

GIS Exercise 1:

Acquiring Study Area GIS Data & Creating a Field Map

The purpose of this exercise is to build a GIS for a study area in the US, from publicly available data. The US Public Domain system makes data fairly easily available. The best choice of web sites has changed over the years, but tends to improve. This lab will primarily use the **National Map Viewer**. We are going to make use of this dataset for watershed assessment, initially for creating a field map, but later to run some analytical models using this and field data.

Assumption: Students are familiar with the use of ArcGIS (ArcMap etc.). There will be no introduction to this software here.

Part 1: Acquiring Study Area GIS Data

With the increasing availability of free data, GISci is starting to provide the means to initiate an area study in very little time, as long as you possess the skills. In this exercise, we will download data to cover a small watershed of your choice. We'll limit this to a small watershed that you can see on your computer screen at finer than 1:100,000 scale.

- Start by setting up your file area (on the network, zip disk, or in the workspace folder on your local computer) to store the data you need. I recommend working on the local hard drive, by creating a folder on D:\workspace with your name. So if your name is Jennifer, create a D:\workspace\jennifer folder.

Introduction to Commonly Available Datasets

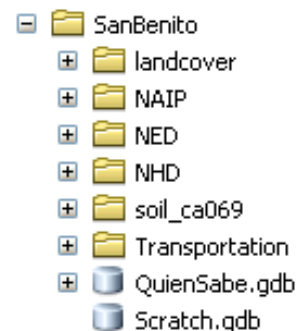
The USGS and other agencies maintain digital datasets of a variety of types. Data that may be of use in studying a watershed, some of which we'll work with in this exercise, are:

1. **Vector Data** of transportation and hydrography. These have been compiled by various agencies for different purposes: the USGS for general purpose, geologic and hydrographic mapping, and the Census Bureau for census tabulation, but now many agencies use GIS data. These have been compiled in the online **National Map**, which makes it much easier to access.
2. **Digital Elevation Models (DEMs)**: use a raster GIS format for storing elevation data. They come in a variety of cell sizes, including 10m and 30m cell sizes for 1:24k quads, with the 10m product the most useful for watershed studies.
3. **Digital Orthophotos** and other orthorectified aerial and satellite imagery. The USGS stores black & white quarter quads (Digital Ortho Quarter Quads), but there are many other sources, including color and false-color imagery that we'll find useful.
4. **Scanned Topographic Maps**. These can help when you need to pick a study area or watershed, or find the names of features. You can either look for a printed map, or download a map of the area to display this map overlain with your other data.
5. **Soils**. Since soil erosion is a major concern in studying watersheds, having data on soil erodibility factors can be useful. Soil data is complex, however, and is covered in a separate document, *Reading SSURGO Soil Data*, which we will cover in the next exercise. You might want to confirm that the soils data is available for your study area. (For example, Santa Clara County isn't.)

Download watershed data with the National Map Viewer

First select a watershed to study. You can use a topographic map from the map library, or go use the on-screen view in the National Map Viewer. This is the best place to download USGS and other data we need for watershed GIS or other purposes: <http://viewer.nationalmap.gov>

- Use the viewer to navigate to and zoom into an area of interest, a complete watershed. Zoom levels appear set to specific scales; for the purpose of this exercise, you should use 1:72,224 or 1:144,448; this should allow you do grab a complete interesting watershed, without having to wait an extremely long time to get large amounts of data. Ideally limit it to one USGS 7.5' quad. You should be able to identify a watershed by studying the stream patterns and hillshades of the landscape. *Note the name of the stream, since this may not be easy to find later. We'll use it to name a file geodatabase for output data.*
- Click the Download Data button to bring up the dialog, and then either (a) use *current map extent* and make sure the entire watershed is included in the view on screen (reduce the window size to reduce the volume of data to just the area containing the watershed), or (b) select *Index 24K* and click on a USGS 7.5' quad labeled on the map. Get US Topo, Transportation, Hydrography, Land Cover, Elevation and Orthoimagery with the following settings set in the subsequent dialogs:
 - Elevation: 1/3 arc second
 - Orthoimagery: 4-band NAIP, most recent.
 - Land Cover: most recent **land cover**, and **impervious surfaces** if they have it.
- You may need to click the right edge to get to the Download Cart to **Checkout** (but it's all free), give your email address, etc.
- Check your email. You'll probably see the elevation come through first; that's good, since that's what you need the most. Orthoimagery and land cover will come through pretty soon, if not right away. Transportation may take a bit longer. Follow instructions for downloading. Note that you need to click the link in your email to start a new process for downloading. What will be returned will be a numbered zip file: *Write these down so you'll know what they are.*
- You will need to extract the zipped data into a project folder you should name with the location. Then for each data layer, create a folder with a good name (like 'NHD', 'NED', etc.) according to the table below, within your workspace folder. You may be renaming a folder with a random number name – it's better to use a meaningful name. Otherwise, it's easy to end up with a bunch of random-number-named folders, so make sure to name things.
- In this project folder, create a *file* geodatabase, and name it with the stream name or other study area name you want to use (I used QuienSabe, a creek name in San Benito County I chose.) You should also create a Scratch file geodatabase for putting temporary data.



