Building Blocks of GIS

Shapefiles, Geodatabases, and Services

Tim Haithcoat

University of Missouri – GRC/MSDIS

with resources from Rowan University

and from ESRI – Jack Dangermond NGAC presentation

Geodatabase vs Other Formats

- Coverages and Shapefiles store geospatial and attribute data in different locations in different formats
  - .shp (proprietary binary format)
  - .dbf (dBase database format)
- Geodatabases store both geospatial and attribute data in the same structure

Benefits and Drawbacks

- Benefits
  - GIS data can now be handled like most other data, and stored in a RDBMS
  - Greater flexibility and functionality
  - “Enterprise” level of managing data
- Drawbacks
  - Speed hit
  - Even more rope to hang yourself with

ESRI Geodatabases

- File Geodatabase
  - Introduced in 9.2, the File Geodatabase is the latest, greatest file-based format from ESRI
- Personal Geodatabase
  - Introduced in 8.x
  - Based on Microsoft Access/Jet Engine
- ArcSDE
  - Software (now part of ArcGIS core) that allows RDBMSs to act as GIS data stores.

Personal Geodatabase

- Based on Microsoft Access
- Great for bringing outside data into ArcGIS
- Limited to 2GB
- Becomes slow as amount of data increases
- Stores data in one file called *geodatabase*.mdb

File Geodatabase

- Latest format
- Best modern format for large datasets
- Very efficient use of storage space
- What you should be using for significant work
- Stores data on disk in several files within a directory named *geodatabase*.gdb

ArcSDE/Enterprise database

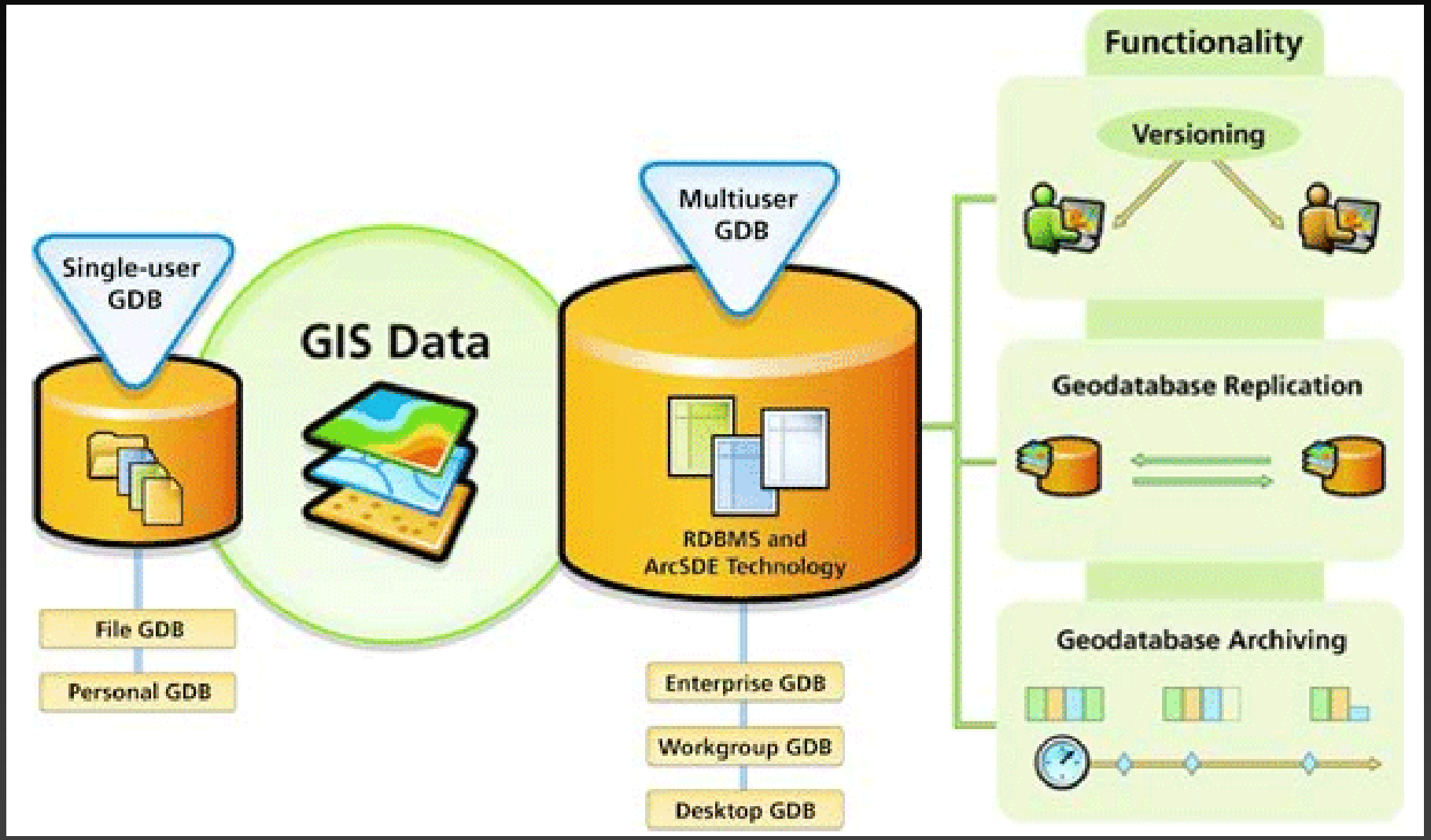
- Most likely stored on a different machine from the one you're using ArcGIS on
- Same basic functionality as other GDBs
- Can store versions of the GIS features, allowing you to see changes over time
- Concurrent users (multi-user and replication)
- Managed (hopefully) by a DB administrator

Transactions

- Geodatabase edits are either committed or rolled back
- Edits performed in a multi-user environment are integrity checked
- Atomic-level editing and revisioning
- Needed to prevent a race condition

