Project for the Strengthening of Spatial Data Infrastructures in Member States and Territories of the Association of Caribbean States

Capacity Building Program

Geographic Information Systems

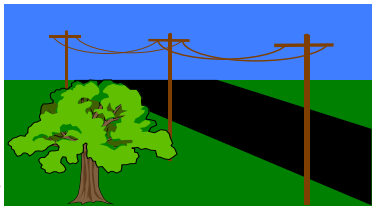
Dr Earl Edwards
Ms. Valrie Grant, MSc.

UN-GGIM:Américas
COMITÉ REGIONAL DE LAS NACIONES UNIDAS SOBRE LA GESTIÓN GLOBAL DE INFORMACIÓN GEOESPACIAL PARA LAS AMÉRICAS
ACS AEC
Association of Caribbean States
Asociación de Estados del Caribe
Association des Etats de la Caraïbe
AMEXCID
AGENCIA MEXICANA DE COOPERACIÓN INTERNACIONAL PARA EL DESARROLLO
INEGI
INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Topic 5:
Geospatial Data Management

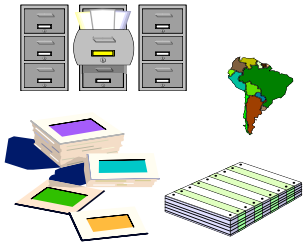
The GIS Process

Geospatial Information processing begins and ends with the real world



Real World

Data Collection



Data Sources

Data Input

Data Management

Data Retrieval & Analysis

Analysis

Information for Decision Making

Users

Implementation

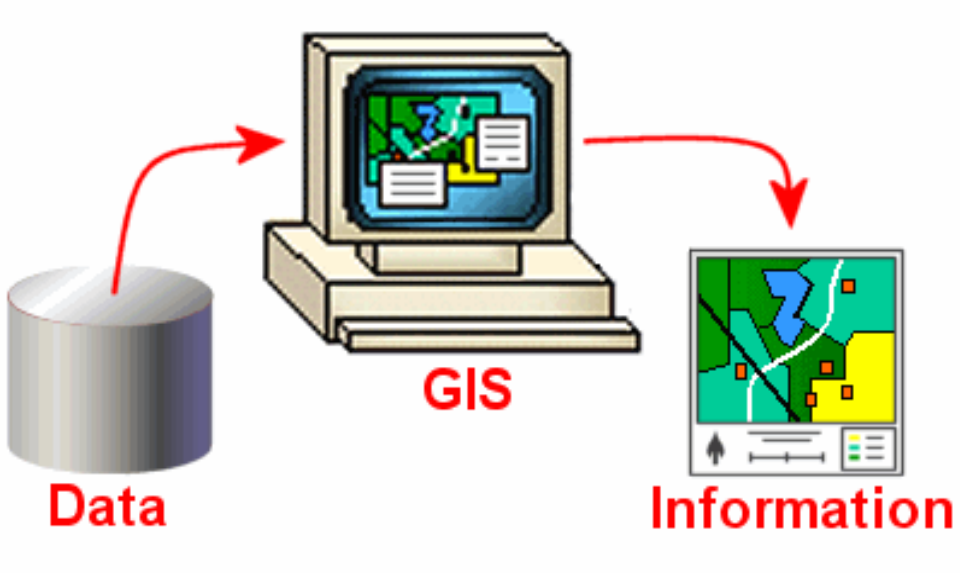
Source: Aronoff (1991)

Topic Outline

- Data and Information
- Database Management Systems (DBMS)
- Relational Database Model
- Spatial Data Storage Formats
- Single User vs Multiuser Geodatabases
- Versioning and Replication
- Distributed Geodatabases

Data and Information

- Simply put, data is what goes into the GIS; information is what comes out.
- Information is data, which when processed, would remove the level of uncertainty in decision making.



Database

- An organized collection related data items
- Stored in a highly structured way
- Represents a model of reality
- Data can be stored in a single location or HDD
- Or distributed across large networks

Database Management Systems

- A DBMS is a computer software application that interacts with the user, other applications, and the database itself to capture and analyze data.
- An application that allows users to interact with data.
- A general-purpose DBMS is designed to allow the definition, creation, querying, update, and administration of databases.

Examples of DBMS

- Server DBMS
  - Microsoft SQL Server
  - Oracle
  - DB2
  - PostgreSQL
  - MySQL
- Desktop DBMS
  - Microsoft Access

Database Types

Type	Typical number of users	Typical architecture	Typical size
Personal	1	Desktop/Laptop/ PDA	MB
Workgroup	5-25	Client/server:2 tier	MB-GB
Department	25-100	Client/server:3 tier	GB
Enterprise	>100	Client/server: distributed	GB-TB
Internet	>1000	Web sever & application servers	MB-GB

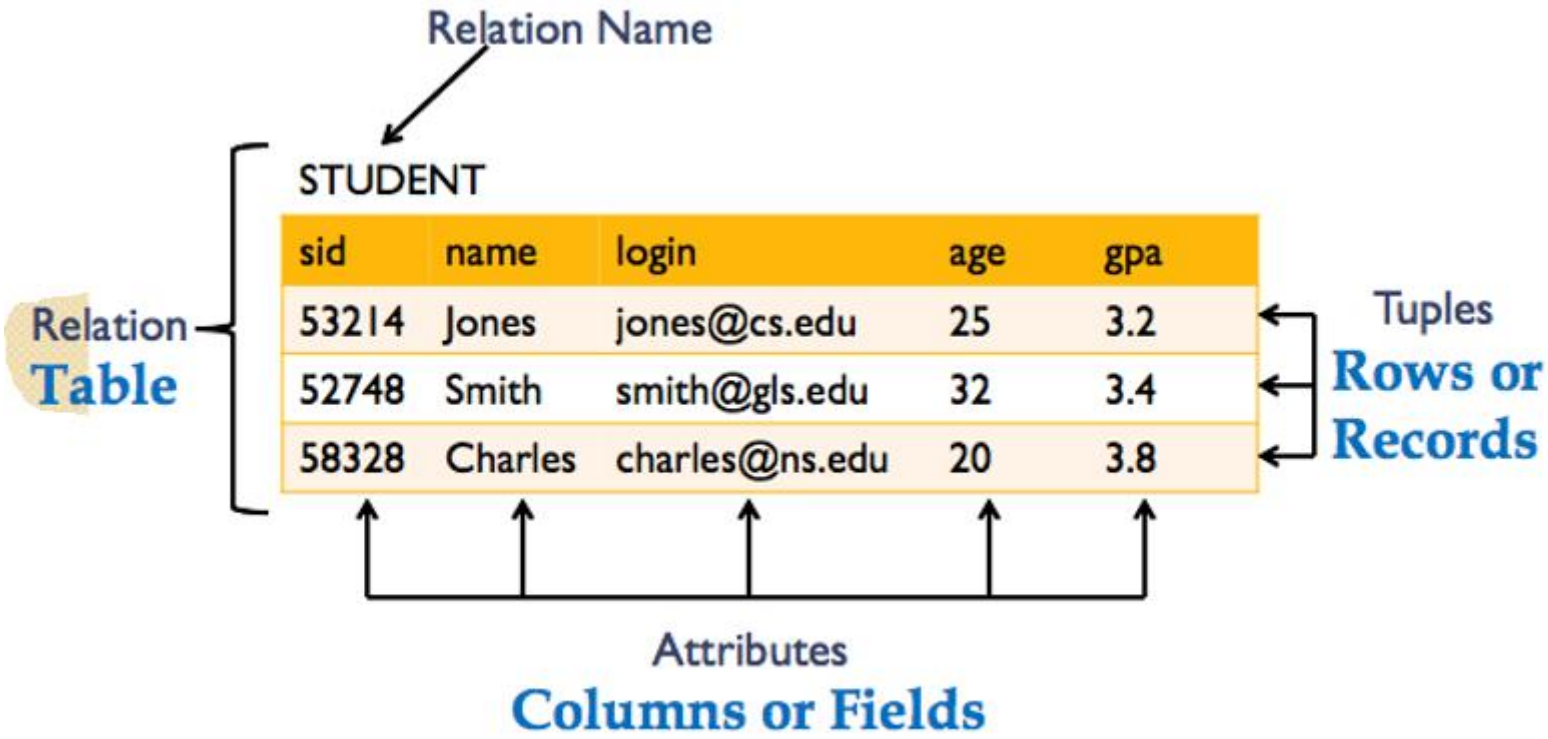
Database Models

- Flat files ‘60
- Hierarchical ‘60
- Network ‘70
- Relational ‘80
- Object oriented ‘90
- Object relational ‘90
- Web enabled ‘90

Relational Database Model

- A method of structuring data in the form of sets of records so that relations between different entities and attributes can be used for data access and transformation
- Proposed by E.F. Codd in 1970 and is based on formal mathematical theory (relational algebra)
- Allows definition of relationships between data tables through common attributes

Relational Terms and Concepts



Monitoring Wells

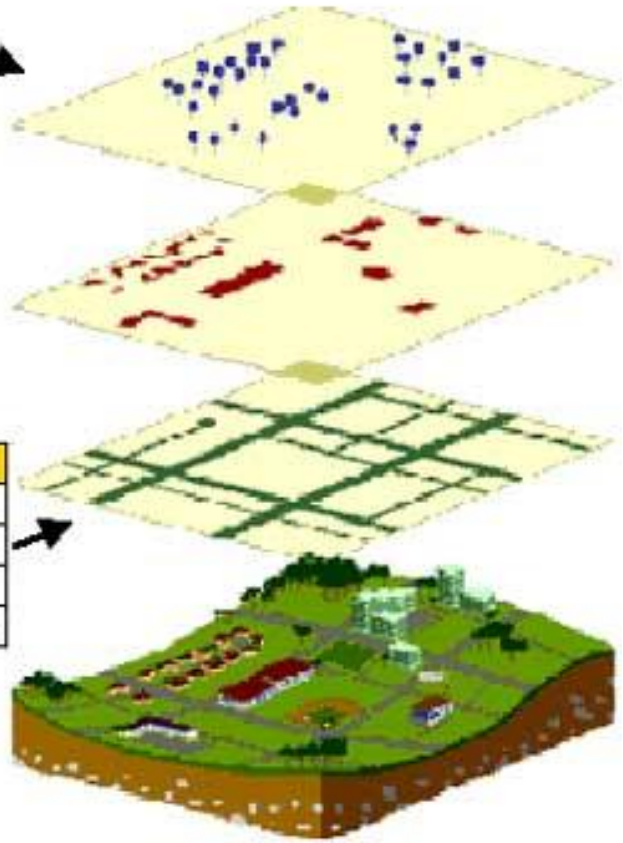
Well ID	Date Sampled	Concentration
C-6A	5/8/94	300
C-8A	5/8/94	20
C-13A	5/8/94	120
C-17A	5/8/94	560

Industries

Facility	Address
Acme	3029 Convington Dr.
Fox	742 West Lake St.
TPC	90 Aspen Dr.

