The Geodatabase

Chase T. Brooke
Ecosystem Science and Management
Texas A&M University
College Station, TX 77843
cbrooke@tamu.edu
Evolution of the acronym GIS

- **GISystems** (1980s)
- **GIScience** (1990s)
- **GIServices** (2000s)

Time line:
- 1980s
- 1990s
- 2000s
Spatial Thinking

“Identifying, analyzing, and understanding the location, scale, patterns, and trends of the geographic and temporal relationships among data, phenomena, and issues.” - Joseph Kerski
Spatial Thinking

What is your definition of spatial thinking?
What is a Geodatabase?

• A spatial relational database management system that stores geographic and related data.

• Central repository for spatial data storage and management.
Evolution of database systems

- File Systems
  - Network DBMS
  - Hierarchical DBMS
  - Relational DBMS
    - Object-Oriented Systems (OODBMS)
    - Object-Relational ORDBMS
Geometry Data Type

• All 2D mapping can generally be accomplished with three geometry types

• Points, polylines, and polygons
Relationships among features

• Topological
  – Determines adjacency and connectivity among features
  – Shared and coincident geometry types

• Spatial

• Network
  – Simple edges and junctions
  – Complex edges and junctions

• General
Topology

Polygons

X: 100.50  Y: 200.50
X: 100.50  Y: 220.00

2 features
8 coordinate pairs
4 coincident
Geometric Network

- The storage of network information
- Several components
  - Simple edges and junctions
  - Complex edges
  - Complex junctions
Geometric Networks

• To build a network
  – Sources and sinks
  – Connectivity rules
    • Edge-junction rules
    • Edge-edge rules

• Perform a network analysis
  – Trace
  – Flow direction
The parts of a geodatabase

• Geographic dataset
  – Feature classes, a collection of features with the same type of geometry
  – Relationship classes, a table that stores the relationships between two feature classes

• Object classes
  – A feature class table
  – A non-spatial attribute table
The parts of a geodatabase
Example Geodatabase

Data source
- Street data
- Buildings data
- Vegetation data
- Integrated data

Data layers

Making your own geodatabase

• Design your data model
  – Evaluate your data
  – Define database structure
  – Add data

• Understand your spatial reference
  – Choose a coordinate system
  – Spatial domain
  – Precision

• Modify your spatial domain if necessary
Benefits of using a geodatabase

• Store a rich collection of spatial data in a centralized location
• Apply sophisticated rules and relationships to the data
• Define advanced geospatial relational models (e.g., topologies, networks)
• Maintain integrity of spatial data with a consistent, accurate database
Benefits of using a geodatabase

• Work within a multiuser access and editing environment
• Integrate spatial data with other IT databases.
• Easily scale your storage
• Support custom features and behavior
Two types of Geodatabases in ArcGIS
Personal geodatabase

• May be read by multiple people
• Only edited by 1 person at a time
• Max size 2 GB
• Currently stores only vector data
• Raster stored as catalogs
Multi-user geodatabase

- Suitable for large workgroups and enterprise GIS implementations
- May be read and edited by multiple users simultaneously
- Requires ArcSDE and a DBMS
- Supports both raster and vector formats
A geodatabase extends a database

- Provides the framework for defining and managing the GCS for a set of data
- Models topologically integrated sets of feature
- Defines general and arbitrary relationships between objects and features
A geodatabase extends a database

• Enforces the integrity of attributes through domains and validation rules

• Binds the natural behavior of features to the tables that store features

• Stores multiple versions so that many users can edit the same data
Spatial Query

• What is a spatial query?

• Name three examples of a spatial query
Spatial Query Operations

• Search
  – Thematic search
  – Search by region
  – Search by classification

• Location analysis
  – Buffer
  – Corridor
  – Overlay
Spatial Query Operations

• Terrain analysis
  – Slope/aspect
  – Catchment
  – Drainage network

• Flow analysis
  – Connectivity
  – Shortest path
Spatial Query Operations

• Distribution
  – Change detection
  – Proximity
  – Nearest neighbor

• Spatial analysis/Statistics
  – Pattern
  – Centrality
  – Autocorrelation
  – Indices of similarity
  – Topology
  – Hole description
Spatial Query Operations

• Measurements
  – Distance
  – Perimeter
  – Shape
  – Adjacency
  – Direction
Architecture of a Spatial Database
Core features of a DBMS

• Persistence
  – Some data are transient
  – Some are persistent

• Transactions
  – Map a database from one consistent state to another
  – All or abort
  – Many transactions executed concurrently
Space Taxonomy

- Orientation, Direction, Topological relationships (adjacent, next to, inside)

- Multitude of descriptions available to organize space

<table>
<thead>
<tr>
<th>Topological</th>
<th>Adjacent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network</td>
<td>Shortest-path</td>
</tr>
<tr>
<td>Directional</td>
<td>North-of</td>
</tr>
<tr>
<td>Euclidean</td>
<td>Distance</td>
</tr>
</tbody>
</table>
Data Model

• **Vector**
  - Points
  - Polylines
  - Polygons

• **Fields**
  - Raster, uniform grid imposed on underlying space
  - TIN, triangulated irregular networks
  - Contour lines
  - Point grids
Data Model Example
Data Model Example

Biodiversity Unit - Conceptual Relationships

Observations - Conceptual Relationships
Data Model Example

Assume each distribution polygon marks a location and distribution for a period of time. If not we need a separate Condition class.