Versioning

- GIS tracks edits made and maintains a journal of all changes to the database
- This record keeping allows for roll backs to any date on record
- Keep one set of records while reverting another
- Same database methodology as Wikipedia



Geodatabase Types

Working with Geodatabases

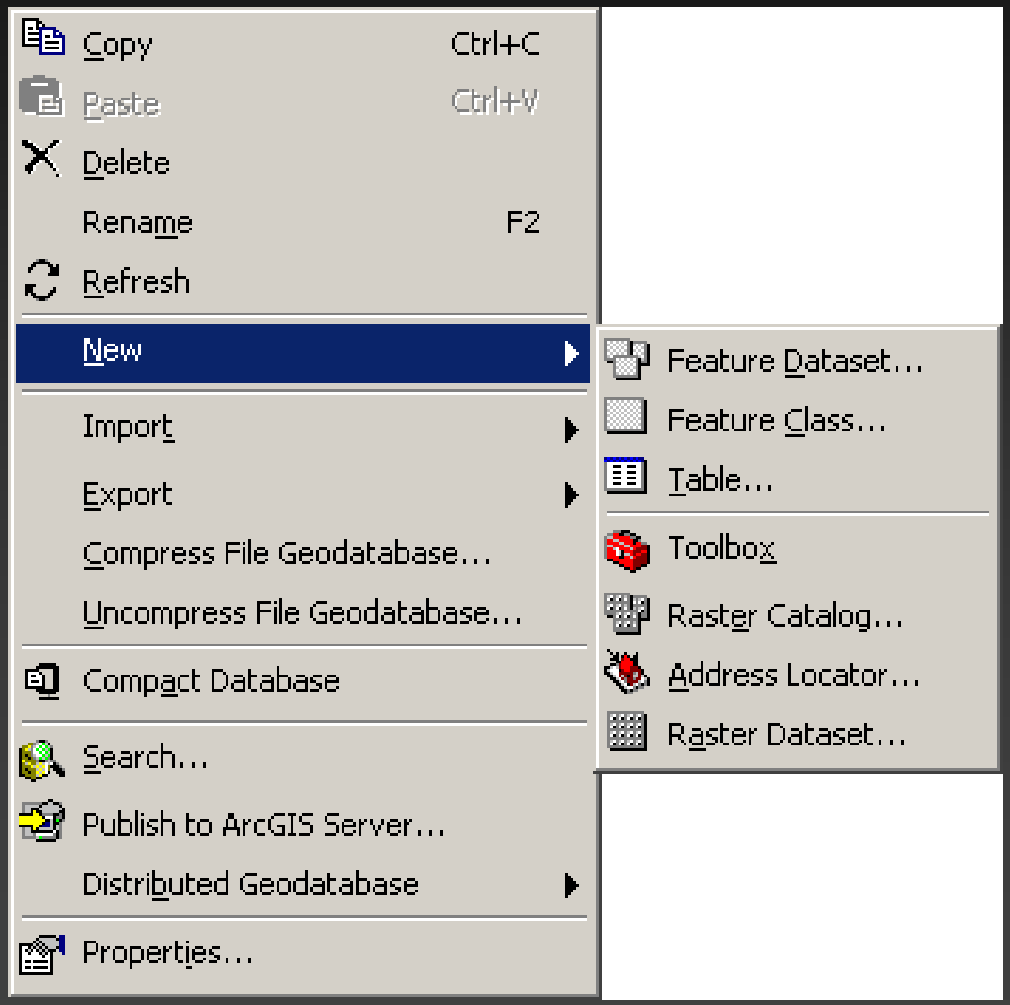
- At a minimum, consider it similar to a subdirectory with shapefiles
- Unlike shapefiles, you can enforce extents, storage types, projections, topology rules, connectivity rules, network-specific rules, and so on
- This additional functionality is implemented through Feature Datasets

Feature Datasets

- A “folder” within the GDB, it preserves projection and extent information for data within the folder (“feature classes”)
- To make it useful, you must set extent and projection information
- Put some forethought into it before specifying projection and extent!

Feature Datasets

- After creating a GDB, right click and choose New >
  Feature Dataset
- The dialog boxes will step you through setting the variables for the Feature Dataset

