Considering Spatial Data

GIS I: Organizing Principles

CEHI
Children's Environmental Health Initiative

Discuss components of spatial data

Explain differences between the raster and vector spatial data models

Present common types of spatial data you may encounter and use

Introduce you to ArcCatalog as a means of organizing and managing your spatial data

Discuss best practices regarding the organization of spatial data, as how you organize your spatial data will either help or hinder your work

What is Spatial Data?

Data with a geographic component:
- Latitude/longitude
- Street address
- Zip code
- County
- State

Data with any location component.

Including but not limited to:

If you really think about it all data are spatial.

Radio call signs are spatial data: the K and W tell us which side of the MS the station is broadcasting from

In GIS your data are stored in a **geographic database**

A geographic database contains spatial information (location) and descriptive information of real world entities.

Data are stored in different models. We will discuss 2 of the main data models, raster and vector.

Modeling the World in GIS

- Two primary methods for storing your spatial data

  - Vector stores shape information using coordinates
    - Example: street

  - Raster stores information in square cells
    - Example: photograph
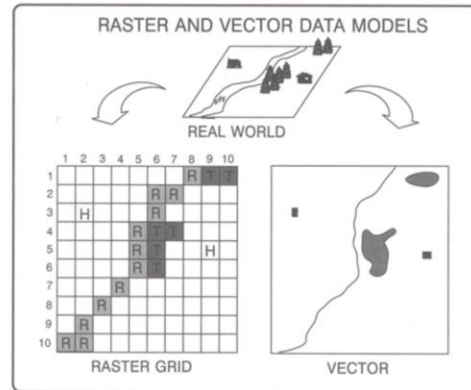
Two data models commonly used to represent spatial data in GIS are the raster and vector data models

Within the **vector data model**, a representation of the world is created using lines, points, and polygons. Vector data is focused on modeling discrete features with precise shapes and boundaries.

-A home becomes a point

-A river becomes a line

-A stand of trees become a polygon

In contrast, **within the raster data model,** a representation of the world is created as a surface divided into a regular grid of cells. Raster data is focused on modeling continuous phenomena and images of the Earth.

Raster data storage uses square cells to model points, lines, or areas continuously. Raster data storage is common for representing: air quality and crime density.

-A home becomes a cell with a value that represents home

-A river becomes a cell with a value that represents river

-A stand of trees becomes a group of cells all with a value that represents tree

Raster data are standard for remotely sensed data, imagery data. In ArcGIS, satellite imagery is commonly stored as raster data.

Vector Abstraction of Real World Entities

- In vector-based spatial data, real world entities are abstracted into three basic shapes
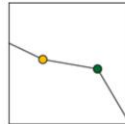
    - Points:
        - facility locations
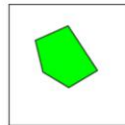   - patient homes
        - fast food restaurants
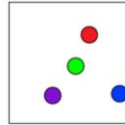
    - Lines:
        - roads
        - rivers/streams
        - movements

    - Polygons:
        - census geographies
        - lakes/ponds
        - parcels/buildings

To represent our world within a computer GIS, creates **geometric representations** of reality.

There are three basic geometric shapes used: points, lines, and polygons.

These shapes are sometimes called geometric objects, geometric features, or feature types.

ArcMap supports a number of vector data formats for storing point, line, and area features. We will focus on the most common, the **shapefile**. But shapefiles are not the only vector format…

Additional vector data formats include the GeoDatabase and KML (keyhole markup language), which is used to display geographic data in a mapping application, such as Google Maps for mobile

Although we will not cover them in detail, ArcMap supports several raster data formats for storing point line and area features.

There are a number of Geoprocessing tools associated with the raster data model within ArcGIS as well.

