

# **Lesson 3: Modelling the Urban World**

**Dr. Kam Tin Seong**

**Assoc. Professor of Information Systems**

**School of Computing and Information Systems,  
Singapore Management University**

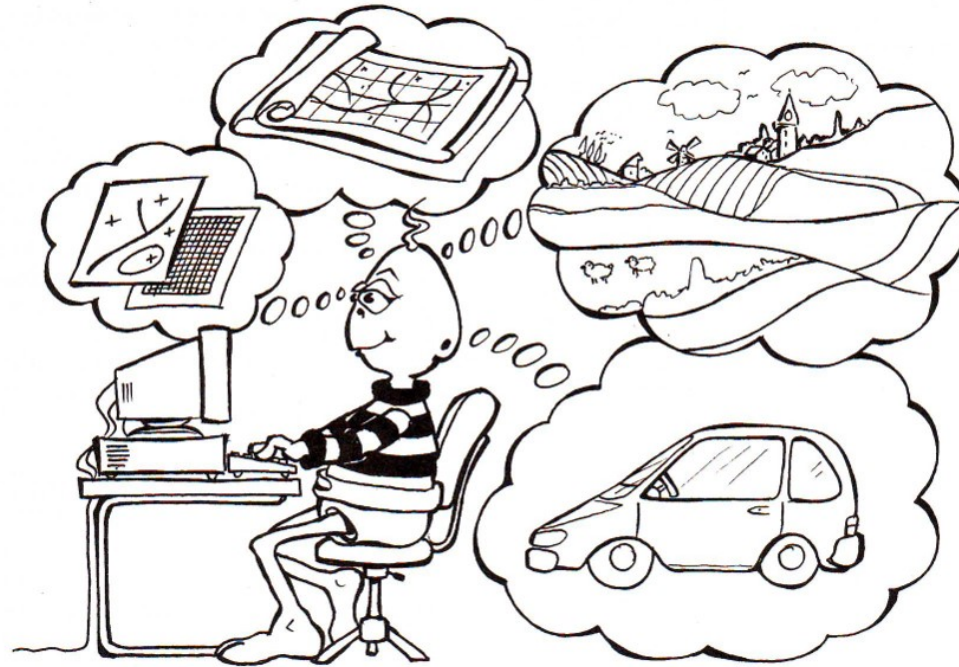
**2019-01-01 (updated: 2021-08-29)**

# Content

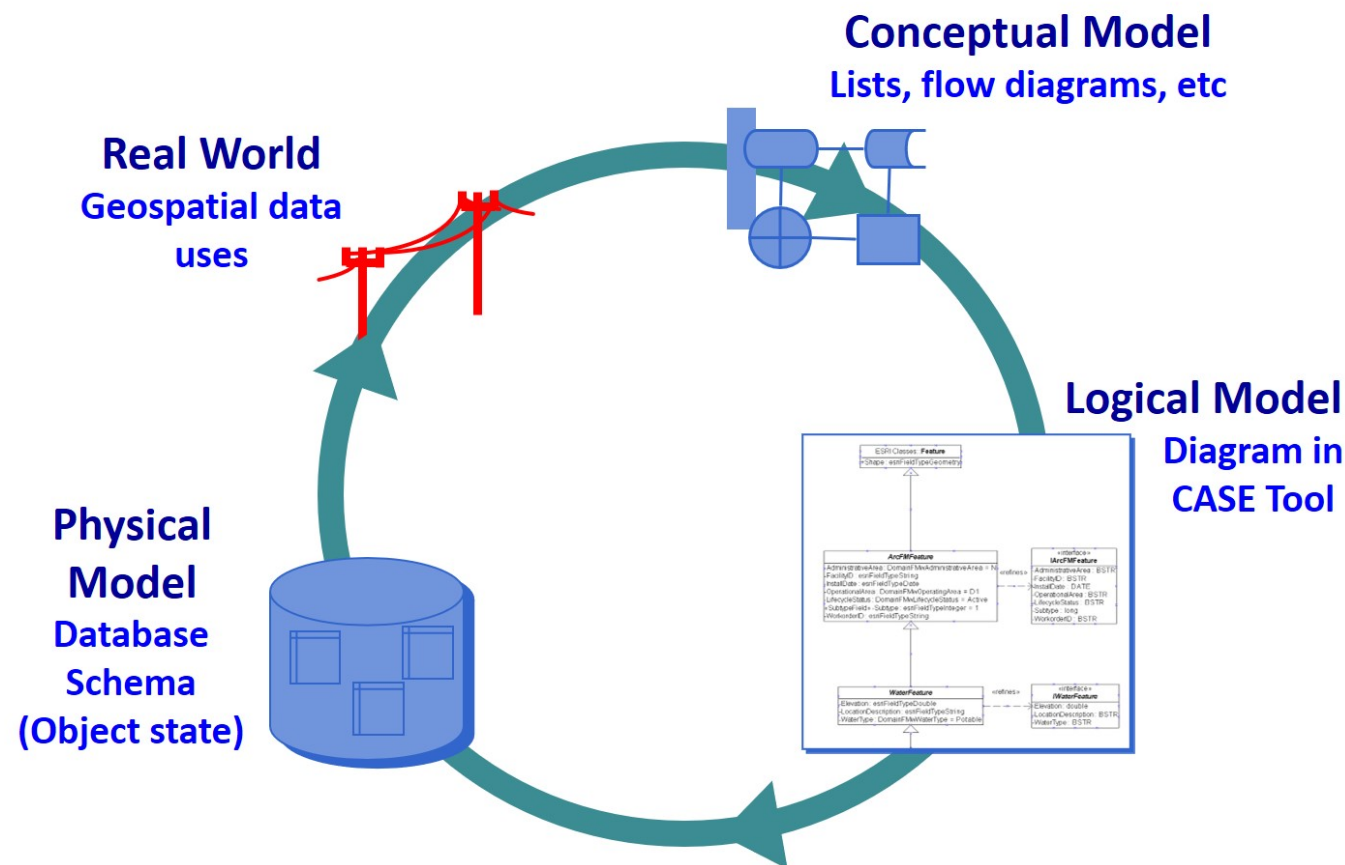
- Geospatial Data Modelling: Basic Concepts
- Geospatial Data Abstraction
- Topology Data Models
- Other Geospatial Spatial Data Properties
- Geospatial Data Repositories
- Metadata of Geospatial Data