Data type	Folder Name	zip file names
Elevation	NED	
Hydrography	NHD	
Imagery (NAIP)	NAIP	
Transportation	Transportation	
Land Cover & Impervious Surfaces	Landcover	

Process things to make them work:

- **US Topo:** This creates a GeoPDF, and unless we figure it out, can only be viewed in Acrobat, but it's still useful. Opening the file will prompt you to download a plug-in, which you won't be able to install on the lab computers. Without the plugin, you can still see the pdf. We could georeferenced it using the corner points, but we won't bother. Consider it a digital view of the map, but we'll mostly be using ArcGIS with the other layers. One purpose will be to get coordinates to download the USGS 7.5' quad from Cal-Atlas (see Appendix).
- **Elevation (NED):** is in meters vertically, and decimal degrees.
 - Use Project Raster (in Data Management toolbox) to create rasters in UTM meters (NAD_1983_UTM_Zone_10N for California west of 120°W (Tahoe), zone 11 east of this meridian (primarily southern California.) Set the output to be 'elevUTM' in your creek-named file geodatabase. (If you have more than one, you might use a different naming convention – just make it something you can keep track of later.) Use BILINEAR Resampling technique and output cell size of 10 m (corresponds to 1/3 arc second).
 - Create a **hillshade** of the UTM elevation raster(s), using the tool in the Surface toolset of Spatial Analyst. This will help to make sure you have good data, and also is a useful display.
 - Create 5-m **contours** (tool also in the Surface toolset) from the same UTM elevations.
- **Orthoimagery (NAIP):** should be ready to go, but come in multiple sections. There's no need to mosaic or reproject, since these are only used as a basemap backdrop. Note: this is 1-m 4-band imagery, so is by far the largest of the datasets, probably requiring a few hundred megabytes for the watershed you have chosen.
- **Transportation:** comes in as a geodatabase. Note that not all of the features that appear to be in the Foundation feature dataset will have data included – all possible data structures are included, but most are not populated. In the datasets I tested, **Trans_RoadSegment** seems to contain data.
- **Hydrography (NHD):** is in decimal degrees, where you can leave them if you use them only for display. If needed to for analysis, can be projected later. Similar to transportation, this comes as a geodatabase with a lot of stubs for data, but with very few containing data. NHDFlowline, HYDRO_NET_Junctions, NHDPoint (includes springs and wells, among other things), and NHDWaterbody are commonly provided.
- **Land Cover:** See the legend here for meanings of the codes. These are different from Anderson Land Use codes. Note that the data are created from satellite imagery, and the shape of the file will seem rotated due to the nature of the satellite path. It should still overlay correctly. **Impervious Surfaces** values are % impervious cover, as interpreted from Landsat imagery; see the metadata to understand the process used to create this.

NLCD 2001 Land Cover Classification Legend

11	Open Water
12	Perennial Ice/Snow
21	Developed, Open Space
22	Developed, Low Intensity
23	Developed, Medium Intensity
24	Developed, High Intensity
31	Barren Land
41	Deciduous Forest
42	Evergreen Forest
43	Mixed Forest
51	Dwarf Scrub*
52	Shrub/ Scrub
71	Grassland/ Herbaceous
72	Sedge/ Herbaceous *
74	Moss *
81	Pasture Hay
82	Cultivated Crops
90	Woody Wetlands
95	Emergent Herbaceous Wetlands

* Alaska Only

Part 2: Create field base maps

You should now have several layers of data that can help for field-based watershed assessment: (1) an elevation model and its derived hillshade and contour layer; (2) some orthorectified aerial photographs; (3) any other vector data you have downloaded, especially streams. These can be very useful for field work. Since your data are in UTM NAD83, this is very handy for use with GPS receivers, most all of which can display in UTM WGS84, which is within a meter of UTM NAD83.