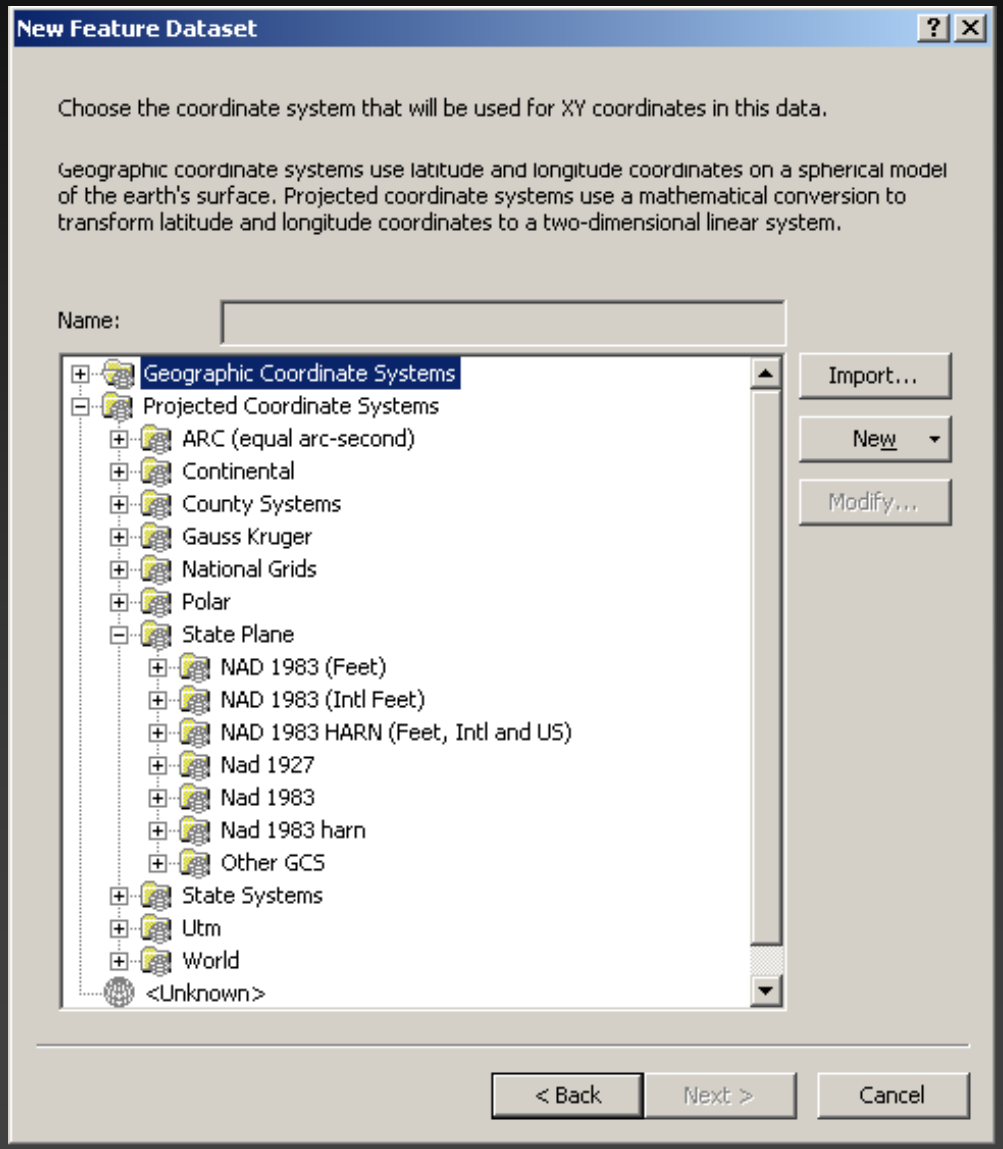


Importance of Extent

- The Geodatabase will only bother with the information within the extent
- It will throw an exception if you attempt to put something that doesn't fit in the box
- ArcGIS can preserve the difference between two points down to the molecular level
- Setting the extent allows you to control the precision at which ArcGIS handles data
- Needlessly too precise, and you'll have errors that'll never show up on the screen, but will still impact your data, analyses, and speed

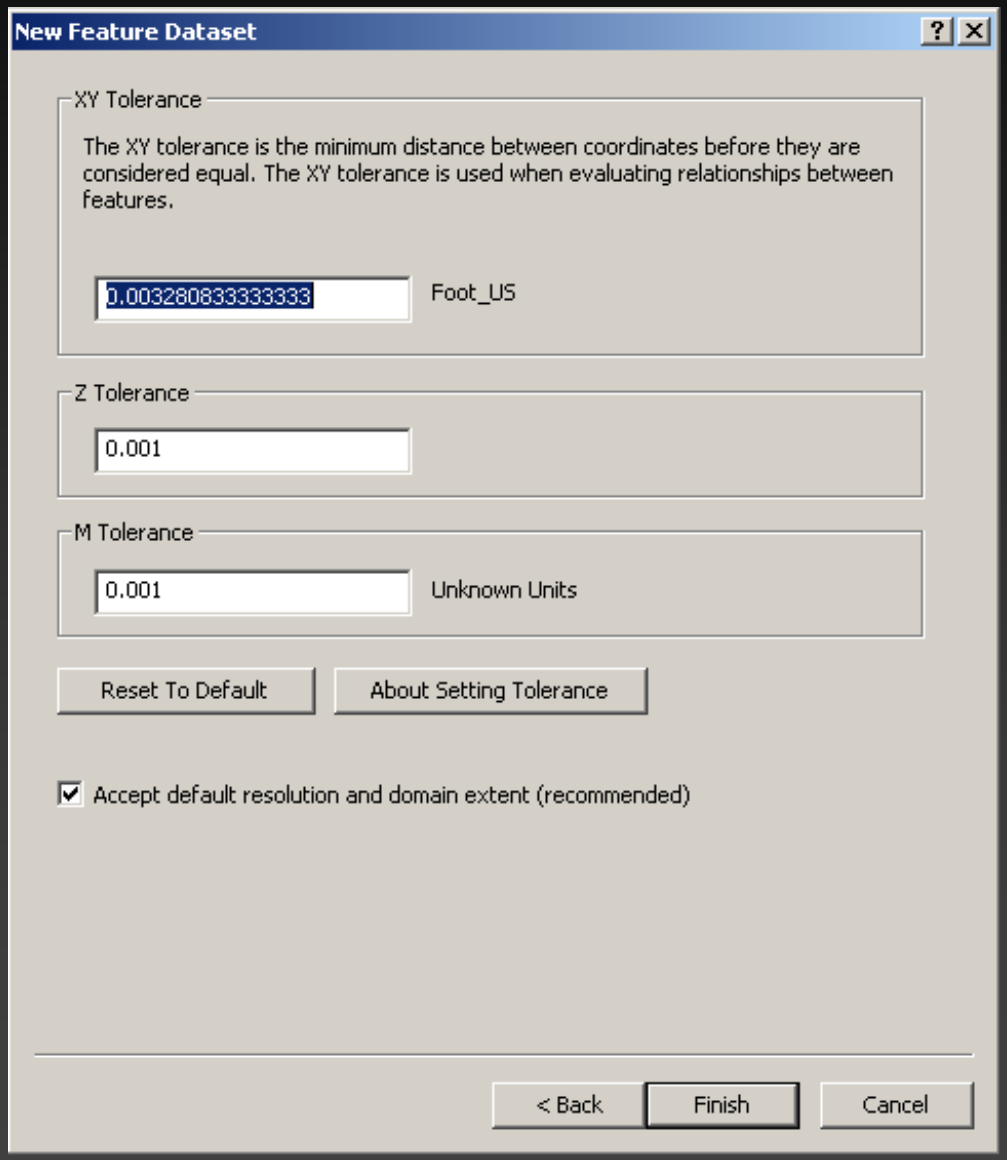
Defining Missouri

- Projection:
  MO State Plane (feet)
- Extent: ?
  - Should it be tight?
  - Should it extend outside the boundaries?