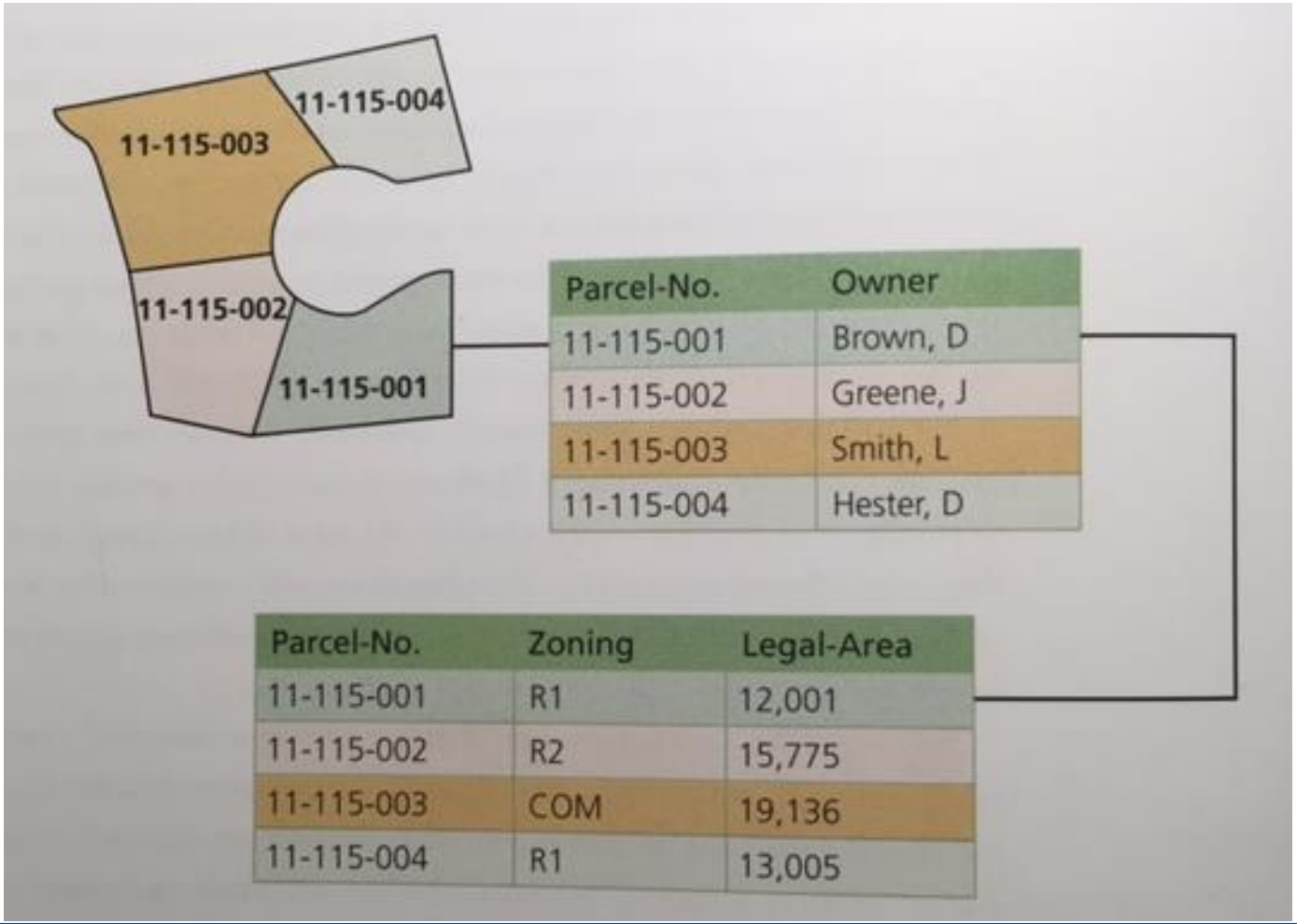
Population

Family Name	Occupants	Addresss
Blake	6	79 Circuit St
Hernandez	2	148 Plain St.
Joy	4	18 Webster St.
Smith	5	4321 Tecumseh Dr.



Linking Tables Using Common Fields

One to Many Relationship

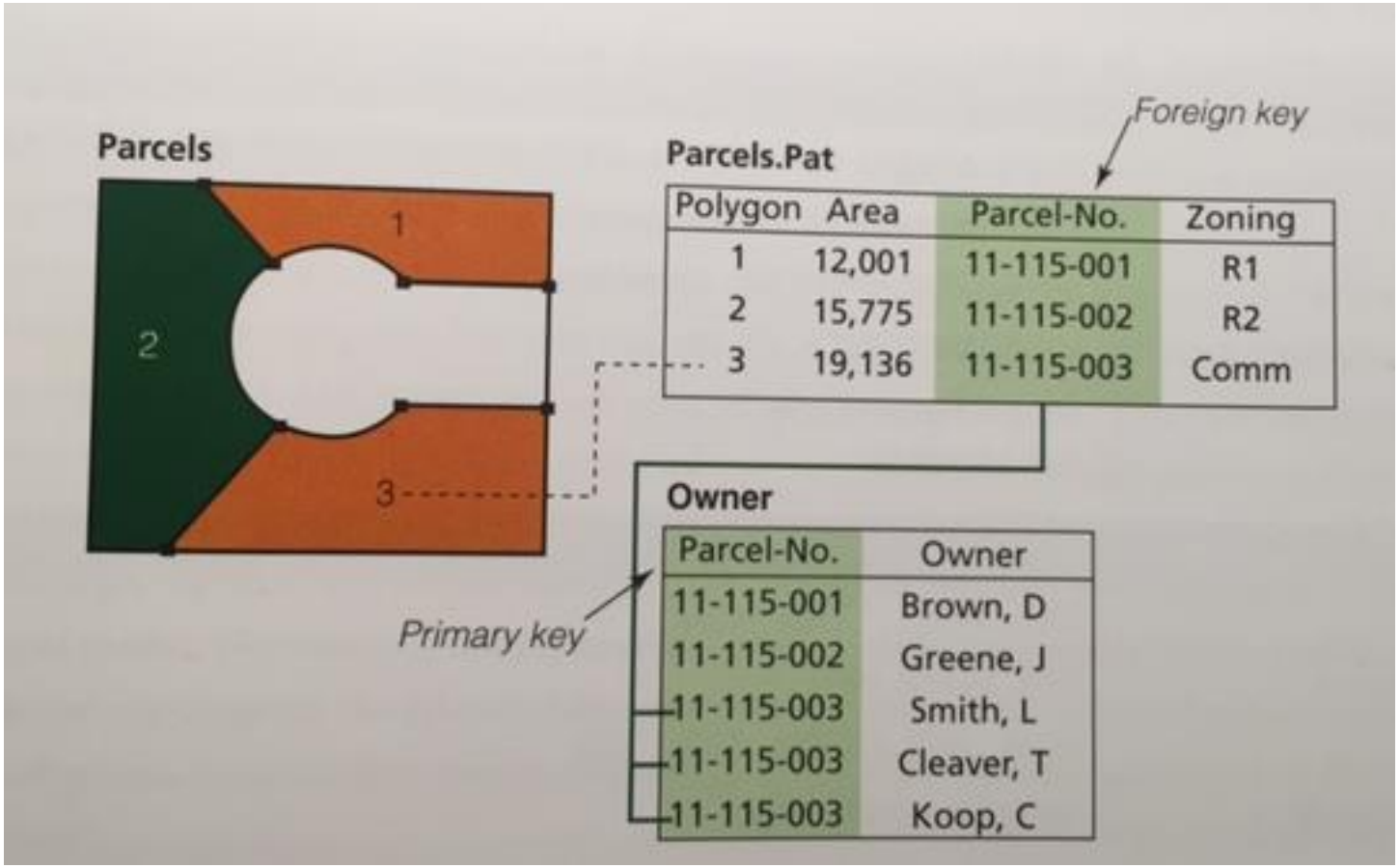


Parcel-No.	Owner
11-115-001	Brown, D
11-115-002	Greene, J
11-115-003	Smith, L
11-115-004	Hester, D

Parcel-No.	Zoning	Legal-Area
11-115-001	R1	12,001
11-115-002	R2	15,775
11-115-003	COM	19,136
11-115-004	R1	13,005

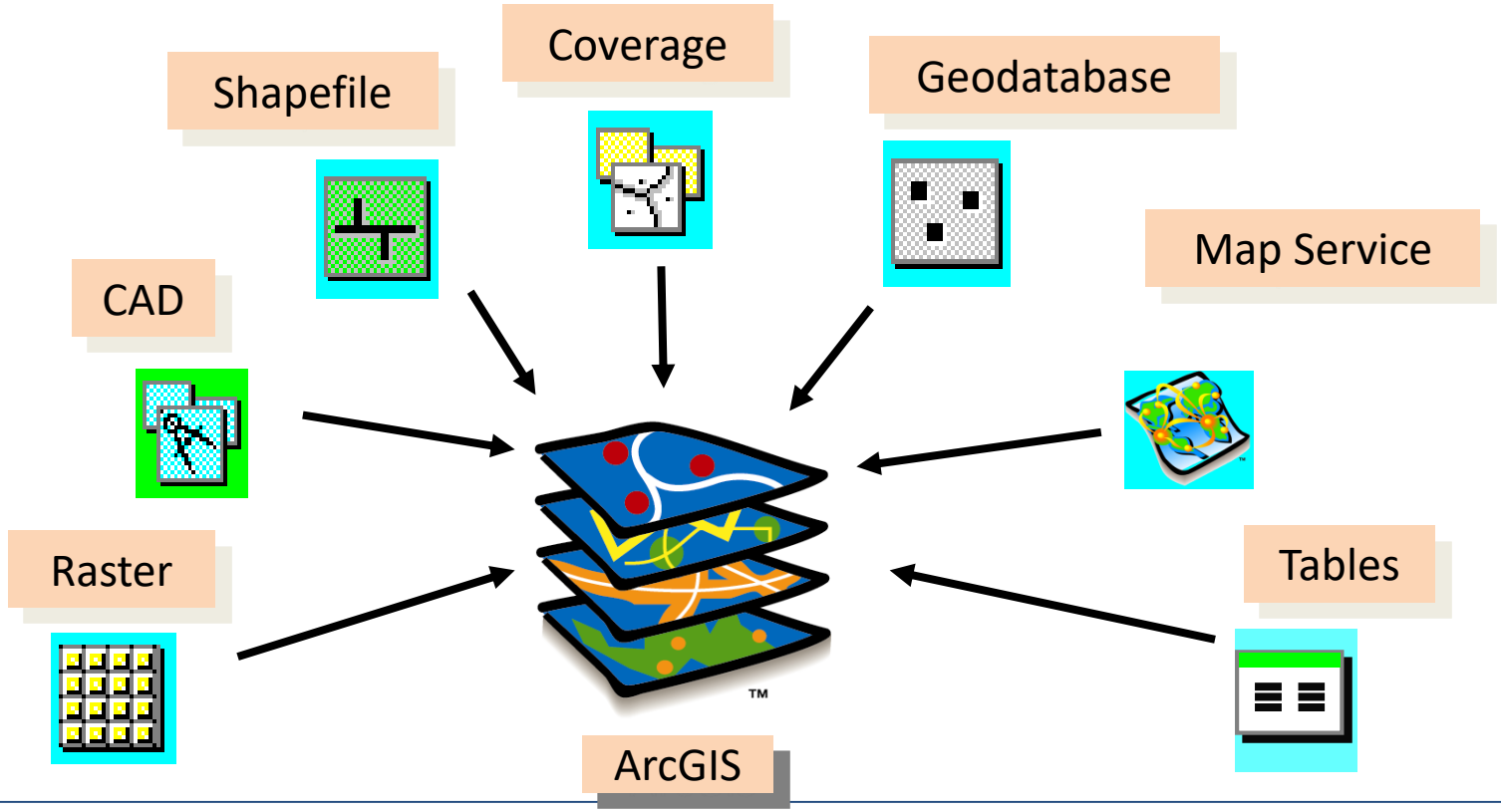
Linking Tables Using Common Fields

One to Many Relationship



Spatial Data Storage Formats

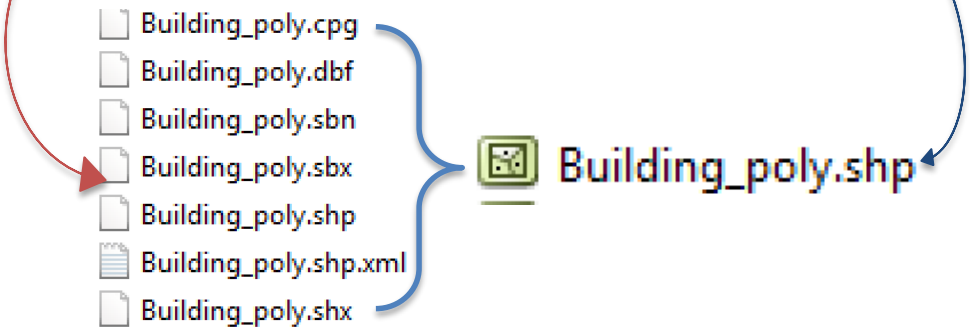
- ArcGIS can work with spatial data in multiple formats



