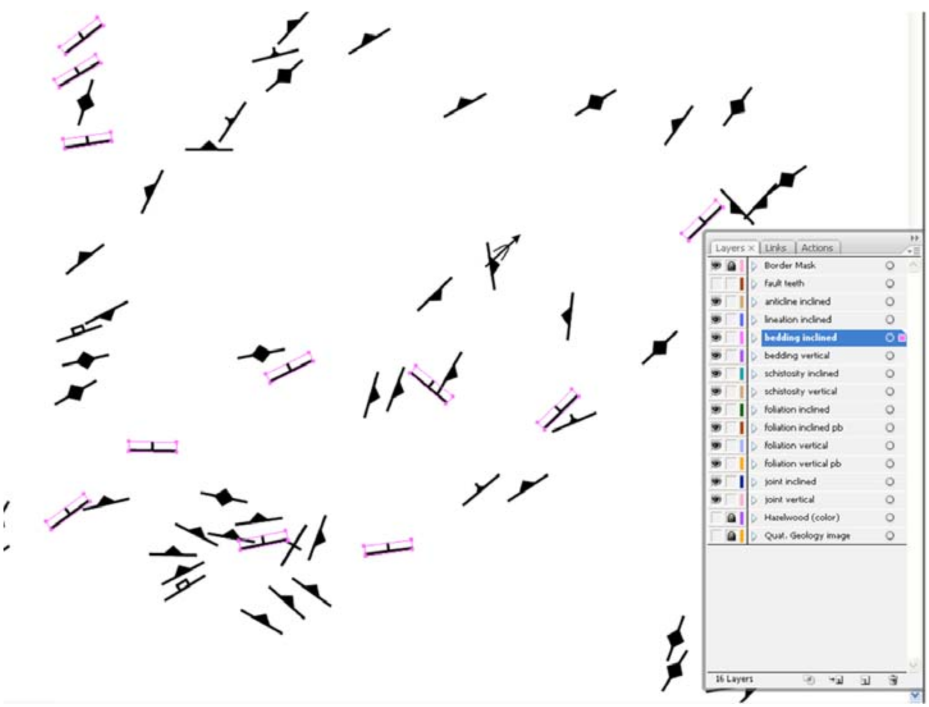

Figure 3. Symbology is separated into layers by type in Adobe Illustrator. If the symbology is built from the Symbol palette, the "Select > Same > Symbol Instance" command can be used to select each type. If the symbols are simply grouped paths, separate them into layers leaving the groups intact.

Another aspect to consider during this process is feature topology. Illustrator does not maintain topological relationships of objects in the same manner as a CAD or GIS system. Points, lines, polygons, text, and other objects can be placed in layers together in Illustrator, and their graphical attributes can be mixed. A classic example of this in Illustrator is the application of a fill to a line that is not closed: it looks and behaves like a polygon, but is not truly so in terms of CAD/GIS feature topology. In CAD and GIS feature classes, topological relationships are explicitly defined and are not interchangeable within that class, i.e., objects in a CAD or GIS feature class can be of only one topological type: point, line, or polygon. Points are a single node or vertex, lines are a set of vertices connected to each other in the order in which they were drawn, and polygons are areas completely enclosed by such a line. This is important to consider when preparing the AI file for export to a CAD drawing, since the Illustrator DWG export engine inspects each object to determine the CAD feature class to which the object will be written, regardless of the layer in which the object resides. Often, objects that look like polygons in Illustrator are just filled lines, and as such will not be written to the Polygon feature class within the DWG file as one might expect. Properly closed paths (polygons) in Illustrator will create an output to both the Line and Polygon CAD classes, because the export engine sees the object as both a path (line) and a closed shape (polygon). If the integrity of polygons is a primary concern during the conversion process, special care will need to be taken to ensure that the shapes are closed properly in Illustrator. Polygons can be recreated in ArcMap (with an ArcEditor or ArcInfo license) from topologically clean lines (lines with no dangles, no self intersections, etc.), so in some cases, such as contact lines that delineate rock unit boundaries, it might be more efficient to ignore the polygons in Illustrator and recreate and attribute them later in ArcMap. Object grouping in Illustrator is also a concern when exporting to a DWG. Grouped artwork will be exported as a composite or "multi-part" shape, which may be topologically undesirable in some cases. Structural symbols (foliations, lineations, fold axes, etc.), on the other hand, may be grouped objects that should be maintained as multi-part features until they can be processed later in ArcMap. Consequently, it is advisable to select and ungroup artwork where topology must be maintained, making sure that the ungrouped objects remain in the appropriate layers.

There are many resources on the web for Illustrator scripts that help clean up and fix artwork. The Swiss Federal Institute of Technology in Zurich (ETHZ) has developed a set of tools for Illustrator that can be helpful in some of these tasks. They are available free of charge from their website (http://www.ika.ethz.ch/plugins/index.html) with one caveat: the tool dialogs are written in German (although there is an English version of the documentation available, along with some helpful examples). For more information about topology in ArcGIS, see the ESRI support website (http://support.esri.com) and search with keyword "topology".

Preparing Artwork for Export

Once the artwork has been cleaned up and separated into layers, it can be readied for export from Illustrator to a DWG file. In this process, the artwork in the AI file is visually assessed and densified appropriately, then straightened to remove any Bezier curves. It is necessary to do this when exporting DWG files for use in ArcGIS, because ArcGIS does not natively support the viewing/importing of CAD curves, which are known as "splines"; it simply ignores objects that have an Entity value of "Spline" in the CAD attribute table. If you have access to AutoCAD or another CAD software package that can fully read/convert DWG files to Shapefiles, the process of densifying and straightening the linework is not necessary in Illustrator, because most of the professional CAD software export engines will do this automatically. During development of the AI to GIS conversion process, we experimented with several other conversion options and software packages. While it is possible to export artwork as a DWG or DXF (AutoCAD drawing exchange format) with the Bezier curves preserved as splines, the subsequent export to a shapefile did not always provide satisfactory results. In some cases the attribute tables in the exported shapefiles did not contain all the original DWG attributes, and in others the density of vertices was not sufficient to maintain the smooth shapes of the original Bezier curves. For these reasons, and the fact that the conversion *can* be accomplished successfully using only Illustrator and ArcGIS, the process outlined here uses the densification and straightening method in Illustrator.

Inspect the artwork in Illustrator and get a sense for the relative density of vertices among the objects in each layer. Artwork may have very different density of vertices or drawing styles, and not all the layers or objects need to be densified to the same extent. For instance, a 7.5-minute quadrangle boundary line is probably a path composed of only four "corner" vertices and does not need to be densified, whereas rock unit contacts are drawn with smooth Bezier curves and will need to be densified several times. Once the layers that need densification are identified, the number of iterations of the "Object > Path > Add Anchor Points" command can be determined. In most cases, 3 to 5 iterations of the command is sufficient. The tool is "dumb" in that it does not inspect the distances between vertices to determine the need for an additional point. The command simply places an anchor point half-way between each existing anchor. This has the effect of making some areas far too dense and others a bit sparse. This is usually not a problem; once the linework is converted to a GIS format, the lines can be generalized with a displacement tolerance that will remove the extraneous points while maintaining the integrity and smoothness of the original artwork. When in doubt, densify again to ensure that there are enough vertices in the sparse areas. If polygons were created from linework using the "Pathfinder > Divide" tool in Illustrator, densify the lines and the polygons the same number of times to ensure they maintain the same density. When each layer has been densified, all the artwork that will be exported to the DWG file should be selected and straightened using the "Object > Path > Simplify…" command. Use the "Straight Lines" option to remove all Bezier curves (Figures 4 and 5).

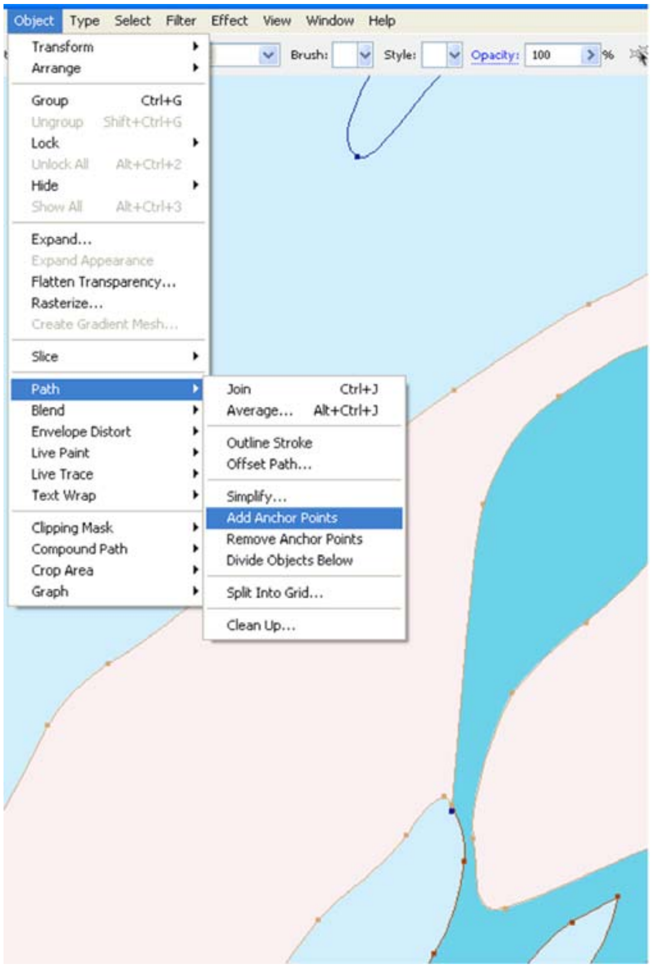

Figure 4. Use the "Object > Path > Add Anchor Points" command 3 to 5 times to densify linework that has been drawn with Bezier curves, to preserve the shapes as closely as possible.