# Geospatial Data Modelling

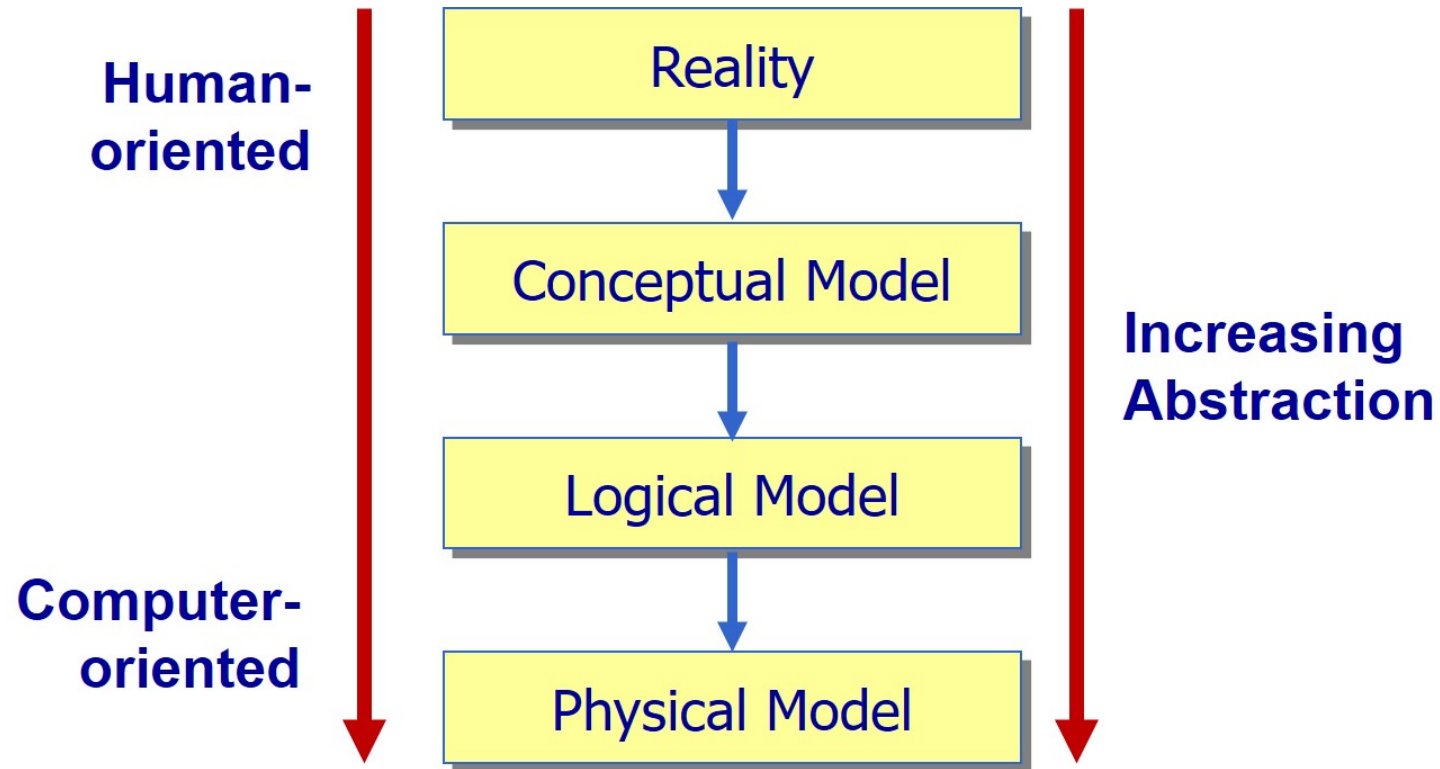
Thinking beyond the data



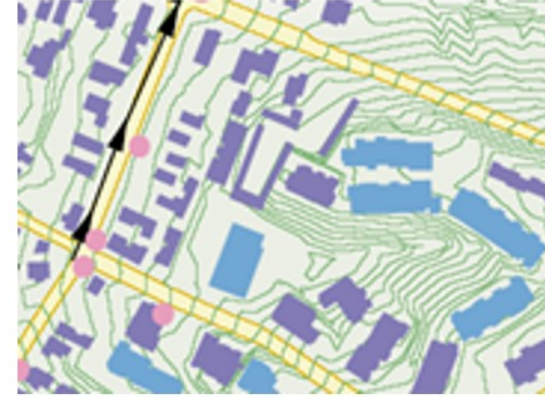
# Geospatial Data Modelling Process



# Levels of Data Model Abstraction



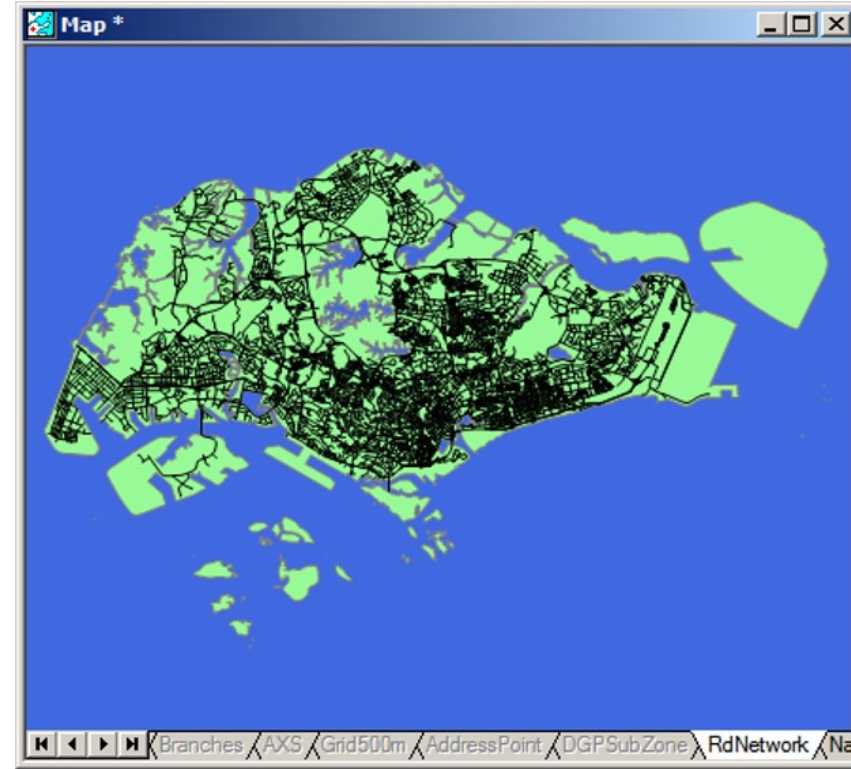
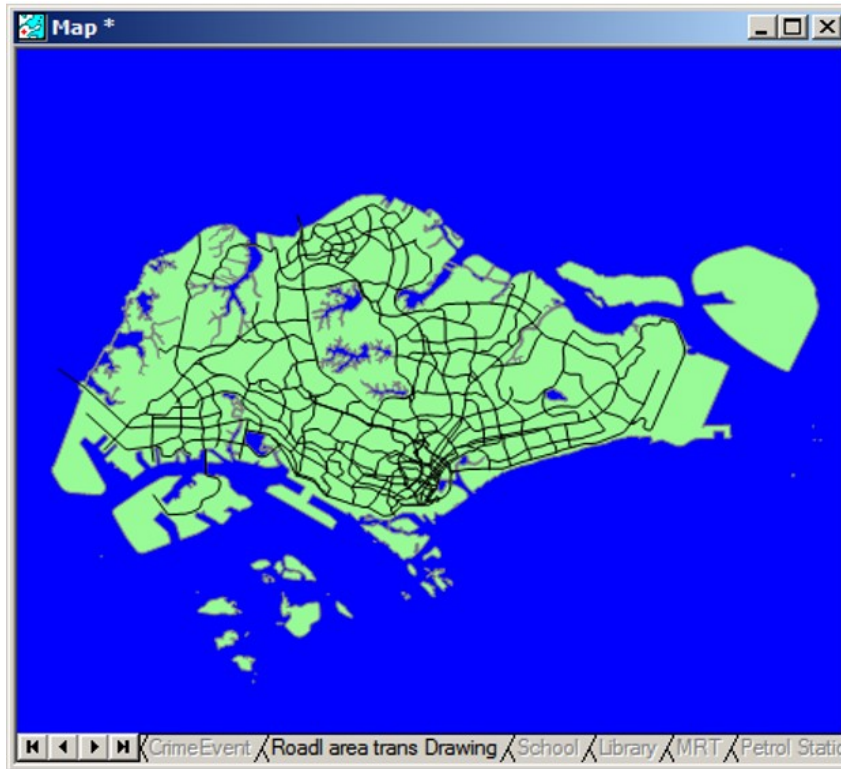
# Why should we worry about?



