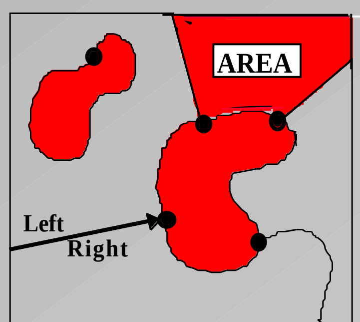
Database Design

presented by:
Tim Haithcoat
University of Missouri
Columbia

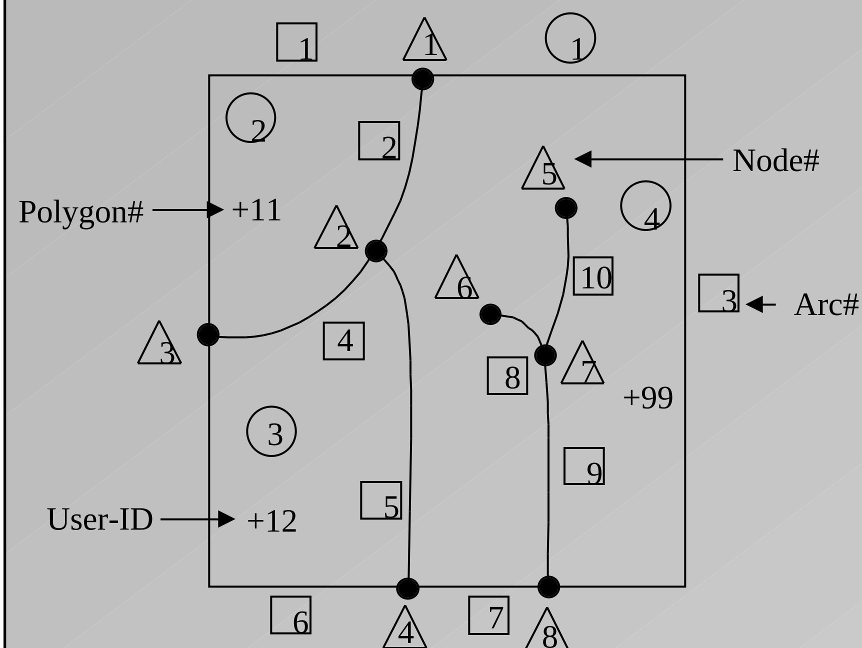
Topology: Defining Spatial Relationships



Perimeter

Spatial properties of topology

Topological Structure



Node Topology

Node#	Arc#
1	1,2,3
2	2,4,5
3	1,4,6
4	5,6,7
5	10
6	8
7	8,9,10
8	3,7,9

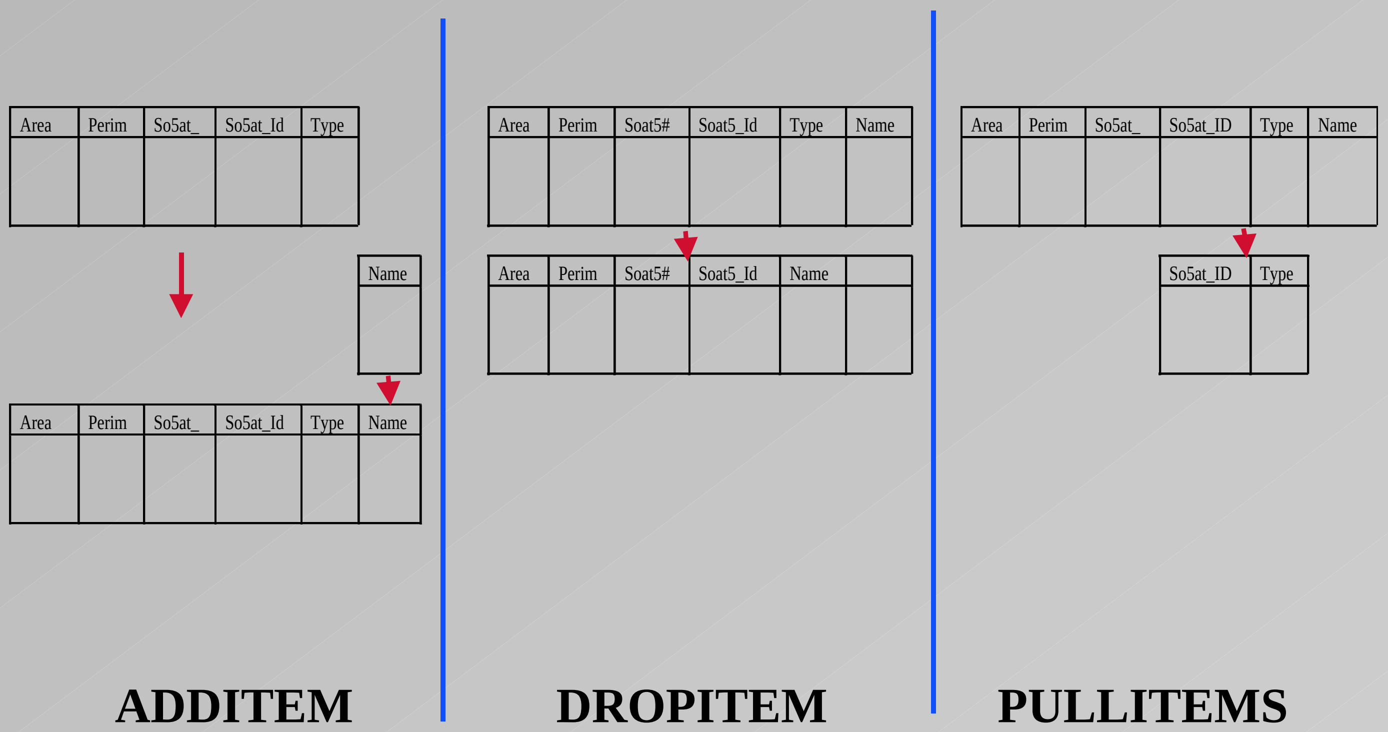
Polygon Topology

Poly #	User-ID	Arc#
1	0	13,6,7
2	11	1,2,4
3	12	4,5,6
4	99	2,3,5,7

Arc Topology

Arc#	From	To	LPoly	RPoly
1	1	3	2	1
2	1	2	4	2
3	1	8	1	4
4	3	2	2	3
5	4	2	3	4
6	4	3	1	4
7	8	4	1	4
8	6	7	4	4
9	8	7	4	4
10	7	5	4	4

Manipulating the Data File Structure



Relating Data Files to the Feature Attribute Table

The JOIN command can be used in PC ARCPLOT, PC ARCEDIT, and TABLES

AREA	PERIM	SO5AT_	SO5AT_ID
-10.0000	40.0000	1	0
5.0000	7.0000	2	298
2.5000	5.0000	3	300
2.5000	5.0000	4	301

SO5AT_ID	SOIL_TYPE
298	30
299	27
300	27
301	32
302	32

Joining Data Files to the Feature Attribute Table

The JOINITEM command permanently joins two data files into one.

Relate Item

Start Item

In_file | **linear | ordered | link** | **Join_file**

AREA	OERUN	SO5AT_	SO5AT_ID

SO5AT_ID	TYPE	NAME	SUIT

AREA	PERIM	SO5AT_	SO5AT_ID	TYPE	NAME	SUIT

GIS Database Design

DESIGN (Dictionary Definition)

A scheme in which means to an end are laid down.

