

Designing a Geodatabase for Your Project: **A Wetlands Delineation Example**

Naomi B. Kalman, Ingrid B. Hogle, Joshua H. Viers, James F. Quinn
Information Center for the Environment
Department of Environmental Science and Policy
University of California, Davis
1 Shields Avenue
Davis, CA 95616

<u>I.</u>	<u>Abstract</u>	<u>1</u>
<u>II.</u>	<u>Introduction</u>	<u>2</u>
<u>III.</u>	<u>Conceptualizing the Geodatabase</u>	<u>2</u>
A.	Data Creation Goals.....	2
B.	Standards.....	5
C.	Dataset Inventory	5
D.	Coordinate Systems.....	6
<u>IV.</u>	<u>Creating the Geodatabase</u>	<u>6</u>
A.	New Data	6
B.	Existing and Supporting Data.....	8
<u>V.</u>	<u>Using the Geodatabase</u>	<u>8</u>
A.	Digitizing.....	9
1.	Setting up the Editing Environment	9
2.	Creating New Data	9
3.	Tools and Customization	10
B.	Topology	10
C.	Workflow	12
<u>VI.</u>	<u>Conclusion</u>	<u>14</u>
<u>VII.</u>	<u>Appendix A.....</u>	<u>14</u>
<u>VIII.</u>	<u>References.....</u>	<u>15</u>

I. Abstract

The key to successful geodatabase design in ArcGIS 8.3 is laying out the project objectives before you begin. Our team at the Information Center for the Environment at UC Davis is mapping for the US Fish and Wildlife Service, as part of the National Wetlands Inventory (NWI). Our geodatabase must facilitate heads-up digitizing by multiple people and integrate seamlessly with the rest of the national inventory. We delineate wetlands using DOQQs combined with ancillary information such as hydric soils and topography. A personal geodatabase contains our wetlands polygons in a feature dataset and other geodata as stand-alone feature classes. Wetlands are exclusively a polygon feature class, as per national standards. We use a topological rule set within

the feature dataset to maintain feature integrity, and an existing NWI tool manages attribution and error checking. We assigned geographic work areas and outlined procedures for synchronizing the data and creating backups.

II. Introduction

The Information Center for the Environment (ICE), an information brokerage within the Department of Environmental Sciences & Policy at the University of California, Davis, is currently engaged in the development of environmental data, analysis techniques, and innovative communication media to support informed natural resource decision making. ICE has undertaken a partnership with the National Wetlands Inventory (NWI) – a significant natural resource management program within the United States Fish and Wildlife Service (USFWS) of the United States Department of the Interior – to update a portion of their inventory in California. As an integral part of our program, we have considered a number of design criteria in the initiation of our National Wetlands Inventory North Coast Update. Foremost in our design criteria is the clear formulation of our project objectives and our resource needs. A primary consideration in our project initiation is determining relevant components of successful environmental data creation. These components include methods for ensuring data integrity and accuracy, addressing existing attribution standards, and maintaining archival and portability requirements.

We approached the process of geodatabase design in three steps. First, we conceptualized the database by identifying our data creation goals, data standards, and necessary ancillary data. Next, we used the internal logic of the geodatabase to guide the organization of our data and create the editing environment. Finally, we refined our database and workflow process as the project progressed.

III. Conceptualizing the Geodatabase

The National Wetlands Inventory provides a convenient framework to discuss geodatabase development and the use of geodatabase technology as an integral part of geospatial data creation and analysis. The needs of the Inventory, with respect to wetland feature geometry and attribute complexity, are simple by most standards. However, successful implementation of geodatabase technology requires thoughtful implementation. To design a geodatabase that would meet the needs of our project, we considered our data creation goals, reviewed the USFWS standards, inventoried our datasets, chose a projection, then organized our new database to hold the existing data and provide an appropriate environment for data creation. This process should serve as a framework from which to expand other environmental data creation activities into more complex geodatabase configurations.

A. Data Creation Goals

Our project is to facilitate the delineation and attribute update for wetland features along California's North Coast (see Figure 1 for a map of the study region). ICE is