Defining Missouri

- In this case, Arc defaults to a grid of 0.00328 feet
- Roughly 4/100ths of an inch
- About a hair's width
- 0.02 feet is slightly smaller than 1/4"

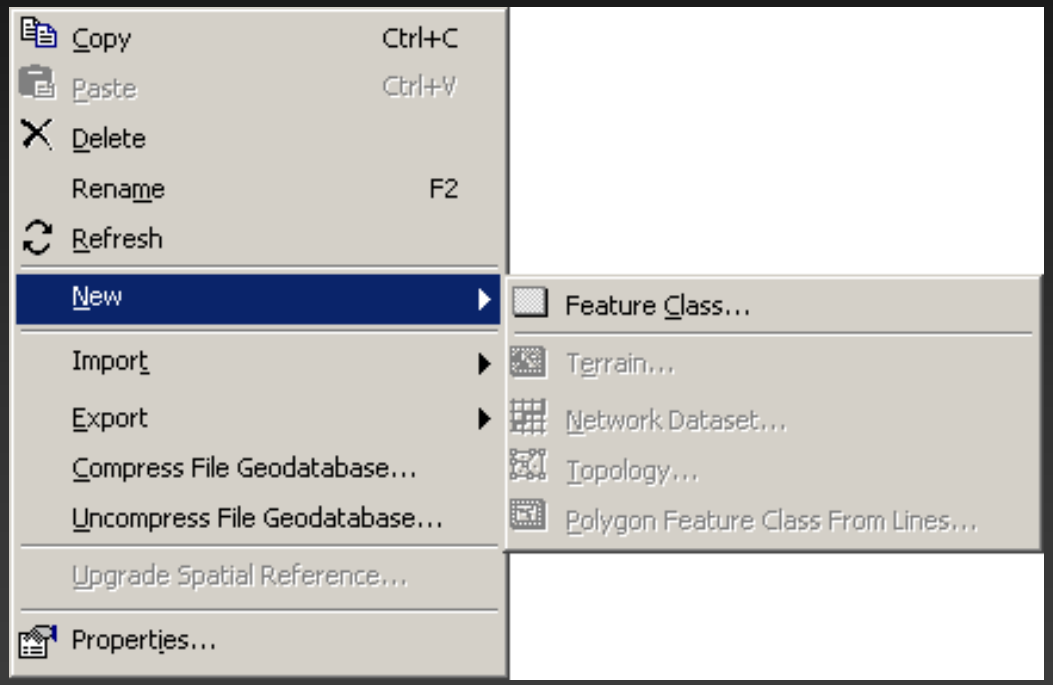


Balancing Precision and Functionality

- Have your extent match the scale at which you are / plan to work
- Leave a little wiggle room
- Working in Missouri? Some of IL, KY, TN, AR, OK, KS, NE, and IA should fall into your box.
- If Greenland fits? Your box is a bit too big.

Additional Functionality

- In your Feature Dataset, right click and see what pops up under New >
- Topology
- Geometric Network
- Network Dataset
- Terrain
- Etc…



Geodatabase as a Container

- Each of these “special” datasets uses the GDB to store data specific to its framework
- Topology stores associated attribute tables, rules, and error information
- Network stores network edge attributes, turn tables, and driving/routing directions

Normalization

- A normalized database is one that has little redundancy within its tables
- Record ID or some other key links to a table with those values
- Instead of storing “Modified Agricultural Wetlands” numerous times as text, store it once as text and refer to it using a key (2140)

Normalization

- Work in a normalized environment
- Analogs:
  - Non-normalized: Excel Spreadsheet
  - Normalized: well made Access DB (lookups)
- When distributing for the public, “flatten” the database out to one table per layer and make it a shapefile

Geodatabase Environment

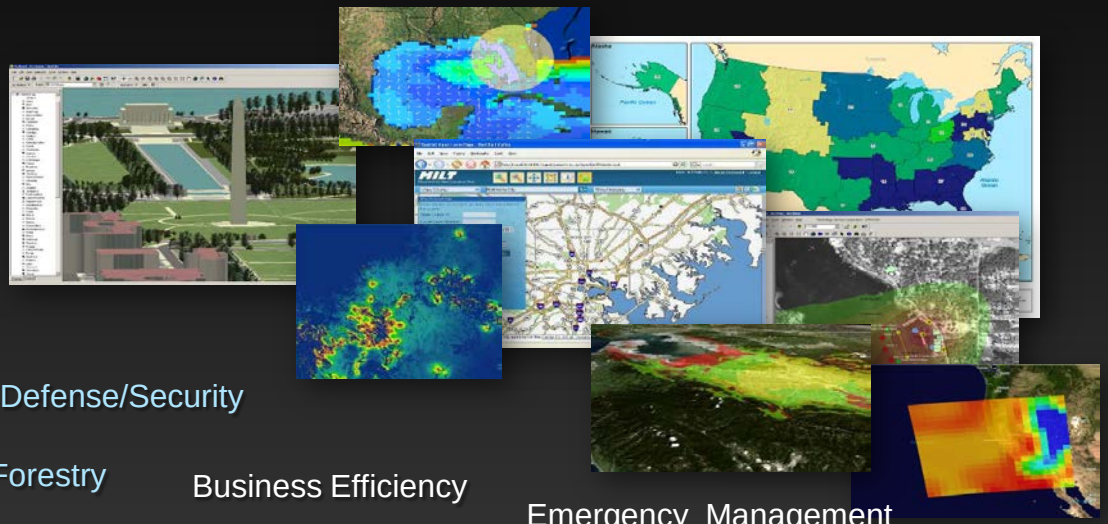
- Important to work in a GDB whenever possible
  - Assured extents, projections, etc
  - Quality control
  - Greater number of tools at your disposal
- Export to other format (.shp) for distribution