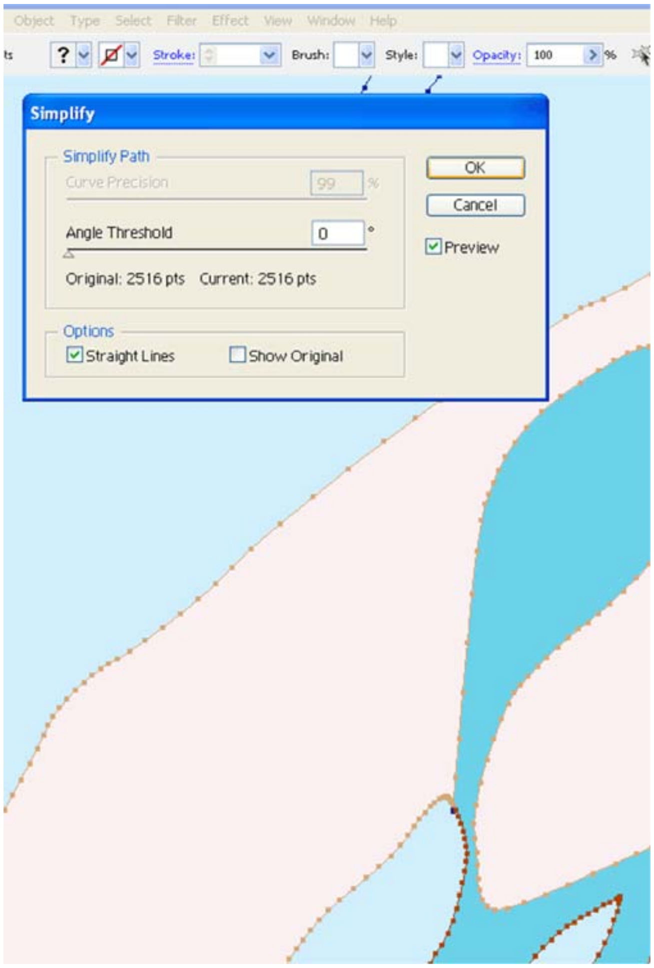

Figure 5. The "Object > Path > Simplify…" command dialog has options to smooth and generalize lines. Checking the "Straight Lines" option removes Bezier handles and readies artwork for export to an AutoCAD DWG file (without splines).

Export to AutoCAD DWG

After preparing the artwork, it can be exported to the DWG file. When a DWG file is exported from Illustrator, objects are written to the output file in the order in which they were drawn (or subsequently reordered) in the layer, from the lowest layer in the stack to the highest; i.e., the bottommost object (first to be drawn) in the bottommost layer is written first, then the next object in the bottommost layer, and so on, layer by layer. With points and lines, the export order is of little consequence, but with polygons, this can have the effect of one polygon obscuring another when previewed in ArcCatalog or when loaded into ArcMap and viewed with the default color fill symbology. This may give the impression that there are polygons missing or that the export did not work correctly when in actuality the polygons are just obscured. The most common example of this phenomenon is a map border polygon stored in an upper layer of the map obscuring polygons, such as rock units, drawn in lower layers. As a general rule, move the border or mask layer to the bottom of the layer stack in Illustrator prior to export so that it will be drawn before the polygons of other layers when viewed in ArcCatalog or ArcMap.

Exporting to DWG is usually fast, depending on the complexity of the artwork being exported. One way to speed the export is to process only a selection of artwork. Unlock all the layers that have artwork to be exported. Use the "Select > All" command or Ctrl+A to select all the artwork in the unlocked layers. Next, navigate to the "File > Export…" command, name the output in the Export dialog box, and click "Save". In the DXF/DWG Options dialog, choose the lowest version of DWG file available (R13/LT95) and check the option to "Export Selected Art Only". The remaining default export settings can be accepted (Figure 6). In testing, the version of AutoCAD drawing did not seem to matter, but the assumption that older formats are generally simpler and thus more interchangeable was the motivation for using the oldest version available. If you want the DWG files to be used in or exported to shapefiles from AutoCAD, export from Illustrator using the file version that matches your release. To better organize the different types of features being exported from Illustrator and to help make the content checking easier, export several DWG files, one for each type or group of features. For a typical geologic map, that means exporting three files: one for map symbology, one for linework, and one for polygons.

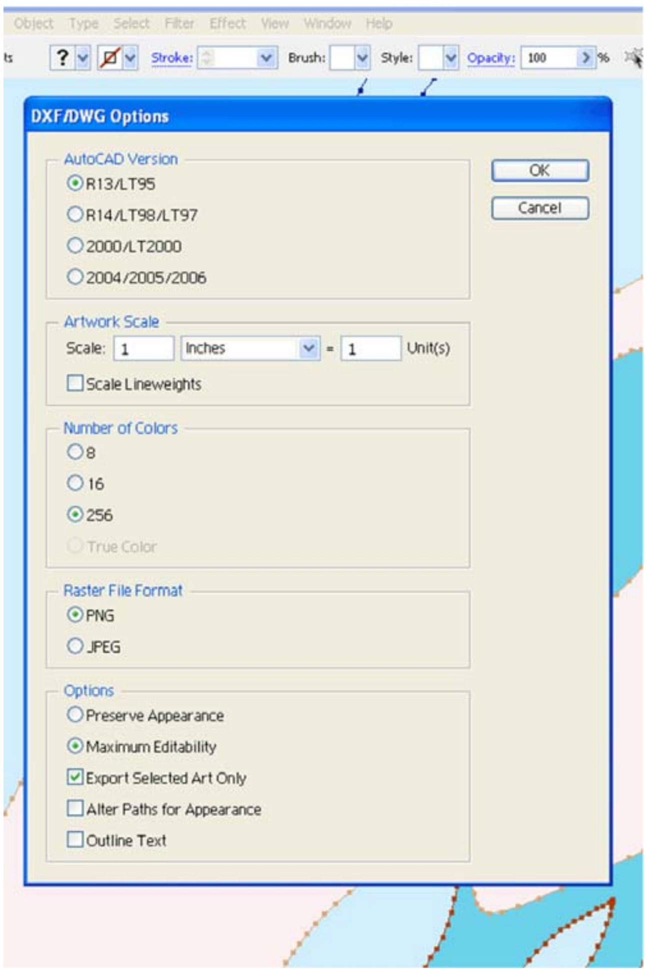

Figure 6. Adobe Illustrator DXF/DWG export options dialog. Use the lowest available file version and choose to export only the selected artwork to maximize export efficiency. All other options can remain the default values.

This is also the point at which a graphic of the complete map or individual layers (station labels, symbology, and dip-strike data, etc.) can be exported as an image to be georeferenced in ArcMap and used as a guide to check the DWG file for errors or omissions, and for feature attribution. Generally, a 300 dpi TIFF image is ideal, but there may be instances where the combination of artwork size and resolution exceeds Illustrator's capacity to write the output file. If so, ensure there are no extraneous points or text beyond the expected outer extent of the content that is to be exported, as this is the leading cause of (unexpected) failure of the TIFF export. Also, turn off any unnecessary layers. If the export continues to fail after removal of unnecessary artwork and/or turning off layers, uncheck the "Anti-Alias" option if it is not already, since this also reduces the complexity of the output. As a last resort, reduce the output dpi a small increment, e.g., 250 instead of 300 dpi, until the export succeeds (Figure 7). This completes the Illustrator portion of the conversion process.