- Create two field maps:
 1. A map of the entire watershed, with these layers, set to a scale that will fit on a page.
 - a. NHDFlowline streams (Ftype) in blue
 - b. NHDWaterBody
 - c. Trans_RoadSegment (Road_Class)
 - d. hillshade
 2. A map of a zoomed-in area, at about 1:10,000 or finer (e.g. 1:5,000), with the layers:
 - a. NHDFlowline streams (Ftype)
 - b. NHDWaterBody
 - c. contours
 - d. NAIP imagery
- Symbolize the layers so you can distinguish roads from trails, streams from aqueducts, etc. Think about the scale you want to work at. A good scale for field work is 1:5000, but you may also want other maps like overview maps of the entire study area, at a coarser scale like 1:8000 or 1:24000. Labeling the contours helps, but make sure you use the right field if you're using hypsography-derived contours, and you'll need to set the label color to something other than black. It helps to zoom into various parts of the map using the layout zoom tool to see what things will look like close-up on the page.
- In Data Frame Properties (View menu), use the Grid tab to create a new Measured Grid (press the New Grid button). Use either the “Tic marks and labels” or “Grid and labels” option. Set the X and Y intervals to an even value that is convenient (e.g. 100 m for a 1:5000 map usually works).
- There are many properties of a grid, most of which you set after you’ve first created it by selecting the name of the grid in the Grids tab (probably called “Measured Grid”), and going to **Properties**. Some common properties:

Subdivisions: (Axes tab) If you use subdivisions, use something simple like 5.

Labels settings like decimal places (Labels tab) While the example here shows a bunch of zeroes after the decimal place, we really don’t need to show these; to remove these, go to the **Labels** tab, then in Label Style go to **Additional Properties** then in Mixed Font Label, go to **Number Format** and set Number of decimal

One size doesn't fit all

Any map has a purpose and an audience. Even in a field map created for yourself needs attention to what the audience (you!) will need when using it. *Scale* is an important consideration here. What works at one scale may not work at another. For example, you may not want to show all of the roads, and you may not want to show roads at all if you can see them better on the imagery. Contours are useful to include, and need to be visible, but don't make them so bold as to obscure the imagery. The same goes for grid lines. Note that your point sizes don't have to be integers; you can use a 0.7 point size. A good idea is to use the layout zoom tool in the layout view to zoom in – only then can you see the effect of the line sizes when you print.

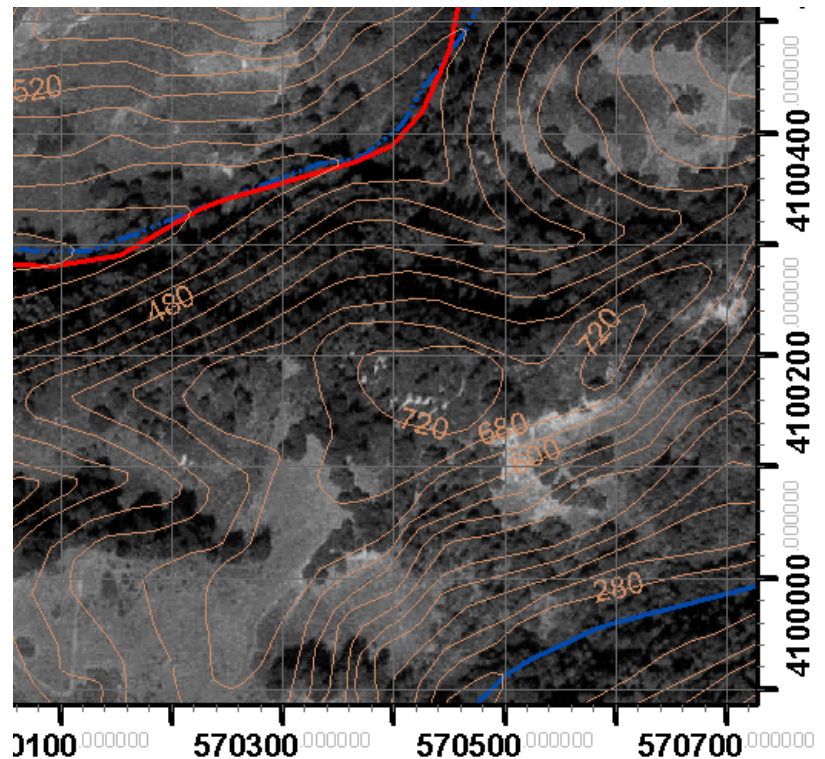
places to 0. Other label settings to consider are label orientation and labeling only two of the four axes (e.g. right and bottom).