Source: ESRI

Shapefile Vector File Format

- Shapefiles are vector composite files made up of 3-13 separate files
- In Windows Explorer all shapefile components are shown, in ArcCatalog the entire shapefile is shown as one item.



✸ Basic shapefile elements, without which, a shapefile is defunct

Possible composite file extensions:

✸**.dbf**- dBase table (database) file, containing attributes

✸**.shp**- the filewhich stores feature geometry (x,y coordinates)

✸**.shx**- file that stores the index connecting .dbf and .shp files

.prj- projection file

.shp.xml- metadata file

.sbn, .sbx- spatial index files- sometimes present

.ain, .aih- attribute index files

.atx- new, ArcGIS, attribute index file

.lxs, .mxs- geocoding index files

.cpg- specifies character set code page

- All components of a shapefile need to be present together, otherwise the shapefile can be defunct or incomplete.
- The projection file is a useful, but not necessary addition to a shapefile. It can be added or changed accordingly.
- All elements will have the same file name but different extensions e.g. Building_poly.
- Shapefiles can only contain a single shape type!

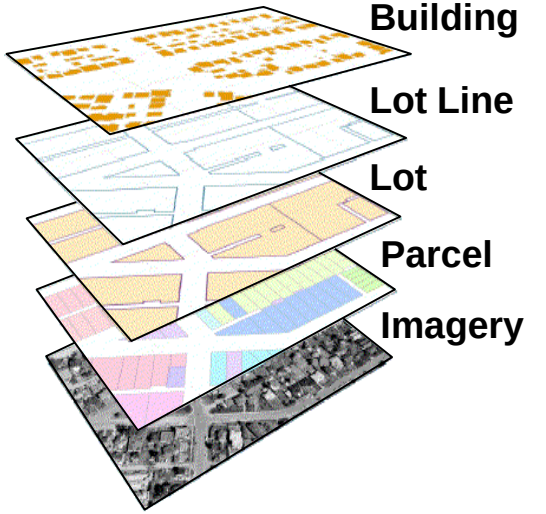
Modelling Reality with Geodatabase

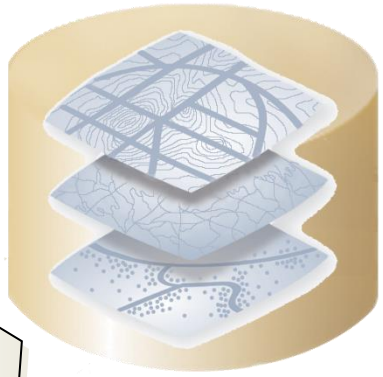
Geodatabase design

Reality

Design process

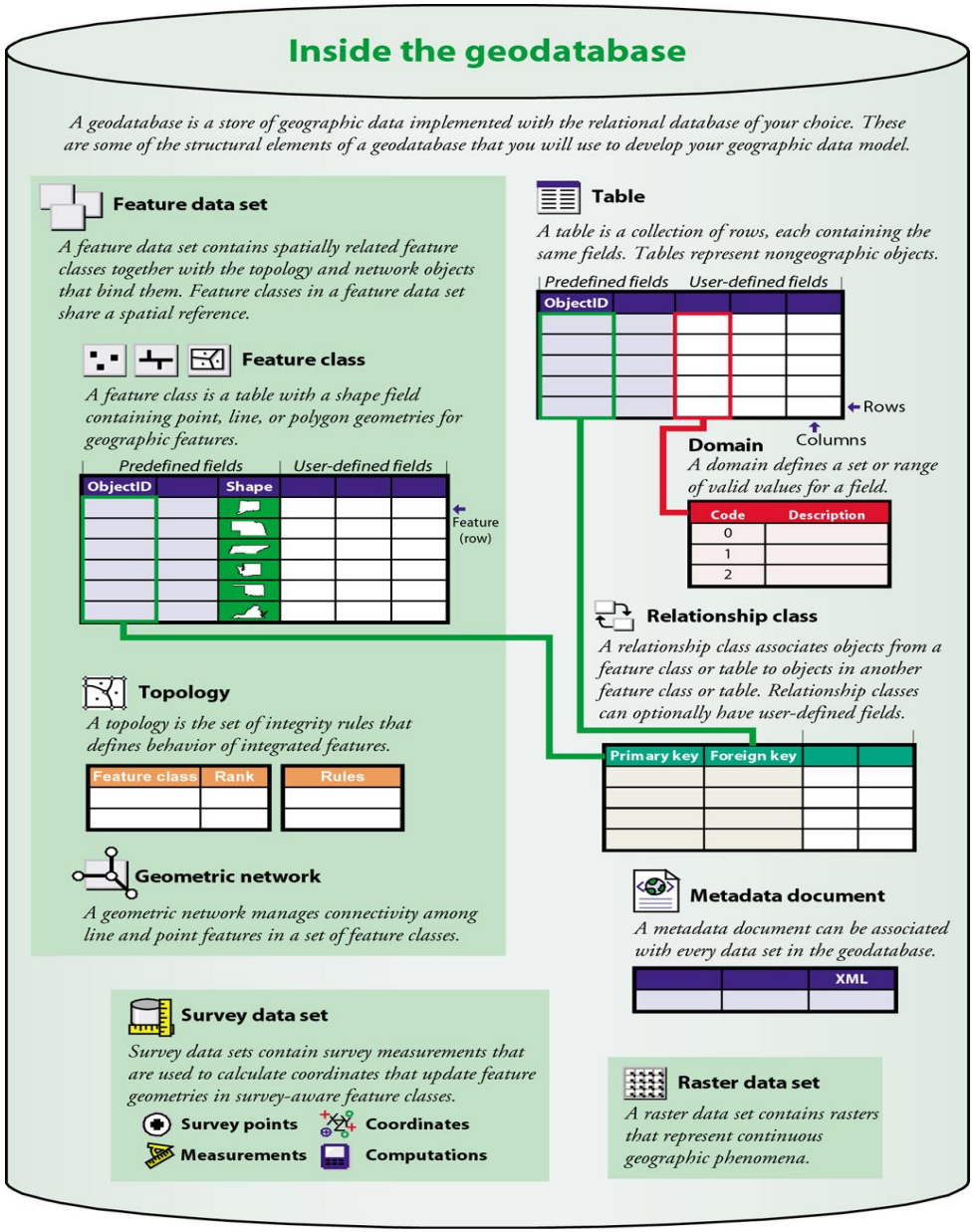
Data Model