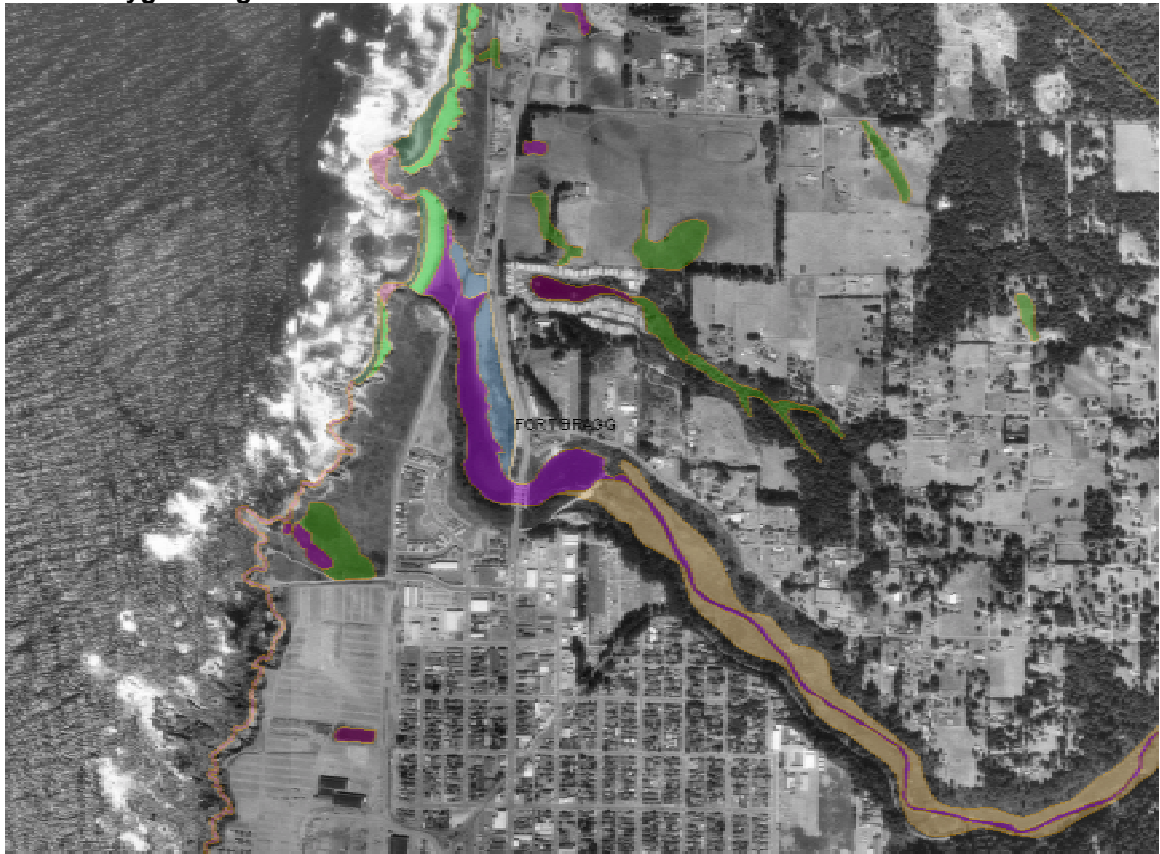
currently addressing the update needs of approximately 50 quadrangles (1:24000 scale) located in Sonoma, Mendocino, and Humboldt Counties.

Figure 1 : Map of study area



Wetland features documented as part of NWI include Palustrine, Estuarine, Riverine, and Marine types, with numerous subtypes and modifiers. This classification system, developed by Lewis M. Cowardin (1979), is a Federal Geographic Data Committee national standard for cataloging wetlands of the United States. The process we employed involves using stereo-pairs of color infrared aerial photos (1:58000) to identify the wetlands using a stereoscope, which are then located and digitized on-screen using Digital Orthophoto Quarter Quadrangles (DOQQs) as a background. Figure 2 shows an example of polygons digitized over DOQQs in a part of our study area.

Figure 2 : Polygons digitized over DOQQs



In addition to the DOQQs, data layers such as soils, topography, major roads, and the coastline serve as ancillary information to aid delineation. Photo-interpretation is supplemented by field work to verify wetland locations and signatures. We delineate wetlands as polygons in a solitary polygon feature class without a background polygon. Each feature is attributed with both a Cowardin Classification code (Cowardin 1979) and a hydrogeomorphic (HGM) code, developed by Sutula et al. (2003) for California. NWI subsequently integrates these data into their seamless geodatabase repository after an internal quality assurance and control process.

After considering several options for spatial data format, we chose the Environmental Systems Research Institute (ESRI) Personal Geodatabase as our format and repository. Geodatabases are the newest ESRI data storage format available, and with the introduction of topology modification tools, they more closely match the capabilities of the older 'coverage' data storage format. The ArcGIS 8.3 desktop platform provides a convenient graphical user interface, as well as tools for maintaining data integrity. Domains make it possible to create a controlled attribute list, which reduces the potential for typographical errors and incorrect attributes, especially when data creation is distributed among several users. A geodatabase provides a single-container framework to organize all project and ancillary data, enabling transfer of the project and inherent data as a discrete unit. Transitioning to the geodatabase data storage format also keeps our project in step with advances in GIS technology.

B. Standards

Our final project deliverables - wetland features for 53 quadrangles in the North Coast region - will be integrated into a national dataset. Therefore, we created feature geometry and attributes in accordance with classification and digitizing conventions. The USFWS is making new digital data and data updates available in personal geodatabase format. NWI no longer supports vector lines and points to delineate streams and springs, respectively. Instead, all wetland features are delineated as polygons. Features, such as streams, that were previously represented by lines have a standard buffer applied instead. There are also standards for the attribute table definition, including field name, length, and format.

C. Dataset Inventory

Delineating wetlands, like many natural features, is greatly aided by existing datasets that describe the natural landscape. There are many of these datasets available; an initial project step was to create an inventory of each dataset, noting its scale and spatial extent. We chose ancillary datasets based on information content, geographic coverage, effective scale, latest update, and relevance. Vegetation, soils, and hydrography are just a few examples of the datasets we considered.