# What to represent?

The world is infinitely complex, but computer systems are finite.

- Representations must somehow limit the amount of details captured.





At what level of generalisation?

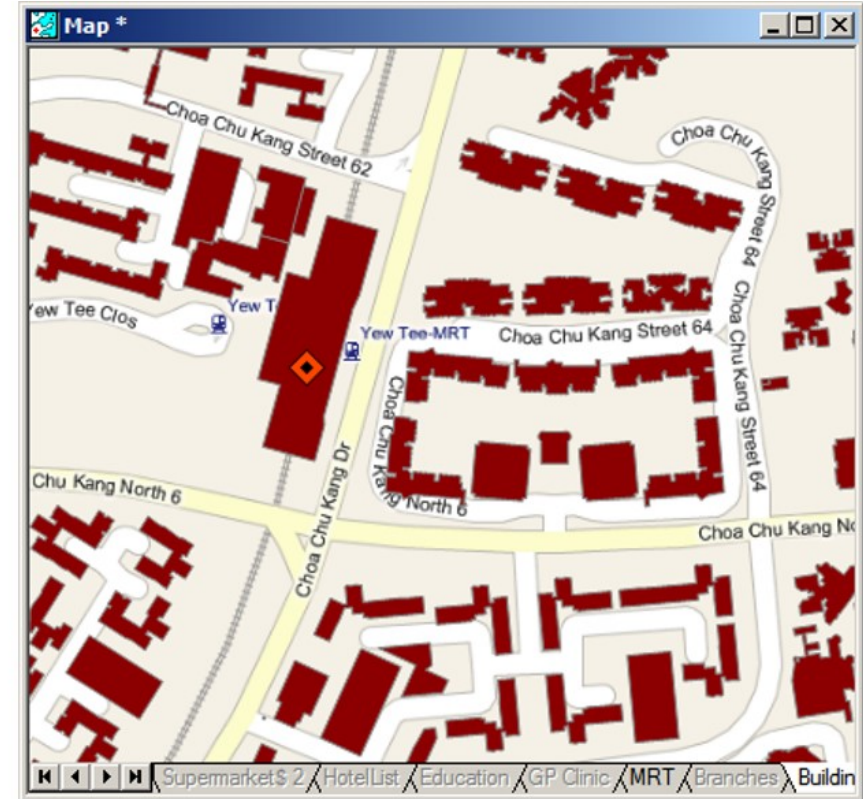
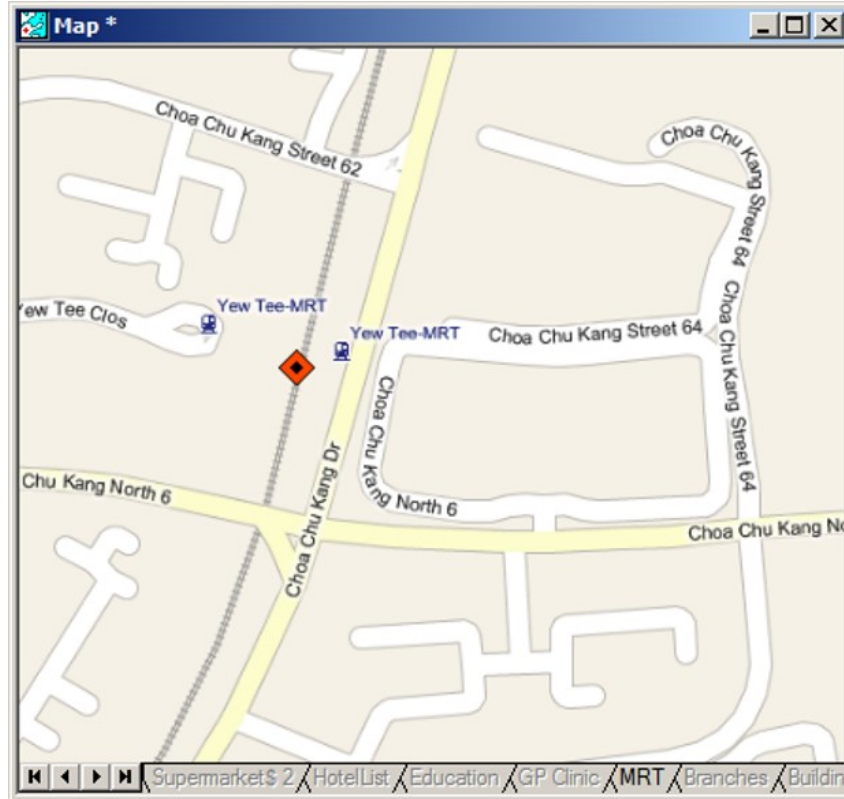
Little White Lies and Lots of Them