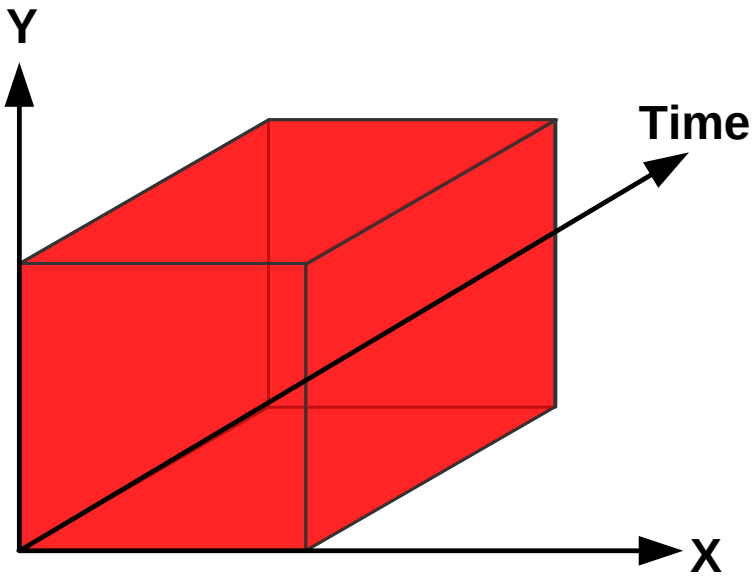


Geodatabase

- A Geodatabase is a store of geographic data implemented with a relational DBMS



Multidimensional Data



Data Cube

141	241	341	441
131	231	331	431
121	221	321	421
111	211	311	411

Time = 1

142	242	342	442
132	232	332	432
122	222	322	422
112	212	312	412

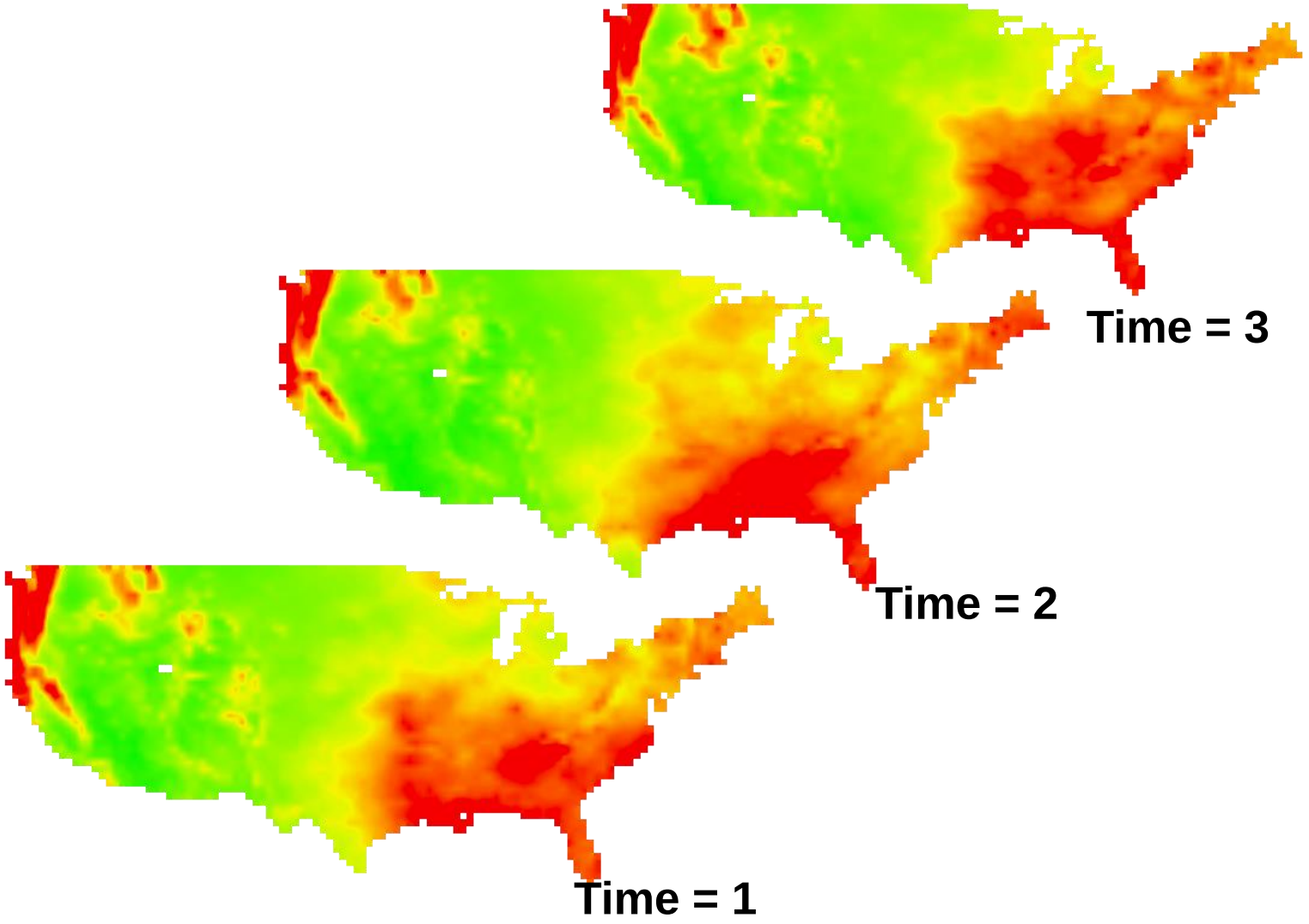
Time = 2

Time Slices

143	243	343	443
133	233	333	433
123	223	323	423
113	213	313	413

Time = 3

Multidimensional Data



Multidimensional Data Formats

- NetCDF (network Common Data Form)
- HDF (4.x and previous releases), HDF-EOS, HDF5 (Hierarchical Data Format)
- GRIB, GRIB II (GRIdded Binary)

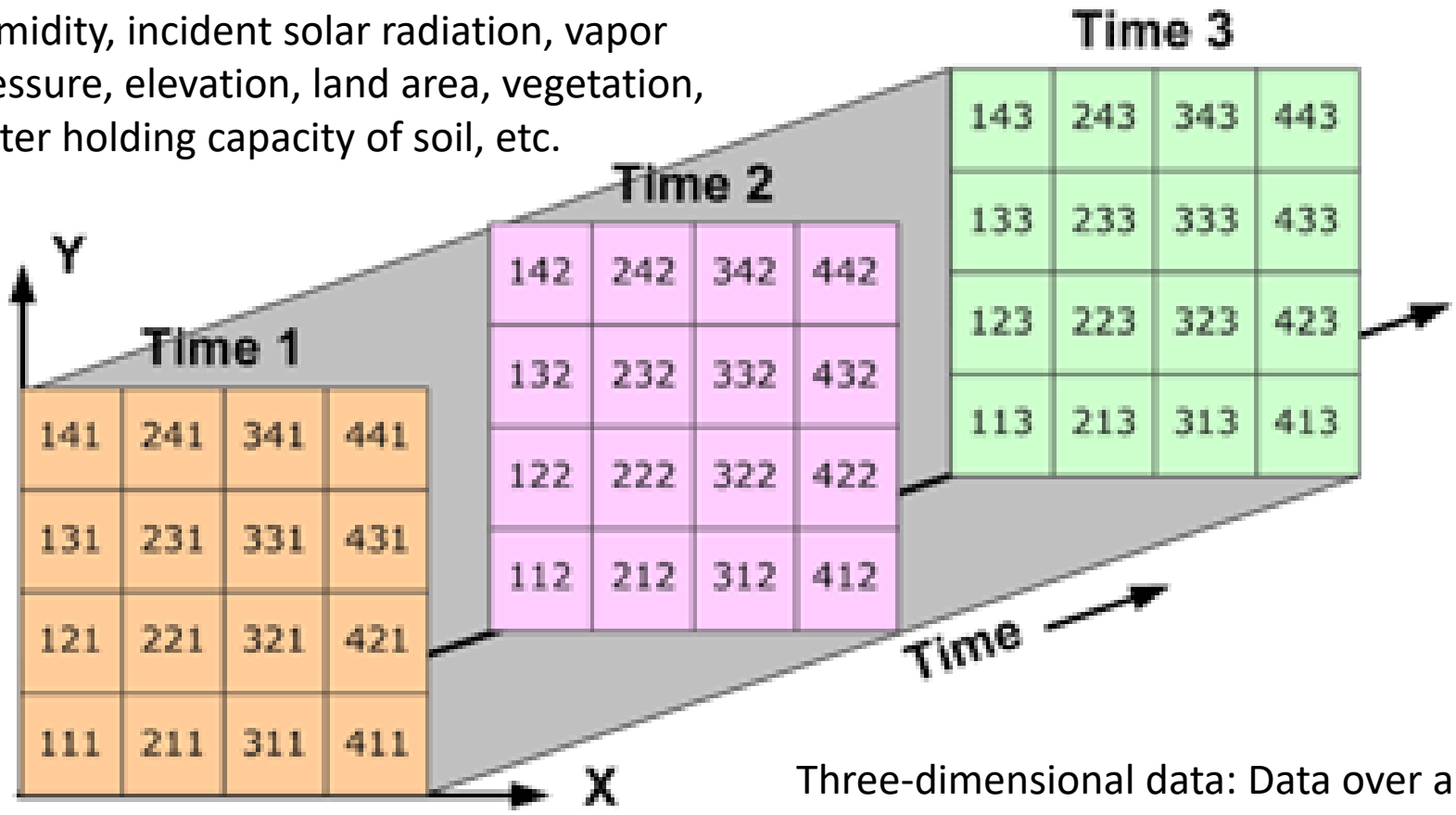
What is NetCDF?

- NetCDF (network Common Data Form) is a data format designed to support the creation, access, and sharing of array-oriented scientific data.

- It is used extensively in the atmospheric and oceanographic communities to store variables, such as temperature, pressure, wind speed, and wave height.

The storage of netCDF data

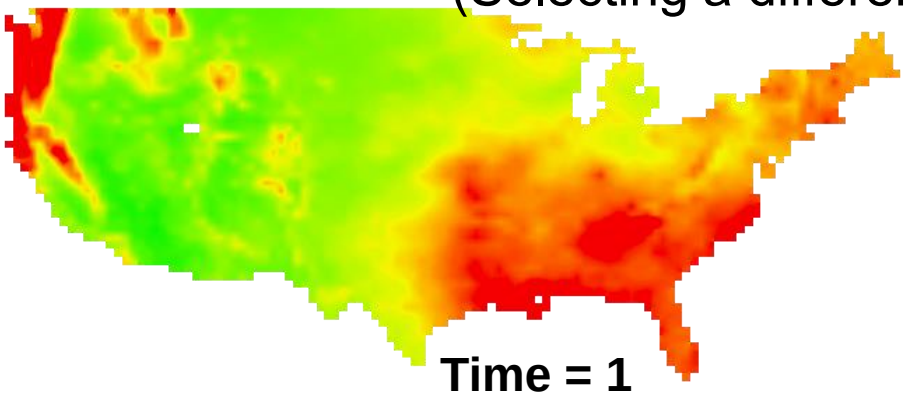
Data: Temperature, irradiance, precipitation, humidity, incident solar radiation, vapor pressure, elevation, land area, vegetation, water holding capacity of soil, etc.

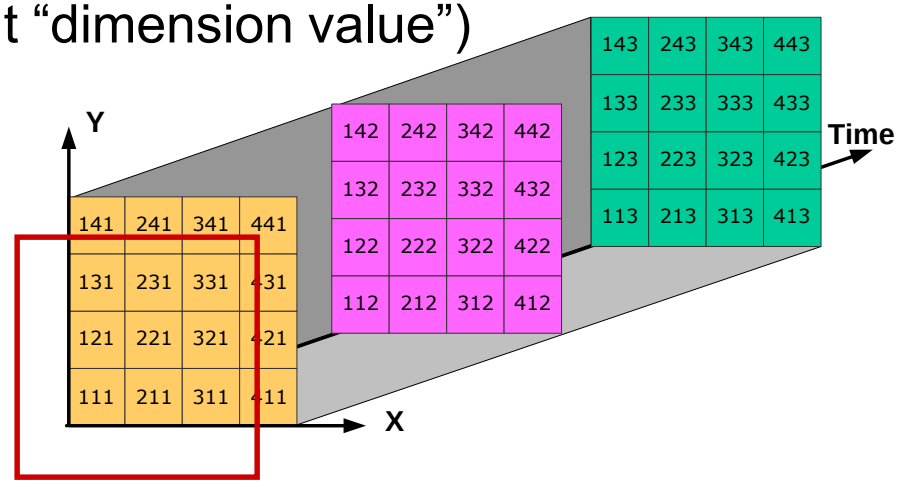


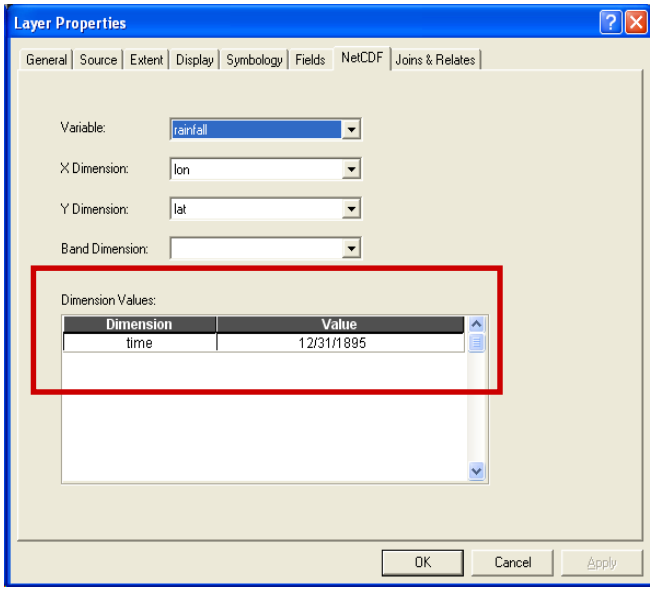
Three-dimensional data: Data over an area varying with time

