Data Sources for GIS in Water Resources
GIS in Water Resources, Fall 2016

GIS is all about working with data and maps, so it's important to know where to obtain this information. A number of repositories of GIS data have been established at the national level in the United States, in Texas, Utah and for the world. ArcGIS Online offers a new way of finding maps and data services that you may find helpful for your studies in this class. This synopsis is accompanied by a separate document that lists data sources and provides internet links to them, so the scope of this synopsis is confined to a brief description of some of the key information sets that we will work with. There are many other information sources that you may find as you do research for your term projects. If you find an information source that you think would be helpful to others in the class, it would be helpful if you could make a map that illustrates this information source, and post your map into ArcGIS Online so that other students in the class could be informed about this information source as well. When you do this, please let your instructor know about that so that we can highlight additional information sources as they emerge during the semester.

The hierarchical system of watersheds and basins is indexed with a system of *Hydrologic Unit Codes (HUCs)*, and the 8-digit HUC's, or *sub-basins* are the most widely referenced drainage areas for GIS data – there are a little more than two thousand of them in the continental US, about the same as the number of counties. The *Watershed Boundary Dataset* is the official set of watersheds and basins of the nation.

The most important information sets we need are those that describe the water system of the United States and its connection to the land system, in other words, what are the rivers, streams, lakes, bays and aquifers of the US and how are they connected with one another and with the landscape in which they are imbedded? A point of departure in this exploration is the *National Hydrography Dataset (NHD)*, which is a GIS dataset of the "blue lines" on topographic maps of the US. It includes streams, rivers, and water bodies of various kinds, an in particular a feature class called *NHDFlowline* that can be made into a geometric network of the stream and river system, including linear segments through lakes so that water movement can be traced on the network through water bodies as well as along rivers. There are two NHD datasets for the US – at 1:100,000 scale, and at 1:24,000 scale, called *medium resolution* and *high resolution* NHD, respectively. These correspond to the scales of the two major series of paper topographic maps that the USGS has published with national coverage of the United States. The EPA supplied additional information to the NHD concerning the upstream and downstream connectively of the surface water flow system. Each distinct feature on the flow system is called a *reach*, and a *river addressing* system exists for location points and lines using *reachcodes*. For the medium resolution NHD, a new dataset has been compiled called *NHDPlus*, in which the local area whose water drains to that reach has been delineated, termed a *reach catchment*.

The land surface elevation of the United States is described by the *National Elevation Dataset* (NED), which is distributed as a grid in geographic coordinates with cells at 1, 1/3 or 1/9 arc second spacing. These cell sizes correspond roughly to 30 m, 10 m and 3 m when projected, depending somewhat on latitude as you will learn in the class on Geodesy. It is important in working with NED data in hydrology to first project the data to a coordinate system appropriate for the area you are working so that cells are square and slope and area information can be properly calculated. There is 1 and 1/3 arc second coverage for all the continental US, 1/9 arc second coverage for some parts of the country where more detailed terrain information has been collected, such as from LIDAR. Perhaps the greatest innovation that GIS has contributed to water resources knowledge is the capacity to automatically process digital elevation model (DEM) data to a delineate watersheds and stream networks. This is accomplished by defining the *flow direction* from one cell to the next, and then tracing these flow patterns, cell to cell, across the entire landscape. In this way, all the upstream cells whose drainage flows through a given cell can be counted, called *flow accumulation*. DEM streams may be defined where the flow accumulation exceeds a threshold value, and watershed areas are identified as the zone of cells whose drainage passes through a distinct outlet cell.

Globally, the land surface terrain of the earth is similarly defined on a DEM of 1km resolution called GTOPO30 – 30 seconds of arc corresponds to approximately 1km on the land surface, and by a DEM of 90 resolution produced from the *Shuttle Radar Topography Mission* (SRTM). The *Hydro1K* dataset is a global delineation of watersheds and stream networks produced from GTOPO30, and *HydroSheds* is a similar dataset derived from the SRTM data.

Besides terrain and hydrography, a number of other themes are important for GIS in Water Resources. Among them are the *SSURGO* soils description of the nation at 1:24,000 scale, a digital compilation of the county soil maps compiled by the US Dept of Agriculture, and the *STATSGO* soil coverage at 1:250,000 scale, a set of state-wide digital soil datasets for the nation. Land cover is described by the *National Land Cover Dataset* an interpretation of Landsat remote sensing imagery to define a set of distinct land cover types. Water observations data, such as streamflow, groundwater levels and water quality, are obtained from the USGS *National Water Information System*. The maps describing observation sites are well suited for GIS representation but the time-indexed observational datasets are less so, and incorporating time series data within GIS is still a challenge. Forecasts of flow on the continental stream network are obtained from the *National Water Model*. Besides physical information, GIS is also useful for *demographic information*, such as from the US Census and the American Community Survey. These data are cataloged by Census blocks and block groups.

A trend in providing access to data on the internet is *data services*. These get the browser out of the way and deliver data directly to the software being used. In this class, we will learn how to access geographic data sources as data *services* rather than as datasets. An *image service* is a rapidly accessible picture of a spatially continuous dataset such as elevation, land cover or soils, where what you see on the screen is just an image, but behind that, in the computing cloud, are the data themselves, aggregated across the United States, and in some cases across the world, so that *geoprocessing services* can be performed upon them, as if they were on your local computer. We'll use these services to do watershed delineation and to summarize watershed attributes. A *dynamic map service* similarly creates an image map of vector GIS data and permits extraction of vector datasets ("Extract Landscape Source Data") for a limited region of space. We'll use this approach to obtain data for the National Hydrography Dataset in this class.

The transition from datasets to data services that is currently going on in GIS is based on *cloud computing* where computation is done on remote computer systems that are highly standardized computational and data storage machines. As in other fields, aggregation of information in a few, centralized systems is becoming a trend that is facilitating integration of information from disparate fields.