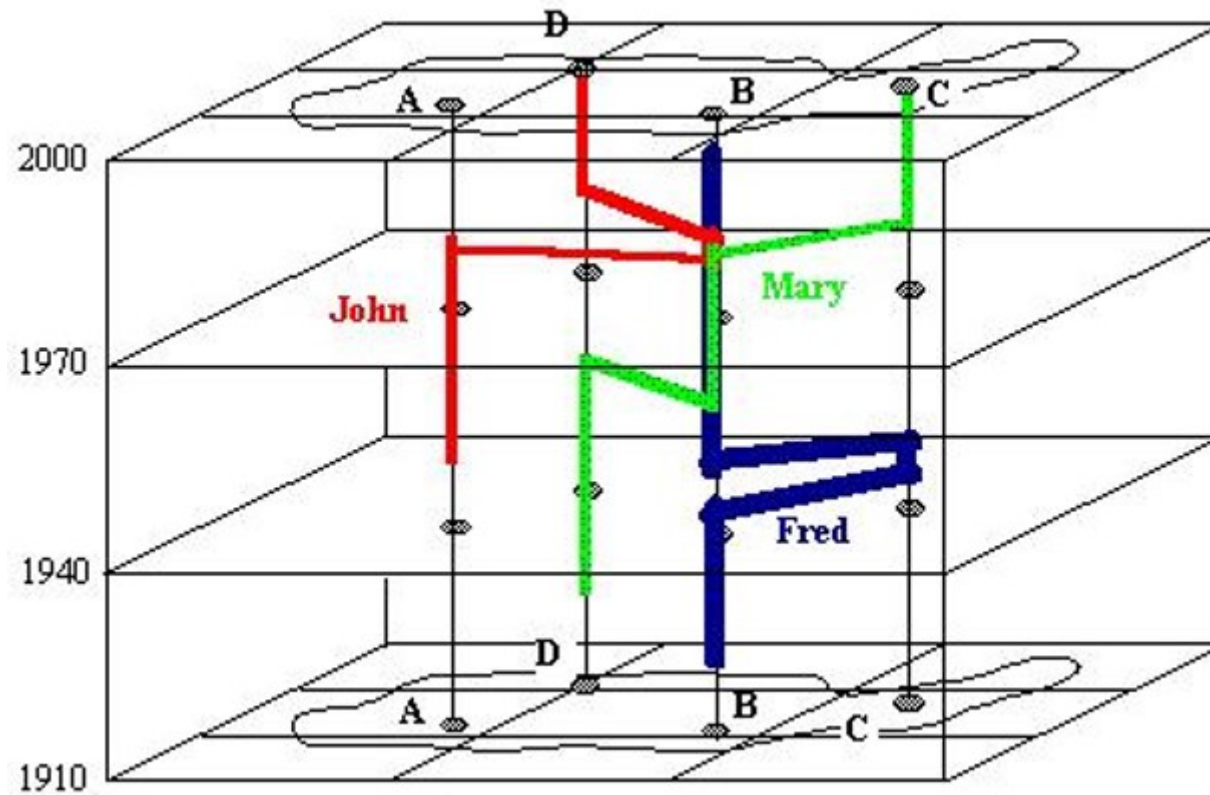
# How to represent?

- A MRT station can be represented by a point feature or a polygon feature.



# How to capture multiple contexts?

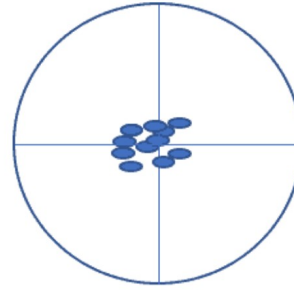
- Geographical data link place, time, and attribute.



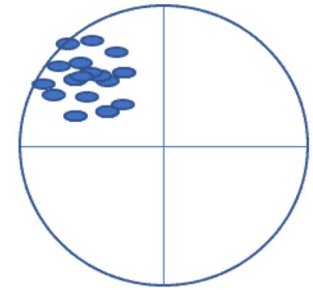
# Accuracy and Precision

- Accuracy can be defined as the degree or closeness to which the information on a map matches the values in the real world.
  - In GIS data, accuracy can be referred to a geographic position, but it can be referred also to attribute, or conceptual accuracy .
- Precision refers how exact is the description of data.
  - Precise data may be inaccurate, because it may be exactly described but inaccurately gathered. (Maybe the surveyor made a mistake, or the data was recorded wrongly into the database).

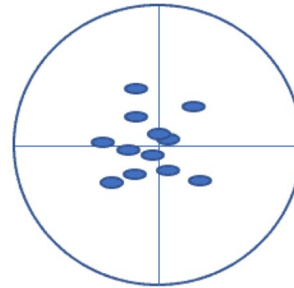
A: precise and accurate



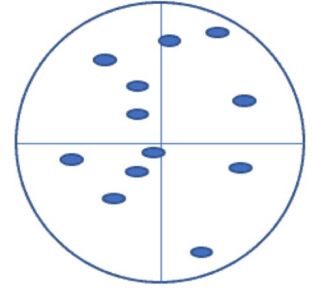
B: precise but not accurate



C: Not precise but accurate



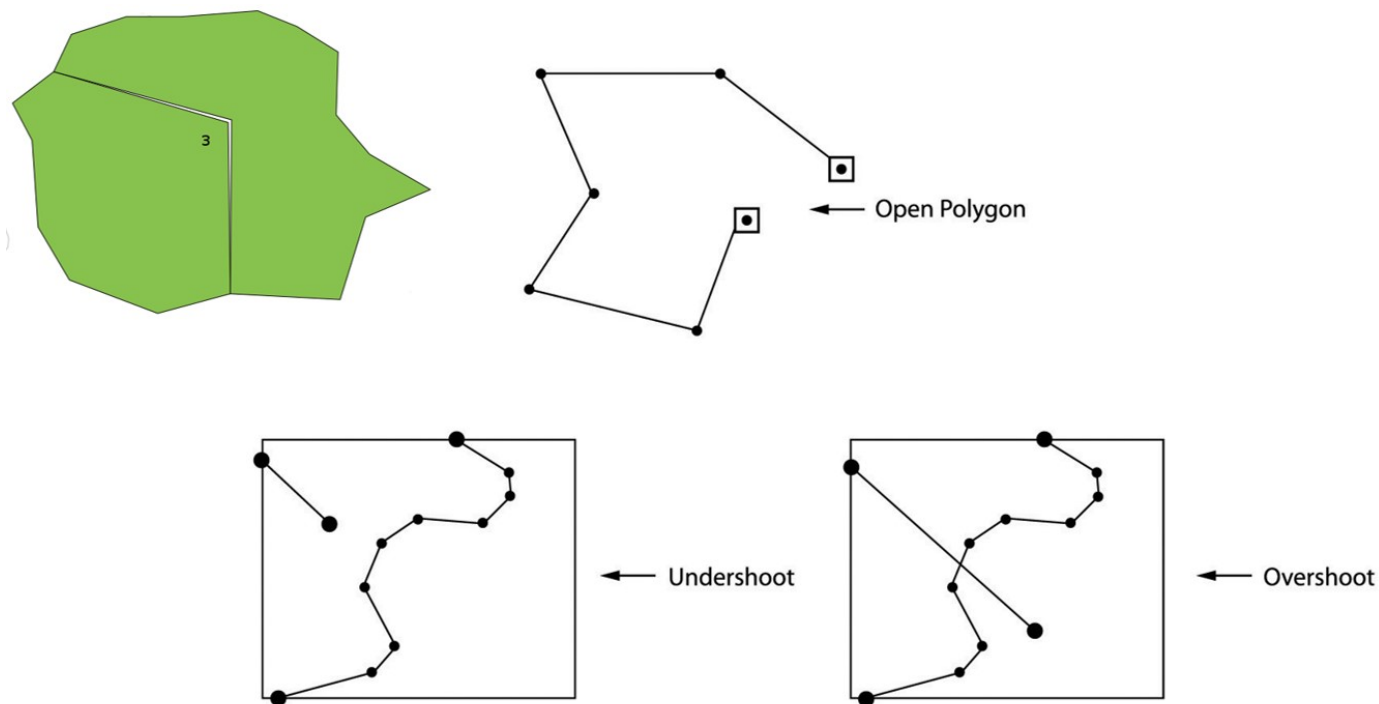
D: Not precise and not accurate



Precision versus Accuracy. Image: Caitlin Dempsey

# Motivation of topological data models

- Topological errors



# What is Topological Errors?

- Topological errors occur often during the digitizing process. Errors of the operator may result in polygon knots, and loops, and there may be some errors associated with damaged source maps as well.