We used California Department of Forestry and Fire Protection (CDF) hydrography (2001) at the 1:24000 scale to guide initial placement of stream and river features within NWI. These data served as our primary base because they are current, detailed, and largely complete for our study area. We buffered CDF vector hydrography lines at a 2.5m radius to create Riverine wetland features (Cowardin Classes R3/R4) within our master NWI Feature Class.

Hydric soils can be another physical indicator of wetlands. We used the Natural Resource Conservation Service Soil Survey Geographic (SSURGO) Database (see Appendix A) for Mendocino and Sonoma counties to derive a layer representing hydric soils. In areas that have an indeterminate photographic signature, hydric soils provide interpreters with another spatial data reference for classifying wetlands. Unfortunately the SSURGO data is not available Humboldt County, which is a large part of our study area, and we also identified a few discrepancies between the published soil survey and our field observations. For example, based on our field surveys and examining the published soil survey for Mendocino County (U.S. Department of Agriculture 2003) we determined that the river valley adjacent to Fort Bragg was mislabeled as a non-hydric soil type. The soil surveys nonetheless proved a valuable resource in the field.

Framework datasets are of the utmost necessity when undertaking regional planning projects (California Resources Agency 2002). We used the following datasets as frameworks for geographic reference, coding procedures, and advanced planning: USGS quadrangles, roads, county boundaries, city point locations, and the coastline. See Appendix A for a description of these data.

Of obvious import to heads-up digitizing is imagery. We used the Digital Orthophoto Quarter Quadrangles from the USGS, and imagery representing the scanned Digital Raster Graphics (DRGs) from USGS Topographic Maps. See Appendix A for a description of these data.

D. Coordinate Systems

Before constructing the geodatabase we had to choose a coordinate system to represent the data. Although ArcGIS 8.3 is capable of coordinate projection on-the-fly, it does not record projection and transformation procedures in a log file, reducing our confidence in spatial accuracy. To minimize the potential to introduce errors, we converted all of our data to the same coordinate system before we commenced digitizing. Universal Transverse Mercator (UTM) with a North American Datum (1983) is the standard Coordinate Projection for both NWI and our DOQQs, making this the best choice for our project. Our study area is in UTM Zone 10 North.

Ancillary data came in a variety of native formats and coordinate systems. Compiling each ancillary dataset into one geodatabase with one uniform projection required advanced planning. While it is possible to simultaneously convert data types and project coordinate systems, ArcGIS does not preserve step-wise documentation of the process in a log file, making these data difficult to recreate if necessary. Instead, we used an explicit, step-wise procedure of data conversion to ensure data integrity and our ability to catalog processing techniques for metadata generation. The procedure is as follows: we created an intermediate geodatabase representing each coordinate system, imported each dataset into geodatabase format, then projected from the intermediate geodatabase to the final one. This procedure allowed us to control any necessary settings for Datum transformations, and enabled creation of all future datasets following the same procedure.

Imagery representing the scanned Digital Raster Graphics (DRGs) from USGS Topographic Maps (1:24000) is in Albers Equal-Area Projection with a North American Datum (1927). We therefore projected each 1 degree image for our study area into the correct projection using ArcINFO GRID commands.

IV. Creating the Geodatabase

After assembling all of our data, deciding on what projection to use and what we generally wanted from the database, we organized the layers so as to create the best possible editing environment. A productive editing environment allows us to create new data with topological integrity, attribute the new data with ease and minimum potential for error, and have any necessary ancillary data at our convenient disposal.