Data, data, everywhere

- In the Internet age, massive amounts of data are compiled, transmitted and analyzed every second
- Understanding the storage and retrieval methods are critical
- Difference between drinking and drowning

Today GIS *Is Very Valuable to Government*

Used in Many Integrated Applications



Conservation

Science/Modeling

Land Management

Natural Resources

Congestion

Crime

Defense/Security

Pollution

Water

Agriculture/Forestry

Business Efficiency

Emergency Management

Economic Recovery

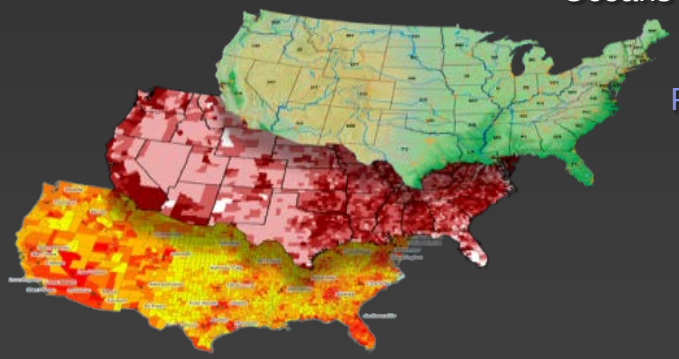
Energy

Biodiversity

Global Warming

Human Health

Oceans



Land Use

Population

Education

Logistics

Energy

Public Safety

Development

Urbanization

Humanitarian Relief

Facility Management

National Security

Environmental Protection

Law Enforcement

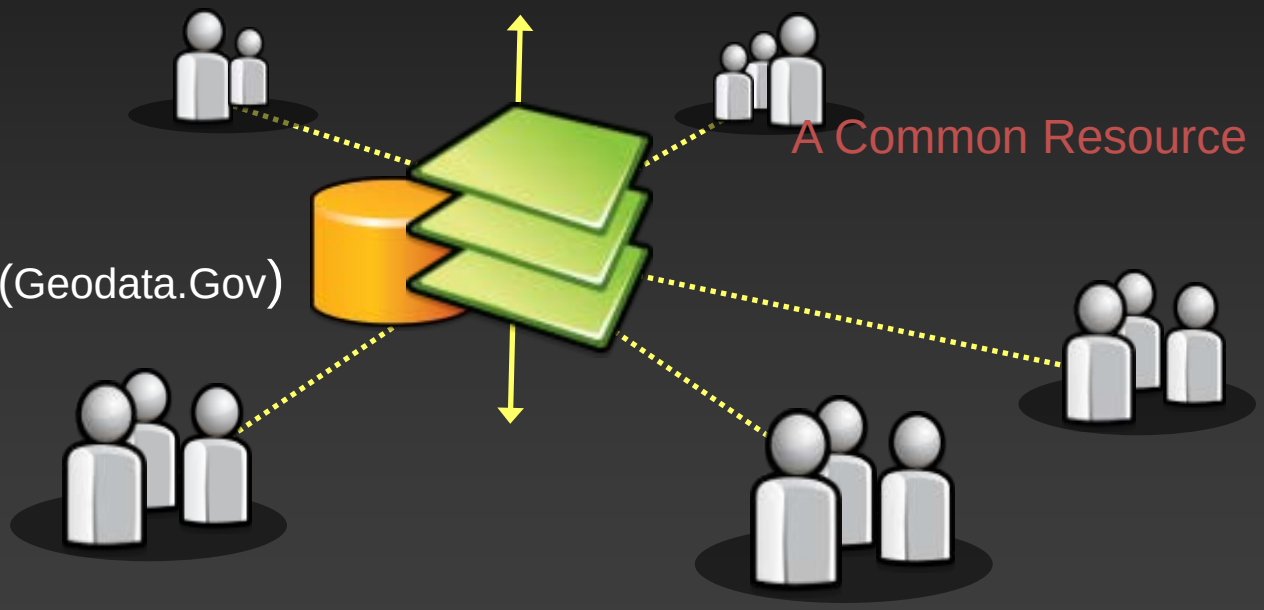
Improving Planning, Management and Decision Making . . .

. . . Providing Critical Infrastructure

Systematic Data Creation & Sharing . . .
. . . Has Been a Key to This Success

Many Agencies Share Data as Part of their Core Mission

(Census Bureau / USGS / USDA . . .)