Dangles



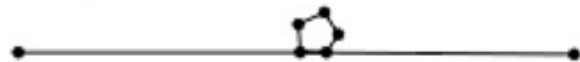
Switchbacks



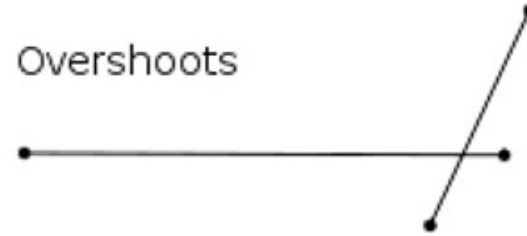
Knots



Loops



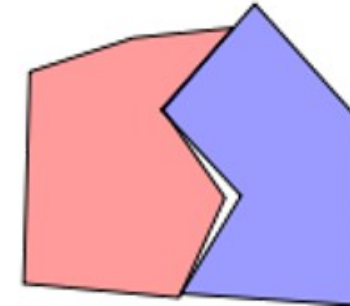
Overshoots



Undershoots

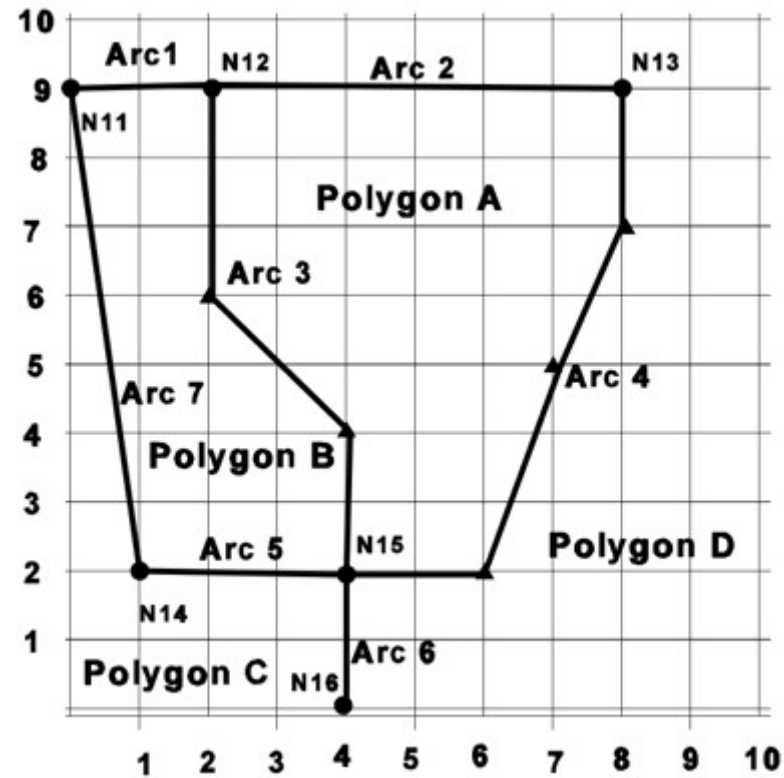


Slivers



# Topological data models

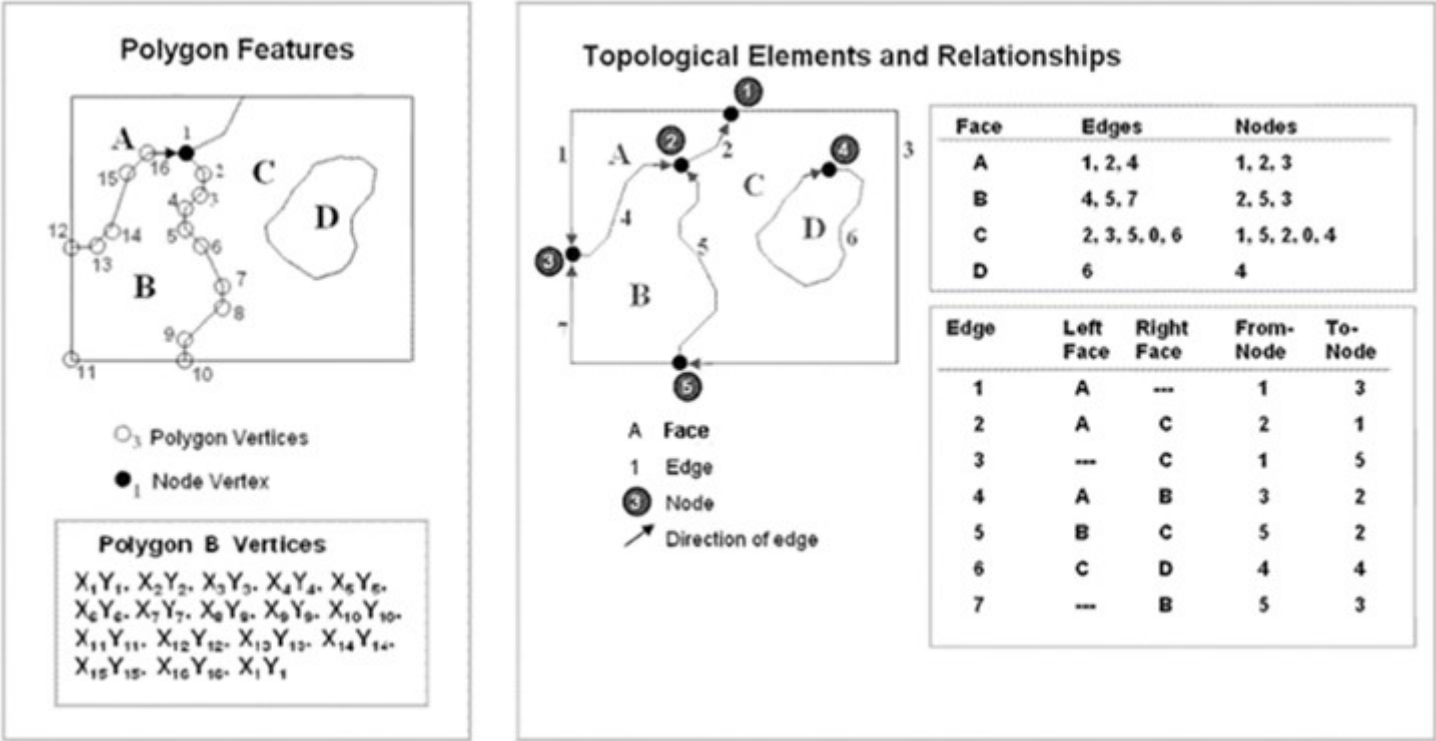
- Data redundancy





# Intelligence GIS Data Model: Topology

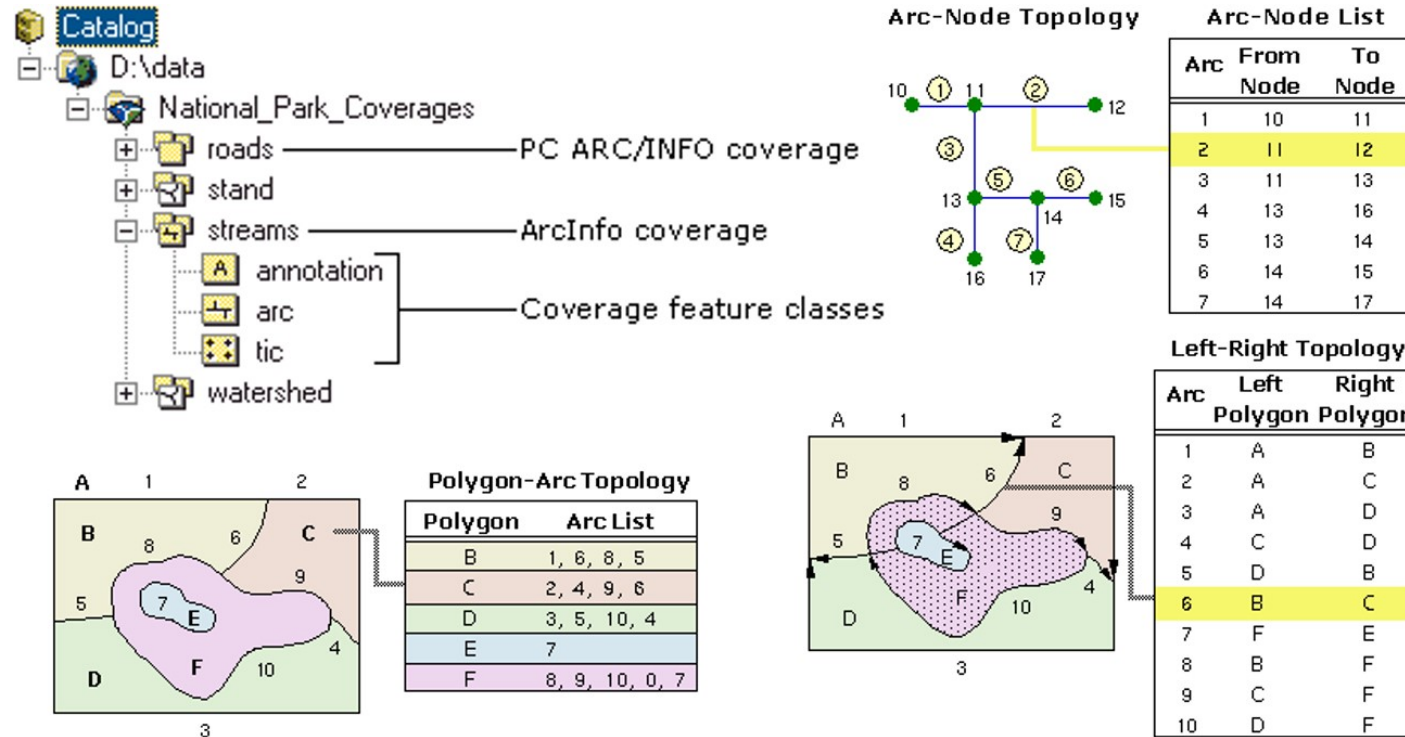
- It manages spatial relationships by representing spatial objects (point, line, and area features) as an underlying graph of topological primitives?nodes, faces, and edges.



Source: [ArcGIS Topology basics](#)

# Topological Data Models: ArcGIS Coverage













- A coverage is a georelational data model that stores vector data and their topology.



Reference: [What is a coverage?](#)

# Rule-based Topology

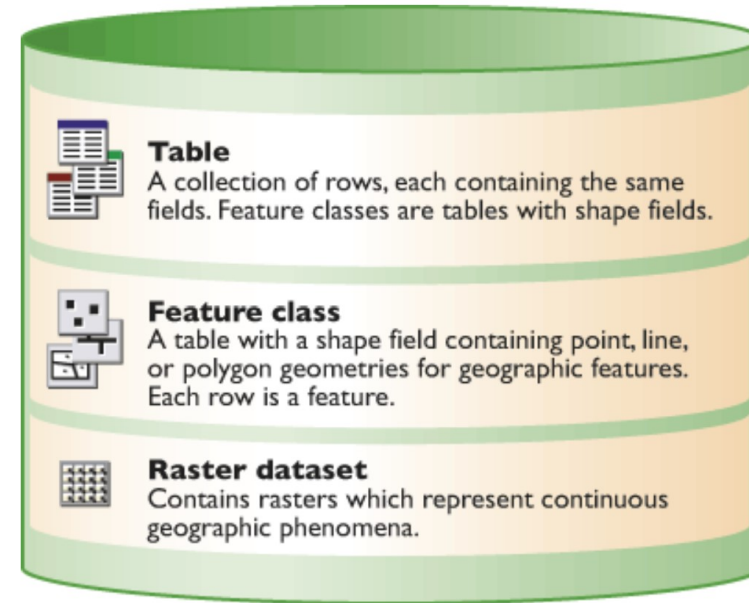
- Topology rules define the permissible spatial relationships between features.
  - The rules you define for a topology control the relationships between features within a feature class, between features in different feature classes, or between subtypes of features.

 Points	 Points on points	 Points on lines <i>Must be covered by endpoint of</i> <i>Point must be covered by line</i>	 Points on polygons <i>Must be properly inside polygons</i> <i>Must be covered by boundary of</i>
 Lines <i>Must not have dangles</i> <i>Must not have pseudo-nodes</i> <i>Must not overlap</i> <i>Must not self overlap</i> <i>Must not intersect</i> <i>Must not self intersect</i> <i>Must not intersect or touch interior</i> <i>Must be single part</i>	 Lines on points <i>Endpoint must be covered by</i>	 Lines on lines <i>Must not overlap with</i> <i>Must be covered by feature class of</i>	 Lines on polygons <i>Must be covered by boundary of</i>
 Polygons <i>Must not overlap</i> <i>Must not have gaps</i>	 Polygons on points <i>Contains point</i>	 Polygons on lines <i>Boundary must be covered by</i>	 Polygons on polygons <i>Must be covered by feature class of</i> <i>Must be covered by</i> <i>Must not overlap with</i> <i>Must cover each other</i>

Source: [Topology in ArcGIS](#)