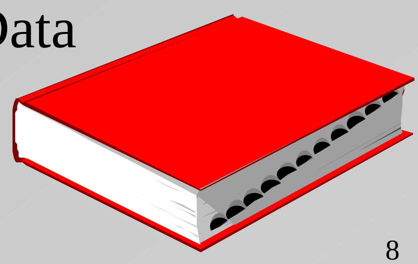
DATABASE DESIGN

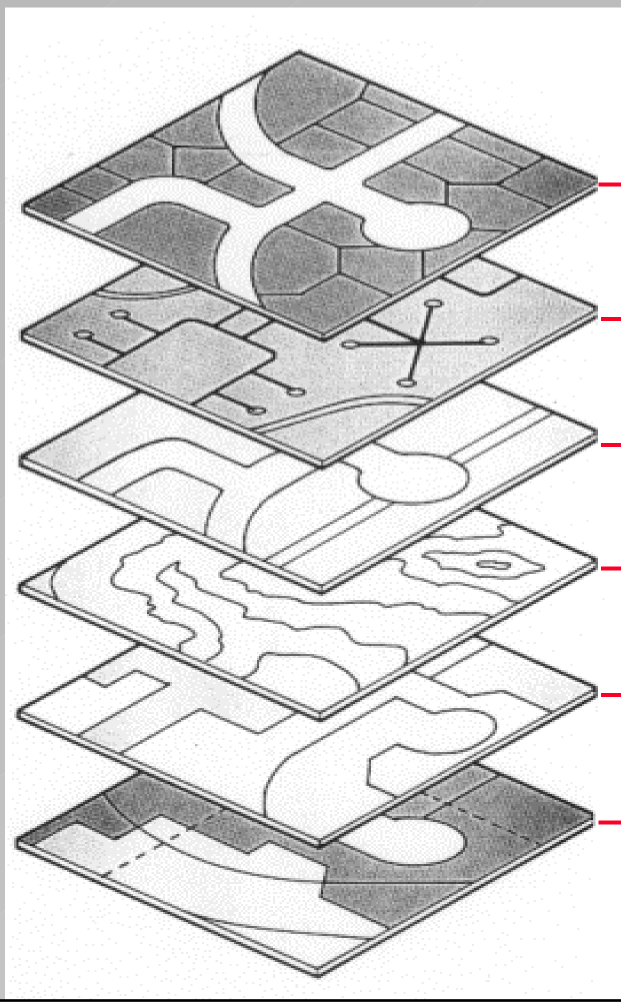
A scheme in which database goals are defined and database alternatives are laid out, analyzed, and an alternative is selected or recommended. Implies:

- That goals are identified and set.
- A structured decision making process:
  - identify goals of database
  - identify issues or problems in the way of goal achievement
  - identify alternatives to resolve issues
  - select or recommend an alternative
  - test alternative
  - act or implement

Data Dictionary

- **Definition**
  - Detailed description of data within a database (documentation)
  - “Users Guide” to the Database
- **Functions**
  - Focal Point for Data Standardization
  - Facilitates Development of Application Programs
  - Assists with Database Design Process
  - Improves Understanding of the Data
  - Reduces Data Redundancy
  - Improves Data Integrity





Layers of Information

Essential data provided by various department of government become layers of information in a GIS. Typical layers include:

Land Records - parcels, subdivisions, easements, right-of-way, ownership, valuation and assessment information.

Public Works and Utilities - sewers, storm drains, street lamps, trees, telecommunications systems and utilities.

Roads - centerlines, intersections, street rights-of-way and emergency access routes.

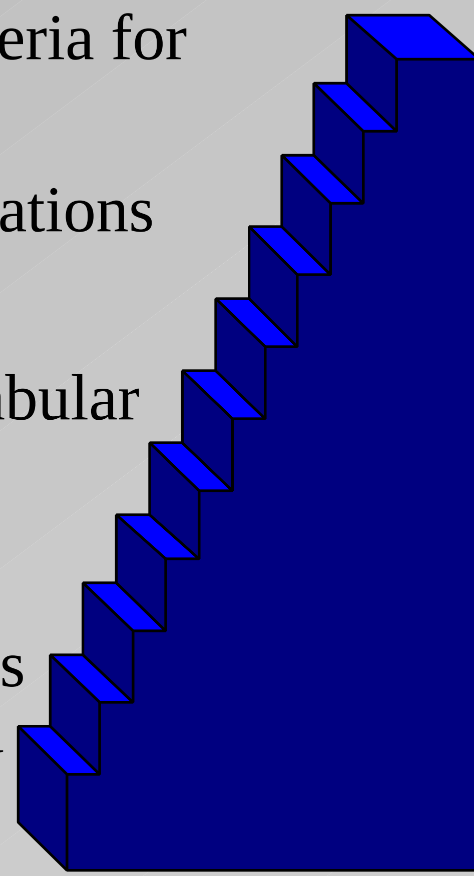
Environmental - geology, soil, vegetation, floodplains, wetlands, hazard areas and noise contours.

Planning - zoning, development proposals and land use plans.

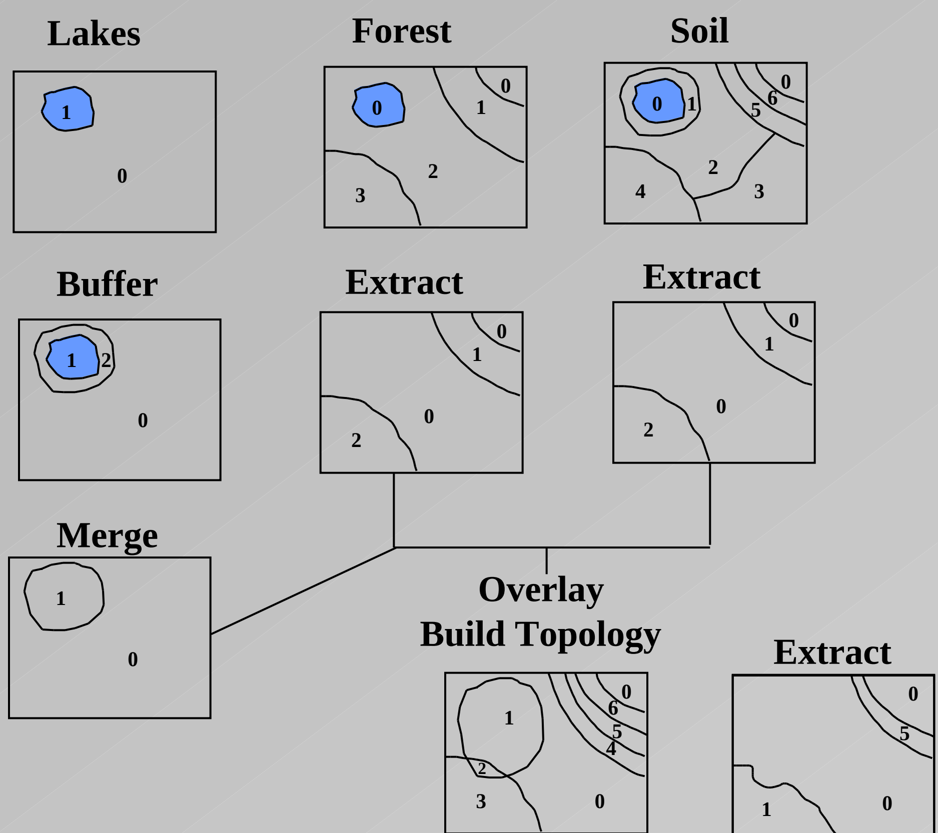
Administrative - city limits, voting precincts, census tracts, fire and police areas, assessment districts and zip codes.

Steps for Performing Geographic Analysis

- Establish the objectives and criteria for analysis
- Prepare the data for spatial operations
- Perform the spatial operations
- Prepare the derived data from tabular analysis
- Perform the tabular analysis
- Evaluate and interpret the results
- Refine the analysis as necessary



Analysis Steps



ID	Buffer	Species	Soil	…
1	1	0	0	
2	1	0	C	
3	0	JP	C	
4	0	JP	0	
5	0	JP	C	
6	0	JP	0	

Presenting the Results

TOPICS

- Designing the components of a map
- Using symbols effectively
- Determining the purpose of map
- Defining the map parameters (size & scale)
- Designing the map layout
- Preparing symbol data
- Creating a final map
- Generating a tabular report

Table Design

- •Primary Key (Unique identifier of what is present graphically)
- •Foreign Key (Unique identifier in another table)

This gives 2 types of tables:

- •Primary Table (Feature Attribute Table) (FAT)
- •Lookup Table (LUT)

There are a set of rules or guidelines to help you design these tables so as to minimize redundancy among all tables. These rules help you decide what tables should exist and what they should contain. These rules are called **normalization.**

Scale

Definition

The accuracy with which the location & shape of map features can be depicted for a given scale.