Raster formats:

- Esri Grid
- ERDAS Imagine
- GeoTiff
- MrSid

Data Format Focus: Shapefile

A Shapefile is a set of related files:

Mandatory:
shapefile.dbf-stores attribute data
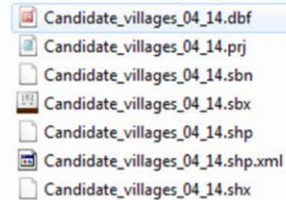shapfile.shx- index file
shapefile.shp- describes shape

Optional but common:
Shapefile.prj- stores coordinate system information
Shapefile.sbn- spatial index part 1
Shapefile.sbx- spatial index part 2
Shapefile.xml- stores metadata

What you see in the ArcGIS environment:

Candidate_villages_04_14.shp

Shapefile

Files shown in Windows environment:
- Candidate_villages_04_14.dbf
- Candidate_villages_04_14.prj
- Candidate_villages_04_14.sbn
- Candidate_villages_04_14.sbx
- Candidate_villages_04_14.shp
- Candidate_villages_04_14.shp.xml
- Candidate_villages_04_14.shx

As stated previously, the shapefile is the most widespread. So, what is a shapefile? It's a collection of related files.

On the right we see what one shapefile: Candidate_villages looks like in the windows environment

On the bottom of the slide we see what the same shapefile looks like in the ArcGIS environment

Index files allows for accurate linking between attribute and shape files (data) (via a unique identifier)

This common spatial data format is widely used and recognized making it useful for collaborative/cooperative projects.

Spatial Data Sources

Established spatial data providers:
•**US Census**
http://www.census.gov/geo/maps-data/data/tiger-line.html

•**Health Resources and Services Admin**
http://datawarehouse.hrsa.gov/

•**Local GIS resources**
http://www.mngeo.state.mn.us/chouse/
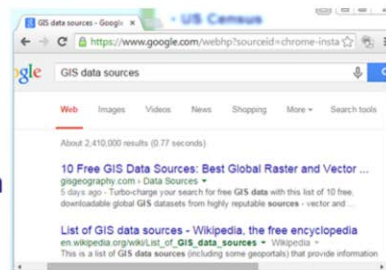https://gis.ny.gov/
http://imap.maryland.gov/
http://planning.hawaii.gov/gis/

Where do you find spatial data?

Screen shot of a google search on the right which came up with over 2 million hits in less than a second when we searched "GIS data sources", there are many places to obtain spatial data. However, not all are reliable.

It is important to use those that are reliable and accurate such as the US census, Health resources and services admin, or local GIS resources.

The are many options; start with those you know….
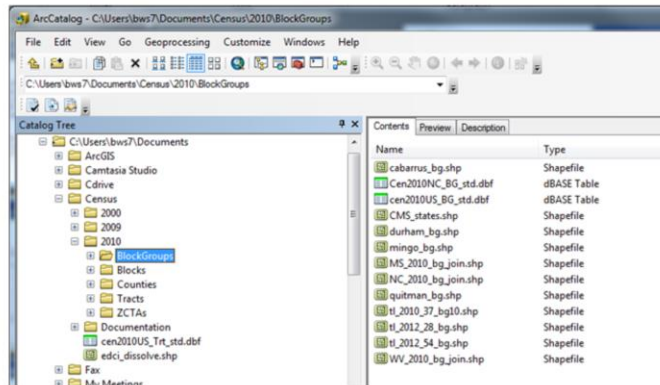
On the topic of obtaining and storing spatial data such as Shapefiles, we must mention and discuss ArcCatalog.

ArcCatalog is part of ESRI's desktop suite of tools that allows you to browse, organize, distribute, and document GIS data. The ArcCatalog interface allows you to learn about your database contents quickly and easily.

Similar to Windows Explorer, ArcCatalog permits easy browsing of your geographic, text, image, and tabular datasets.

Further, ArcCatalog allows for easy data management such as moving, copying and deleting files. You saw above how many files can make up a shapefile. Due to the number of files, it can be difficult to move, copy, and delete all the files at once.

Additionally, if one of the mandatory files is missing/doesn't make the move, your shapefile will not work in ArcGIS. Thus, we recommend using ArcCatalog to manage, move, copy, and delete your files as only one file appears as a shapefile (and thus all associated are deleted/moved at once).

Lastly, ArcCatalog allows you to document your data (i.e., metadata).

There are four ways to view your data in ArcCatalog: the **Contents**, **Preview Geography, Preview table,** and **Description** tabs.

You will get a chance to see how ArcCatalog works during the exercise!