Interval settings in case you decide the intervals are too busy.

Lines settings to choose between lines and ticks and to change the symbol. It helps to use the *layout* zoom to look close at a small section of the map to see what it will look like. Depending on the background image, you may find color a challenge to select. Line thickness may also take experimentation; it might either obscure too much if too thick, or be so thin as to disappear.

- What to turn in: pdf files of your two maps. To create the pdf, from ArcMap, File/Export Map... and select pdf as the output format. Make sure your map includes an explanatory title that will help you use the map in the field.

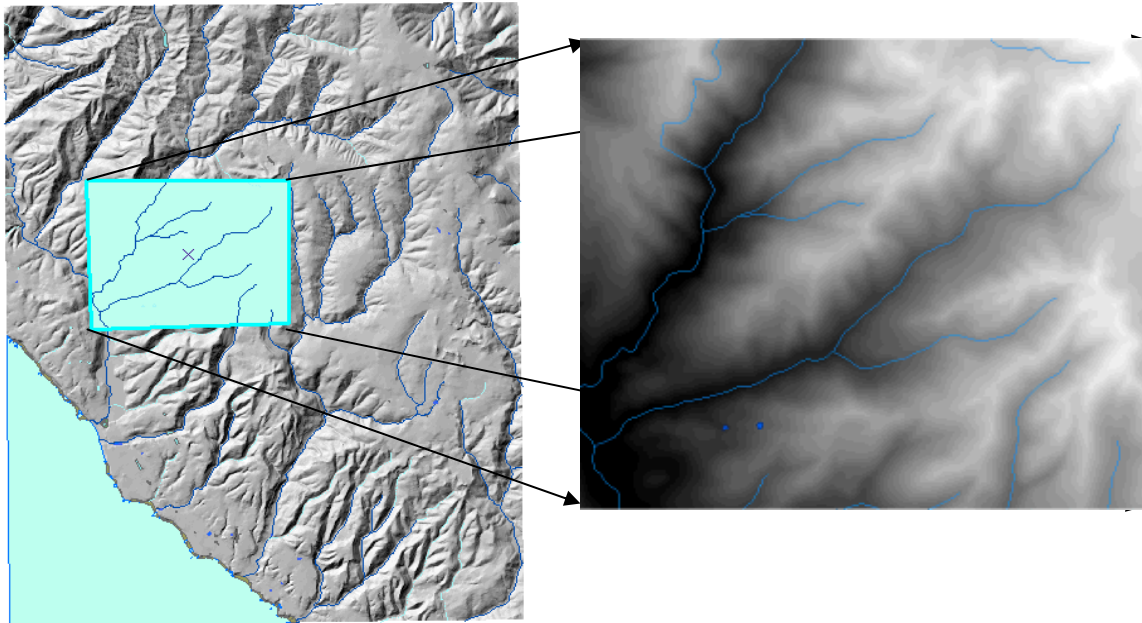
Figure 1: Portion of a map showing the UTM grid labels



Appendix 1: Clipping your data to a study area

It is often useful to clip a smaller study area. To clip your themes to a smaller area, try the following:

1. In ArcCatalog, right-click your workspace and use New... Shapefile... to create a **polygon** shapefile. Name it **StudyArea**. Make sure to establish its Spatial Reference (projection coordinate system) by using the Edit button to select a predefined projected coordinate system: UTM NAD83 Zone 10 (the zone for most of California, including all of northern California).
2. In ArcMap, add StudyArea.shp and use the Editing toolbar to, first, Start Editing, specifying StudyArea.shp as what you want to edit; then:
3. Use the Create New Feature tool to digitize a polygon of a study area. To create a rectangular area, find a rectangle tool in the Editor menu>More tools>Advanced Tools.
4. When you're done, select "Stop Editing" in the Editing menu.
5. Use the Clip tool (Analysis Tools>Extract Tools) to clip your vector datasets (the DLG-derived datasets like roads and streams) to the StudyArea shapefile.
6. Use Spatial Analyst options to set the Working Directory to the workspace you created (like **f:/davenport** in my example), the Analysis Extent to the StudyArea shapefile, and in that same Extent tab set the **snap** extent to elev (to avoid muddying pixels).
7. With these settings in force, use the Raster Calculator to clip the raster datasets (like elev) to the StudyArea, by entering a simple statement: `elevclip = elev`. If your original elev raster is in feet, you could do two things at once to create one in meters, with `elevm = elev`.