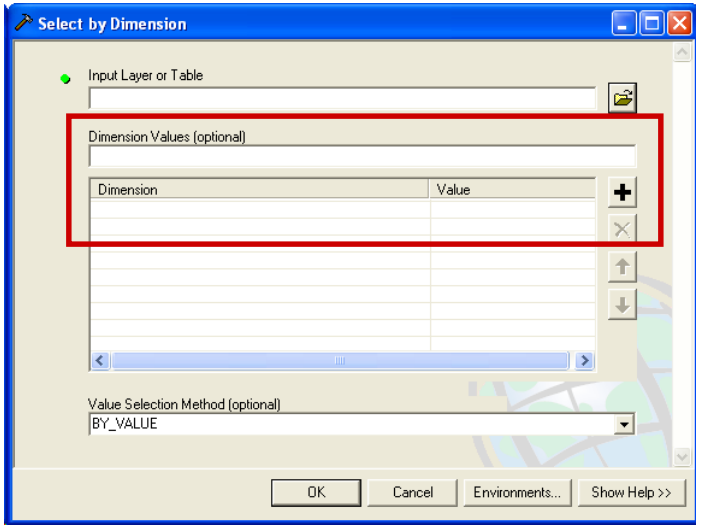
Changing Time Slice

(Selecting a different “dimension value”)

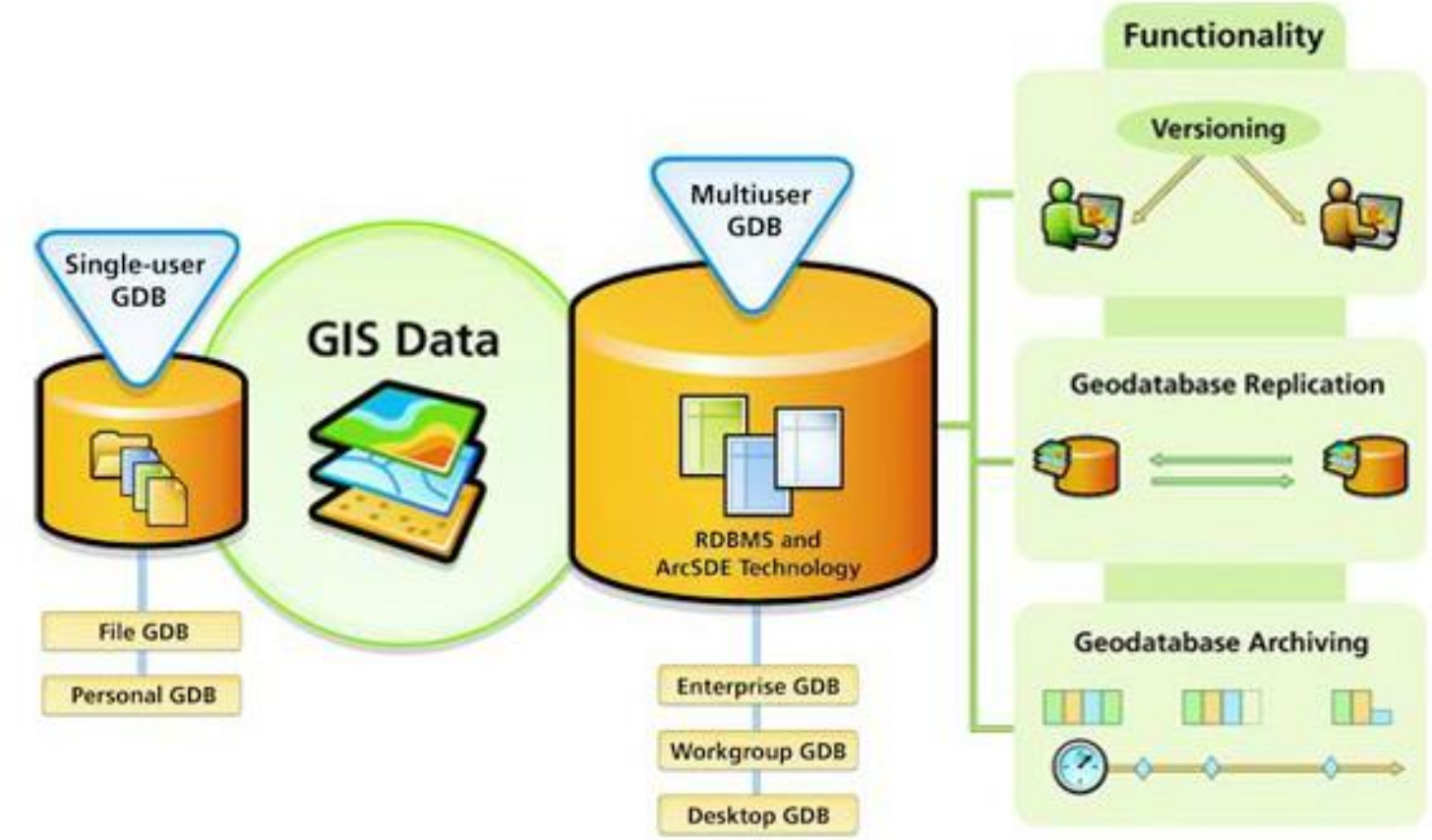




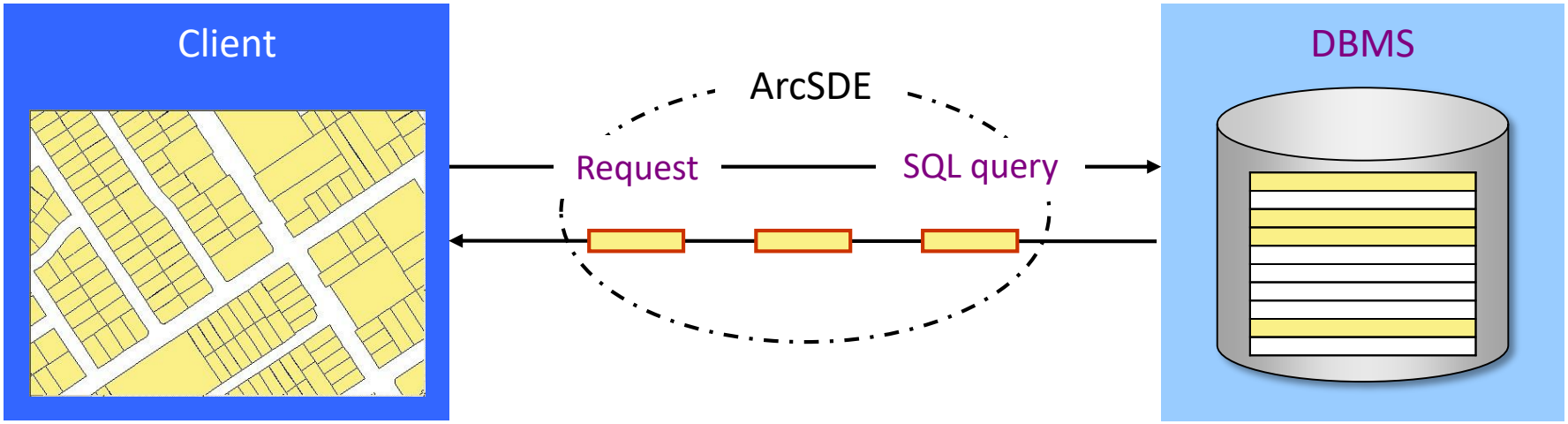




Single User vs. Multiuser Geodatabase



ArcSDE technology client-server model

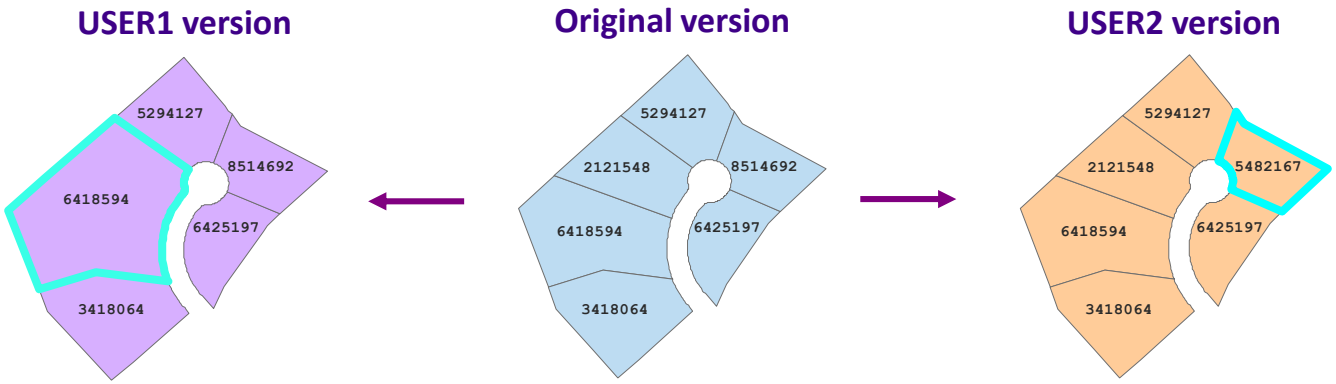


- All data accessed over TCP/IP network
- ArcSDE translates (acts as a gateway to the DBMS)
  - Spatial and attribute filters limit rows returned
- ArcSDE performs spatial filtering

What is a version?

- A version represents a snapshot in time of the entire geodatabase and contains all the datasets in the geodatabase.
- Versions are not separate copies of the geodatabase.
- Instead, versions and the transactions that take place within them are tracked in system tables.
- This isolates a user's work across multiple edit sessions, allowing users to edit without locking features in the production version or immediately impacting other users and without having to make copies of the data.
- Versions are only supported in ArcSDE geodatabases

Overview of versioned editing



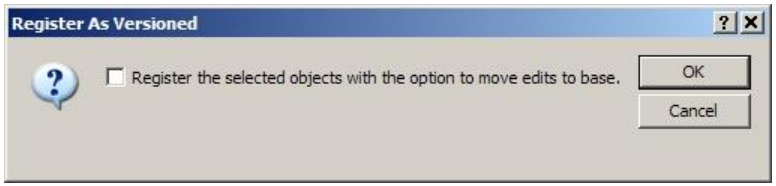
- Method for presenting and tracking changes to tables
- Multiple, alternate versions may co-exist
  - Appears to users as if they have their own copy of a table
- Includes mechanisms for merging changes
  - ArcGIS offers tools to resolve conflicts

Versioning

- Versioning allows multiple users in a multiuser geodatabase to edit the same data without applying feature locks or duplicating data.
- When you start editing, you are working with your own representation of the version.
- Other users who are connected to the same version cannot see any of your changes until you save the edits.
- When you are ready to apply your edits, you will merge your changes through a process of reconciling edits, resolving conflicts, and posting your changes to the parent version of the geodatabase.