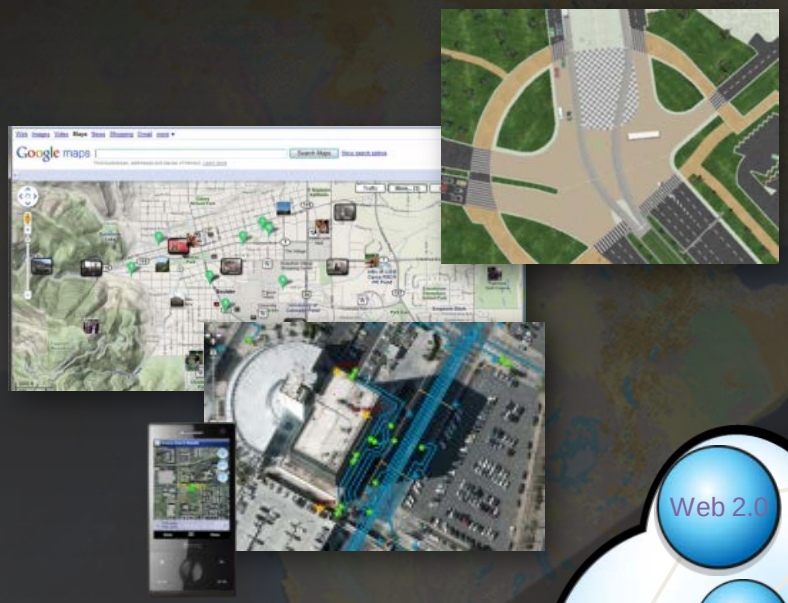


Data Clearinghouses Have Improved Access . . .

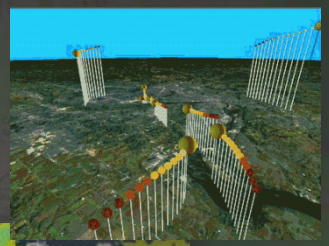
. . . But Haven't Resulted in a Geospatial Framework

GIS Technology is Changing - Becoming Internet Focused

Richer, Easier & More Pervasive
New Styles, Patterns & Techniques



New Services



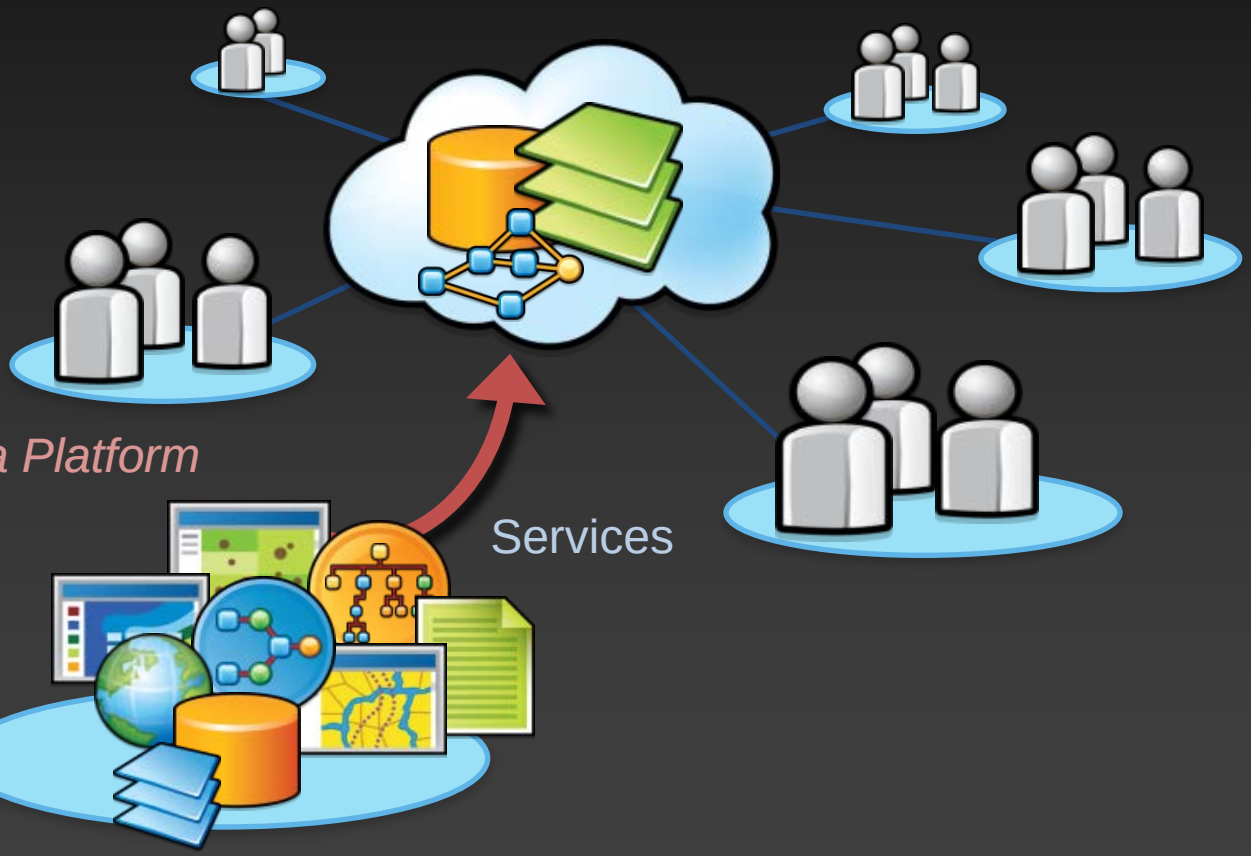
New Media

Many Opportunities for Spreading Geography Everywhere

Governments Are Beginning to Create Geo-Services

Providing Building Blocks of NSDI

Standardized
Open & Interoperable
Easily Integrated
Easy to Use
Application Ready