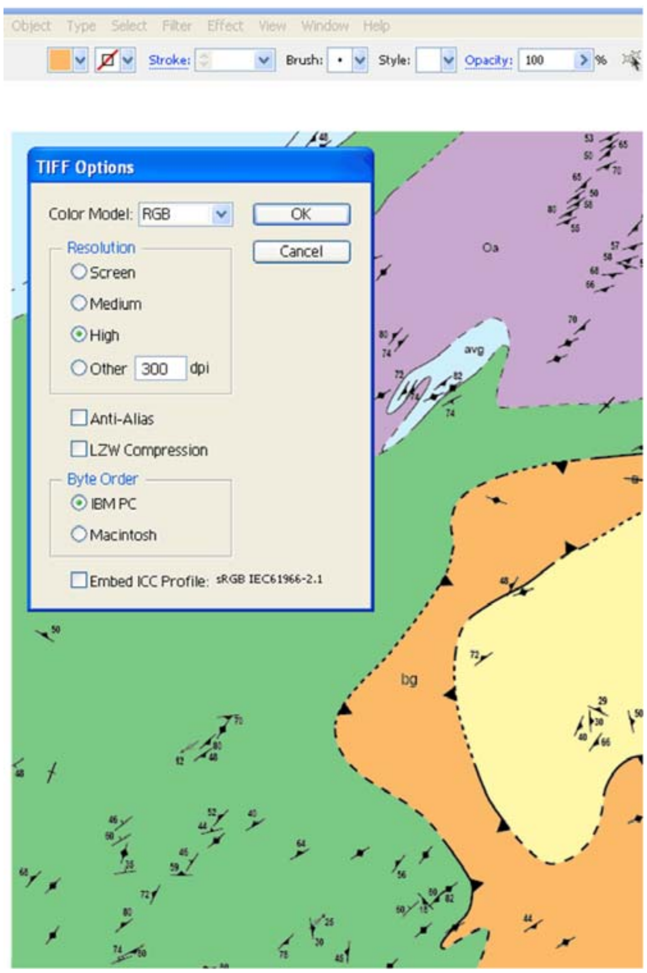

Figure 7. Adobe Illustrator TIFF export options dialog. Use RGB color mode and if necessary turn off the "Anti-Alias" option to reduce the complexity of the export while keeping the resolution high in order to maintain legibility in the image. Optionally, enable LZW compression to reduce file size.

DWG FILES TO FEATURE CLASSES

When the DWG export is complete, the contents of the file can be viewed in ArcCatalog (or ArcMap). Check the Polygon and Polyline classes, as appropriate, for content (Figure 8). If there are features missing, inspect them in the original Illustrator file to make sure that they were straightened. Occasionally, there will be one or more features that will not export properly. Make note of these and use the reference imagery as a guide to digitize these features into the database at some later time. When the exported DWG files have been sufficiently vetted, the process of putting the features into feature classes can begin.

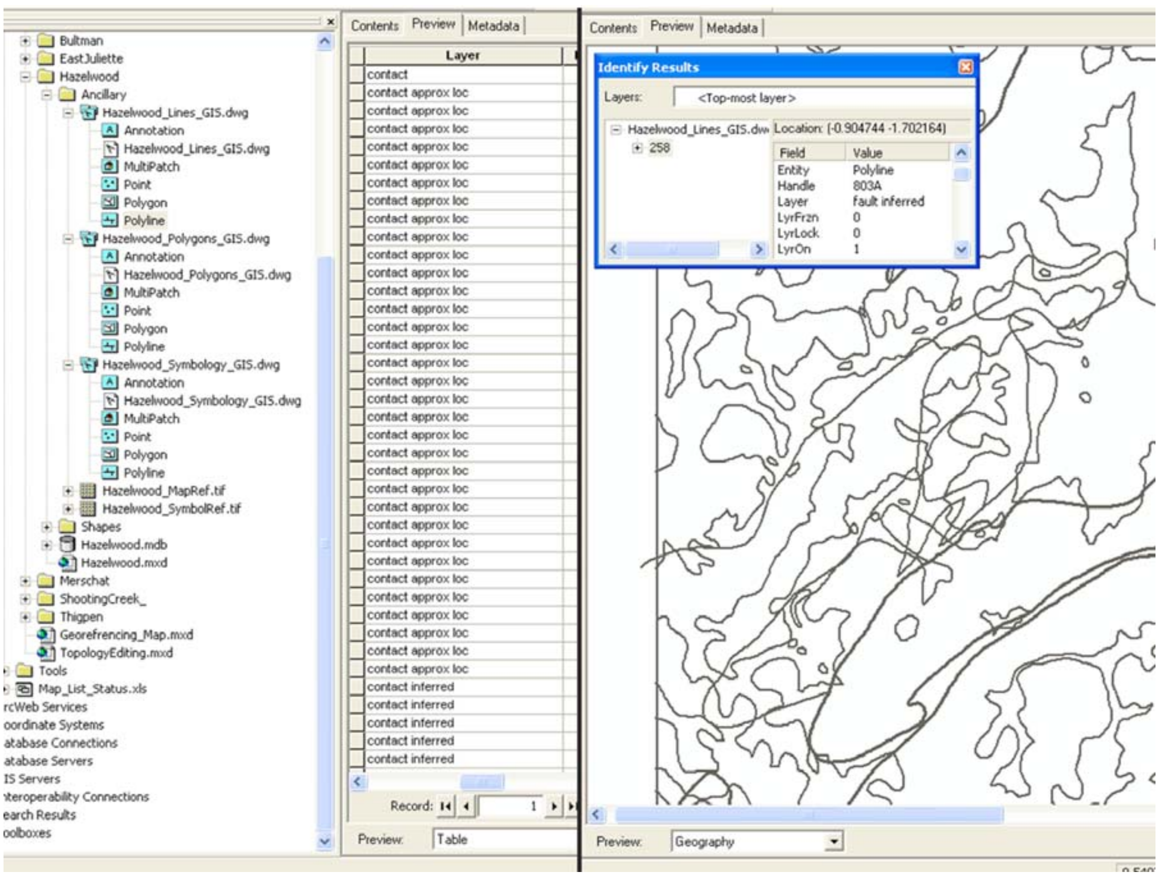

Figure 8. ArcCatalog is used to inspect the contents of the DWG file. Check the Polygon and Polyline classes of each DWG file for content by selecting the class and previewing the attribute table. The Identify tool can also be used to check the properties of individual features.

DWG Conversion to Shapefile

AutoCAD DWG files cannot be edited directly in ArcGIS; in order to manipulate the features, the CAD feature classes will need to be stored in an ArcGIS-editable format.

When we first developed this process, the preferred method was exporting to shapefiles, spatially adjusting the shapefiles, and appending them to a geodatabase. Since that time, more efficient methods have been established to get the classes directly into a geodatabase. But the shapefile, which for many organizations may be a staple of their geospatial data storage structure, is a good choice for storing and sharing GIS datasets because it is a format that is widely accepted and read/written by many CAD and GIS software packages. For these reasons, the process to convert and spatially adjust the CAD feature classes as shapefiles will be outlined here.

There are two common ways to export CAD feature classes to shapefiles: one uses ArcMap and the other ArcCatalog. The options and dialogs for these two methods are quite different. In ArcMap, export options, aside from the output location and file name, are virtually non-existent, so this method is not recommend. ArcCatalog utilizes the “Feature Class to Feature Class” ArcToolbox script, which is more advanced and has more robust options for the output. Of particular importance are the Field Map and SQL query, which allow the user to export selected fields and selected features, respectively. To initiate the export to shapefile in ArcCatalog, simply right-click the appropriate CAD feature class and choose “Export > To Shapefile (single)…” (Figure 9). Once the dialog appears, set the output location and file name for the shapefile. A SQL query can be created in the “Expression” field to export only a selection of features. The bottom pane of the dialog window is the Field Map, where the user can select the attribute table fields and their names in the output. This is the great advantage of using the Feature Class to Feature Class tool in ArcCatalog to export the shapefiles. Unnecessary fields in the CAD feature class can be dropped from the output, and fields that are to be kept can be renamed and re-dimensioned to match existing schemas (e.g., “Layer” [*type* String, *length* 255] can be renamed to “Label” [*type* String, *length* 50]). Generally, the only field that is kept is the “Layer” field, which is renamed and dimensioned as in the previous example.