Distribution
- DistributionID
- DistributionType
- BiodiversityName
- BiodiversityUNITED
- AreaName
- AreaDescription
- LocationMethod
- Confidence
- NumIndividuals
- ReproductiveCondition
- AgeSexRatio
- SuccessionalStage
- StartDateTime
- EndDateTime

Feature

Object

DistributionObject
- ID
- StartDateTime
- EndDateTime

DistributionTrend
- DeltaNumIndividuals
- DeltaReproductiveCondition
- DeltaAgeSexRatio
- DeltaSuccessionalStage

PhysicalContext
- Substrate
- Elevation
- Microclimate
- Process

BiologicalContext
- LandscapeConfig
- Habitat
- Process

ConservationContext
- LandscapeConfig
- Habitat
- Process

Most of the distribution-related objects describe the environment present with the distribution. The Trend object is different because it describes a summary of differences in distribution over time.
Structured Query Language (SQL)

• Declarative language

• User specifies the result

• Example: Find all land parcels next to my house.

SELECT M.address
FROM land_parcel L, city_name M
WHERE Adjacent (L,M) AND M.address = “MYHOUSE”
Types of queries

- Range (select all points within 10 km of the city limits)
- Join (combine two tables on a common attribute)
SQL Examples

• Join

```sql
SELECT S.name
FROM Senator S, Business B
WHERE S.soc-sec = B.soc-sec
S gender = 'Female'
```

• Spatial Join

```sql
SELECT S.name
FROM Senator S, Business B
WHERE S.district.Area() > 300
Within(B.location, S.district)
```
File organization and indices

Programmer’s Point of View

Database Manager’s Point of View
Data mining

• Systematic search of potentially useful information embedded in digital data

• Hot topic of research inside and outside of academia

• BIG DATA
Big Data Landscape

Acquisition → Marshalling → Analysis → Action

**Data Acquisition**
- IBM
- SAP
- SAS
- TIBCO
- HP
- Ab Initio
- Microsoft
- Oracle
- Informatica
- Splunk
- Numenta
- Syncsort

**VLDW and BI Appliances**
- SAS
- Oracle
- IBM
- SAP
- HP
- Microsoft
- TIBCO
- Parcels
- Teradata

**Analytics**
- EMC
- IBM
- SAS
- HP
- Mzinga
- GoodData
- Microsoft
- Kxen
- Oracle
- Teradata

**BPM & Action**
- TIBCO
- EMC
- Oracle
- Progress Software
- Adobe
- Pega
- Infosys
- IBM
- Open Text
- Salesforce
- Oracle
- Informatica

Including Complex Event Processing (CEP) tools

**Data Providers**
- LexisNexis
- comScore
- Inrix
- Microsoft
- Nielsen
- Experian
- GfK
- Clarivate
- Nielsen
- ipsofacts
- IRI
- FACT
- DATA
- SAAS
- Knoema
- Open Data
- GfK

**No SQL**
- hadoop
- Oracle
- cloudera
- EMC
- IBM
- Google
- SAP
- Teradata
- MarkLogic

**Data Virtualization**
- Composite
- Oracle
- Teradata

**Content Management**
- ORACLE
- EMC
- IBM

**BI Tools**
- ORACLE
- Microsoft
- GoodData
- Panorama
- Log XML
- Information Builders

**Data Governance**
- IBM
- SAP
- TIBCO
- HP
- Oracle

And all your own data
And your partners data

Capgemini
- Capping IT off
Manuel Sevilla
- 2012

(c) 2014, A. Michelle Lawing
Spatial data are Big Data
Spatial data mining

• Find patterns in data with respect to geography

• Not new in GIS, but methods are still in their infancy

• Up to 80% of digital data is spatial in nature!
Spatial data mining

• A spatial database specialist can leverage expertise to...
  – Design search algorithms to handle large amounts of data
  – Extend SQL with “mining” methods

• Random sampling commonly used in data mining

• Because of spatial autocorrelation, other techniques might need to be employed (spatial statistics)
Spatial databases

- Important for management of spatial and related data

- Important for the query of spatial data to extract both spatial and related data