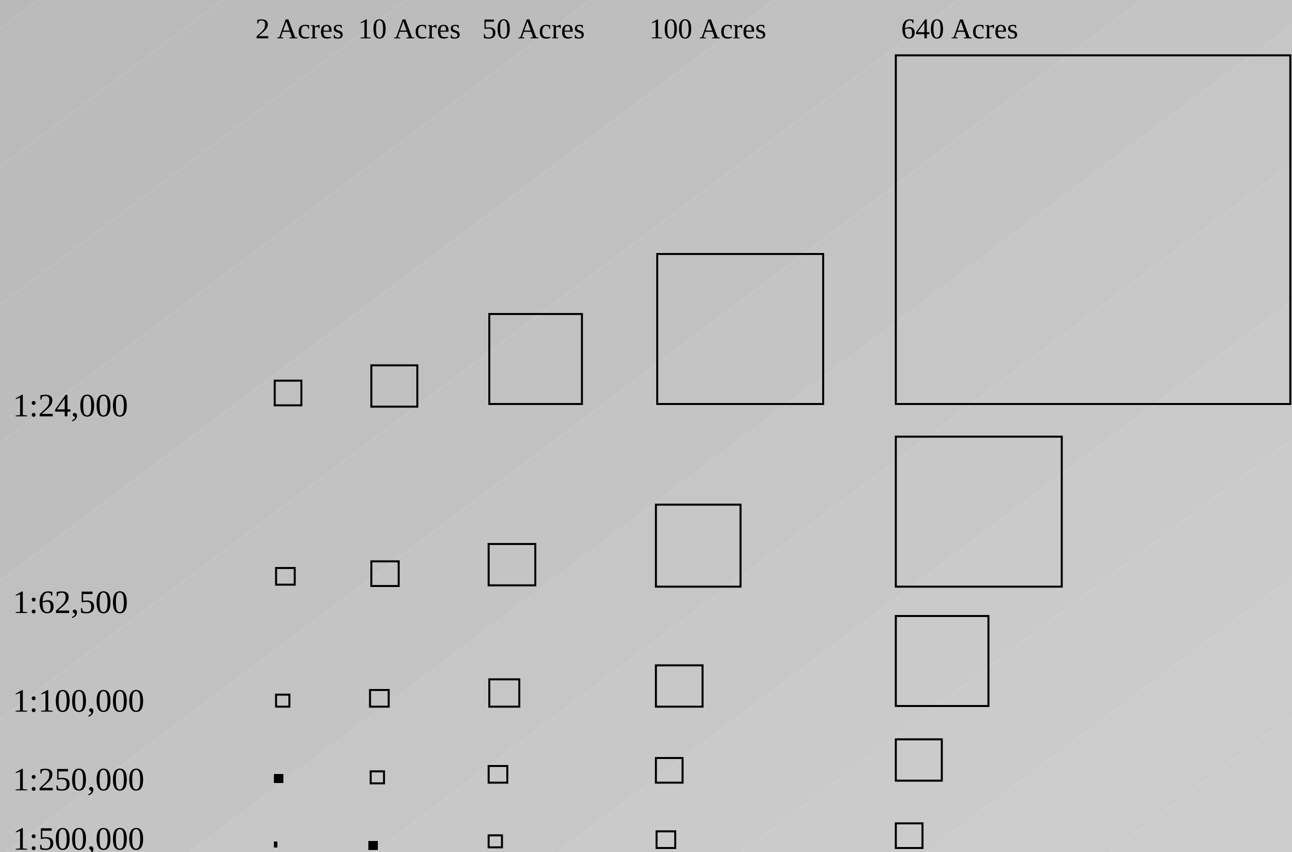
- ✗ Decreasing map scale results in lower map resolution as selected features are:
  - Smoothed
  - Simplified
  - Aggregated
  - Eliminated
  - Reduced in Dimension
- ✗ Understand/document all GIS data source resolutions
- ✗ Categorize sources by resolution groups
- ✗ Make careful choices regarding upward & lower bounds of resolution groups.
- ✗ Begin to define scales of the GIS database

Scale - Data Resolution

	Polygon	Lines	
	Acres	Mile	Feet
1:24,000	2-3	.05	250
1:62,500	5-10	.12	650
1:100,000	25-50	.2	1050
1:250,000	250-500	.5	2600
1:500,000	500-1000	1.0	5280

Data below these resolutions are generally merged into surrounding data, converted to a point or deleted.

Data Resolution



Why Database Design?

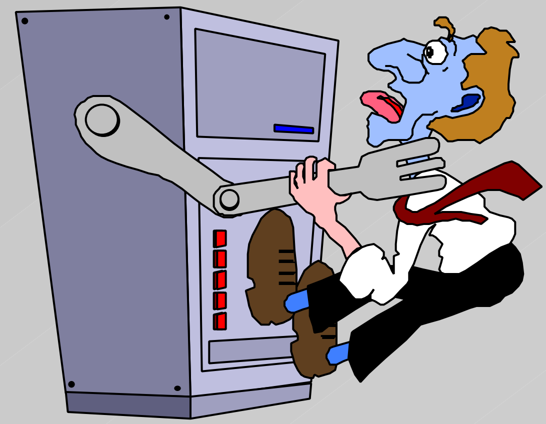
- The design provides the comprehensive mental framework & organization of the database.
  - It allows the database to be viewed in its entirety and evaluate how the various aspects of the database need to interact.
  - It allows for the early identification of major issues, potential problems, and design alternatives.
- To determine answers that define GIS efficiency constraints.
  - about users
  - about data
  - about input
  - security
  - management
  - hardware

GIS components

What happens if you do NOT conduct a database design?

Not conducting a comprehensive design can lead to a poorly constructed and functionally unsatisfactory database. The lack of a design can result in:

- unsupported applications
- an uncorrectable database
- lack of coherency within the database



Building the Database (Database Design)

- ✓ Determine study area bounds
- ✓ Determine what coordinate system
- ✓ Identify the required spatial data layers
- ✓ Determine the required feature attributes
- ✓ Define each attribute and its codes
- ✓ Coordinate registration
- ✓ Code schemes
- ✓ Allocate storage
- ✓ Project organization

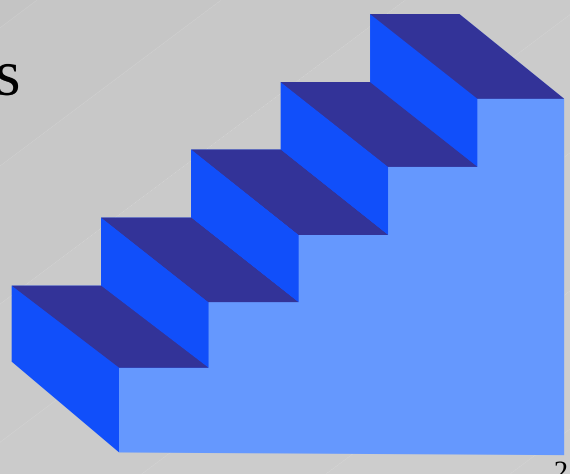
Stages in Database Design

- **CONCEPTUAL**
  - Software & hardware independent
  - Describes & defines included entities
  - Identifies how entities will be represented in the database
  - Requires decisions about how real-world dimensionality & relationships will be represented
- **LOGICAL**
  - Software specific but hardware independent
  - Sets out the logical structure of the database elements, determined by the database management system used by the software
- **PHYSICAL**
  - Both hardware & software specific
  - Requires consideration of how files will be structured for access

Starting Your Project

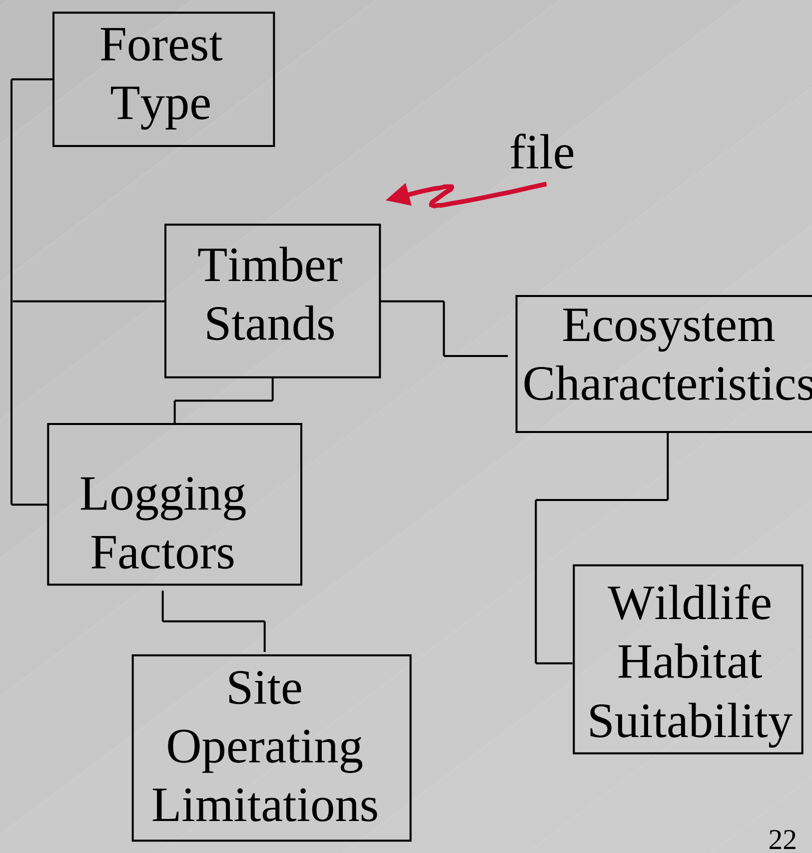
Four Major Steps in conducting a typical GIS project