Registering object as versioned

- Enables versioned edits
  - Feature classes, feature datasets, tables
- Must register entire feature dataset
  - Registers all feature classes
- Creates delta tables
  - Adds (A) and Deletes (D)
  - DBMS statistics created
  - DBTUNE controls storage



Move Edits to Base

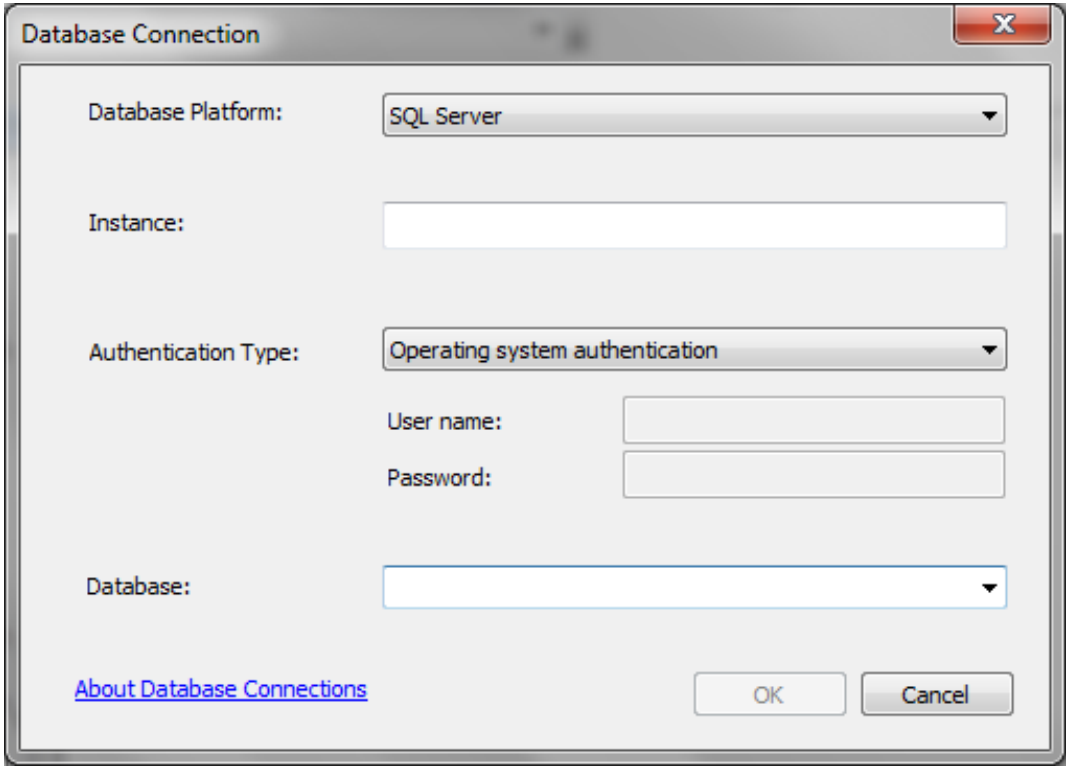
- This is an option available when registering data as versioned. It allows edits made to the DEFAULT version of the geodatabase to be immediately moved from the delta tables to the base tables.
- Specifying this option when you register the data as versioned can be useful if the modifications you are making will take only a few minutes to complete and if you are connecting to a versioned geodatabase with a third-party application.
- You cannot use the move edits to base option on datasets that contain a topology or network, are archived, or participate in replication.

Editing Versioned Data

The general workflow for editing a **versioned** ArcSDE geodatabase is as follows:

1. Establish a connection with the geodatabase.
2. Register the data as versioned.
3. Add the data to ArcMap.
4. Start editing and make your edits.
5. Review and reconcile any conflicts between the version being edited and the target version.
6. Post changes to the parent database.

Creating a database connection



Editing Non-Versioned Data

The general workflow for editing **non-versioned** data is as follows:

1. Make sure the data is not registered as versioned.
2. Use the Editing Options dialog box to configure ArcMap to perform nonversioned editing.
3. Add the data to ArcMap.
4. Start editing and make your edits.
5. Save edits and stop the edit session.

Delta Tables

- The adds and deletes tables for a dataset are collectively referred to as the delta tables because they store changes made to the dataset.
- Registering a feature class as versioned creates an adds and a deletes table. These tables track edits made to the dataset and allow you to edit a dataset without blocking other users from accessing or editing it.
- When you register a dataset as versioned, you can register it as fully versioned (the default option) or with the option to move edits to base.

How versioned edits are stored

ROADS

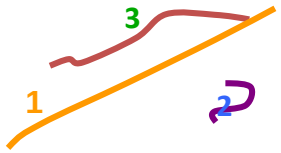
Shape	ObjectID	Name
	1	OAK
	2	ELM

A91

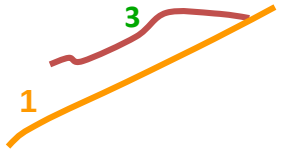
Shape	ObjectID	Name	StateID
	3	PLUM	1
	1	OAK	3

D91

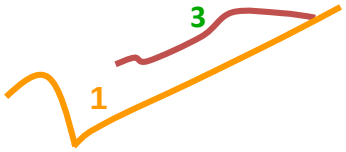
ObjectID	StateID
2	2
1	3



Adding



Deleting

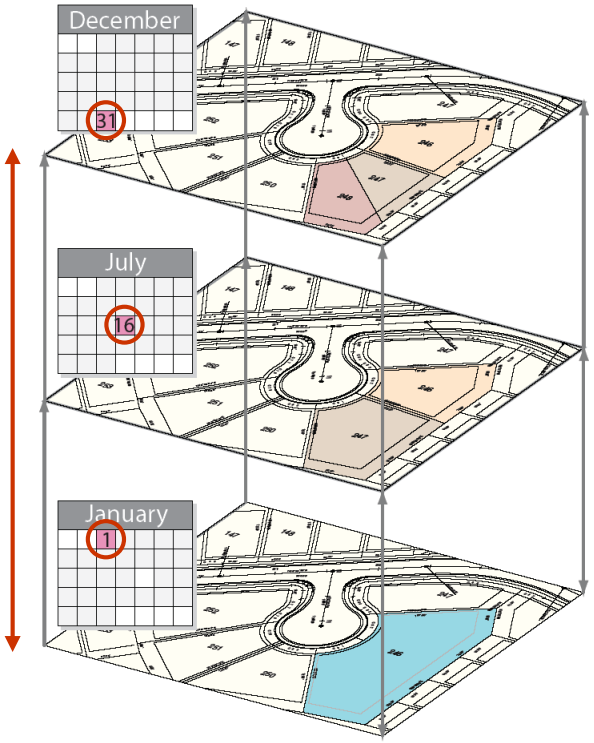


Updating

A

Overview of geodatabase archiving

- Maintain record of edit transactions
- Edits are preserved in a history class
  - Denoted with FROM and TO dates
    - Transaction time is recorded
- Built on versioning architecture

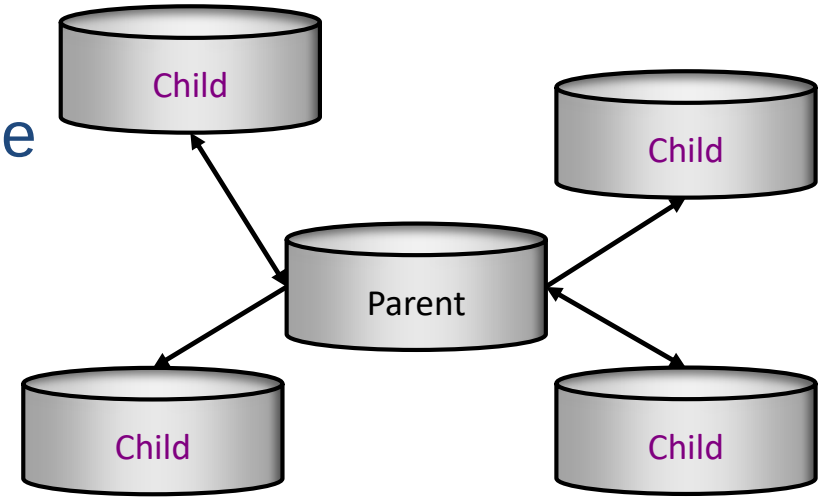


Distributed Data

- Copies of data distributed in multiple locations
- Can provide
  - Improved data availability with poor networks
  - Load balancing: Separate offices can work on same data
  - Field projects
  - Fail over
- Options
  - Copy/Paste: Hard to synchronize edits
  - **Geodatabase replication**
  - DBMS replication: Limited support for geodatabases

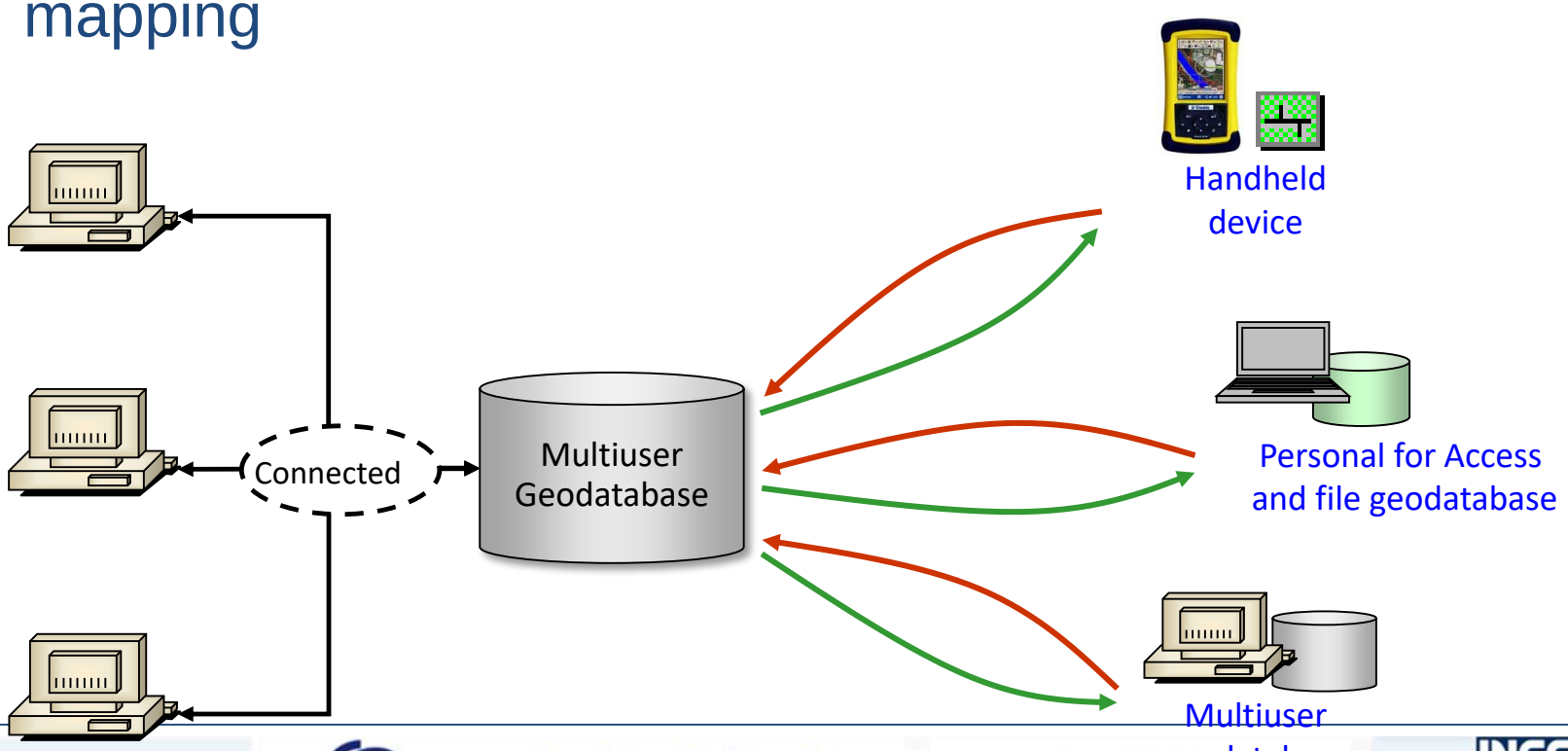
Fundamentals of geodatabase replication

- User defines data to replicate from source geodatabase
- Replica describes data and how to synchronize changes
  - Parent replica
  - Child replica
- Data edited in versioned environment
- Synchronize changes
  - Send to related geodatabase