Mosaicking multiple quads

If your study area covers more than one quad: first, to download another quad of data, *you must create new USGS data folders and output folders*. For this class, there is no need to merge feature data like roads and streams; they can come from different folders (e.g. Davenport and Felton quads). There is also no need in mosaicking DOQQs. Mosaic your elevation rasters with the **mosaic to new raster** tool in ArcToolbox (Data Management/Raster). An important setting is pixel type: "32_BIT_FLOAT". Make sure the input rasters all have the same vertical units.

... Then you can clip the result to your study area.

Appendix 2: Other Data Sources

Census Bureau TIGER/Line files

- Advantages: pretty easy to download, one county at a time, easy to use shapefile format, includes road names and address ranges, and stream names.
- Disadvantages: sometimes poor spatial accuracy, sometimes crudely digitized, sometimes applies topological rules that extend features inappropriately for our purposes.
- Source:
 1. Search for Census TIGER files, and make sure to use 2008 or later files – the earlier ones have poor accuracy.
 2. The files are organized by county: select California, then the county;
 3. When you download the file, you'll probably need to use 7zip, winzip or other program to unzip. The default Windows XP unzip utility won't work, as of this writing (2009). Note that the files you want are buried several layers deep in the zip file.
 4. The attributes are pretty easy to use, including address ranges for streets and other useful attributes:
Select streams with "HYDROFLG" = 'Y', roads with "ROADFLG" = 'Y', railroads with "RAILFLG" = 'Y'. There are other linear features selected with "OLFFLG" = 'Y'.
- Requirements: any ArcGIS license should work.

California State Data: Cal-Atlas

- Advantages: just for the state, so large scale is a bit easier to find
- Disadvantages: a bit confusing...
- Internet Source: Search 'Cal-Atlas' or go to <http://www.atlas.ca.gov/> then into **Download**.
- CERES Geofinder. At Cal-Atlas, go to **Geofinder, then GIS datasets**
<http://casil.ucdavis.edu/cgi-bin/gf/geofinder#>

ESRI Data (on the installation disks, also on our server: ESRIdata/DataFolder

USDA Geospatial Data Gateway

- Soils data mostly, but there are other things.
- Source: <http://datagateway.nrcs.usda.gov/GatewayHome.html>

National Geologic Map Database (USGS)

A good resource, though most of the maps are not georeferenced. Many large scale maps are USGS Open File Reports, hand drawn, etc. A wonderful resource, but you would need to georeference them. Most are drawn on a topographic base, so this isn't very difficult since there are registration points visible (mostly NAD27).

http://ngmdb.usgs.gov/ngmdb/ngm_catalog.ora.html

Scanned USGS quad DRGs from Cal-Atlas

- Using either a printed topo quad, or the US Topo(s) you downloaded, get the latitude and longitude of the *southeast corner* of your quad to derive the five digit number that combines the integer values: e.g. latitude 38 something, longitude 122 something makes 38122. Then use this diagram to work out the grid cell code for the particular quad, one of 64 in the one degree rectangle, again relating to the southeast corner.

latitude _____ ° _____ ' _____ ''
 longitude _____ ° _____ ' _____ '' :
 5 digit: _____ grid cell: _____

- Go to the Cal-Atlas download center <http://www.atlas.ca.gov/>, then **Download**, then **Imagery, Basemaps, Land Cover**. There we'll go to **Basemaps** for DRGs.
- Start with the **DRG**: In the **Base Maps** folder, go to **drg**, and then **7.5_minute_series_albers_nad83_trimmed**. Then go into the appropriate 5-digit code folder and using the grid cell you've determined for , download the .tif and .aux file for your quad by **right-click-Save As...** to put it in your folder. For example, for Bolinas, with a latitude 37°52'30" and longitude 122°37'30", it's o37122h6.tif and o37122h6.aux. You should see the name of the quad shown, which will help avoid selecting the wrong file.
- Repeat for other quads you need. These are all projected to California Teale Albers, but we won't need to reproject them since we'll use them as a backdrop the same way we use imagery. They are useful for getting names of features not easily found or not existing in the NHD and transportation data. We can also read elevations (though they're usually in feet) and see other topographic mapping patterns like a forest vs grass vs scrub vegetation.