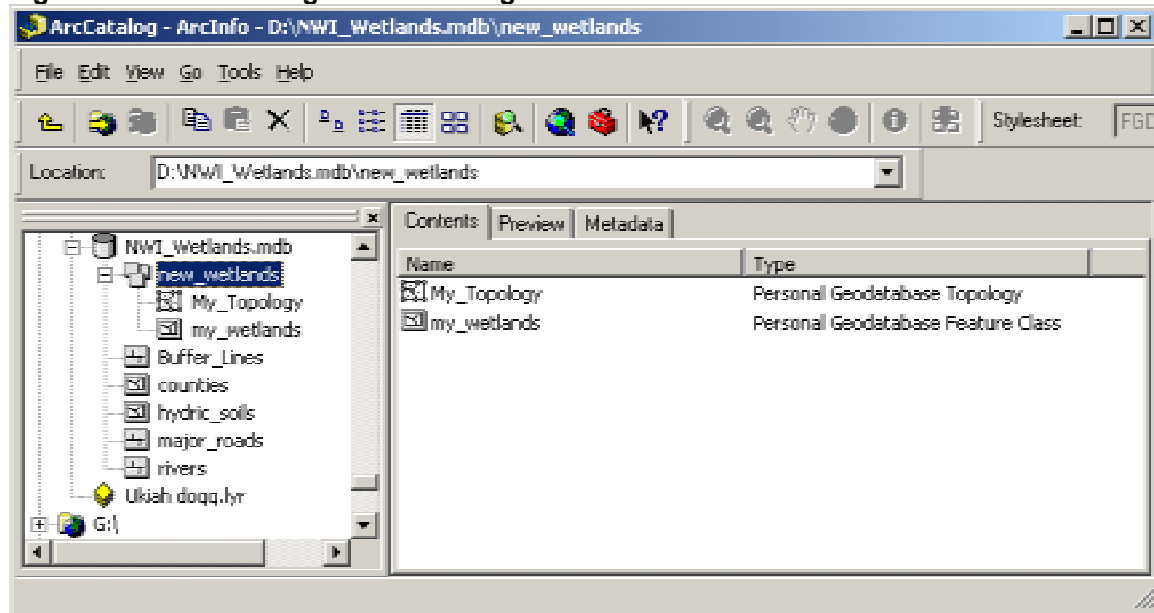
A. New Data

In order to use GIS data properly for querying and analysis, it must be clean and integral – that is, the features must not have duplications, gaps, overlaps, or other flaws. Overlapping polygons will cause errors when converting the data layer into raster format, and small gaps between polygons can cause holes that perpetuate through an analysis. Topology rules were created to guard against these kinds of data flaws.

All newly created data should have a topological reference. The rules for setting up an ArcGIS topology in a geodatabase therefore dictated how we organized our spatial data. An ArcGIS topology can only exist within a feature dataset, which is a group of feature classes that all have the same projection and extent (Environmental Research Systems Institute 2003). Therefore, we created the feature class that represents our wetlands within a feature dataset. This allows us both to develop rules that maintain

feature integrity and keep the most important feature class separate and easily accessible. Figure 3 provides a schematic example of our database setup, with the wetlands feature class in a feature dataset and the ancillary data as stand-alone feature classes.

Figure 3: Schematic of geodatabase organization

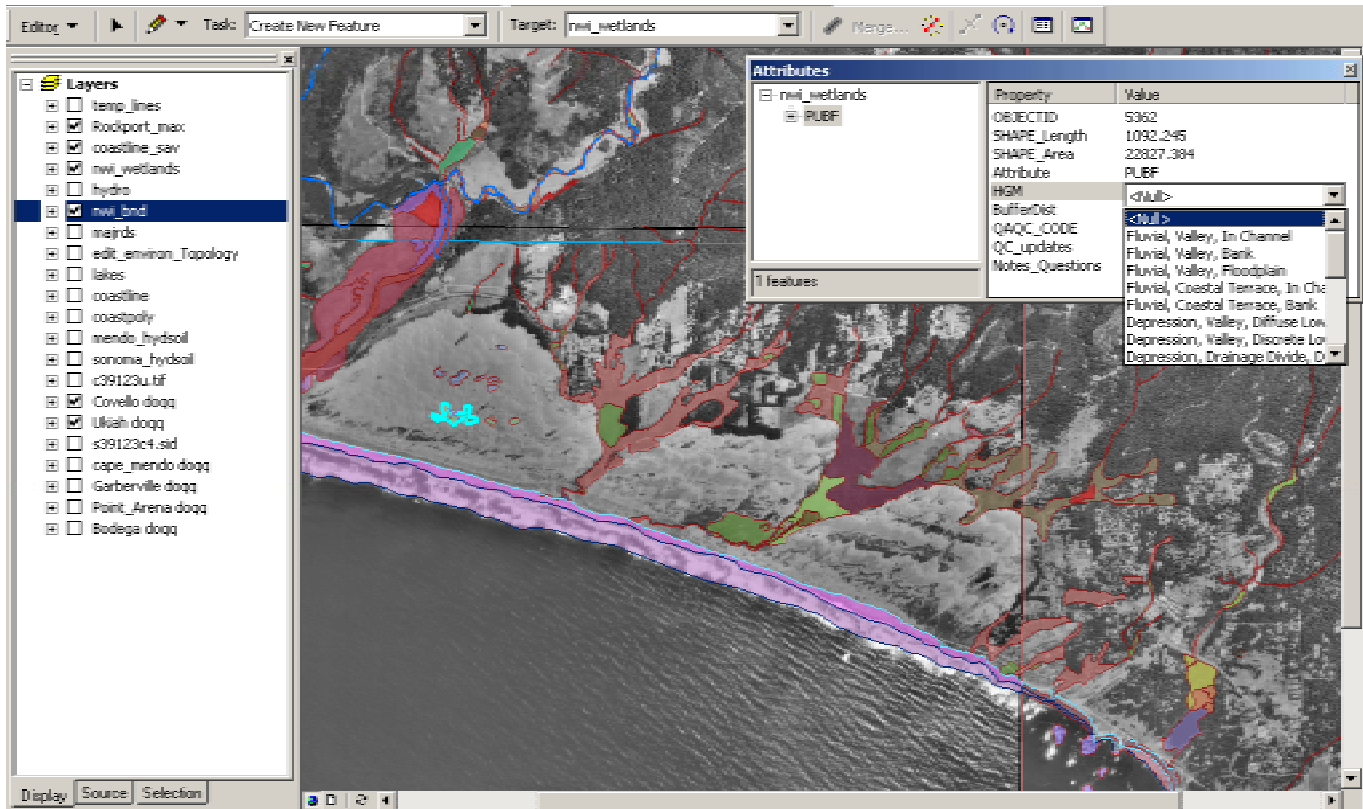


Our wetlands all have polygon feature geometry, so we chose from the ‘polygon-to-polygon’ rule options. The rules we used are: polygons must not overlap and polygons must not have gaps. That is, two wetlands types cannot exist in the same place at the same time, and features must share a complete edge with other features. The combined rules require that each feature must touch and only touch the other features around it, creating a seamless representation in two dimensions over a surface.