Organizing Data: Folder Structure
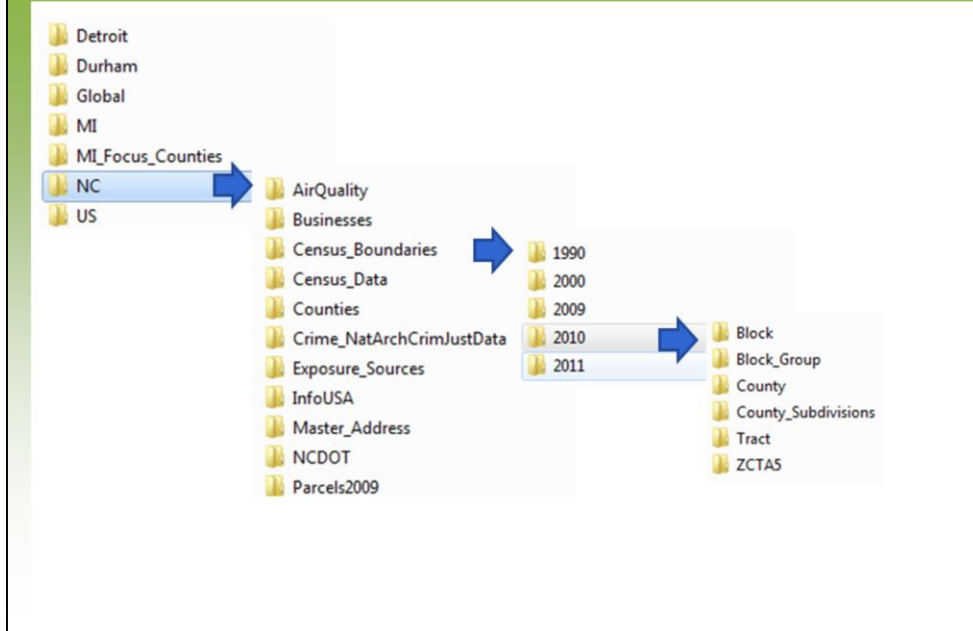
As you work with spatial data, you will begin to accumulate "reference" shapefiles, such as counties in your state, ZCTAs, in your state, block groups, blocks, etc. With regard to your folder structure, we strongly suggest that you consider an organizational logic that takes advantage of your spatial data. It is crucial that you organize your reference spatial data (i.e., shapefiles) in such a way that allows for easy access and understanding.

We organize our reference spatial data based on geography. Data can be at different geographies. Shapefiles can be at the state, county, zip code, block, or neighborhood level.

We have air quality, business, census information, etc. for the state of North Carolina. Within the NC folder, we have a folder named "Census_Boundaries".

Within the "Census_Boundaries" folder are….

Different years, as census boundaries may change over time.

Then, organized within each year are different geographies. Each folder contains the shapefile that corresponds to the year and geography.

This organization allows for CEHI employees and collaborators to know exactly where to find data and shapefiles! There is no time lost searching.

Naming Folders and Files

- **Keep names short; include key information:**
  - **Geography (NC, Durham County, Tanzania)**
  - **Time stamp if applicable**
  - **Subject (Parcels, PrmryRds)**

- **Consider standard abbreviations (County=cty)**

- **Avoid using the words "final" or "new"**

- **Do not accept ArcGIS default names (Export_Output, Geocoding_result)**

- **Establish standards for the order of elements**

Establishing a standard naming convention can be a major contributor to the file's usability. Taking a careless approach to naming and organizing files and folders can add unnecessary complication to the tasks of data sharing and collaboration.

Keep names as short as possible but include key information. Consider including following information in name (while keeping it short and simple): geography, time, and subject.

Consider using abbreviations (Coutny=cty, Tract=trt, etc.).

Avoid using the words final or new, as any future updates make these names confusing.

Do not accept ArcGIS default names. Include key information instead.

Consider establishing standards for the order of the elements.

Here are some examples.

Remember: the naming convention allows you to keep track of files. Putting in the time and effort up front to ensure organization will result in less complications when data sharing and collaborating on files.

For further reference about CEHI's standards for file and folder naming conventions see the Resources folder on the Box site.