Spatial Data

1. 1. Vector and rastar data PREPARED BY:- 13BCL041 13BCL043 13BCL044 13BCL045 13BCL046 13BCL047
2. [2.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-2-638.jpg?cb=1431408009)The data or information that identifies the geographic location of features and boundries. On earth, such as natural and construted features like Ocean, lake, pond etc. Spatial data is usually stored as coordinate and topology, and is data that can be mapped. s
3. [3.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-3-638.jpg?cb=1431408009)IN GIS THEARE ARE TWO BASIC SPATIAL DATA TYPES RASTER DATA VECTOR DATA IN GIS THERE ARE TWO TYPES OF SPATIAL DATA TYPES
4. [4.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-4-638.jpg?cb=1431408009)Vector Data Vector data provide a way to represent real world features within the GIS environment. A vector feature has its shape represented using geometry. The geometry is made up of one or more interconnected vertices. A vertex describe a position in space using an x, y and optionally z axis. In the vector data model, features on the earth are represented as: • points • lines / routes • polygons / regions • TINs (triangulated irregular networks)
5. [5.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-5-638.jpg?cb=1431408009)Vector Data This system of recording features is based on the interaction between arcs and nodes, represented by points, lines and polygons. A point is a single node, a line is two nodes with an arc between them, and a polygon is a closed group of three or more arcs. With these three elements , it is possible to record most all necessary information. Points Lines Polygons
6. [6.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-6-638.jpg?cb=1431408009)Vector Data Advantages • accurately representing true shape and size • representing non-continuous data (e.g., rivers, political boundaries, road lines) • Vectors can store information About topology • A vector data model uses points stored by their real (earth) coordinates and so requires a • precise coordinate system. • Geographic Coordinate System Latitude/Longitude • Cartesian Coordinate Systems X,Y Coordinate system
7. [7.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-7-638.jpg?cb=1431408009)Vector Data Disadvantages: • The location of each vertex needs to be stored explicitly. • Vector data must be converted into a topological structure. • This is often processing intensive and usually requires extensive data cleaning. • Updating or editing of the vector data requires re-building of the topology.
8. [8.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-8-638.jpg?cb=1431408009)Vector Data
9. [9.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-9-638.jpg?cb=1431408009)Vector Data
10. [10.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-10-638.jpg?cb=1431408009)Raster Data Raster Data – cell –based data such as aerial imagery and digital elevation models. Raster data is characterized by pixel values. Basically, a raster file is a giant table, where each pixel is assigned a specific value from 0 to 255. The meaning behind these values is specified by the user – they can represent elevations, temperature, hydrology and etc.
11. [11.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-11-638.jpg?cb=1431408009)Copyright © 2006 by Maribeth H. Price 1-11 \*Portraying large areas at high precision is problematic 90m resolution 10m resolution \*Storage space increases by the square of the resolution
12. [12.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-12-638.jpg?cb=1431408009)Advantages: • Raster is the best way to store continuously changing values such as elevation, slope. • Analysis faster and more flexible then vector for many application. • Rapid computations ("map algebra") in which raster layers are treated as elements in mathematical expressions
13. [13.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-13-638.jpg?cb=1431408009)Copyright © 2006 by Maribeth H. Price \* Density Least cost path Distance Interpolation Viewshed Buffers
14. [14.](https://image.slidesharecdn.com/presentationspatialdata1-150512051852-lva1-app6892/95/ppt-spatial-data-14-638.jpg?cb=1431408009)Raster Data Disadvantages: • It is especially difficult to adequately represent linear features depending on the cell resolution. • Network linkages are difficult to establish. • Processing of associated attribute data may be cumbersome if large amounts of data exists. • Raster maps inherently reflect only one attribute or characteristic for an area. • Most output maps from grid-cell systems do not conform to high-quality cartographic needs.