After creating the wetlands feature class, we created attribute fields. The standard NWI method of storing the Cowardin classes is a text field named ‘Attribute’. We also added fields to document hydrogeomorphic modifiers and for general comments about confusing or difficult feature interpretation. Because we have multiple technicians creating wetland features, we added a field to indicate who created each feature.

Domains provide an easy way to control attribute values and reduce typographical errors. Although we considered creating domains for attributing the polygons with Cowardin classes, a set of custom NWI tools made it unnecessary. Developed by the United States Geological Survey (USGS), the tools attribute polygons with standard Cowardin classification codes and check for attribution errors. For the hydrogeomorphic notation, a domain became an ideal way to create a uniform list of attribute codes and decrease the potential for data entry error. We downloaded a utility from the ESRI developer sample website (Environmental Research Systems Institute 2004) that reads values from a table into the domain, making it possible to create a master list of codes and distribute it to each technician. Figure 4 illustrates the use of domain attribute pick-lists in the ArcGIS editing environment.

Figure 4: Using domains to assign HGM attributes



B. Existing and Supporting Data

The ancillary data reside as stand-alone feature classes in the geodatabase. As mentioned before, each dataset was first imported into feature class format, and subsequently projected into the master database. A description of these data can be found in see Appendix A for a description of these data.

Small seasonal streams, previously represented as lines, are delineated by a buffering a line feature to a width of 5 meters. Where available and accurate, the CDF vector hydrography lines (California Department of Forestry and Fire Protection 2001) were used for this purpose. Although these data do not require topological reference rules, we stored them in their own feature dataset for ease of organization.

In some instances we created our own linear hydrography features which we stored in a separate feature class. Because these intermediate lines do not require any topological reference rules, we stored them as a standalone feature class to keep our feature dataset uncluttered.

Images cannot be stored in a personal geodatabase in ArcGIS version 8.3. To keep them organized and easily accessible, we collected the images into the same directory as the geodatabase and organized them by 250k quadrangle. We stored DRGs in a similar manner.

V. Using the Geodatabase

After constructing the geodatabase, we encountered new challenges with work flow patterns and with our production tools. We found many ways of setting up the editing environment to streamline our digitizing process, and we envisioned some potential avenues for future customization. The following section describes some of these issues and the ways that we adapted our methodology to meet new challenges.

A. Digitizing

ArcMap 8.3 introduced a new suite of digitizing tools and options, and there was an initial period of navigating through the menus and toolbars to learn the new functionality. Two ESRI virtual campus courses were helpful understanding this environment:

“Creating and Editing Geodatabase Features (for ArcEditor and ArcInfo)” (Environmental Research Systems Institute 2003), and *“Creating and Editing Geodatabase Topology (for ArcEditor and ArcInfo)”* (Environmental Research Systems Institute 2003).

1. Setting up the Editing Environment

The ArcGIS graphical user interface provides many built-in avenues for customization, and a thoughtful setup of the editing environment can streamline the digitizing process. Display and organization of key buttons and thoughtful symbolizing features helps maintain visual organization and easy access. In particular, we utilized the ability to group layers, create and save symbol sets, change transparency, add buttons to toolbars, and display and dock toolbars such as snapping and topology for ready and convenient access.

The Data Frame setup provided a solution to a particular work-flow issue. The aerial photographs are oriented relative to flight direction, and the flight lines were flown north to south. Therefore, the photographs can only be oriented along this north-south axis for use with the stereoscope. We rotated the data frame in ArcMap so that our features were oriented in the same direction as our aerial photographs.

Keyboard shortcuts are an indispensable tool to streamline digitizing; we used them to zoom in (key z), zoom out (key x), pan (key c), show vertices for snapping (key v) and to isolate topology nodes for selection (key n).

2. Creating New Data

Choosing the correct data editing tool requires an understanding of how the features relate to each other within the ArcGIS topological framework. The tools are organized as feature tasks and topology tasks. The feature tasks operate on the individual spatial objects, while the topology tasks modify the boundaries between objects.

For example, there are two methods for creating new polygons: ‘Create New Feature’ and ‘Auto-Complete Polygon’. ‘Create New Feature’ is a feature task, and it creates a new, autonomous spatial object; ‘Auto-Complete Polygon’ is a topology task, and incorporates the edges of existing polygons into the new polygon. Similarly, the ‘Edit Tool’ edits a single polygon feature, while the ‘Topology Edit Tool’ edits the edge that two features share. The ‘Auto-Complete Polygon’ tool was the most appropriate in almost all cases, and became virtually indispensable to our work.

3. Tools and Customization

The USFWS provided a suite of tools designed by the USGS (2002) to help us conform to digitizing and attributing standards. The tools work inside the ArcMap environment, and are accessed via a custom menu. The Attribution Tool aids attribution by allowing the user to select from the Cowardin classes hierarchically. The Favorites Tool allows the user to enter their most frequently used codes for quick access in a pull-down menu. The Verification Tool performs a quality review of the data, checking for invalid codes, mislabeled lakes, and adjacent polygons with the same attribute, among other things.

