

UNIT 3

GIS Data Sources & Structures

Knowledge Management

Geographical Information System (CACS454) — BCA IV/VIII

Data is the single most valuable component of a GIS, and how that data is captured, sourced, structured, modeled and organized into a database determines the accuracy, cost and usability of every product the system produces. This unit examines the process of data capture, the four main sources of GIS data, the core data structures (vector, raster and TIN), and the principles of data modeling and database design that turn raw spatial data into well-managed geographic knowledge.

1. Capturing GIS Data

Data capture is the process of encoding existing spatial and attribute data — whether from paper maps, field observations or imagery — into a digital, computer-usable form that a GIS can store and analyze. It is typically the most time-consuming and expensive stage of any GIS project.

Common Data Capture Methods

Method	Description
Manual Digitizing	Tracing features from a paper map fixed on a digitizing tablet using a hand-held cursor to record coordinates.
Heads-Up (On-Screen) Digitizing	Tracing features directly on a scanned or georeferenced image displayed on a computer screen.
Scanning	Converting a paper map into a raster image using an optical scanner, often followed by vectorization.
Keyboard / Coordinate Entry	Typing in known coordinate values directly, e.g. from survey field books or COGO records.
Direct Field Capture	Recording feature locations directly in the field using GPS receivers or total stations.
Automated Conversion	Importing already-digital data such as satellite imagery or CAD drawings, followed by format conversion.

Typical Data Capture Workflow

- Pre-processing — cleaning and preparing the source map or image, including scale checks and quality assessment.

- Geometric (Spatial) Correction — registering or georeferencing the source to real-world coordinates using known control points.
- Feature Capture — digitizing or importing points, lines and polygons representing real-world features.
- Editing & Topology Building — correcting errors such as gaps, overshoots and overlaps, and establishing spatial relationships.
- Attribute Linking — attaching descriptive (non-spatial) data to each captured feature.

Note: Errors introduced during data capture — such as digitizing slivers or mismatched registration points — propagate through every later stage of analysis, so quality control at this stage is critical.

2. Sources: Maps, GPS, Images and Databases

GIS data is assembled from four principal source categories, each with distinct strengths, costs and levels of accuracy.

Source	How It Is Used
Maps	Existing topographic, cadastral or thematic maps are scanned and digitized (or heads-up digitized) to extract vector features and attribute categories already compiled by earlier surveys.
GPS / GNSS	Global Positioning System receivers record precise field coordinates for points, tracks (lines) and area boundaries in real time, ideal for ground-truthing and rapid field mapping.
Images	Satellite imagery, aerial photographs and orthophotos provide continuous, up-to-date raster coverage used directly for analysis or as a backdrop for digitizing new features.
Databases	Existing tabular or relational databases — such as census records, land-ownership registers or utility billing systems — supply attribute data that is joined to spatial features through a common key (e.g. parcel ID).

Example: Combining Multiple Sources

A municipality builds a drinking-water GIS by digitizing pipe networks from old engineering maps, capturing new valve locations with handheld GPS units, using a recent satellite image to verify road alignments, and joining household connection records from the billing database to each building's parcel ID — merging four different sources into one consistent spatial database.

3. Structures: Vector, Raster and TIN Data Structures

Once captured, spatial data must be organized using a data structure that a computer can store, query and analyze efficiently. GIS software relies on three principal structures.

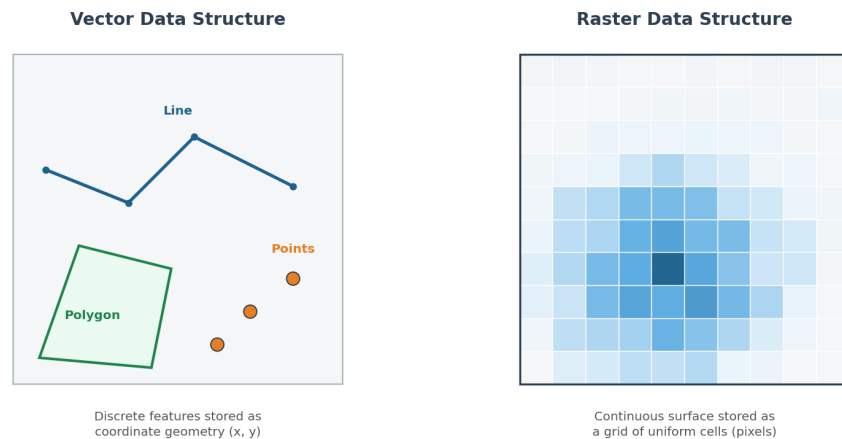


Figure 3.1 — Vector data represents discrete features; raster data represents continuous surfaces as a grid of cells.

Vector Data Structure

Vector data represents discrete real-world features as points (e.g. a well), lines (e.g. a road) and polygons (e.g. a land parcel), each defined by precise x, y coordinate geometry. Vector structures preserve exact shape and boundary and can encode topology — explicit spatial relationships such as adjacency and connectivity between features.