GIS Data Types

1. 1. GIS Data Types GIS Topics and Applications John Reiser Rowan University
2. [2.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-2-728.jpg?cb=1295965741)Data Types <ul><li>Two Major Families of GIS Data </li></ul><ul><ul><li>Raster </li></ul></ul><ul><ul><li>Vector </li></ul></ul><ul><li>Raster is grid based </li></ul><ul><li>Vector is: </li></ul><ul><ul><li>coordinate based (cartesian, polar, 3D, linear) </li></ul></ul><ul><ul><li>topological </li></ul></ul><ul><ul><li>object oriented </li></ul></ul>
3. [3.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-3-728.jpg?cb=1295965741)Coding Raster GIS Data 1 1 1 1 2 3 4 4 1 1 1 2 2 3 4 4 1 2 2 2 3 3 4 4 2 2 2 3 3 4 4 4 3 3 3 3 5 5 5 5 1 1 1 1 6 5 5 5 1 1 1 1 1 5 5 5 1 1 1 1 1 1 5 5 Reality Raster Mode Model of Reality
4. [4.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-4-728.jpg?cb=1295965741)Raster Limitations <ul><li>Raster grid must cover the entire study area. </li></ul><ul><li>Files can grow to enormous sizes for large study areas with small cell sizes. </li></ul><ul><li>Attributes are limited and linking to tabular data is impractical. </li></ul><ul><li>Adjacency is easy to determine, but topology is lacking. </li></ul><ul><li>Raster grid cells are not “aware” and cannot have actions attached to them in the geodatabase. </li></ul>
5. [5.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-5-728.jpg?cb=1295965741)Vector Data <ul><li>We represent points, lines and polygons with vector data. </li></ul><ul><li>Moving away from shapefiles, we can store multipoints, annotation, dimensions, coordinate geometry, cadastral fabrics, networks and multipatches (3D objects) in the geodatabase as vector data. </li></ul>
6. [6.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-6-728.jpg?cb=1295965741)Points Points in the World Out There Vector Encoding Resulting Image 1 4 3 2 Point X Y 1 X 1 Y 1 3 X 3 Y 3 2 X 2 Y 2 4 X 4 Y 4 X Y
7. [7.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-7-728.jpg?cb=1295965741)Lines Lines in the World Out There Vector Encoding Resulting Image Line X Y 1 X 11 Y 11 X 12 Y 12 . . . . . . X 1n Y 1n 2 X 21 Y 21 X 2n Y 2n 1 2 3 3 X 31 Y 31 X 32 Y 32 . . . . . . X 3n Y 3n 4 4 X 41 Y 41 X 4n Y 4n X Y Vertex Node
8. [8.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-8-728.jpg?cb=1295965741)Polygons and Multiparts <ul><li>Polygons must close upon themselves, so that the first and last vertex are the same. </li></ul><ul><li>Polygons may include internal rings that act as “donut holes” – areas excluded from the surrounding polygon. </li></ul><ul><li>Polygons (and points and lines) may also include multiple parts. </li></ul>
9. [9.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-9-728.jpg?cb=1295965741)Attributes <ul><li>Each vector feature has an accompanying record in the database. </li></ul><ul><li>In its simplest form, one feature has one record in the attribute table. </li></ul><ul><li>We can store multiple vector objects in one database record. </li></ul><ul><li>Multipoints, polylines and multipolygons store a series of vector objects as one feature and one attribute table record. </li></ul>
10. [10.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-10-728.jpg?cb=1295965741)Multipart Features <ul><li>When we store multiple geometries as one record, we refer to them as multipart features . </li></ul><ul><li>ArcGIS will – by default – create multipart features as outputs from ArcToolbox. </li></ul><ul><li>Let ’s look at some real world examples… </li></ul>
11. [11.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-11-728.jpg?cb=1295965741)Enclaves and Exclaves <ul><li>Long Beach is a Township in Ocean County. It ’s in four separate pieces on LBI. </li></ul><ul><li>Storing it as one record makes sense – it ’s one township, not four. </li></ul><ul><li>Nearby Tuckerton is an enclave within Little Egg Harbor Township. </li></ul>