The custom tools have many advantages, although some minor improvements would make them even better. For example, the Favorites Tool was exceptionally helpful, and it would be even more so if we could sort and organize the entries, and if the toolbar was dockable instead of floating – it hides behind the application every time you click away from it.

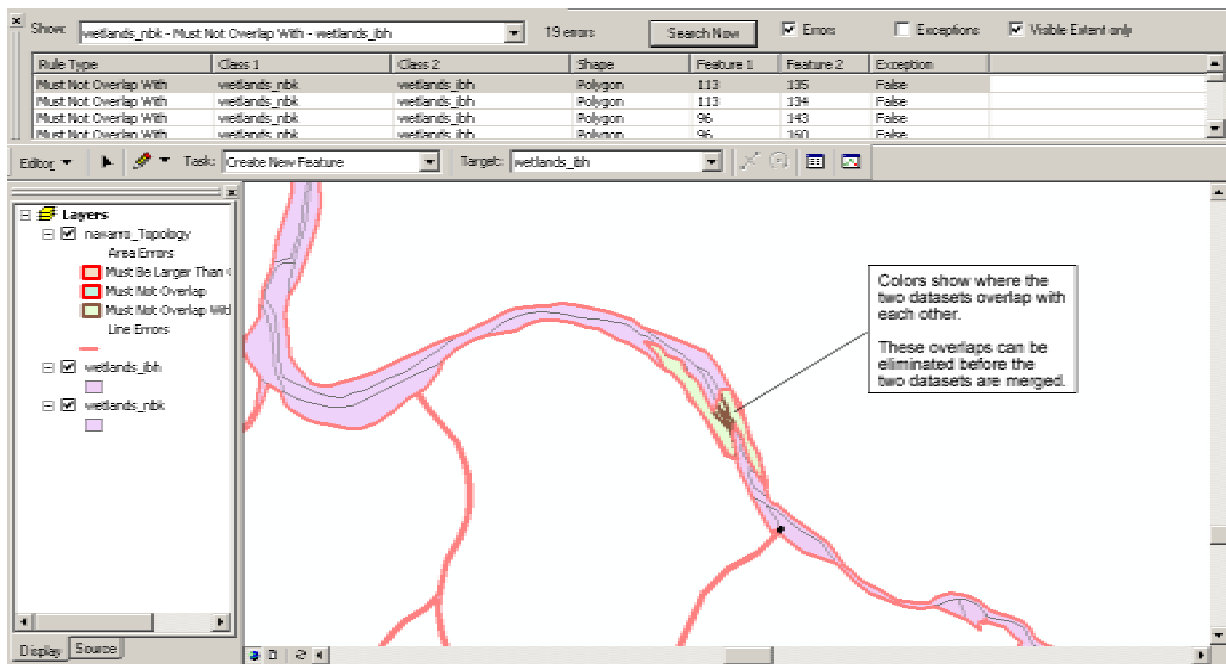
As we digitized, some other avenues for customization became apparent. Many of the digitizing functions require multiple mouse-clicks to execute – for example, to create a new polygon it is necessary to select the edit task and the edit tool. To cut a polygon is a four-step process of activating the selection tool, selecting the polygon, selecting the “cut feature” edit task, and then selecting the tool. These processes could be streamlined with the use of buttons that combine some of these functions. For example, with the click of one button you could select “auto-complete polygon” and “draw tool” at the same time. Or you could create a two-step button for cutting polygons; the first step is selecting and then you automatically move to the next step which is to draw the cutting line.

B. Topology

The addition of ArcGIS topology to the geodatabase framework was the most compelling reason to adopt the new data storage format. There are several differences between ArcGIS topological rules and the rules from previous software versions. As opposed to earlier topological rule sets in ArcINFO - where polygons are a connection of arcs, vertices and nodes - in ArcGIS geodatabase topology each feature is considered a separate spatial object and has topological relationships with other features (Environmental Research Systems Institute 2003).

One very inconvenient aspect of ArcGIS topology is that island features are considered gaps under the “must not have gaps” rule, as none of the feature edges touch another feature. Since many wetland features, such as ponds, do in fact exist in isolation, it is necessary to mark these features as exceptions in order to find the true problem gaps in the dataset. Figure 5 illustrates the difference between the gaps that must be eliminated and the acceptable island polygons.

Figure 5 : Different types of polygon "gaps"

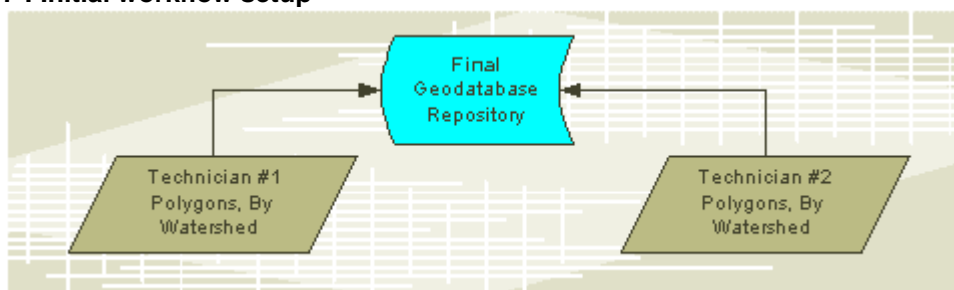


This process allowed us to mend any overlaps and gaps between the two datasets while the features were still whole, separate, and obviously attributed. Once these changes were made, it was possible to paste the two datasets together with very few errors.