- Determine the Objectives
- Building the Database
- Performing the Analysis
- Presenting the Results



Define Database Structure

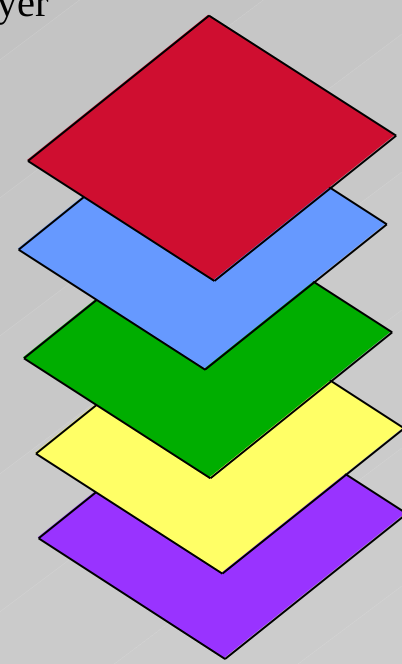
- The Physical Structure of the Database
  - Data Elements Assigned to Files
  - Relationships Among Files



Organizing Data into Layers

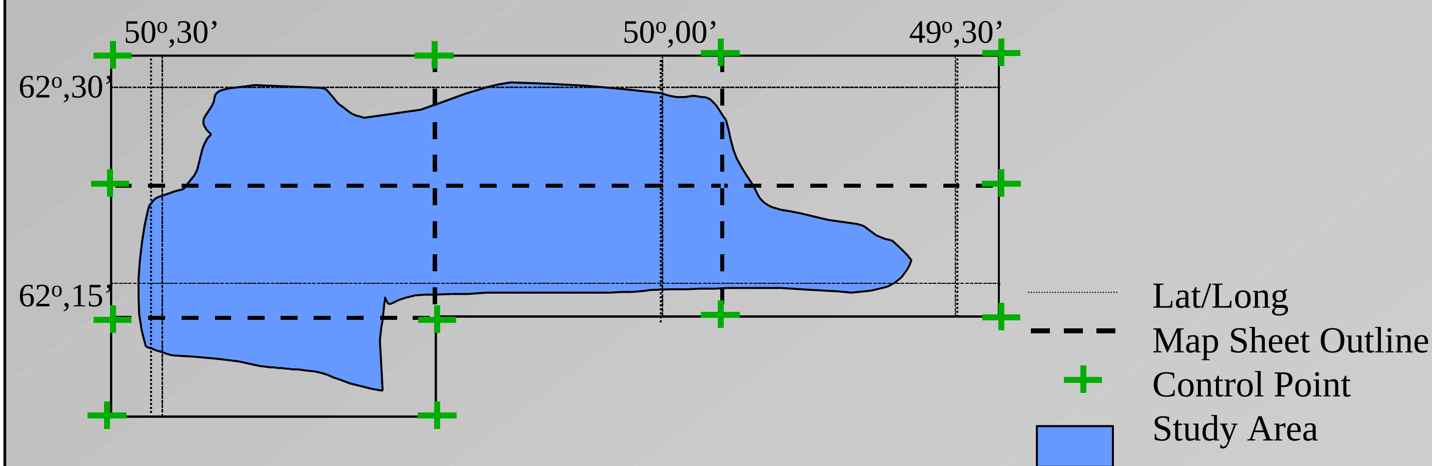
RELATED TO:

- Type of Feature
  - Less restrictions on putting points into same layer
- Data - Data relationships
  - Common lines
  - Hierarchical relationships
- Data - Function Relationships
  - Some data entities tend to be used together
  - Some functions are dominate - database can favor these
  - Creating continuous coverages
- Variety of user views
  - More divergent user requirements, more primitive data organization
- Kinds and Quantities of Updates
  - Heavily updated entities should be isolated



Define Study Area

- To define the GIS database study area, you must determine:
  - The geographic extent of the data elements required to meet user needs (boundaries)
  - Several points within the extent of the database for which the coordinate value are known in a common Cartesian coordinate system (Geographic Reference/Control Points)



Special Aspects of GIS Database Design

- Data model is hybrid
- Location is a special kind of key
- Issues connected with coordinates
- Issues connected with cartography
- Standard database literature and theory do not take account of these special characteristics

Spatial element normally not dealt within database

- ✗ 1773 Eucalyptus, Perris, CA
- ✗ 117 15 00 33 30 10
- ✗ UTM Zone 11S; N 6487529 m; E 455322 m

- All refer to the same location on the earth. In a regular database, there is no way to relate these 3 ways of describing location.
- GIS forces you to deal with this issue so that location truly becomes a special kind of key.
- The explicitly definition of location allows diverse data sets to be related that ordinarily we could not related.
- Georelational Model

Organize & Describe Data Elements

- Data elements should be organized into “layers” according to the logical relationships among them.

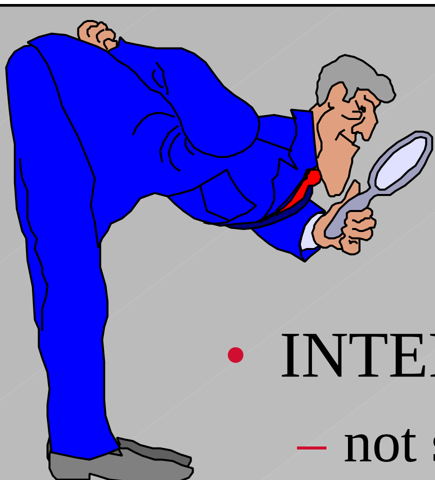
Terrain Layer	Infrastructure Layer
Soil	Roads
Slope	Railroads
Geology	Transmission Lines
Aspect	Pipelines
Vegetation	

- Data elements should be described according to a well-defined classification and coding schema.

Vegetation	Description	Code
Forest	Pine	(P) or (100)
	Spruce	(SP) or (101)
	Cedar	(c) or (200)
Wetland	Emergent Marsh	(EM) or (402)
	Spring Bong	(SB) or (403)

Advantages of Database Approach

- ☑ reduction in data redundancy
- ☑ maintenance of data integrity & quality
- ☑ data are self-documented or self-descriptive (metadata)
- ☑ avoidance of inconsistencies
- ☑ reduced cost of software development
- ☑ security restrictions

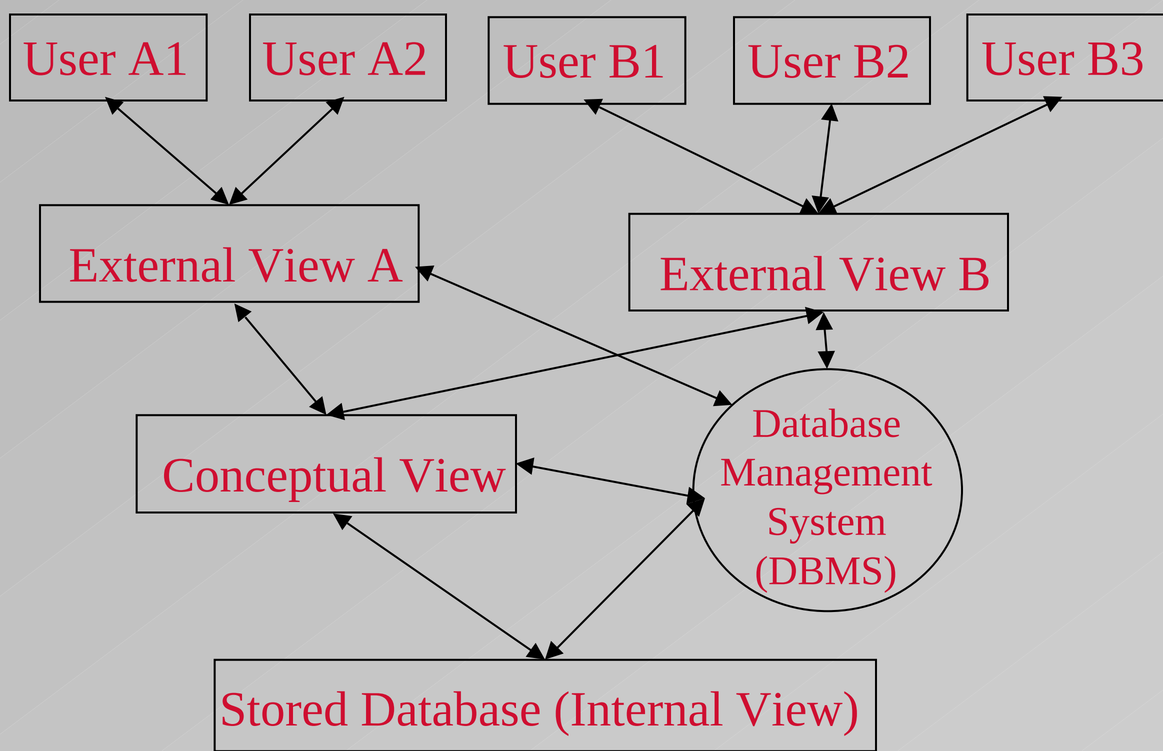


Views of the Database

- INTERNAL VIEW
  - not seen by the user or applications developer
- CONCEPTUAL VIEW
  - where the database administrator builds and manages the database
- EXTERNAL VIEW
  - what the user or programmer sees - can be different to different users and applications

Views of the Database

Adapted from: Date, G.J. 1987. An Introduction to Database Systems, Addison-Wesley. Reading, MA, p. 32



Common Design Problems

- ☑ You have one table with a large number of fields that don't all relate to the same subject.
- ☑ You have fields that are intentionally left blank in many records because they aren't applicable to those records.
- ☑ You have a large number of tables which contain the same fields.