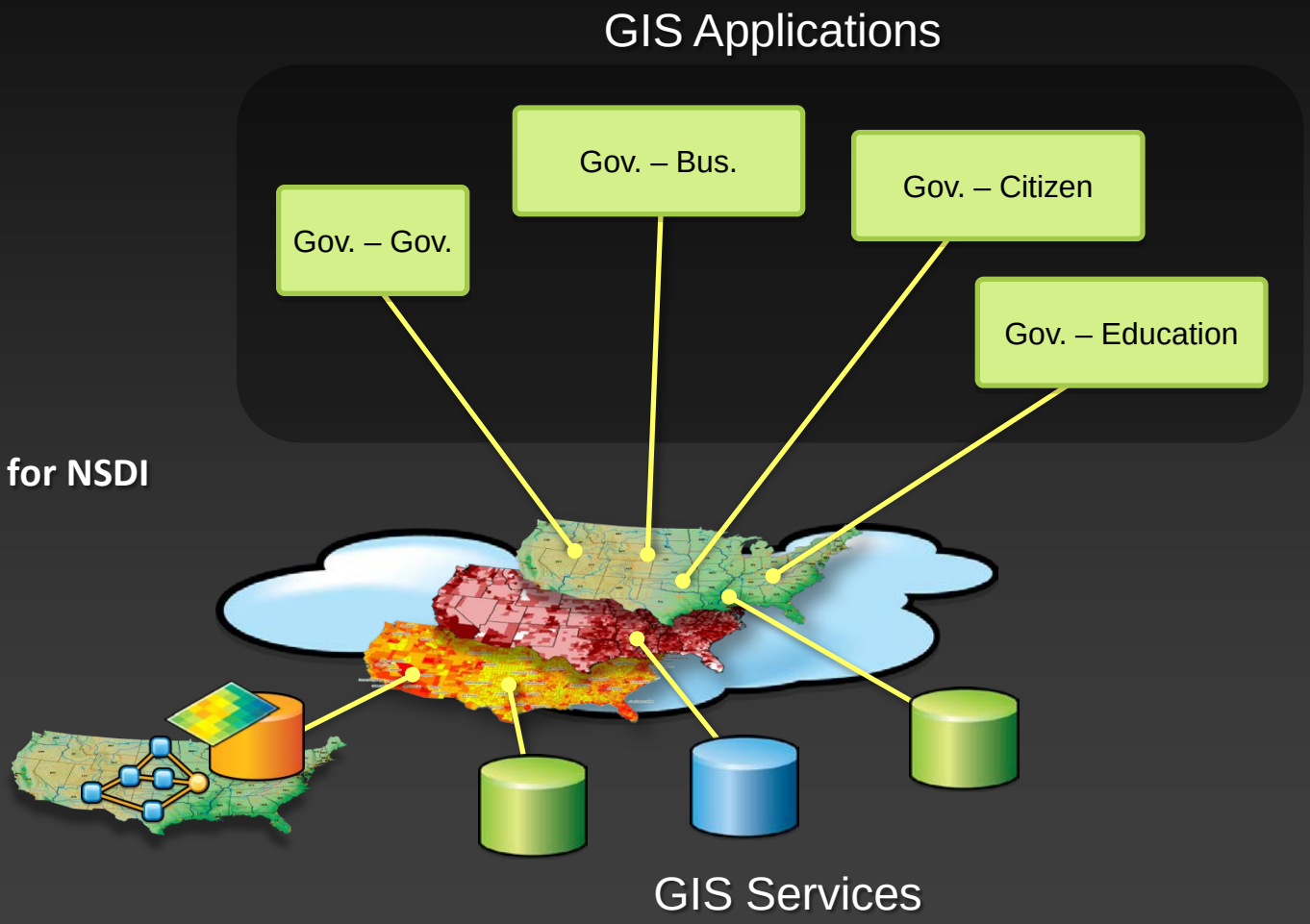


Using the Internet as a Platform

Moving from Data File Sharing to Creating Standards Based Services . . .
. . .Opening Access to Many New Users & Applications

GeoServices will be Implemented Rapidly

Supporting a Large Community of Users & Applications



Providing a Platform for NSDI

Integrating All Levels of Government . . .

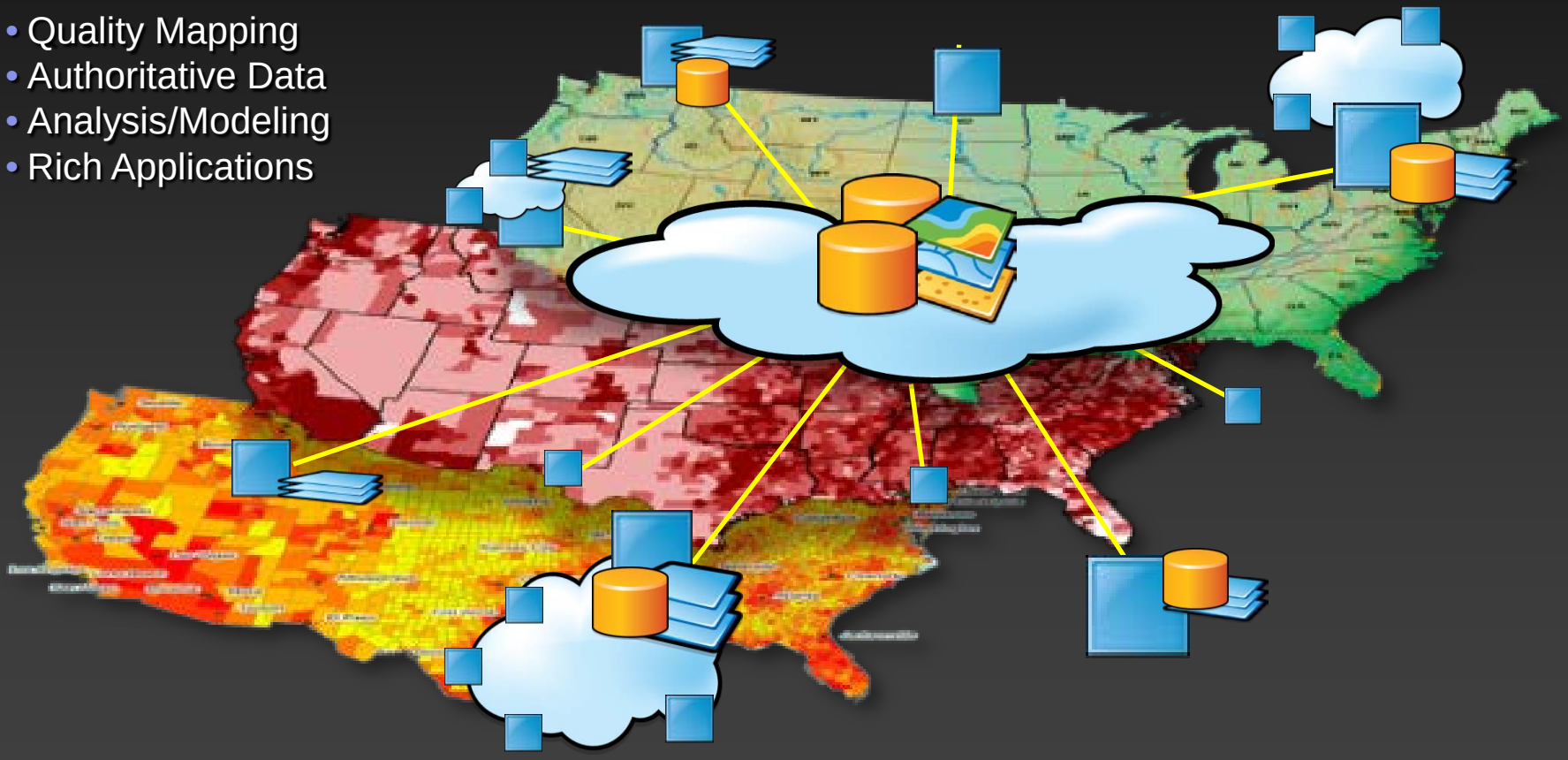
Supporting Open Access, Collaboration & Transparency