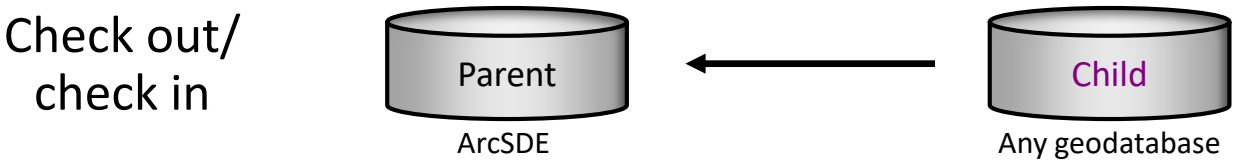
Replicating data

- Replicate to another geodatabase or handheld device
- Implement as part of workflow
  - Enables local/remote data access for editing, analysis, or mapping

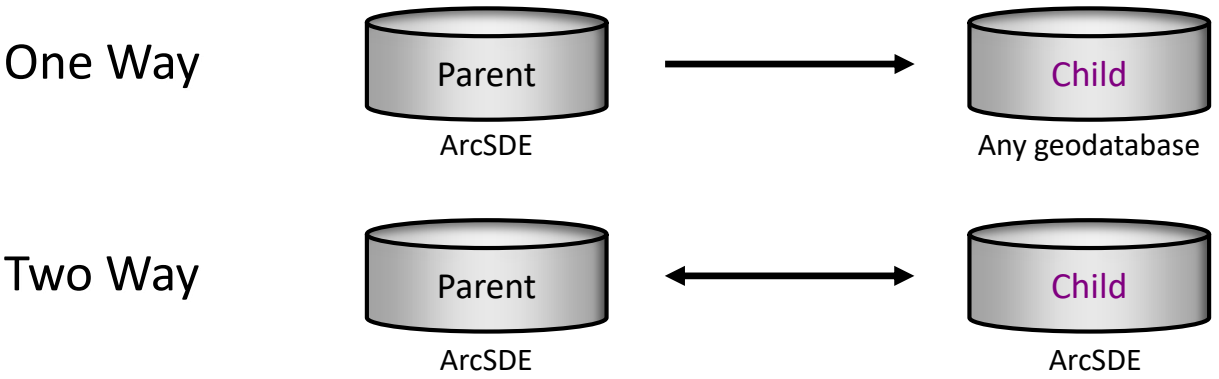


Types of replication

- Single generation
  - One check out/check in operation



- Multigeneration
  - Changes synchronized multiple times

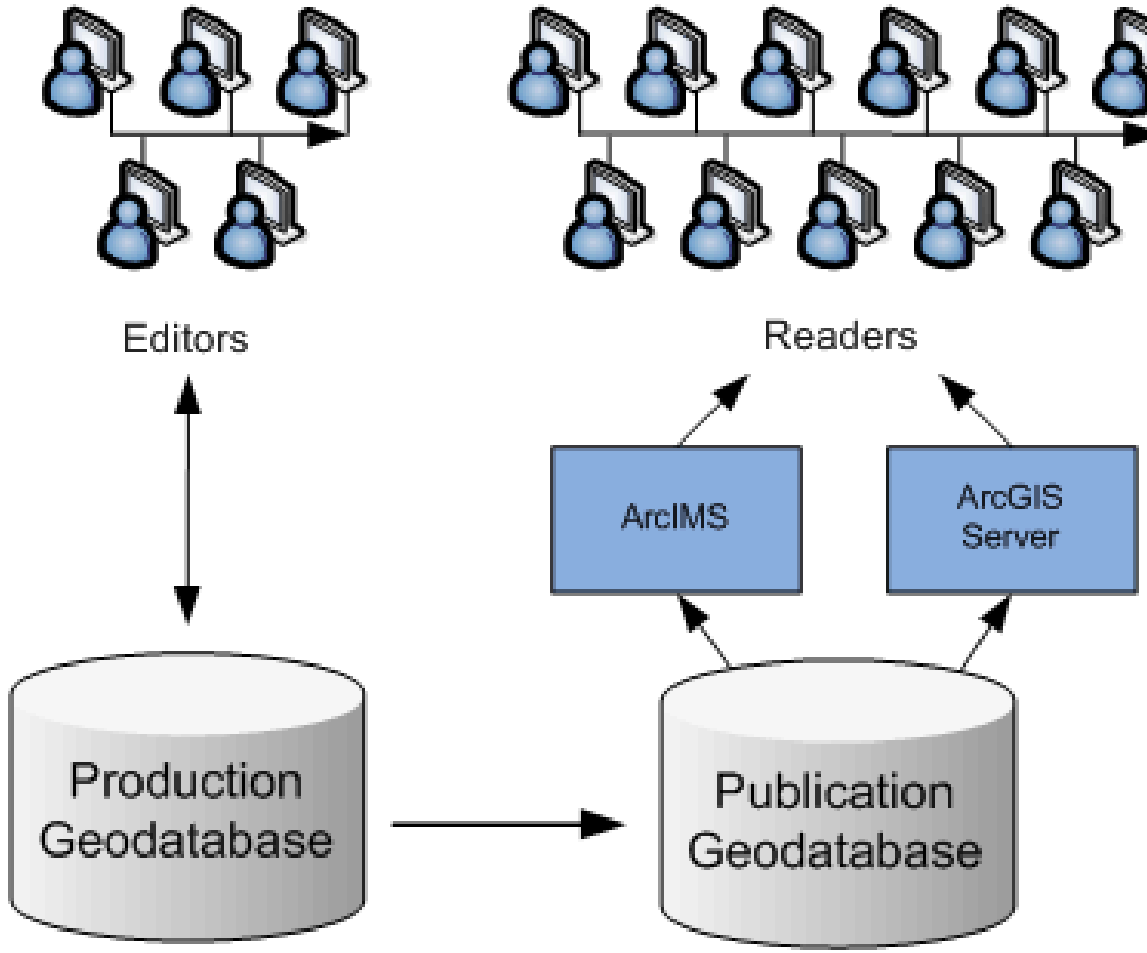


Synchronization

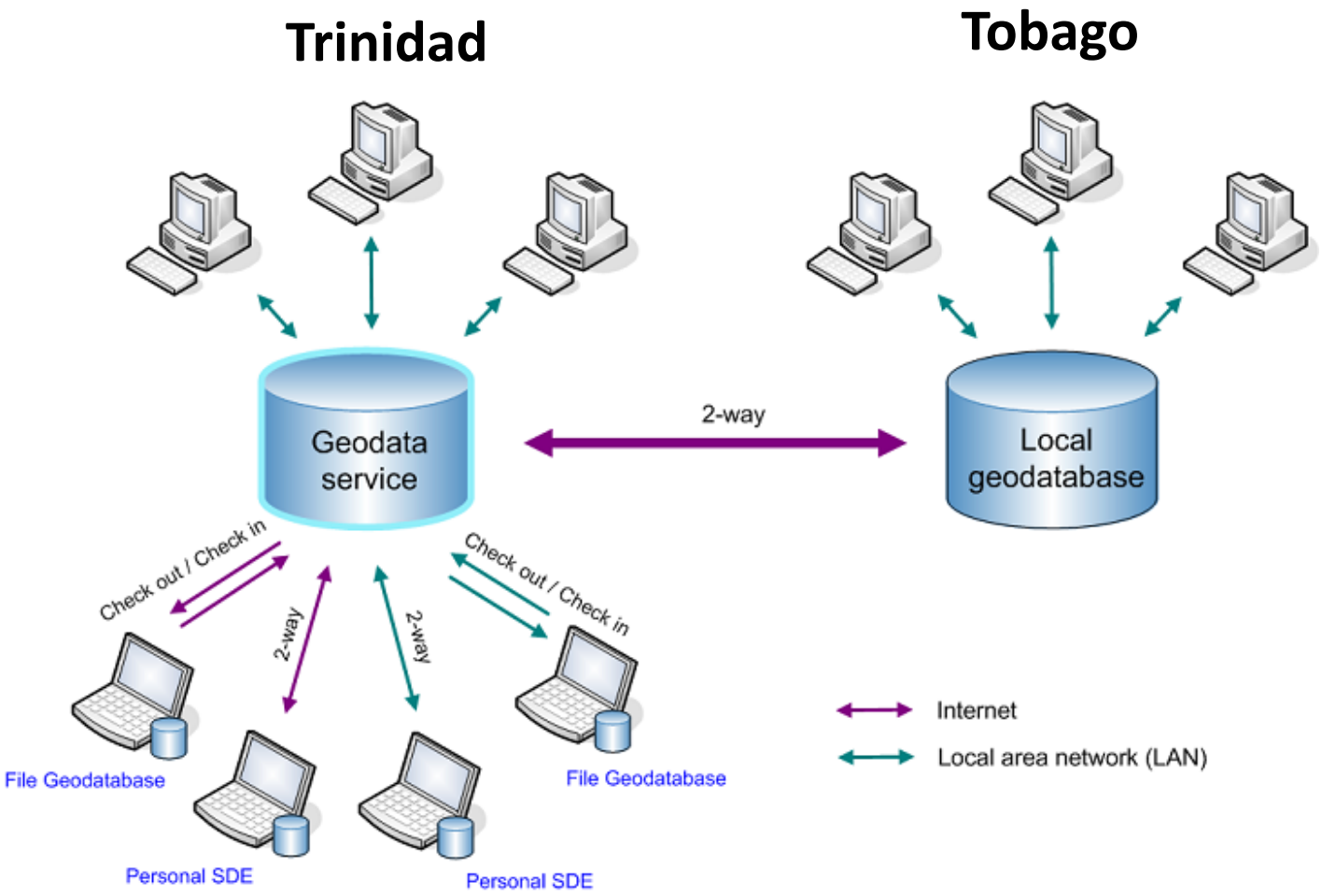
- Connected
  - All replicas accessible on the network
  - Performed in a single process
    - Example: Synchronize wizard in ArcCatalog
- Disconnected
  - Replicas not on the same network
  - Performed by export, file transfer, and import
    - Example: Export changes to a delta XML file and send via snail mail