12. [12.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-12-728.jpg?cb=1295965741)LIDAR Data <ul><li>Lidar data is 3D elevation data recorded from an airplane. </li></ul><ul><li>Stored as “mass points” – even a small area is composed of thousands of point features. </li></ul><ul><li>To lower overhead, the points are stored as multipoints – roughly 3,500 points per attribute table record. </li></ul><ul><li>No real need for attributes, simply XYZ points. </li></ul>
13. [13.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-13-728.jpg?cb=1295965741)Why do we do this? <ul><li>Depends on the structure of our data model. </li></ul><ul><li>Possibly lower overhead; easier processing. </li></ul><ul><li>Normalization </li></ul><ul><ul><li>DB Normalization is a key component of database design. </li></ul></ul><ul><ul><li>Less redundancy </li></ul></ul><ul><ul><li>Update attributes in one place, instead of multiple </li></ul></ul><ul><ul><li>Keeps it simple, students. (Admittedly, it ’s usually more time to set up, but much lower cost to maintain.) </li></ul></ul>
14. [14.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-14-728.jpg?cb=1295965741)Best Practices <ul><li>There are many ways to represent reality digitally. </li></ul><ul><li>GIS and DB “Best Practices” are exactly that – the (or one of) best way to handle a situation, regardless of software and hardware. </li></ul><ul><li>Even though we ’ll primarily deal with ArcGIS, thinking about “the best way” is the only way that you can be sure you’re effectively managing your GIS system. </li></ul>
15. [15.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-15-728.jpg?cb=1295965741)Smart Objects <ul><li>Vector features can follow rules and have “business logic”. </li></ul><ul><li>Still composed of points, lines, and/or polygons, these objects are higher-level models of reality. </li></ul><ul><li>Topology-based Rulemaking </li></ul><ul><li>Cadastral Fabric </li></ul><ul><li>Networks </li></ul><ul><li>Terrain Models </li></ul>
16. [16.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-16-728.jpg?cb=1295965741)What is Topology? <ul><li>Shared Geometries, Adjacency and Overlap </li></ul><ul><li>Where points, lines, and polygons share individual vertices. Move a point and it moves a vertex in a line/polygon, and vice versa. </li></ul><ul><li>Two polygons that share vertices are considered adjacent. </li></ul><ul><li>Overlapping (or non-overlapping) features can be located, and then marked as errors. </li></ul>
17. [17.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-17-728.jpg?cb=1295965741)Simple Example <ul><li>The boundaries of two properties should never overlap, and there should never be a gap between them, unless intentional. </li></ul><ul><li>Clear error in parcel boundaries. </li></ul>
18. [18.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-18-728.jpg?cb=1295965741)Policy-based Topology Rules <ul><li>In the NJ State Plan, CESs and the Environmentally Sensitive Planning Area both represent areas of environmental importance. </li></ul><ul><ul><li>Thus, CESs should never be placed on top of the ES Planning Area. </li></ul></ul><ul><li>In our utility network, poles hold up the transmission lines. </li></ul><ul><ul><li>The transmission line features must always share a vertex with the utility pole point features. </li></ul></ul>
19. [19.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-19-728.jpg?cb=1295965741)Networks <ul><li>Analysis can be performed across a network, represented by a feature dataset of points and lines. </li></ul><ul><li>Road network or water, sewer, utility, rail, etc… </li></ul><ul><li>Optimal route – shortest, lowest cost, avoiding left turns, follow height and weight restrictions, time of day restrictions, include real-time traffic… </li></ul><ul><li>Multi-modal – walk/bike to bus stop, bus to train, walk from train to final destination. </li></ul>