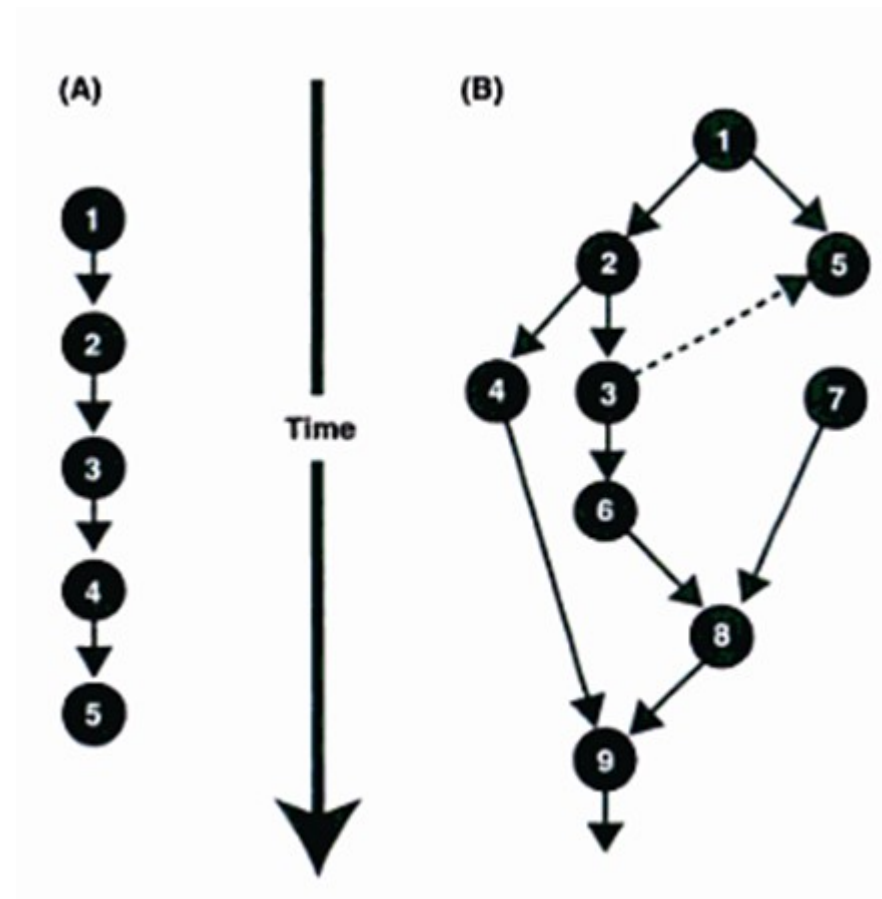
# Rule-based Topology: ArcGIS Geodatabase

- The geodatabase is a collection of geographic datasets of various types used in ArcGIS and managed in either a file folder or a relational database.
  - It is the native data source for ArcGIS and is used for editing and data automation in ArcGIS.
- For more information, visit [What is a geodatabase?](#)



# Issue of Multi-user editing: Transactions

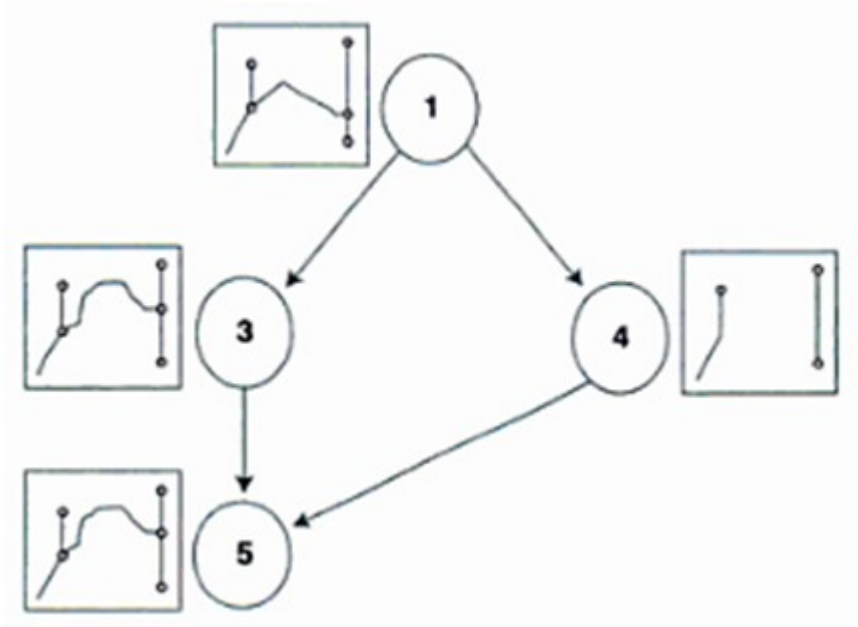
- A transaction is a group of changes that are made to a database as a coherent group. All the changes that form part of a transaction are either committed or the database is rolled back to its initial state
  - The solutions: pessimistic locking vs optimistic versioning



Source: Longley, P.A. et. Al. (2005) 2 nd Edition, Geographic Information Systems and Science, John Wiley & Sons.

# Versioning

- Creating multiple versions for editing and resolving conflict.



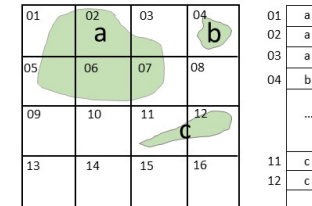
Source: Longley, P.A. et. Al. (2005) 2 nd Edition, Geographic Information Systems and Science, John Wiley & Sons.



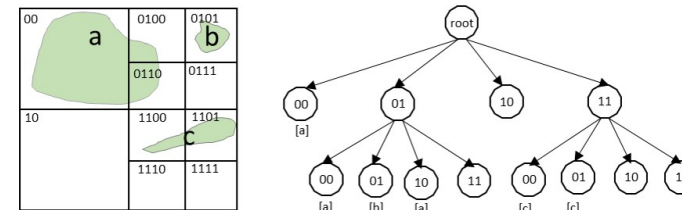
# Spatial Indexing

- Spatial indexes are used by spatial databases to optimise spatial queries.
  - Indexes used by non-spatial databases cannot effectively handle features such as how far two points differ and whether points fall within a spatial area of interest.

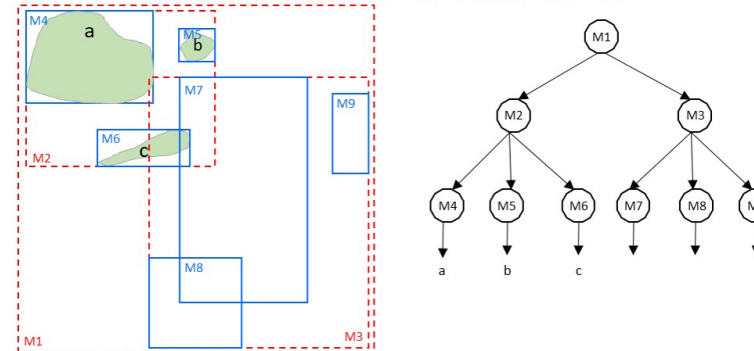
An example of a fixed grid structure.



A representation of a quadtree structure.



R-Tree Hierarchical Index in Minimum Bounding Rectangles (MBRs).