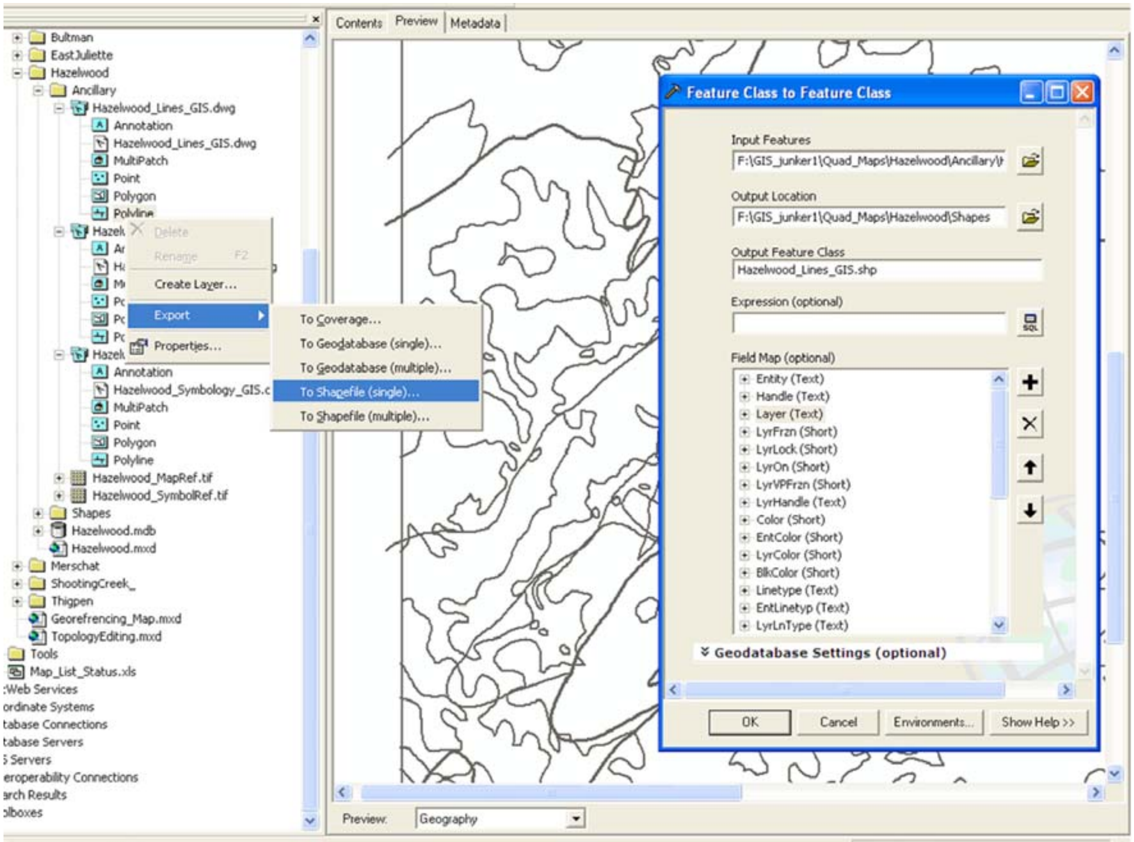

Figure 9. ArcCatalog is used to export the appropriate class from each DWG file to a shapefile via the “Feature Class to Feature Class” tool. This dialog allows the user to set the output location and file name, as well as specify a SQL query and remap and/or re-dimension the fields from the input DWG attribute table to the output shapefile attribute table.

Before viewing and editing the exported shapefiles in ArcMap, check them for structural problems by running the “Repair Geometry” ArcToolbox script. The script can be accessed through the ArcToolbox “Data Management Tools > Features” toolset. The purpose of running this tool is to repair invalid feature representations in the shapefile structure that may exist after export from CAD. The most common problems fixed by the script are null or empty geometry, self-intersections, and improperly sorted rings in polygons (“donut” polygons with inner and outer rings drawn in opposite directions). Errors of this type are usually not visible and do not affect map display. These errors usually manifest themselves when you attempt to edit or perform geoprocessing functions on the shapefiles. Fixing these errors helps ensure proper behavior of the features, whether in the shapefile or in a geodatabase. Also, execute the “Simplify Line” or “Simplify Polygon” tool (whichever is appropriate) located in the “Data Management Tools > Generalization” toolset. These scripts help clean up the shapefiles by removing small “hiccups” in the features and extraneous points that were added during the densification process in Illustrator. The dialogs for both scripts are basically the same, with the option to set a maximum allowable offset for points in the output. The polygon version adds the option to set a minimum area for the simplified shapes. Generally, a small offset around 0.1 m (0.3 ft) is enough to allow removal of most of the extra points while maintaining the shapes. Some errors in feature topology that are introduced in the process are fixed by the tool itself, while others can be fixed later using ArcGIS Topology tools. If simplifying a line class and a polygon class that have shared geometry (i.e., the polygons were created from lines using the "Pathfinder > Divide" tool in Illustrator), be sure to simplify both classes using the same tolerance. This will help reduce the number of errors later when validating the feature topology.

Spatial Adjustment and Append to Geodatabase

Once the shapefiles have been exported they can be loaded into ArcMap and spatially adjusted. If you are familiar with how to georeference imagery in ArcMap, the process of spatial adjustment is very similar. By adding control points that link the coordinates in the shapefiles (which maintain the CAD file coordinate system) to a properly projected grid or "footprint" from another feature class that matches the extent of the original map, the contents of the shapefile "snap" to the new, projected coordinates. Once the features have been snapped, the projection of the shapefile can be defined in ArcCatalog. It is important to remember that spatial adjustment only adjusts the coordinate space and updates the extent of the features; it *does not* set the projection. USGS 7.5-minute quadrangle map-based features are the easiest to spatially adjust, as the USGS quadrangle index grid can be obtained in shapefile format free of charge from a number of web sources (e.g., GeoCommunity- http://data.geocomm.com/quadindex). Most of the maps that we have converted have been partially or fully quadrangle-based, so most of the spatial adjustments are as easy as linking the four corners of the quadrangle feature in the shapefile to the appropriately projected quadrangle polygon (Figure 10). For more information on the spatial adjustment tools, see the ArcGIS Desktop Help.

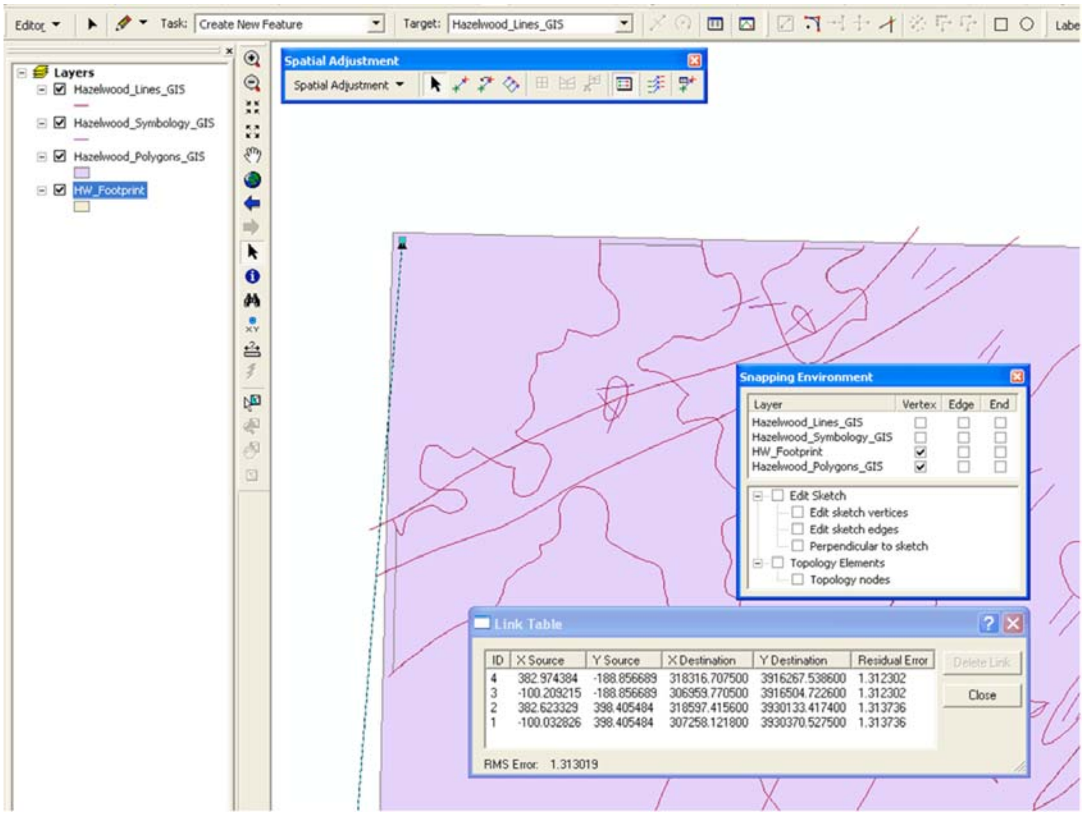

Figure 10. Exported shapefiles are loaded into ArcMap and spatially adjusted by snapping the corners of the "footprint" feature in the shapefile to the corners of a projected quadrangle boundary feature already stored in the geodatabase. Note the small root mean square (RMS) error in meters. If the error is more than a few meters, verify that the footprint shape is correct.