20. [20.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-20-728.jpg?cb=1295965741)GIS is extensible <ul><li>With modern GIS, a polygon is not just a polygon. </li></ul><ul><li>Software can be adapted to fit your model of reality. </li></ul><ul><li>The software can be easily extended to create new data types and perform new analyses. </li></ul><ul><li>GIS can be adapted to store, model, and display data about any observable phenomenon on the Earth. </li></ul>
21. [21.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-21-728.jpg?cb=1295965741)Different GIS Models <ul><li>Almost every GIS defines features on the Earth using one of two methods: </li></ul><ul><ul><li>Point or series of points (Vector) </li></ul></ul><ul><ul><li>Grid cells or group of cells (Raster) </li></ul></ul><ul><li>Other than that, how those pieces fit together to represent reality is entirely dependent on the GIS model. </li></ul><ul><li>Let ’s look at OpenStreetMap. </li></ul>
22. [22.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-22-728.jpg?cb=1295965741)OpenStreetMap <ul><li>OSM was developed by British developers. </li></ul><ul><li>The Ordinance Survey (analogue to the USGS) completely restricts the use of their data through high licensing costs. </li></ul><ul><li>OSM was developed to be an open, license-free map of the World. </li></ul><ul><li>You are allowed to edit and update the map, provided your contributions are not encumbered by licensing issues. </li></ul>
23. [23.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-23-728.jpg?cb=1295965741)OSM Data Primitives <ul><li>Unlike what we ’ve used before, OSM has an entirely different way of storing GIS data. </li></ul><ul><li>With ArcGIS, we store a set of layers </li></ul><ul><ul><li>Polygon shapefile: lakes, parks, etc… </li></ul></ul><ul><ul><li>Line feature class: roads, rails, streams, etc… </li></ul></ul><ul><ul><li>Point shapefile: points of interest, water towers, etc… </li></ul></ul><ul><ul><li>Annotation: place name labels, and so on… </li></ul></ul><ul><li>OSM stores everything in one big database. </li></ul><ul><li>All features can share parts – topology. </li></ul>
24. [24.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-24-728.jpg?cb=1295965741)Data Primitives <ul><li>Nodes </li></ul><ul><ul><li>Most basic unit – building block of all others. </li></ul></ul><ul><ul><li>One latitude/longitude coordinate pair. </li></ul></ul><ul><ul><li>Can have its own attributes. </li></ul></ul><ul><li>Ways </li></ul><ul><ul><li>Series of nodes denoting a linear feature or an area. </li></ul></ul><ul><ul><li>Open way – linear feature – roadway, rail, etc. </li></ul></ul><ul><ul><li>Closed way – areal feature – lake, building, etc. </li></ul></ul><ul><li>Relations </li></ul><ul><ul><li>Grouping ways and nodes. </li></ul></ul>
25. [25.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-25-728.jpg?cb=1295965741)Attributes <ul><li>No attribute table with defined field names. </li></ul><ul><li>Attributes are free form – want to tag something where “rowan”=“awesome”? Nothing’s stopping you! </li></ul><ul><li>Unknown tags won ’t influence rendering of the map. </li></ul><ul><li>Huge list of well known, renderable tags. </li></ul>
26. [26.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-26-728.jpg?cb=1295965741)Extensibility <ul><li>OSMs open data structure and licensing has allowed for many adaptations of the data. </li></ul><ul><li>Rendering: </li></ul><ul><ul><li>Mapnik and Osmarender </li></ul></ul><ul><ul><li>CycleMap , Hike and Bike </li></ul></ul><ul><li>Web Services </li></ul><ul><ul><li>European WMS Service </li></ul></ul><ul><ul><li>Routing </li></ul></ul>
27. [27.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-27-728.jpg?cb=1295965741)Extensibility <ul><li>University of Maryland – College Park uses OSM for its campus map. </li></ul><ul><li>Map is routable, giving you walking directions between buildings. </li></ul><ul><li>Routing can be &quot;accessiblity-aware&quot; routing you around stairs, unstriped crosswalks, etc. </li></ul>