A National GIS / NSDI Could Emerge

A Distributed Network of Systems & Services

Providing

- Quality Mapping
- Authoritative Data
- Analysis/Modeling
- Rich Applications



. . . Integrating National, Regional & Municipal Geospatial Knowledge

The Foundation will be New Computing Paradigms

GIS Servers / Services

Sharing

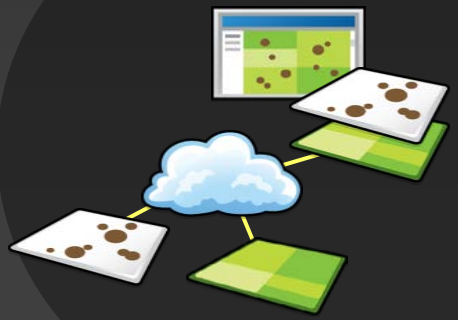
Services Based Architecture

Agility

Mobility

Extend enterprise & support real time

Web 2.0s—mash-ups



Integrate diverse content

Cloud computing

Gov 2.0

Efficiency and Reliability

QUESTIONS?

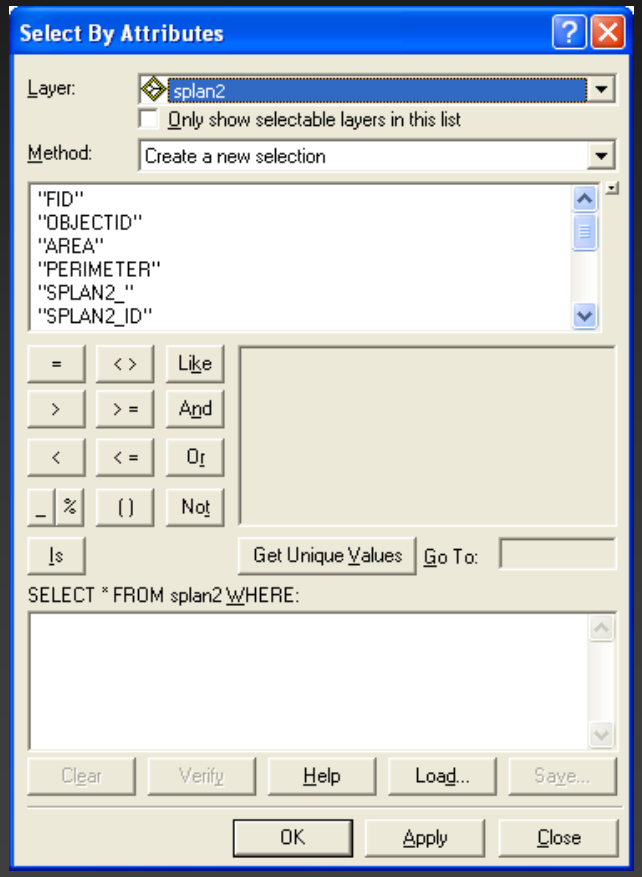
Going Further

Standard Query Language

- SQL is the standardized method of interacting with a database
- Even Access allows you to use SQL
- Insert (new records into a DBMS)
- Update (existing records in DBMS)
- Delete (remove records from DBMS)
- Where (limits your results)

Select Statements

- Most common SQL query you will encounter
- “Select By Attributes” has this as the foundation
- Nothing more than “SELECT * FROM *gis_layer* WHERE…”



Joins

- In ArcGIS or Access, you join two (or more) tables together using a primary key.
- If the keys match, the secondary tables are tacked on to the first
- Again, geospatial is special, so GIS has another type of join

Spatial Joins

- Relationship not determined by key, but by proximity or connectivity
- Contains/Within/Overlaps
  - One feature falls entirely within another
- Touches/Intersects/Crosses
  - One feature touches another
- Equals or Disjoint
- List of spatial relationships.

Relations

- Joins work for one-to-one relationships, where one record in a table matches to one (and only one) record in a foreign table.
- Often, data requires the use of a one-to-many or many-to-many relationship.
- In GIS, joins are strictly 1-to-1. Relations allow the GIS user to access more complicated relationships in the database.