Refining the Design

Questions to Ask:

- Did you forget any fields? Is there information you need that isn’t included?
- Did you choose a good primary key for each table? Is it easy to type and remember?
- Are you repeatedly entering duplicate information in one of your tables?
- Do you have tables with many fields, a limited number of records, and many empty fields in individual records?

Issues in Database Design

- What storage media should be used?
  - How big is the database?
  - How will database grow over time?
  - What are the required access speeds?
- Should data be partitioned by location or theme?
- Distributed or centralized? If distributed, on what host does what data go?
- Who is responsible for making sure the data standards are adhered to?
- Who is responsible for enforcing the data standards?

Database System Concepts

Definition

A database is a collection of non-redundant data which can be shared by different application systems.

- ✗ Stresses the importance of multiple applications and data sharing.
- ✗ The database becomes a common resource for agency.
- ✗ Implies separation of physical storage from use of the data by an application program (i.e., program/data independence).
- ✗ Changes can be made to the data without affecting other components of the system.

Build the Database

This is the most critical and time consuming part of the project.
The completeness and accuracy of the database will determine the quality of the analysis and final products.

STEPS:

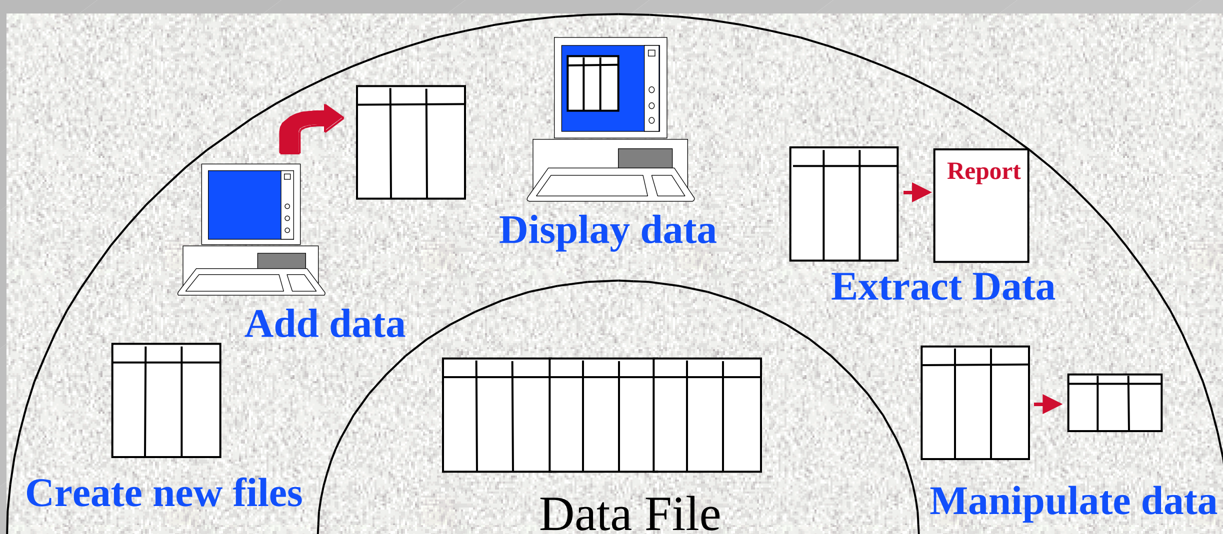
- Designing the database
- Automating the data
- Getting the spatial data in
- Making the spatial data usable
- Getting the attribute data in
- Managing the database

Determining Objectives

Questions to Consider:

? What is the problem to be solved?
? How is it solved now?
? Are there alternative methods for solving the problem?
? What are the final products of the project - reports, working maps, presentation quality maps?
? How frequently will products be generated?
? Who is the intended audience for the final products - technicians, planners, managers, the general public?
? Are there, or will there be, other uses for this data?
? If there are other uses, what will be the specific requirements?

Database Management System (DBMS) Functions



Relational Database Systems

- ✓ Most popular DBMS model for GIS
- ✓ Flexible approach to linkages between records comes the closest to modeling the complexity of spatial relationships between objects.

Relational Database Management System (RDBMS)

Table A

Name	Address	Parcel #
J. Smith	18 Lawers Dr.	216554
T. Brown	14 Seminole Tr.	887419

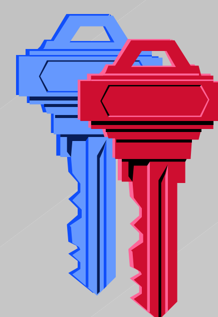
Table B

Parcel #	Assessed Value
887419	120,000
216397	100,000

Keys

Definition: A key of a relation is a subset of attributes with the following attributes:

- Unique identification
- Non-redundancy



Types of Keys

✓ PRIMARY KEY

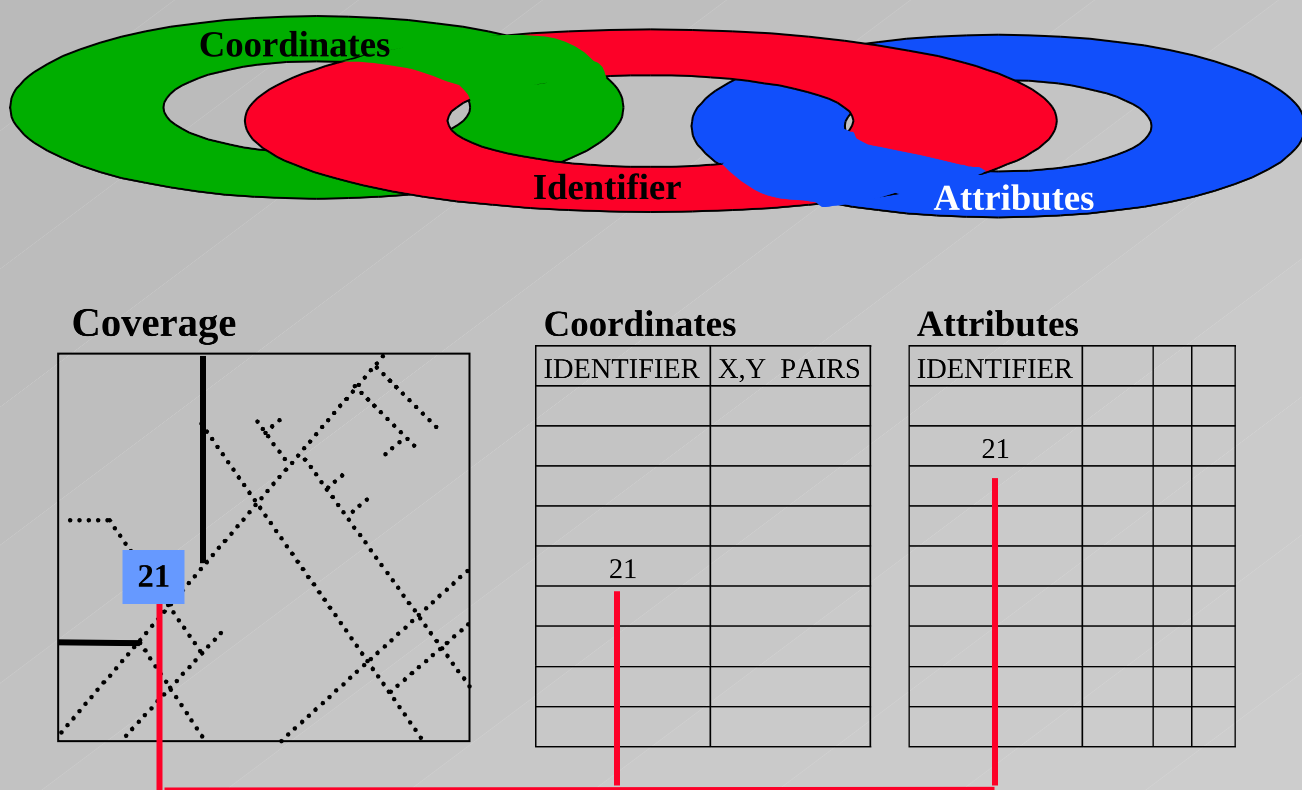
- Serves as the sole row level addressing mechanism in the relational database model.
- It can be formed through the combination of several items.