With the shapefiles properly spatially referenced in ArcMap and their projections defined in ArcCatalog, the features can be loaded into the appropriate classes in the geodatabase, or used as is. We prefer to load the adjusted shapefiles into a geodatabase to take advantage of the organizational and editorial advantages of the geodatabase, such as the Topology tools. Use the Append tool located in the "Data Management Tools > General" toolset to append features to the appropriate classes in a geodatabase. Generally, there should be at least four feature classes in a geologic feature geodatabase: 1) a line or polygon class that defines the map extent ("footprint"); 2) a polygon class to store rock units; 3) a line class to store faults and contacts; and 4) a point class to store stations with structural and lithologic information. Additional classes can be added to store surficial deposit polygons, cross-section lines, annotation, and other cartographic or geologic features. The Append tool has a dialog similar to that of the Feature Class to Feature Class tool, and has the ability to map the fields from the input to the output when you select the "NO_TEST" Schema Type option (Figure 11).

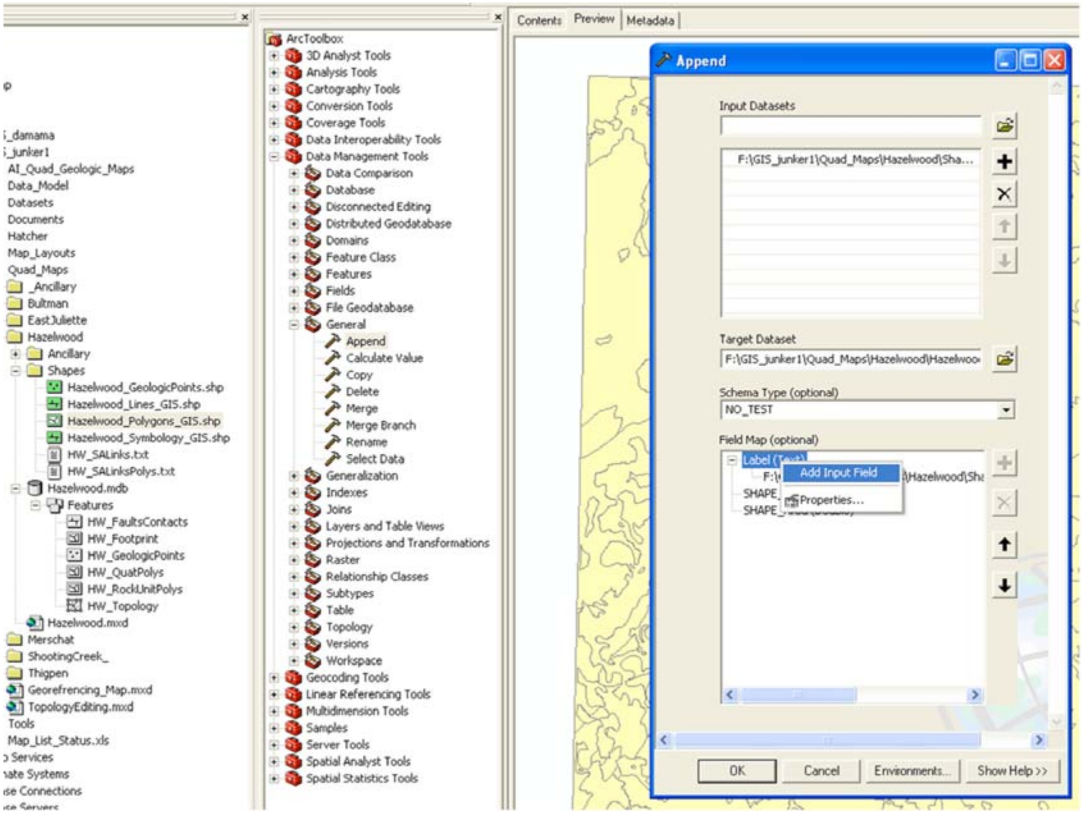

Figure 11. The Append tool can be used to add the features from the spatially adjusted shapefiles to the appropriate class in the geodatabase. Use the "NO_TEST" schema option to map fields from the shapefile attribute table to the geodatabase feature class attribute table.

Geologic Symbology Processing

If one of the adjusted shapefiles contains features that represent geologic structural symbology, it will need additional processing to convert it to point features in a geodatabase feature class or shapefile, with appropriate attribution. This step in the conversion process is the least certain in terms of the specific tasks needed to get to a finished product. Depending on the nature of the symbology that was used in Illustrator, the conversion can take a number of different paths. Ideally, the original map in Illustrator has a layer that contains the station locations, labeled with the station ID, as well as a stand-alone table (e.g., an Excel spreadsheet or a tabular text file) that has the rest of the station information based on that same ID. These station "points", which probably exist as small polygons (circles or squares) in the symbology shapefile, can easily be converted to point features and then manually attributed with their station ID from a background reference image in ArcMap. The station point feature class' attribute table can then be joined on the station ID to the table containing the additional station data. With the tables joined, the attributes from the stand-alone table can be referenced to define the symbol type, color, text label, etc., for the ArcMap layer, to be used in the finished map. This is, of course, the ideal situation. For many older paper or Mylar maps that have been redrafted in Illustrator, it is usually not possible, because the station data exists only in a field book and the stations were marked on a paper or Mylar map. If this is the case, some additional scanning, georeferencing, and digitizing in ArcMap, and transcription of field book data may be necessary to bring the station features up to par with their newer, digitally captured counterparts.

Most of the AI files that were created by our group use a relatively consistent symbol set that was propagated in a template file from map to map as each old paper map was redrafted, or as new maps were drawn in Illustrator. The symbols were constructed of simple lines and polygons, and with some clever querying and attribution in ArcMap, the meaningful parts of the symbol could be isolated and converted to points with a couple of basic attributes. First, start an edit session in ArcMap and select all the features in the symbology class and "explode" the multi-part features so that each part of the symbol can be selected individually. If the features are in a shapefile, the "Length" field will have to be added to the table manually and the values calculated using the Field Calculator or Calculate Geometry command in the table view. The basic process of flagging features to retain entails sorting through the features and identifying ways to isolate and attribute different parts of the symbology (Figures 12 through 17). In this example, the "Length" attribute is used to help separate the symbol parts (Figure 12), and an attribute called "Keep" is created to flag the selections in the table. After calculating the "Keep" attribute (Figure 13), systematically pan around the map searching for problems, correcting the "Keep" attribute as necessary. Once all the features have been inspected, features with "Keep" = 1 are selected and an ArcMap Field Calculator expression is used to determine the azimuthal rotation (strike) of the feature (Figure 14). The "keeper" features are now ready to be converted to points and appended to the geodatabase (Figure 15). Note that some of the points may need adjustment in order to be located accurately. A custom tool for ArcMap called "Attribute Features" that was developed with Visual Basic for Applications is used to assign the Dip (or Station) value to each feature using the reference imagery that was exported from Illustrator as a guide (Figure 16). Symbology is applied to the points by importing an ArcMap Layer file (e.g., GeologicPoints.lyr) through the Symbology tab in the layer Properties dialog. Finally, systematically pan around the map again looking for features that have the wrong symbol, inverted rotation, etc., and make corrections (Figure 17).