C. Workflow

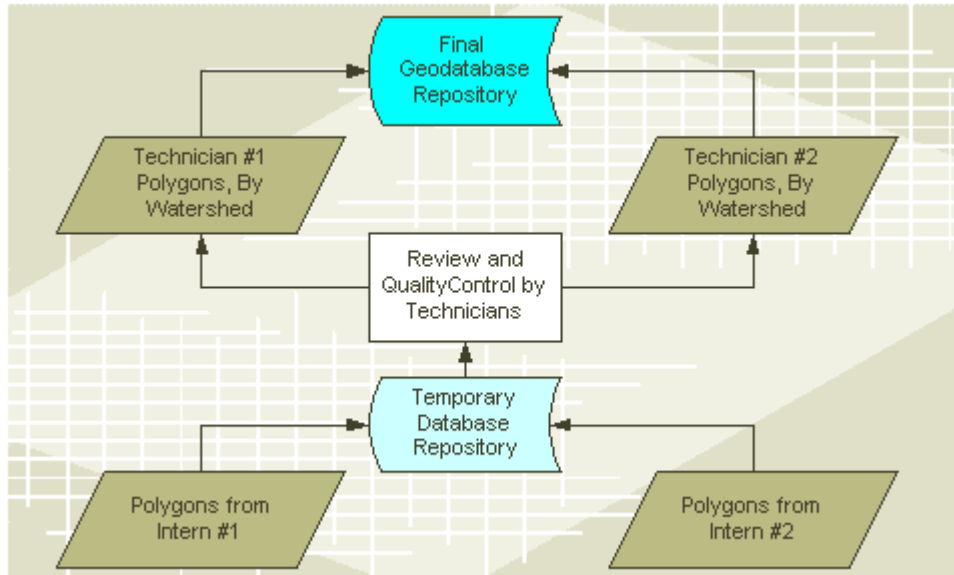
At the outset of the digitizing process, we had only two technicians creating polygons. We chose to use a personal geodatabase and divide the work into geographically separate areas. Wetland feature data within the geographically separated areas could then be merged into the master geodatabase. Initially, we chose to divide the project area by quadrangle because that is the unit by which NWI catalogs the wetland inventory. However, we quickly realized that the quadrangle boundaries cut across many natural features which later had to be reconstructed during the merging process. We therefore redistributed the work areas by watershed, defined by the CDF hydrography, since watersheds are inherently non-overlapping. Figure 7 diagrams our initial workflow setup.

Figure 7 : Initial workflow setup



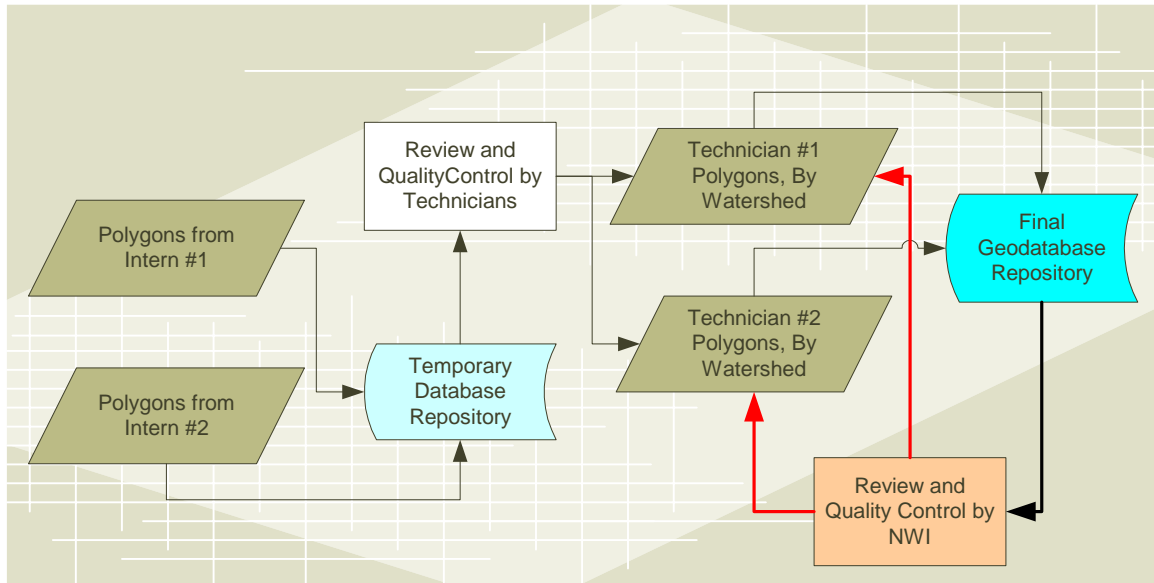
A few months into the project we added two student interns to our digitizing team. Their work required review and quality control before being incorporated into the master geodatabase. The workflow was easy to modify: the students put their work in a central location where the technicians could access it as needed and incorporate it into their respective datasets to perform the review. Figure 8 shows our modified workflow including the two interns.