✓ FOREIGN KEY

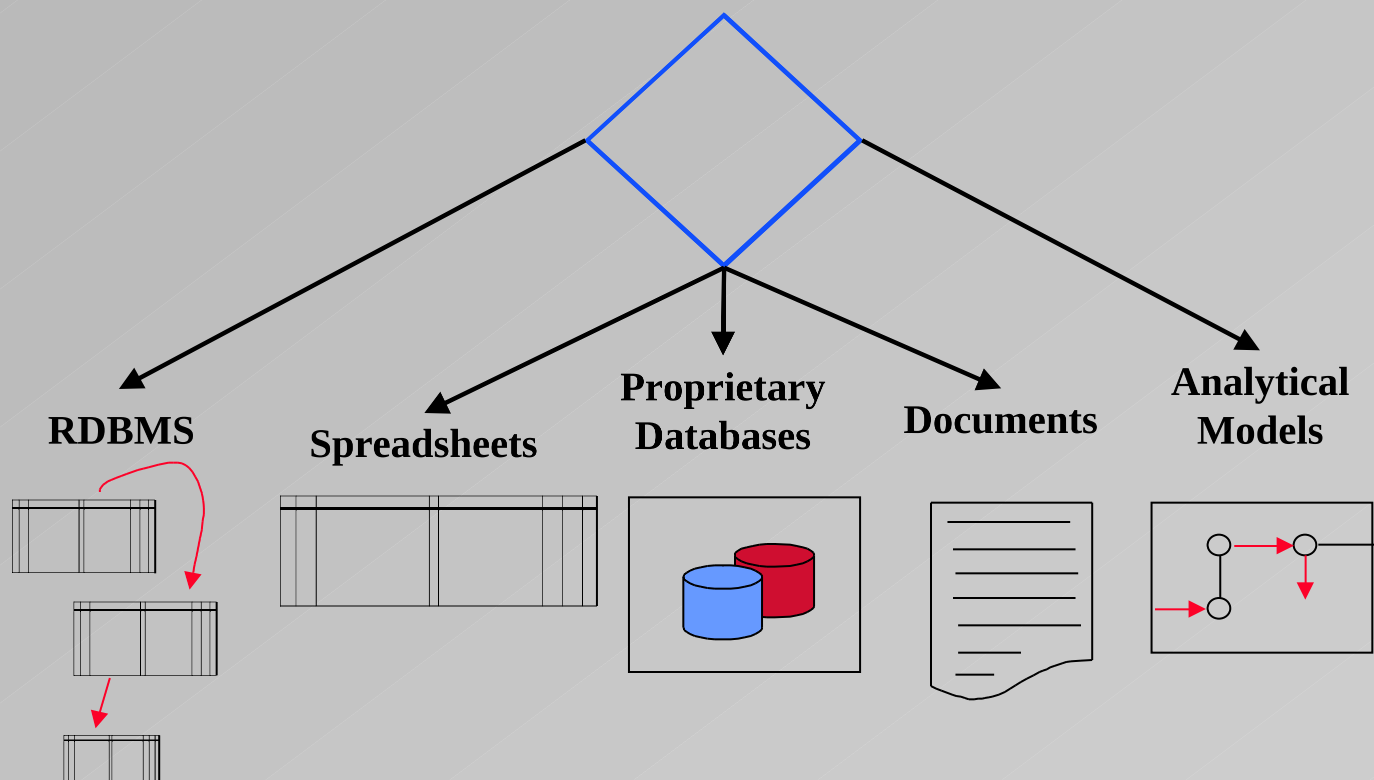
- A column or set of columns within a table that are required to match those of a primary key of a second table.

✓ Once linked, these keys form a RELATIONAL JOIN - thereby connecting row to row across the databases within which they reside.

Coverage Data Model



GIS as a Gateway to External Data



Assess User Needs

To design a GIS database, you must know:

- ✓ *Users:* Who will use the database?
- ✓ *Data Elements:* What kinds of data will they need to have in the database?
- ✓ *Purpose:* Why do they need the data and how will they use it?
- ✓ *Sources:* Where are the data elements they need?
- ✓ *Quality:* What is the resolution, accuracy, and reliability of the data available?
- ✓ *Precision:* Does the input data warrant the storage of double precision coordinates?

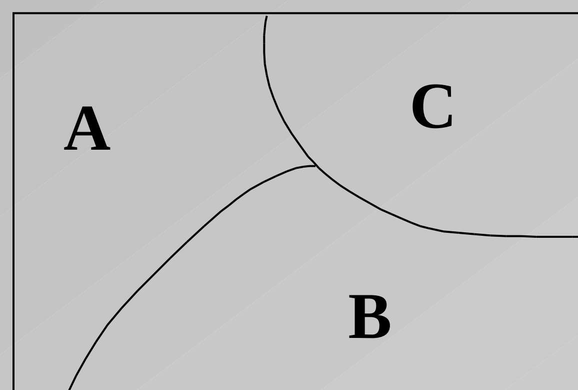
Item Types

EXAMPLE	TYPE	ABBREVIATION
Main Street	Character	C
10/15/90 10/15/1990	Date	D
23675	Integer	I
347.22	Numeric	N

Major GIS Data Models

A	A	A	A	C	C	C	C
A	A	A	A	C	C	C	C
A	A	A	B	B	C	C	C
A	A	B	B	B	B	B	B
A	B	B	B	B	B	B	B

RASTER



VECTOR

RASTER DATA MODEL

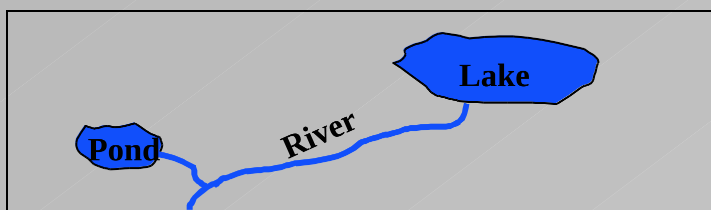
ADVANTAGES

- Location of cell implied by position in matrix, no need to store
- Neighboring cells in neighboring locations, convenient for analysis of neighborhood relationships
- Accommodates both discrete and continuous data
- Process algorithms much simpler
- Map unit boundaries inherently present
- Compatibility with output devices

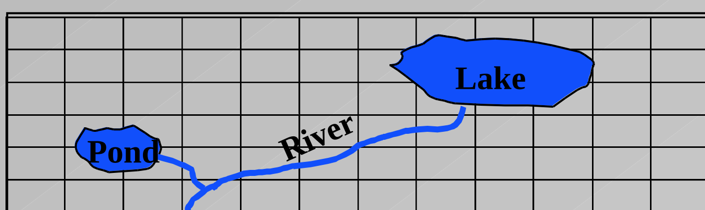
DISADVANTAGES

- More storage required - data redundancy
- Cell size limits resolution
- Image access mostly sequential
- Accommodation of attribute data cumbersome
- Input and output problematic