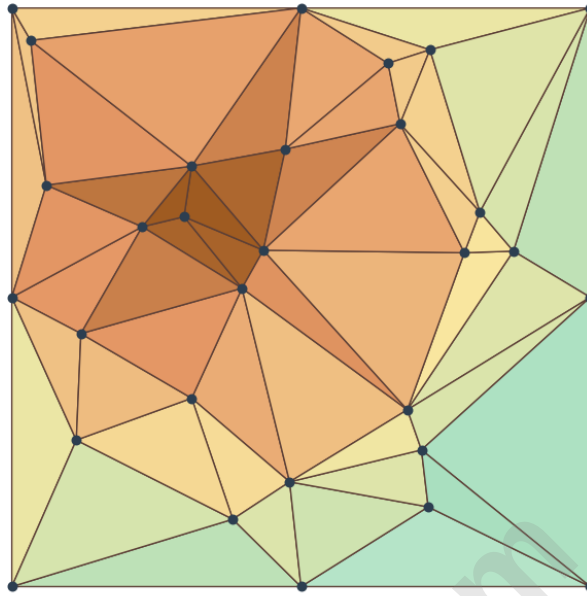
Raster Data Structure

Raster data divides space into a regular grid of equally sized cells (pixels), with each cell holding a single value representing the feature or condition present at that location — such as elevation, temperature or land-cover class. Raster structures represent continuously varying phenomena naturally and are the native format for satellite and aerial imagery.

TIN (Triangulated Irregular Network)

A TIN represents a continuous surface — most commonly terrain elevation — as a network of non-overlapping triangles formed by connecting irregularly spaced sample points using Delaunay triangulation. Because point density can vary, a TIN can capture fine detail in areas of rapid change (e.g. cliffs) while using fewer points in flat, uniform areas — making it more efficient than a uniform raster grid for terrain modeling.

TIN — Triangulated Irregular Network



Irregularly spaced elevation points connected into non-overlapping triangles

Figure 3.2 — A TIN surface built from irregularly spaced elevation points.

Comparing the Three Structures

Property	Vector	Raster	TIN
Basic Unit	Point, line, polygon	Grid cell (pixel)	Triangle facet
Best For	Discrete features (roads, parcels)	Continuous surfaces (rainfall, imagery)	Terrain / elevation surfaces
Storage Efficiency	Compact for sparse features	Can be large at fine resolution	Adapts density to terrain complexity
Topology Support	Strong (native)	Implicit (cell adjacency)	Strong (triangle adjacency)

4. GIS Data Modeling

A data model is the conceptual framework that defines how real-world geographic phenomena are abstracted, organized and represented within a GIS. Moving from the real world to a working database typically passes through four levels of abstraction.

- Reality — the actual, infinitely complex geographic phenomena on the Earth's surface.
- Conceptual Model — a simplified, human-oriented view identifying the entities of interest (e.g. “roads”, “parcels”) and their relationships, independent of any software.

- **Logical Model** — organizes the conceptual entities into a specific structure (e.g. vector or raster, relational tables) without regard to a particular software product.
- **Physical Model** — the actual implementation in a specific GIS software or database management system, including file formats and storage structures.

Common Data Model Types

Model Type	Approach
Object-Based (Entity) Model	Represents the world as discrete, identifiable objects (points, lines, polygons) each with its own attributes — matches the vector view of the world.
Field-Based Model	Represents the world as one or more continuous surfaces (fields) where a value exists at every location — matches the raster view of the world.
Network Model	Represents interconnected linear features (roads, pipelines, streams) with explicit connectivity for tracing and routing analysis.

5. GIS Database Design

GIS database design is the structured process of translating a data model into an efficient, reliable and maintainable spatial database, typically managed by a Spatial Database Management System (SDBMS) such as PostGIS, Esri ArcSDE or Oracle Spatial.

Stages of Database Design

- **Requirements Analysis** — identifying what questions the database must answer and which features and attributes are needed.
- **Conceptual Design** — defining entities, attributes and relationships, often using an Entity-Relationship (ER) diagram.
- **Logical Design** — translating the conceptual design into a schema of feature classes, tables and relationships (e.g. normalized relational tables).
- **Physical Design** — implementing the schema in the chosen SDBMS, including file structure, indexing and storage allocation.
- **Testing & Maintenance** — validating data integrity, performance and topology rules, with ongoing updates as data changes.

Key Design Considerations

- **Normalization** — organizing attribute tables to minimize redundancy and maintain consistency.

- Spatial Indexing — using structures such as R-trees to speed up spatial queries over large datasets.
- Topology Rules — enforcing constraints such as “parcels must not overlap” or “roads must connect at intersections”.
- Metadata — documenting data source, accuracy, projection and update date so the database remains usable and trustworthy over time.

Key Takeaways

- ✓ Data capture converts paper maps, field measurements and imagery into digital form through digitizing, scanning, GPS survey or automated import.
- ✓ GIS data is sourced from four categories — maps, GPS, images and existing databases — each contributing different strengths.
- ✓ Vector structures suit discrete features, raster structures suit continuous surfaces, and TIN efficiently models terrain using irregularly spaced points.
- ✓ Data modeling moves through four levels of abstraction — reality, conceptual, logical and physical — and follows an object-based, field-based or network approach.
- ✓ GIS database design proceeds through requirements analysis, conceptual, logical and physical design, guided by normalization, spatial indexing, topology rules and metadata.

— *End of Unit 3* —