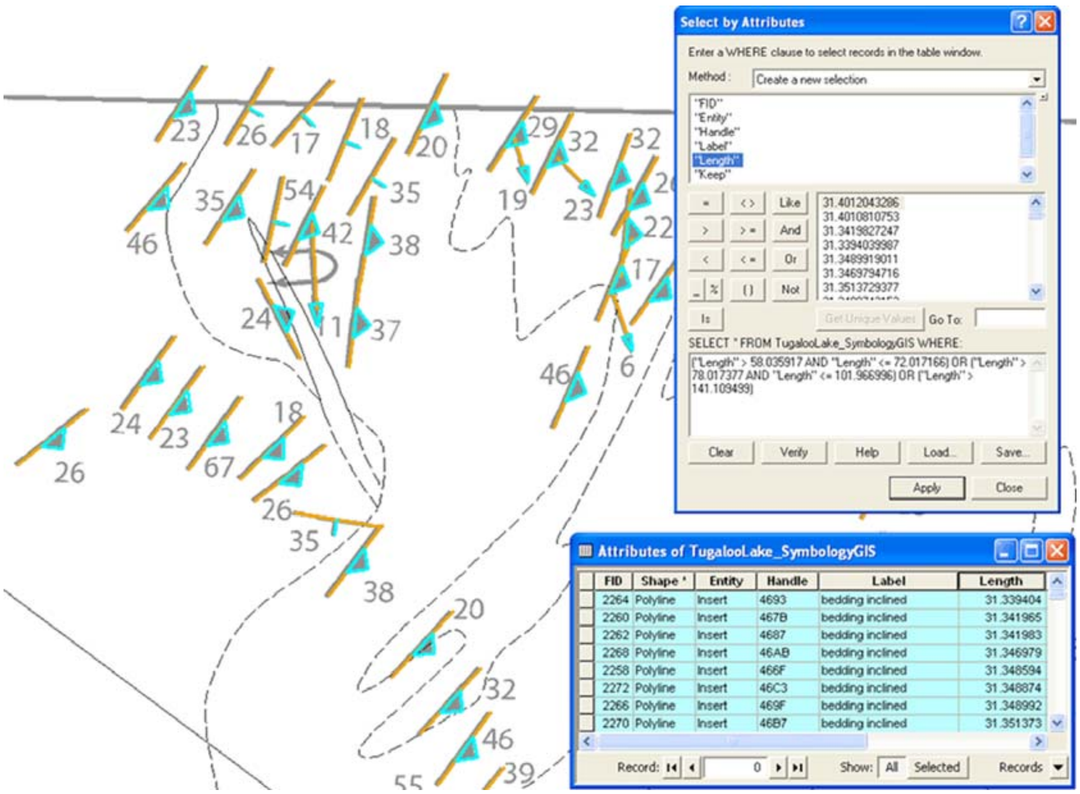

Figure 12. The lengths of the different parts of the symbology features are assessed and a query is developed to select the unwanted parts. Systematically pan around the map to make sure that the query is selecting the appropriate features before continuing.

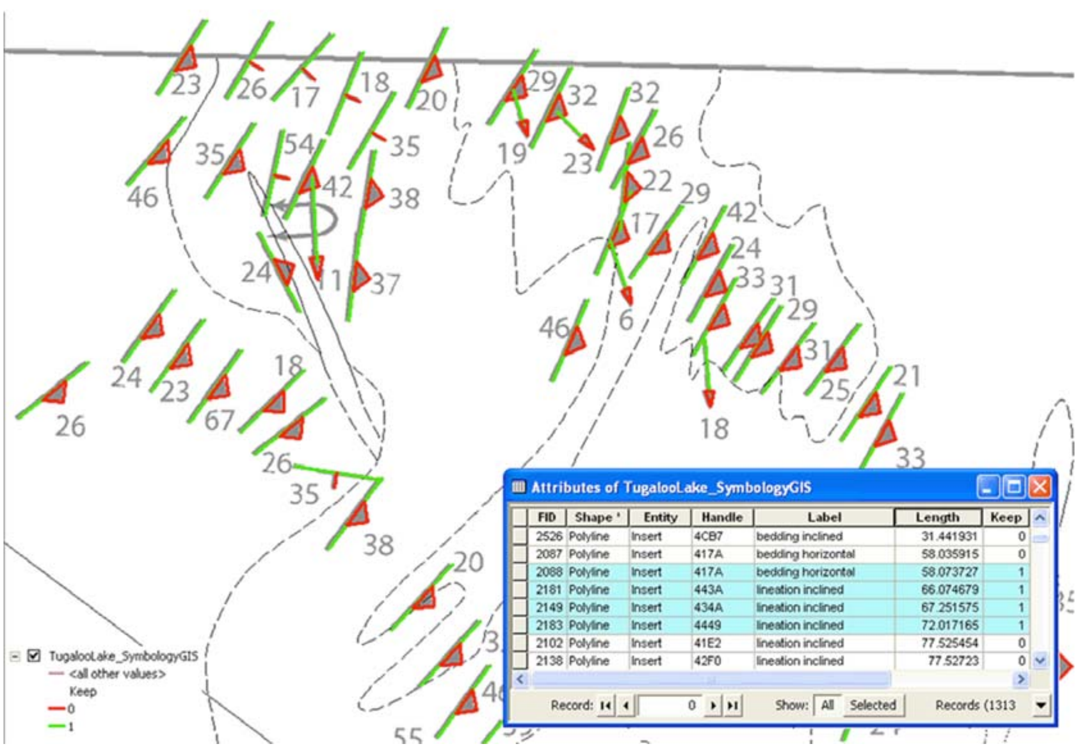

Figure 13. The selection made in the previous step is inverted and the “Keep” attribute is calculated to “1”.

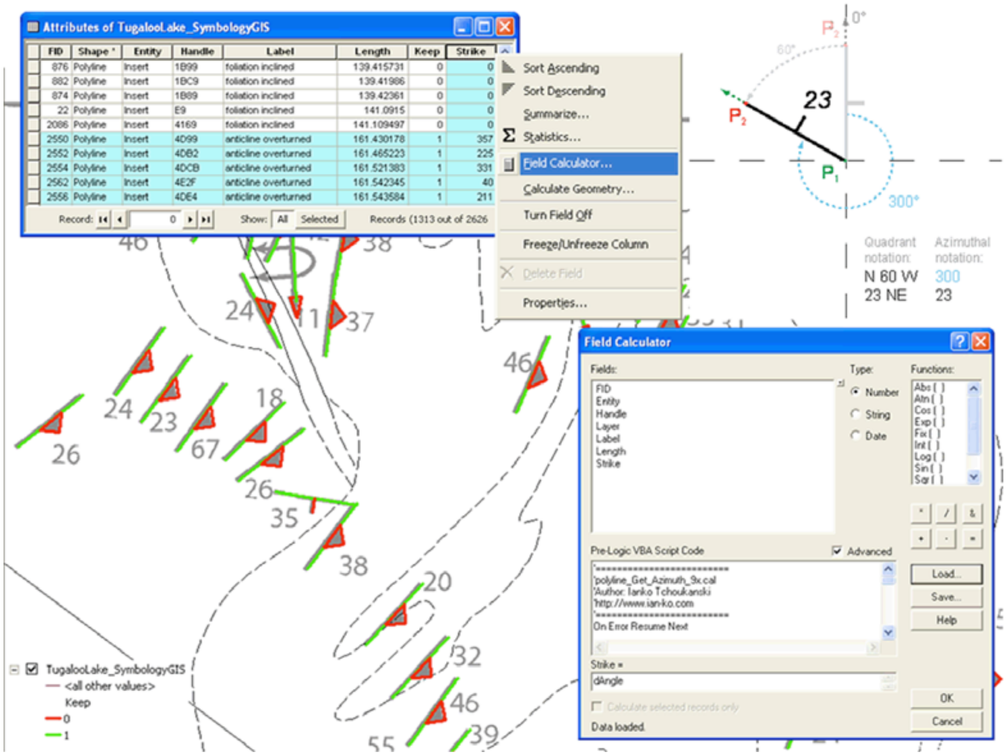

Figure 14. Features with a “Keep” attribute equal to “1” are selected and the azimuthal rotation (strike) of the feature is calculated using a custom Field Calculator expression.

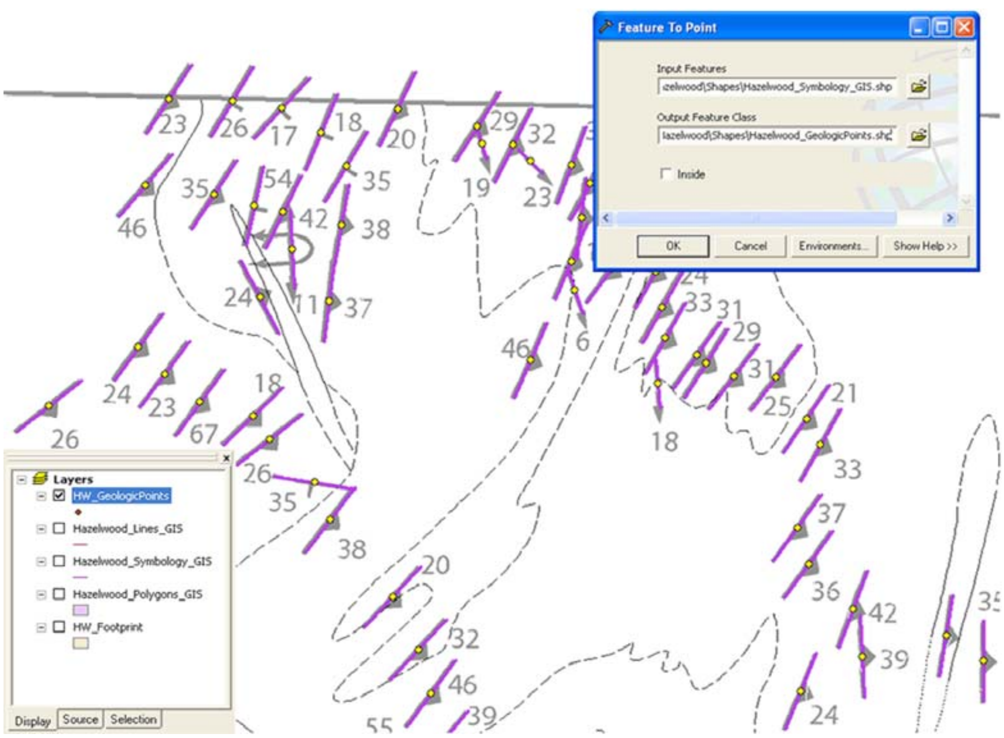

Figure 15. The “keeper” features are now ready to be converted to points using the “Feature To Point” tool. The points can be appended to the geodatabase and their locations edited if necessary.

