Introduction to ArcGIS Software
Synopsis of Class 2, GIS in Water Resources, Fall 2012

The ArcGIS Geographic Information System is developed by the Environmental Systems Research Institute (ESRI) of Redlands, CA, and it is by far the most widely used GIS in the world. Key people in the development of ESRI are Jack Dangermond, the President, and Scott Morehouse, the Chief Software Engineer. Over the past 30 years, they and their colleagues have created a GIS software and data framework that continues to evolve and develop – the latest evolution being ArcGIS Online, a web-based repository for geospatial information and mapping that is accessible from many devices besides ArcGIS Desktop. The videos from the 2012 ESRI User Conference Plenary session give a vivid sense of the current state of GIS and also of its future direction:
http://www.esri.com/events/user-conference/agenda/plenary-videos.html

Central to the vision of GIS is the idea of story-telling using maps. *Data* in tables are made more compelling *information* by being visualized in maps, thematic integration of map information from many sources creates *knowledge*, sharing this knowledge and collaboration among people improves *understanding*. Good geoinformation products provide timely information, disseminate knowledge, communicate importance, show status and performance, support decision making, illustrate change, and show the future. A philosophy of *Geodesign* is emerging to leverage geographic information, resulting in designs that more closely follow natural systems. There might in the future be a similar approach for *Hydrodesign*, leveraging geospatial and temporal information to design sustainable homes, communities, industries and societies with respect to their dependence on water.

The intellectual architect of ArcGIS at ESRI is Scott Morehouse, and he writes "all geographic information systems are built using formal models that describe how things are located in space. A formal model is an abstract and well defined system of concepts. A geographic data model defines the vocabulary for describing and reasoning about the things that are located on the earth. Geographic data models serve as the foundation on which all geographic information systems are built." (from the Preface to "Modeling our World", Zeiler, 1999).

ESRI has developed several such models for describing geographic information. The first, which originated with ArcInfo, is the *coverage* model that stores geographic data in binary files (Arc), and attributes in the Info database. Some GIS data is still distributed in this form as files with a .e00 extension. The next, introduced with ArcView, is the *shapefile* model, which is a looser description of points, lines and areas, and whose file format is open source. This has become the most widely used mechanism for distributing GIS data. The third, introduced with ArcGIS, is the *geodatabase*, which is a much more highly structured data storage system that includes feature classes, tables, grids, geometric networks, and relationships. A geodatabase is a customized version of a relational database system, such as Access, SQL/Server, or Oracle, that are widely used for storing purely tabular information. Finally, a *file geodatabase* format has been developed that is the standard for sharing GIS data through ArcGIS Online. This has the same structure as the regular geodatabase but does not require that the local computer have a relational database resident on it in which to store the geographic information.

The fundamental form of storing information in a relational database is a *table*, which has rows or *records*, and columns or *fields*. Each field describes a particular item of information, or *attribute*, and it has a column header with a *field name* to identify that attribute. Each record is a collection of values of the attributes, indexed by the attribute value in the *key field* that serves as the unique identifier of that collection of values. The geospatial coordinates of a point, line or area are stored in a field named *shape*, and the resulting table describes a *feature class*, with each record describing a single *feature*. If the table has no shape field, it is called an *object class*. Information in two tables can be *related* by associating the common values of their key fields. Note that this concept of a database table is very different from a spreadsheet table that can contain formulas in the cells and has implicit computational capabilities. Database tables just have data stored in them – all the computation is done by external functions operating on these data. Data storage and data computation are kept strictly separate in GIS.

You interact with ArcGIS Desktop through *ArcMap*, which is an interface that allows you to view data in maps and query the associated information in attribute tables. Linked within ArcMap is *ArcCatalog*, which provides an overview of the contents of a geodatabase, and mechanisms to add new data to it, and *ArcToolbox*, which provides access to folders of tools that perform geographic operations. As you become familiar with working with GIS data, you'll see that the mechanism of dealing with these data is quite different than other computational methods that you may have used before. In GIS, data processing is incremental. Each tool carries out a single, specific operation that requires a particular kind of input dataset, and produces a modified version of this dataset as its output. Tools can be executed in a sequence called *geoprocessing* that can be carried out through a visual interface called *Model Builder*. We'll use Model Builder in an exercise later in the semester. ArcGIS Desktop is the primary GIS that you'll use this semester. Some of the information that you will use will be drawn from ArcGIS Server, which provides web access to a repository of geospatial information and maps, and we'll also use base maps and geospatial information drawn from ArcGIS Online, which is a centralized cloud repository for maps and geospatial information services. You can interact with ArcGIS Online directly using a web browser at http://www.arcgis.com

ArcGIS has several add-on packages that we will use this semester, including Spatial Analyst and 3D Analyst. *Spatial Analyst* is used for raster analysis of grids, and it includes a toolbar with commonly used functions and a toolbox with more extensive routines, including a set of standard functions for hydrology that delineate watersheds and stream networks from digital elevation models that we'll use in this class. Spatial analyst also includes tools for interpolation of point observations into surfaces that many students in this class find useful for constructing maps of spatially distributed hydrologic phenomena, such as rainfall maps, and maps of contaminant concentrations in water bodies. In ArcGIS, grid, or raster information is comprised of square cells only, arranged in rows and columns, where the number of rows and columns can be different so as to create a rectangular domain of study interest. If your study domain, such as a watershed, does not have a rectangular shape, it can be bounded with a rectangular box, and the cells outside the watershed designated as *NODATA* cells that have no valid value, and do not participate in geoprocessing operations. A special case of raster data are *Images*, which have integer values over a defined range, such as 0-255, used in reporting information from remote sensing from satellites and aircraft. *3DAnalyst* is used for working with surfaces constructed using Triangulated Irregular Networks, as arise from LIDAR (LIght Detection And Ranging) measurement of land surface terrain, and also from land surveying and aerial photogrammetry.

Reference: Zeiler, M., *Modeling our World: The ESRI Guide to Geodatabase Concepts*, ESRI Press, Redlands CA, 1999.