Creating a Raster



REALITY - HYDROGRAPHY



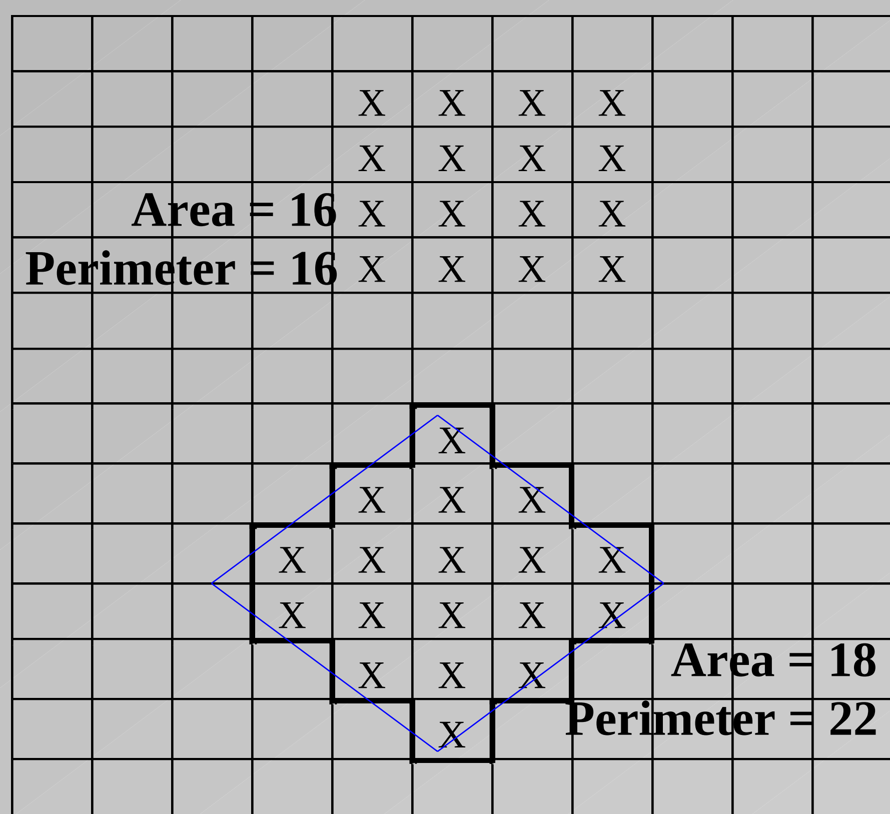
REALITY OVERLAID WITH A GRID

0	0	0	0	0	0	0	1	1	1	0	0
0	0	0	0	0	0	1	1	1	1	1	0
0	0	0	0	0	0	0	1	1	1	0	0
0	1	1	0	0	0	2	2	0	0	0	0
0	1	1	2	2	2	0	0	0	0	0	0
0	0	0	2	0	0	0	0	0	0	0	0

0 = No Water Feature
1 = Water Body
2 = River

RESULTING RASTER

Area & Perimeter Functions in Rasters



Raster Data Values

0	0	1	1	1	2	3
0	1	1	1	2	2	3
0	0	0	1	1	2	2
0	0	0	1	1	1	2
0	0	0	0	0	0	1

INTERGER VALUES

examples: soil class
# of farms

2.1	2.3	2.4	2.7	2.9	3.2	3.4
1.9	2.1	2.4	2.6	3.0	23.3	3.4
1.8	2.0	2.3	2.5	2.8	3.1	3.2
1.8	1.9	1.9	2.3	2.5	2.8	3.0
1.8	1.8	1.8	1.9	2.1	2.3	2.7

REAL (DECIMAL) VALUES

example: elevation

a	a	b	b	c	c	a
a	b	b	b	c	c	a
a	b	b	b	b	c	c
a	a	b	b	b	c	c
a	a	a	b	c	c	c

NON NUMBERIC VALUES

example: vegetation class

Logical Operations on Raster: LOGICAL AND

WHAT CELLS ARE BOTH A **AND** 7?

A	A	B
A	A	B
C	C	B

6	7	7
6	7	7
8	8	7

RECLASSIFY

1	1	0
1	1	0
0	0	0

RECLASSIFY

0	1	1
0	1	1
0	0	1

MULTIPLY

1x0	1x1	0x1
1x0	1x1	0x1
0x0	0x0	0x1

→

0	1	0
0	1	0
0	0	0

Logical Operations on Raster: LOGICAL OR

WHAT CELLS ARE BOTH A **OR** 7?

A	A	B
A	A	B
C	C	B

6	7	7
6	7	7
8	8	7

RECLASSIFY

1	1	0
1	1	0
0	0	0

RECLASSIFY

0	1	1
0	1	1
0	0	1

ADD

1+0	1+1	0+1
1+0	1+1	0+1
0+0	0+0	0+1

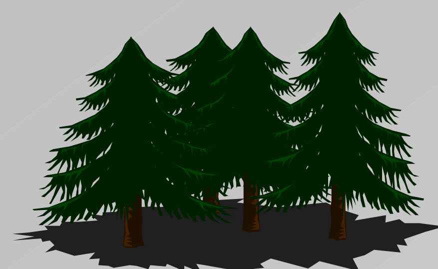
1	2	1
1	2	1
0	0	1

RECLASSIFY

1	1	1
1	1	1
0	0	1

Example Raster Database

Raster Description:
- resolution is 500 m
- entire array is 2.5 km by 2.5 km
- columns are aligned North-South



LAYER 1: LAKE
(1= lake, 0= not lake)

1	1	0	0	0
1	1	0	0	0
1	0	0	0	0
0	0	0	0	0
0	0	0	0	0

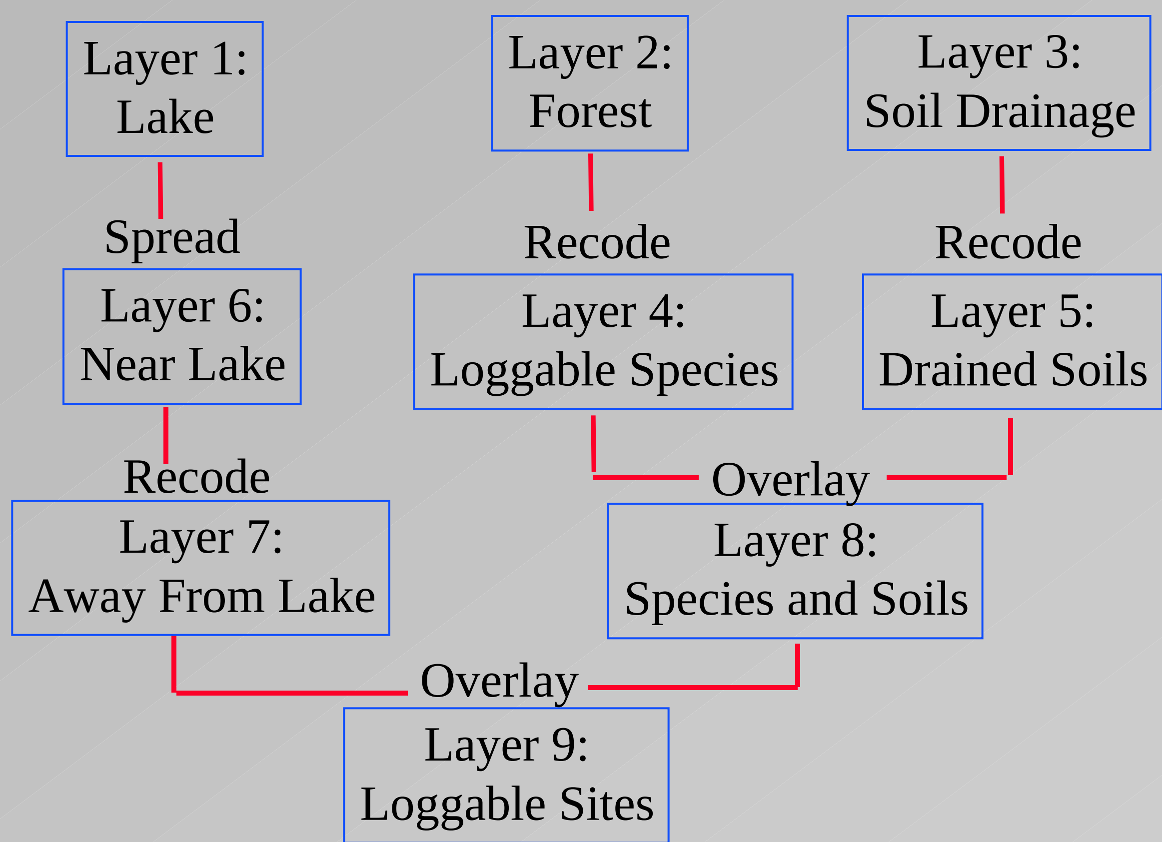
LAYER 2: FOREST SPECIES
(1= black spruce, 2= jack pine, 0=no lake)

0	0	2	2	2
0	0	2	2	2
0	1	1	1	2
1	1	1	1	1
1	1	1	1	1

LAYER 3: SOIL DRAINAGE
(0=waterlogged, 1=poor, 2=good)

0	0	2	2	2
0	0	1	1	2
0	0	1	1	2
0	1	1	1	2
1	1	1	2	2

Example Project Steps



OBJECTIVE:
Identify areas suitable for logging

An area is suitable if it satisfies the following criteria:

- is Jackpine (Black Spruce are not valuable)
- is well drained (poorly drained and waterlogged terrain cannot support equipment, logging causes unacceptable environmental damage)
- is not within 500 m of a lake or watercourse (erosion hazard)