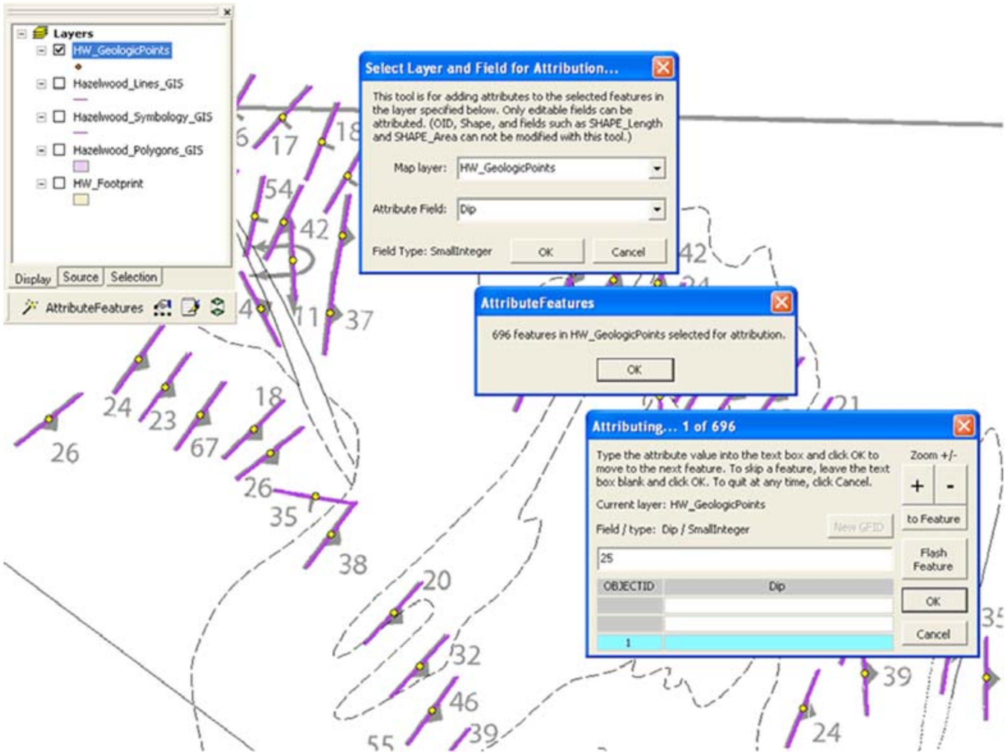

Figure 16. Attributes are added to each feature in turn using the custom "Attribute Features" tool based on the reference imagery that was exported from Illustrator.

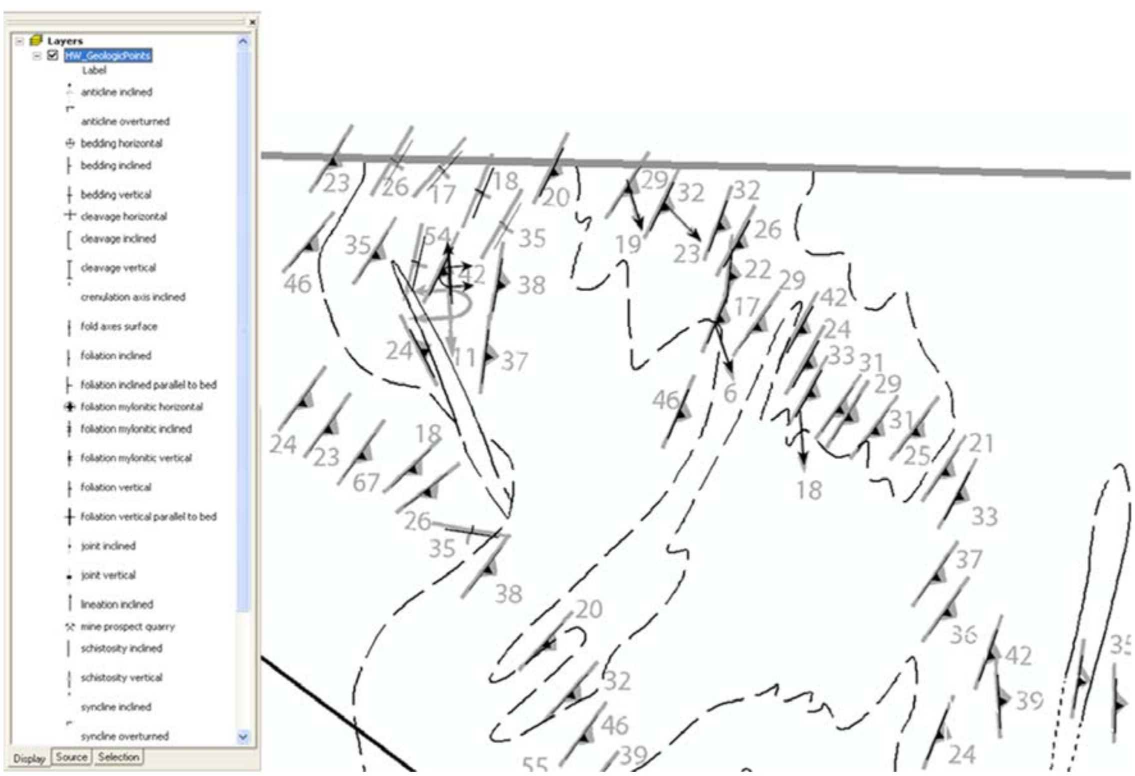

Figure 17. Symbology is applied to the points by importing an ArcMap Layer file. Errors are easily spotted and corrected. (The slight offset of the new symbology is due to incorrect offsets in the layer file, which can be easily corrected.)

TOPOLOGY, EDGE MATCHING, QUALITY CONTROL, AND FINAL STEPS

Once all of the features have been loaded into classes in the geodatabase, some of the tools available exclusively to the geodatabase can be utilized to improve the integrity and quality of the data. The following sections will briefly cover some of the most important data integrity and quality control (QC) measures that can be implemented within the geologic geodatabase, as well as general quality controls that are necessary to assure that the converted maps are as good as possible.

Topology and Edge matching

Within the geodatabase, features in different classes can be related to one another by their geometric relationships using a special type of relationship class called a Topology class. In previous versions of ArcInfo, the coverage data model had topology rules built in to help maintain feature structure validity, tolerances, and the relationships between different geometries. Shapefiles, on the other hand, do not utilize internal feature topology, because they store only one geometry type per dataset (e.g., point, line, polygon). This is one major advantage of the geodatabase; within a feature dataset, topological relationships can be defined and customized to help maintain the integrity of individual features within a feature class, as well as maintain geometric relationships between features in different classes. For more information on how topology works within the geodatabase and how to set up a topology class and define its properties, see the ArcGIS Desktop Help and search with the keywords “Topology in ArcGIS”.

For geologic features, there is a simple set of topology rules that can be implemented to help correct errors that may exist in the converted CAD features, as well as maintain the geometric relationships necessary to ease editing and creation of new features at some later time. In general, the two classes that need to participate in the geodatabase topology are the rock unit polygon and the fault and contact line classes. The rock unit polygons need to be discrete and have their edges covered by a fault or contact. The faults and contacts must also be discrete, and they should not overlap or intersect each other or themselves. Table 1 and Table 2 list a basic set of topology rules for the polygon and line classes, respectively, that are necessary to maintain these essential geometric relationships. Once the rules are established, the topology can be validated. Make a copy of the database *before* validating a topology class, because the validation process makes edits to the features and thus will occasionally produce undesired results that are *not* reversible. Once validated, the topology errors can be assessed and corrected in an ArcMap edit session, or marked as exceptions.

Table 1. List of basic topology rules for polygons representing rock units.

Feature Class	Rule	Feature Class
RockUnitPolys	Must Not Overlap	n/a
RockUnitPolys	Must Not Have Gaps	n/a
RockUnitPolys	Boundary Must Be Covered By	FaultsContacts

Table 2. List of basic topology rules for lines representing faults and contacts.