Geodatabase Replication Use Cases

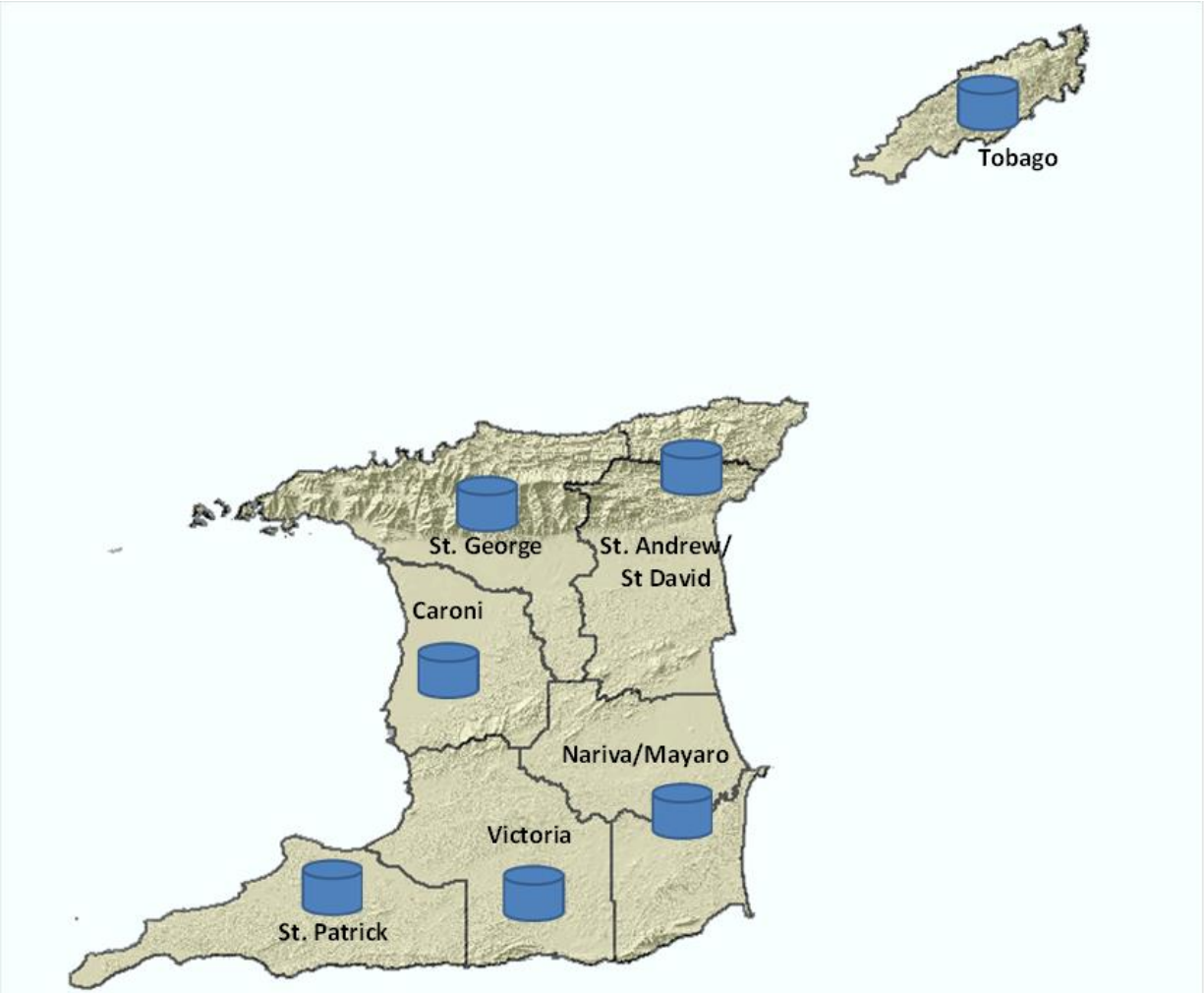
Production-Publication Replication



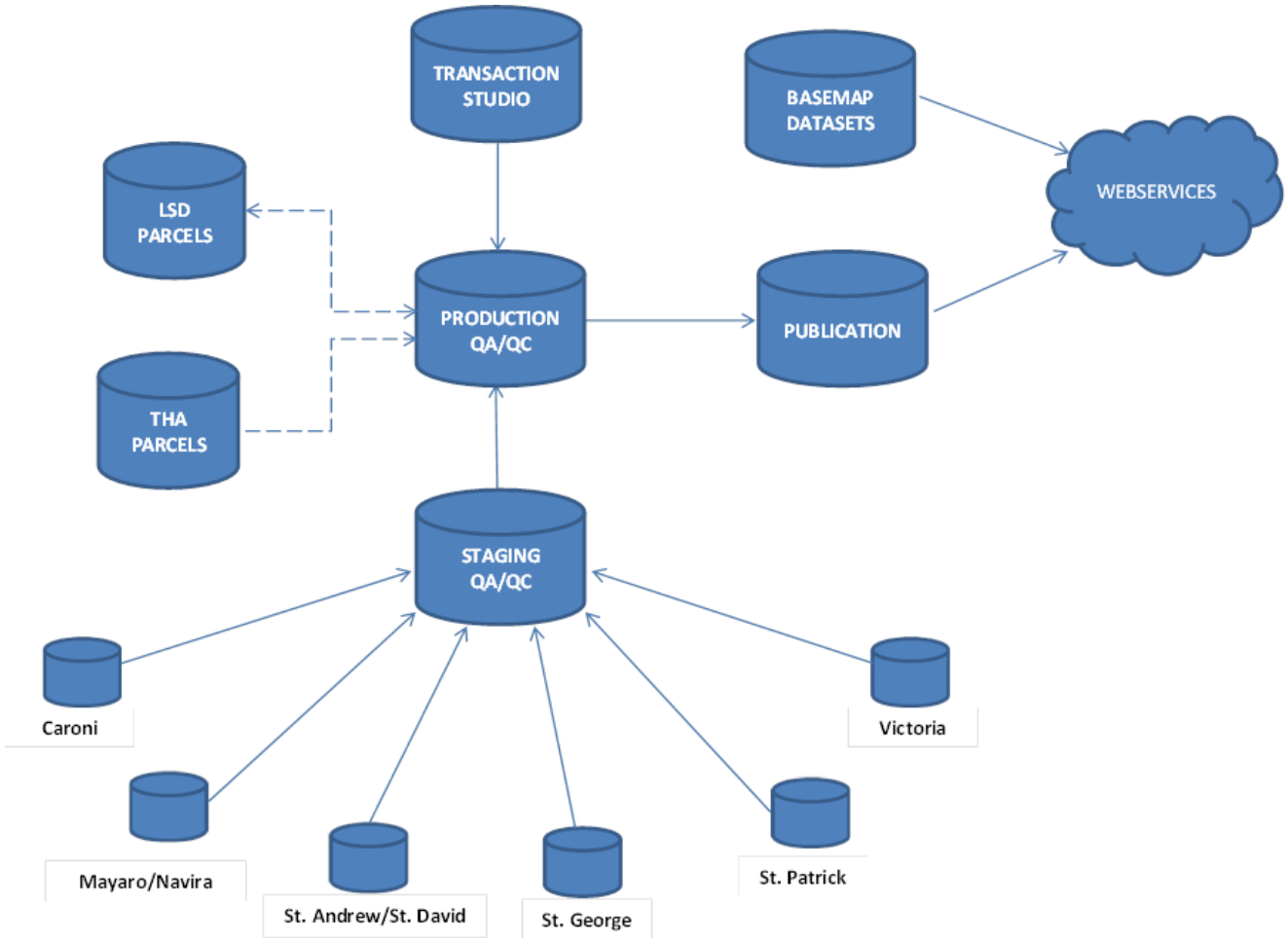
Distributing Data using Replication

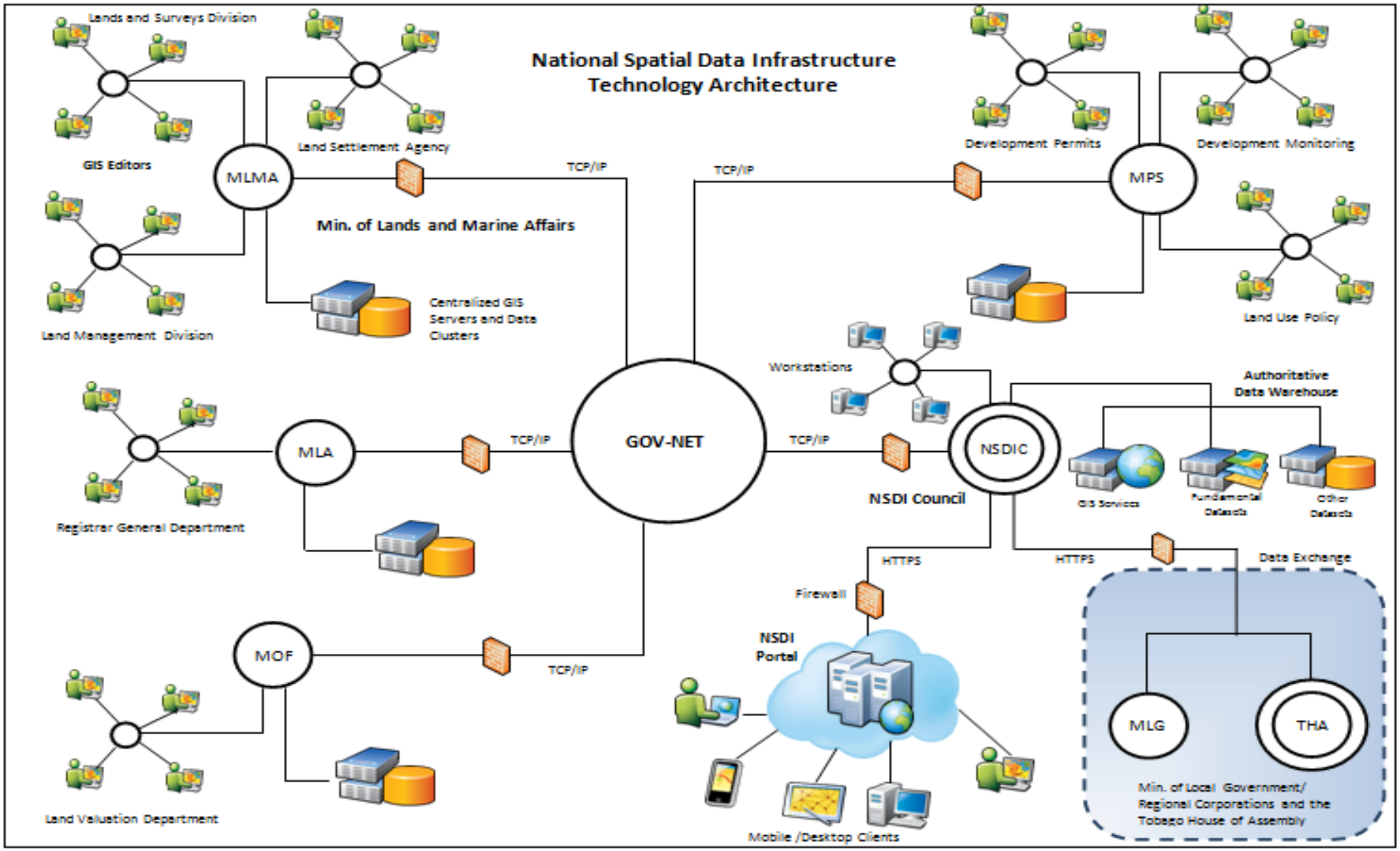


Geodatabase Distribution for Land Management



Geodatabase Replication Design




National Spatial Data Infrastructure
Technology Architecture
Lands and Surveys Division
GIS Editors
Land Settlement Agency
MLMA
Min. of Lands and Marine Affairs
Land Management Division
Centralized GIS Servers and Data Clusters
TCP/IP
TCP/IP
Development Permits
Development Monitoring
MPS
Land Use Policy
Workstations
Authoritative Data Warehouse
Registrar General Department
MLA
TCP/IP
GOV-NET
TCP/IP
NSDIC
NSDI Council
GIS Services
Fundamental Datasets
Other Datasets
HTTPS
HTTPS
Data Exchange
Firewall
NSDI Portal
MOF
TCP/IP
Land Valuation Department
Mobile /Desktop Clients
MLG
THA
Min. of Local Government/ Regional Corporations and the Tobago House of Assembly

UP NEXT …..

Activity: Managing data using Geodatabase