PROCEDURE
Example Project Details (1 of 4)

RECODE LAYER 2 - Y IF VALUE 2 (JACKPINE), N IF OTHER VALUE

LAYER 2: FOREST SPECIES

0	0	2	2	2
0	0	2	2	2
0	1	1	1	2
1	1	1	1	1
1	1	1	1	1

LAYER 4: LOGGABLE SPECIES

n	n	y	y	y
n	n	y	y	y
n	n	n	y	y
n	n	n	n	n
n	n	n	n	n

RECODE LAYER 3 - Y VALUE 2 (GOOD), N IF OTHER VALUE

LAYER 3: SOIL DRAINAGE

0	0	2	2	2
0	0	1	1	2
0	0	1	1	2
0	1	1	1	2
1	1	1	2	2

LAYER 5: DRAINED SOIL

n	n	y	y	y
n	n	n	n	y
n	n	n	n	y
n	n	n	n	y
n	n	n	y	y

PROCEDURE
Example Project Details (2 of 4)

SPREAD THE LAKE ON LAYER 1 BY ONE CELL (500 M)

LAYER 1: LAKE

1	1	0	0	0
1	1	0	0	0
1	0	0	0	0
0	0	0	0	0
0	0	0	0	0

LAYER6: NEAR LAKE

1	1	1	0	0
1	1	1	0	0
1	1	0	0	0
1	0	0	0	0
0	0	0	0	0

RECODE LAYER 3 - Y VALUE 2 (GOOD), N IF OTHER VALUE

LAYER 6: NEAR LAKE

1	1	1	0	0
1	1	1	0	0
1	1	0	0	0
1	0	0	0	0
0	0	0	0	0

LAYER 7: AWAY FROM LAKE

n	n	n	y	y
n	n	n	y	y
n	n	y	y	y
n	y	y	y	y
y	y	y	y	y

PROCEDURE
Example Project Details (3 of 4)

OVERLAY LAYERS 4 & 5 - IF BOTH 4 & 5 ARE Y, N OTHERWISE

LAYER 4: LOGGABLE SPECIES

n	n	y	y	y
n	n	y	y	y
n	n	n	n	y
n	n	n	n	n
n	n	n	n	n

LAYER 5: DRAINED SOIL

n	n	y	y	y
n	n	n	n	y
n	n	n	n	y
n	n	n	n	y
n	n	n	y	y

LAYER 8: SPECIES & SOILS

n	n	y	y	y
n	n	n	n	y
n	n	n	n	y
n	n	n	n	n
n	n	n	n	n

PROCEDURE
Example Project Details (4 of 4)

OVERLAY LAYERS 7 & 8 - IF BOTH 7 & 8 ARE Y, N OTHERWISE

LAYER 7: AWAY FROM LAKE

n	n	n	y	y
n	n	n	y	y
n	n	y	y	y
n	y	y	y	y
y	y	y	y	y

LAYER 8: SPECIES & SOILS

n	n	y	y	y
n	n	n	n	y
n	n	n	n	y
n	n	n	n	n
n	n	n	n	n

LAYER 9: LOGGABLE SITES

n	n	n	y	y
n	n	n	n	y
n	n	n	n	y
n	n	n	n	n
n	n	n	n	n

RESULT:

The loggable cells are
y on layer 9.

VECTOR DATA MODEL

ADVANTAGES

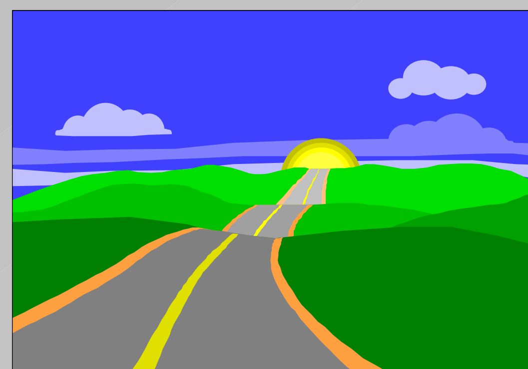
- Requires less storage
- Original map can be represented at near-original resolution
- Resource features such as forest types, roads and streams can be individually retrieved and processed
- Accommodation of attribute data easier
- No need to convert input to raster format
- Better suited for line-type map production

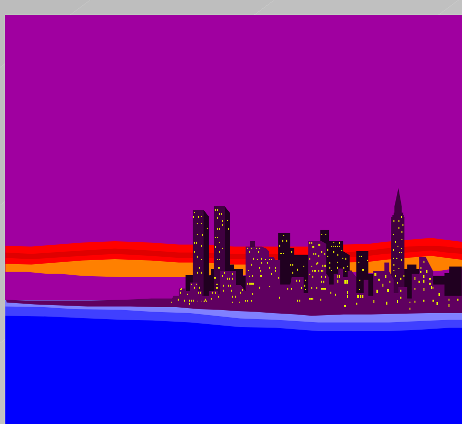
DISADVANTAGES

- Explicit storage of x-y coordinates
- Relationships of points must be formalized in topology
- Algorithms far more complex
- Cannot readily accommodate continuous data

Example: Vector GIS Data

	point	line	area
SCALE	city wells	highway political boundary streams	ag. land urban land city highway airport



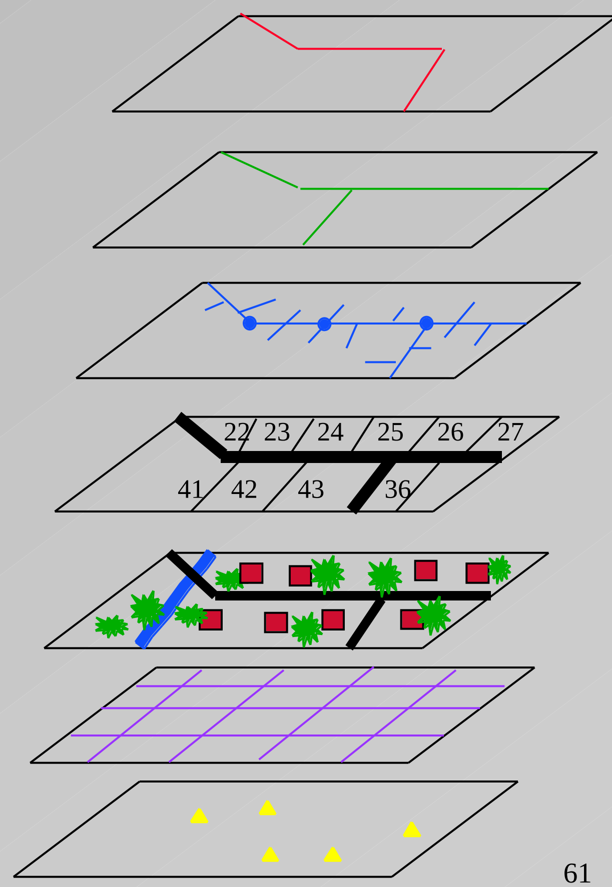


Example Attributes:

city:	population, name
wells:	depth
highway:	number
political boundary:	type
streams:	name
ag. land:	growth potential, acreage
urban land:	urban landuse type, acreage
airport:	name

Data Layers

- Political/Administrative Districts
- Zoning
- Utilities
- Parcels
- Orthophoto Base Map
- State Plane Reference Grid
- Geodetic/Survey Control



What is a Map?

- ✓ a graphic representation of part of the Earth’s surface
- ✓ can show the geographical distribution & positions of various features such as roads, towns, relief, soils, etc.
- ✓ it’s structure conveys information easily & readily to the reader

Map Components

- ✓ GEOGRAPHIC FEATURES: Includes areas, lines, and points drawn from various coverages in the geographic database.
- ✓ CARTOGRAPHIC ELEMENTS: Elements which make the map easier to read, understand and interpret.

How Digital Maps are Created & Stored

TOPICS

- Two basic types of information presented on a map:
  - spatial
  - descriptive
- Major geographic features:
  - points
  - lines
  - polygons
- How geographic features are stored in the computer
- Topology
- How descriptive map information is stored in the computer
- How the spatial data and descriptive data are linked

Data Organization

DATA ELEMENTS

- Parcels
  - Easements
  - Zoning
  - Blocks
- Special Uses
  - Wildlife Habitat
  - Archaeological Sites
  - Flood Zones
  - Drainage Districts
  - Sewer & Water Facilities

- Administrative
  - Census Tracts
  - Census Blocks
  - Fire Response Zones
  - Council Districts
  - Precincts
  - Zip Codes
- Natural Resources
  - Geology
  - Soils
  - Vegetation

  - Hydrology
  - Slope
  - Ecosystems
  - Watersheds