Feature Class	Rule	Feature Class
FaultsContacts	Must Not Overlap	n/a
FaultsContacts	Must Not Intersect	n/a
FaultsContacts	Must Not Have Dangles	n/a
FaultsContacts	Must Not Self-Overlap	n/a
FaultsContacts	Must Not Self-Intersect	n/a
FaultsContacts	Must Be Single Part	n/a

If the maps being converted from AI to geodatabase cover adjoining quads, edge matching must be done to ensure that the rock units and contacts line up properly and have consistent attribution. Add the adjoining rock unit polygon and fault and contact line classes to ArcMap and inspect the shared edge. In an edit session, the Snapping environment can be set up so that the vertices of features from one quad can be snapped to the vertices of the features in the other (Figure 18). Once the adjoining maps have been edge matched and the topology checks and edits are completed, the geodatabase spatial QC is complete.

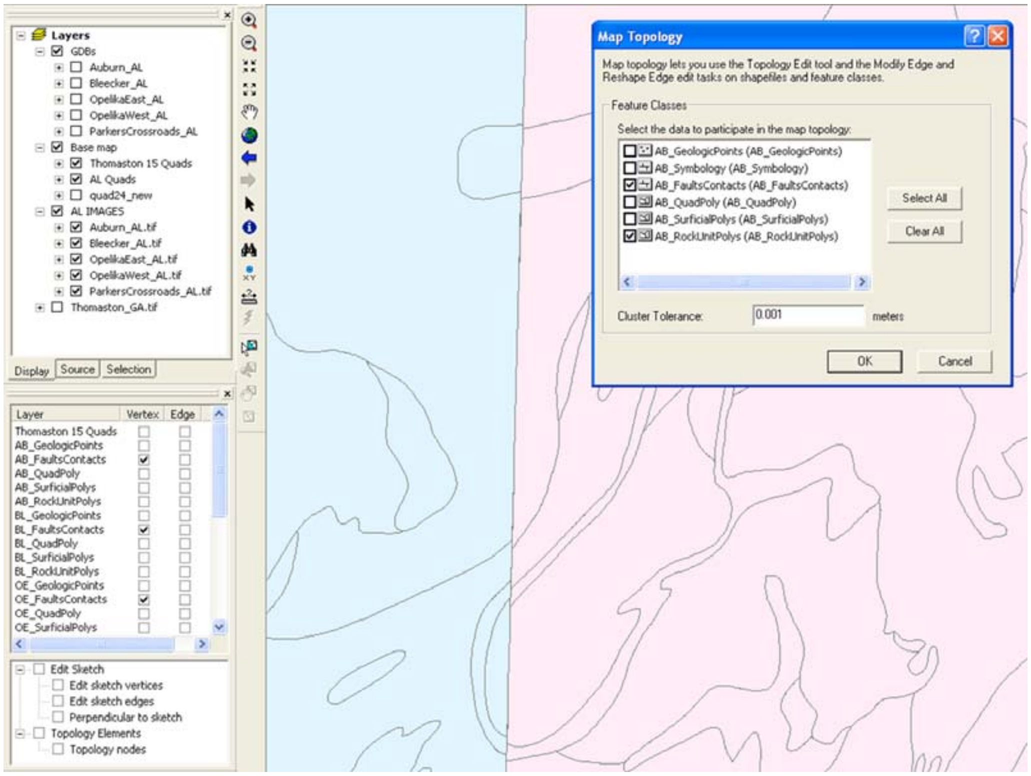

Figure 18. Adjoining maps should have their edges matched. In an edit session, use the Snapping environment (bottom left) to set the features that will take precedence over others and use Map Topology to specify classes with shared geometry.

Other Quality Controls and Final Steps

With the conversion complete, and the geodatabase spatially QCed, any additional QC or final cleanup can be done to finish the process. Generally, there will be some additional work to make sure that all the data have been converted properly, as well as some preparation of the data to be displayed cartographically. Some other QC steps to consider are: 1) cleaning up any <null> or erroneous values in the feature class attribute tables; 2) checking the original source maps and field books or spreadsheets to verify station attributes; 3) cleaning up the workspace, including deleting ancillary or temporary files and compacting the geodatabase in ArcCatalog; and 4) preparing metadata documents for the geodatabase.

For data that are to be displayed on a printed or electronic map, additional feature classes may need to be added to store cartographic features, such as cross-section lines and lines for holding labels or continuous graphical elements like fault teeth. Labels can be converted to annotation and their locations edited and stored permanently in the geodatabase. All the geodatabase feature classes can be added as layers in ArcMap and placed in a layout with appropriate symbology, annotations, a legend, and rock unit descriptions. ArcMap layer files can be applied to layers to keep fills, lines, and points symbolized and labeled consistently and to exact specifications. A topographic base, shaded relief image, photographs, etc. can be added to the layout to enhance the map. See Figure 19. Finally, the map can be exported to any number of different graphics formats to be printed or viewed electronically, such as an Adobe PDF document or a TIFF or JPEG image.

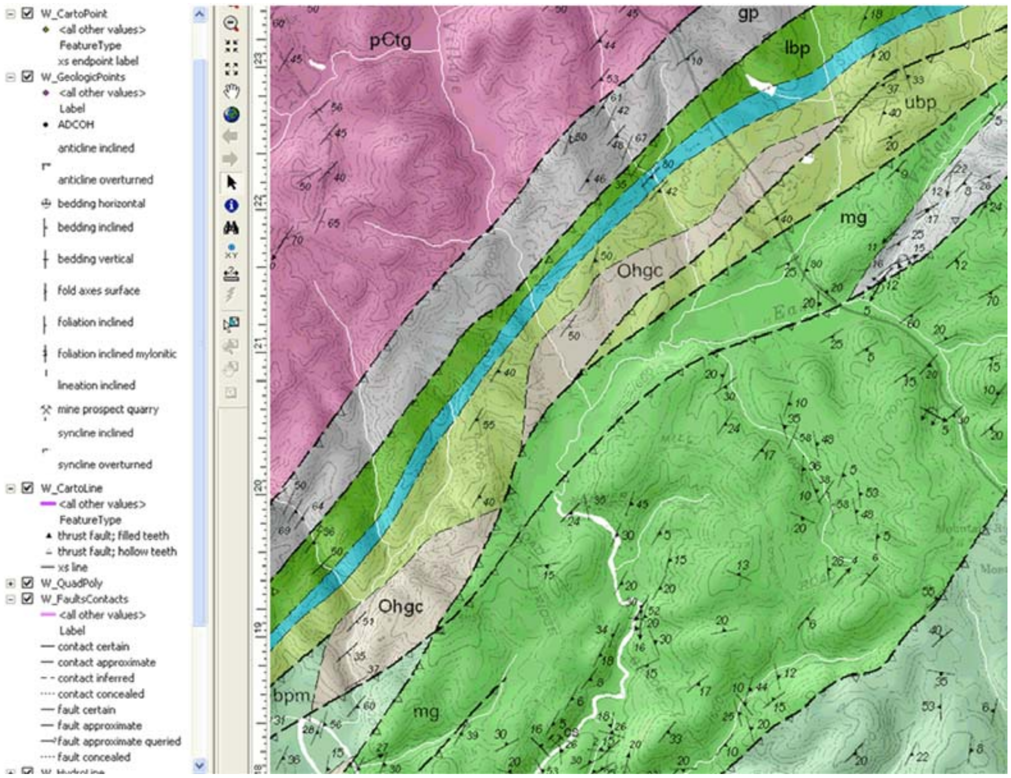

Figure 19. Sample geologic map with proper symbology and a hillshade image, in ArcMap.

CONCLUSION

This paper presents a set of steps and processes for converting maps stored in the popular Adobe Illustrator digital graphics format to GIS-compatible shapefiles and geodatabases. This is a working method, and, as such, suggestions for improvements are encouraged and welcomed. In addition to the presentation from the DMT 08 meeting (available online at http://ngmdb.usgs.gov/Info/dmt/DMT08presentations.html) and this paper, a thread covering the topic of AI file conversion can be found on ESRI's Mapping Center blog at http://blogs.esri.com/Support/blogs/mappingcenter/ (search for "DMT"). Any questions or suggestions regarding the process should be directed to Andrew Wunderlich (gibbon@utk.edu) or posted to the ESRI Mapping Center blog.

REFERENCES CITED

Hatcher, R.D., Jr., 2005, Non-survey, non-digital completed geologic maps in file drawers and theses: How can they be transformed into useful available digital geo-spatial data?, *in* Soller, D.R., ed., Digital Mapping Techniques '05—Workshop Proceedings: U.S. Geological Survey Open-File Report 2005–1428, p. 11-21, http://pubs.usgs.gov/of/2005/1428/hatcher/.

Mauldin, Jennifer, 2006, Converting Adobe Illustrator Maps to ArcMap Format, *in* Soller, D.R., ed., Digital Mapping Techniques '06—Workshop Proceedings: U.S. Geological Survey Open-File Report 2007–1285, p. 165-168, http://pubs.usgs.gov/of/2007/1285/pdf/Mauldin.pdf.