To learn more about Spatial Indexing, refer [DM-66 - Spatial Indexing](#)

# Geospatial Data Repositories

- File-based
  - Vector format
  - Raster format
  - Database management systems
- Middleware
- Geospatial extension
  - Commercial
  - Open source

# About SpatiaLite



- An open source library intended to extend the SQLite core to support fully fledged Spatial SQL capabilities.
  - SQLite is intrinsically simple and lightweight:
    - a single lightweight library implementing the full SQL engine
    - standard SQL implementation: almost complete SQL-92
    - no complex client/server architecture
    - a whole database simply corresponds to a single monolithic file (no size limits)
    - any DB-file can be safely exchanged across different platforms, because the internal architecture is universally portable
    - no installation, no configuration
- SpatiaLite is licensed under the MPL tri-license terms. you are free to choose the best-fit license between:
  - the MPL 1.1
  - the GPL v2.0 or any subsequent version
  - the LGPL v2.1 or any subsequent version
  - For more information
  - [Homepage](#)

# GeoPackage

- An open, non-proprietary, platform-independent and standards-based data format for geographic information system implemented as a SQLite database container.
- To learn more about geopackage, visit [geopackage homepage](#) and at [wiki](#).



## Getting to know GeoPackage

- Use open source SQLite as the storage engine, more reliable and platform independent than MDB, but with the advantage of easy, language independent, read/write access via SQL.
  - The SQLite engine is open source and multi-platform, so no Windows dependency.
  - The SQLite engine stores data in a single file, so no need to zip/unzip all the time .
  - Leverage existing OGC standards like the WKT standard for spatial reference systems, and the WKB standard for binary geometry representation.
- Document the format and include an extension mechanism so it can evolve over time and so third parties can experiment with new extensions.
- To learn more, read [Getting Started With GeoPackage](#)

# Enterprise Geographic DBMS

- Provide spatial database management functions on top of conventional database management systems
  - For example IBM DB2 Spatial Extender, Informix Spatial Datblade



**ORACLE®** **11g** Oracle Spatial & Oracle Locator:  
DATABASE **Location Features for Oracle Database 11g**



## Geographic Middleware Extensions

- Middle-tier application server that extend the DBMS software kernel to manage geographic data
  - For example: Geomedia Transaction Server, SpatialWare from MapInfo, Enterprise Server from Manifold

What about



**ArcSDE**

# About PostGIS

- **PostGIS** spatially enables the **PostgreSQL** open source relational database. The database can then be used to store and query geospatial data (vector and raster).
- To learn more, visit [Introduction to PostGIS](#) and [wiki](#).



# Main Features of PostGIS

- High performance, robust spatial database built on PostgreSQL
  - Simple Features for SQL (SFSQL) compliance
  - Proven reliability and transactional integrity (ACID compliance)
  - Provides spatial representations of geometry types both vector (points , lines, polygons) and raster.
  - Support for common and advanced spatial operations such as geometry creation and conversion, reprojection , buffer, convex hull, generalization, union, and more
  - Geodetic support for measurements across the globe/dateline
  - Command-line and graphical tools for flexible management Yes, PostGIS is on its way!
- To learn more, read [Why should you care about PostGIS? — A gentle introduction to spatial databases](#) and [PostGIS Manual](#).

## Who Use PostGIS

- Proprietary Desktop: ArcGIS, Manifold, Safe FME, CadCorp SIS, MapInfo Professional
  - Proprietary Server: ArcServer , Ionic Enterprise, MapDotNet Server
  - Open Source Server: GeoServer , Mapserver , Mapnik , DeeGree , SharpMap
  - Open Source Desktop: GRASS, QGIS, uDig , gvSIG



## Carto: Database on the cloud

- [Carto](#) is a Software as a Service (SaaS) cloud computing platform that provides GIS and web mapping tools in the browser. (Formally know as CartoDB )
  - Carto was built on open source software including PostGIS and PostgreSQL .
  - The tool uses JavaScript extensively, including use in the front end web application, back end Node.js based APIs, and for the client libraries.
- To learn more, visit [wiki](#) and [How CARTO works](#)

# Meta data: Data about the data

- Identification information
  - Data quality information
  - Spatial data organisation information
  - Spatial reference information
  - Entity and attribute information
  - Distribution information
  - Meta data reference information
  - Citation information
  - Time period information
  - Contact information



# An example of metadata

- Metadata is information about data. Similar to a library catalog record, metadata records document the who, what, when, where, how, and why of a data resource.
- Geospatial metadata describes maps, Geographic Information Systems (GIS) files, imagery, and other location-based data resources.
- To learn more, visit [Geospatial Metadata](#)

## Soil Survey Geographic 2008 (SSURGO) Database for Atlantic County, New Jersey (Projected to NJ State Plane Feet, NAD83)

Metadata also available as

### Metadata:

- [Identification Information](#)
- [Data Quality Information](#)
- [Spatial Data Organization Information](#)
- [Spatial Reference Information](#)
- [Entity and Attribute Information](#)
- [Distribution Information](#)
- [Metadata Reference Information](#)

#### Identification Information:

##### Citation:

##### Citation Information:

*Originator:* New Jersey Department of Environmental Protection (NJDEP)  
*Originator:* U.S. Department of Agriculture, Natural Resources Conservation Service (NRCS)  
*Publication Date:* 20080818

*Title:* Soil Survey Geographic 2008 (SSURGO) Database for Atlantic County, New Jersey (Projected to NJ State Plane Feet, NAD83)  
*Edition:* Version 20080818

*Geospatial Data Presentation Form:* vector digital data

##### Publication Information:

*Publication Place:* Trenton, NJ  
*Publisher:* NJ Department of Environmental Protection (NJDEP)

*Other Citation Details:*  
*Advisory:*

#### Purpose:

SSURGO depicts information about the kinds and distribution of soils on the landscape. The soil map and data used in the SSURGO product were prepared by soil scientists as part of the National Cooperative Soil Survey.

#### Supplemental Information:

Digital versions of hydrography, cultural features, and other associated layers that are not part of the SSURGO data set may be available from the primary organization listed in the Point of Contact.

#### Time Period of Content:

##### Time Period Information:

*Range of Dates/Times:*  
*Beginning Date:* 19991013  
*Ending Date:* 20080815

*Currentness Reference:* publication date

#### Status:

*Progress:* Complete

*Maintenance and Update Frequency:* As needed

#### Spatial Domain:

##### Bounding Coordinates:

*West Bounding Coordinate:* -74.986720  
*East Bounding Coordinate:* -74.306382  
*North Bounding Coordinate:* 39.730072  
*South Bounding Coordinate:* 39.287662

#### Keywords:

##### Theme:

*Theme Keyword Thesaurus:* None  
*Theme Keyword:* soil survey  
*Theme Keyword:* soils  
*Theme Keyword:* Soil Survey Geographic  
*Theme Keyword:* SSURGO  
*Theme Keyword:* survey

##### Theme:

*Theme Keyword Thesaurus:* ISO 19115 Topic Category  
*Theme Keyword:* environment

##### Place:

*Place Keyword Thesaurus:* USGS Topographic Map Names Data Base  
*Place Keyword:* Atlantic County  
*Place Keyword:* New Jersey