Figure 8 : Workflow including student interns



Our work flow includes a review process by NWI staff. After completing each quadrangle, we send our data to the NWI regional headquarters for review and comments. Documenting revisions to the original feature creation efforts is necessary for a variety of reasons, including general effort tracking, point-in-time snapshots, and evaluating benchmarks for quality and accuracy. Modifying original data eliminates the possibility of comparing different versions of the data, and such a comparison could be useful for reporting, analysis, or review for future efforts. Our solution has been to take weekly ‘snapshots’, or permanent back-ups, of the database which could be compared to each other if the need should arise. Figure 9 shows our complete work flow, including the NWI review process.

Figure 9 : Quality Control work flow



VI. Conclusion

Although more complex than the shapefile data storage format, an ArcGIS geodatabase has an inherent structure that guides its own organization. There are many advantages to transitioning to the new data format; primarily, this includes storing and organizing all vector data relevant to the project. The ArcGIS 8.3 desktop platform provides a convenient graphical user interface, as well as tools for maintaining data integrity. Domains make it possible to create a controlled attribute list, which reduces the potential for typographical errors and incorrect attributes, especially when data creation is distributed among several users. A geodatabase provides a single-container framework to organize all project and ancillary data, enabling transfer of the project and inherent data as a discrete unit. Transitioning to the geodatabase data storage format also keeps our project in step with advances in GIS technology.

Geodatabase development is an iterative process, and the design may need modification as work progresses, but we found that advance planning can minimize this necessity. This process should serve as a framework from which to expand other environmental data creation activities into more complex geodatabase configurations.

VII. Appendix A

Data Inventory

Hydrography:

Watershed Assessment 1:24,000 Hydrography
 California Department of Forestry and Fire Protection (CDF)
<http://frap.cdf.ca.gov/data/frapgisdata/select.asp>

Soils:

Soil Survey Geographic (SSURGO) database for Mendocino County, Western Part, California

Soil Survey Geographic (SSURGO) database for Sonoma County, California

Natural Resource Conservation Service Soil Survey Geographic (SSURGO) Database

<http://www.nrcs.usda.gov/branch/ssb/products/ssurgo/>

Quadrangles:

Map Grid

CaSIL <<http://gis.ca.gov/index.epl>> vector data: <http://gis.ca.gov/BrowseCatalog.epl>

Cities:

Geographic Names (GNIS)

CaSIL <<http://gis.ca.gov/index.epl>> vector data: <http://gis.ca.gov/BrowseCatalog.epl>

Counties:

County Boundaries (1:24000)

CaSIL <<http://gis.ca.gov/index.epl>> vector data: <http://gis.ca.gov/BrowseCatalog.epl>

Roads:

Major Roads

CaSIL <<http://gis.ca.gov/index.epl>> vector data: <http://gis.ca.gov/BrowseCatalog.epl>

Coastline:

California State Lands Commission

<http://www.slc.ca.gov/>

Imagery:

Digital Orthophoto Quadrangle GeoTiff (DOQQ)

CaSIL Data Collections: <http://gis.ca.gov/data.epl>

Topography:

California Digital Raster Graphics

CaSIL Data Collections: <http://gis.ca.gov/data.epl>

VIII. References

California Department of Forestry and Fire Protection (2001). Watershed Assessment 1:24,000 Hydrography.

Developed by Watershed Assessment Program, in cooperation with the USDA Forest Service Remote Sensing Lab. Includes separate linear and polygon layers.

California Resources Agency (2002). Initial Assessment of the Health and Condition of California's Lands and Natural Resources. Sacramento, CA, California Resources Agency: 43.

Cowardin, L. M. (1979). Classification of wetlands and deepwater habitats of the United States. Washington, D.C., Fish and Wildlife Service U.S. Dept. of the Interior.

Environmental Research Systems Institute (2003). ArcGIS: Working with Geodatabase Topology.

Environmental Research Systems Institute (2003). Creating and Editing Geodatabase Features (for ArcEditor and ArcInfo).

Environmental Research Systems Institute (2003). Creating and Editing Geodatabase Topology (for ArcEditor and ArcInfo).

Environmental Research Systems Institute (2004). Table to Coded Value Domain Tool. **2004.**

This tool will allow you to convert a table of values to a coded value domain in a GeoDatabase.

Sutula, M., E. Stein, et al. (2003). Draft Hydrologic/Geomorphic Coded for Additional Attribution of Standard NWI Mapping of Wetlands in State of California. Westminster, CA, Southern California Coastal Water Resource Project,: 3.

U.S. Department of Agriculture, N. R. C. S. (2003). Soil Survey Geographic (SSURGO) database for Mendocino County, Western Part, California.

U.S. Geological Survey (2002). Attribution and Verification Tools. **2003.**