28. [28.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-28-728.jpg?cb=1295965741)Versioning <ul><li>Each data element is tracked and each previous version is maintained. </li></ul><ul><li>User accounts required to track edits. </li></ul><ul><li>Mistakes can be easily reverted to their previous version, provided no subsequent changes have been made. </li></ul><ul><li>Versioning & multi-user editing are available in ArcGIS using managed geodatabases. </li></ul>
29. [29.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-29-728.jpg?cb=1295965741)Real World: Haiti <ul><li>In response to the earthquake in Haiti, OSM users from around the world digitized donated satellite imagery taken after the event. </li></ul><ul><li>Users coordinated their efforts on the OSM wiki . </li></ul><ul><li>Relief workers on the ground and coordinators are using the OSM data to help recovery . </li></ul>
30. [30.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-30-728.jpg?cb=1295965741)Contributing to OSM <ul><li>OSM needs updates and refinements to keep the map current and detailed. </li></ul><ul><li>Offer several methods to update: </li></ul><ul><ul><li>Walking Papers (printed maps) </li></ul></ul><ul><ul><li>GPX (GPS eXchange format) uploads </li></ul></ul><ul><ul><li>Potlatch (in browser editing) </li></ul></ul><ul><ul><li>JOSM (standalone editor) </li></ul></ul>
31. [31.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-31-728.jpg?cb=1295965741)JOSM <ul><li>The Java OSM editor is perhaps the most feature-rich editor available for OSM. </li></ul><ul><li>Plugins allow for the verification of topology and tags and the use of WMS services. </li></ul><ul><li>Just like OSM, it is free, open source software. </li></ul>
32. [32.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-32-728.jpg?cb=1295965741)Contributing to OSM <ul><li>Create an account and send me your username. </li></ul><ul><li>Locate an area lacking detail. It can be around your hometown or some place you feel comfortable. </li></ul><ul><li>Research the area. Use Walking Papers, GPS, and the NJ aerial photography to help you update. </li></ul><ul><li>Be creative with what can be mapped: bike racks and pedestrian paths within parks (and the parks themselves). </li></ul>
33. [33.](https://image.slidesharecdn.com/gislecture02-110125142426-phpapp01/95/gis-data-types-33-728.jpg?cb=1295965741)A Note: NJ Land Use <ul><li>I added the Land Use data for NJ from the data available on the DEP's website. </li></ul><ul><li>Despite heavy messaging, it can still be difficult to edit in OSM, so be careful. </li></ul><ul><li>It is from 2002, so if you find ways tagged with &quot;landuse&quot;=&quot;construction&quot; explore them to see what was added in the 2002-2007 time period. </li></ul><ul><li>Don't hesitate to ask for help! </li></ul>

GIS Data Structure

1. 1. GIS Data Structures 1
2. [2.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-2-638.jpg?cb=1446831726)Representing Geographic Features: เราอธิบาย geographical features อยางไร? • โดยอธิบายในรูปแบบของข้อมูลสองแบบ คือ: – Spatial data ซึ่งอธิบายตำแหน่งที่ตั้ง (where) – Attribute data ซึ่งอธิบายลักษณะของสิ่งที่เกิดขึ้น ณ ที่ตั้งนั้น (what, how much, and when) เราแสดงขอมูลใน GIS เปนดิจิตอลอยางไร? • โดยการจัดให้เป็น layers ตามลักษณะของข้อมูลเกี่ยวกับ geographical features (เช่น hydrography, elevation, water lines, sewer lines, grocery sales) โดยใช้: – vector data model (coverage in ARC/INFO, shapeﬁle in ArcView) – raster data model (GRID or Image in ARC/INFO & ArcView) • โดยจัดลักษณะของ data properties ในแต่ละ layer ให้สอดคล้องกับ: – projection, scale, accuracy, and resolution เรานำขอมูลตางๆ เขามาในระบบโปรแกรมในคอมพิวเตอรอยางไร? • ใช้ความสามารถของระบบ GIS ที่สามารถจัดระบบฐานข้อมูลแบบ relational Data Base Management System (DBMS) 2
3. [3.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-3-638.jpg?cb=1446831726)Real World > Data Needed • Basic carrier of information = entity – Real-world phenomenon not divisible into phenomena of the same kind • An entity consists of: – Type Classification – Attributes – Relationships 3
4. [4.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-4-638.jpg?cb=1446831726) Entity type also known as qualitative data – or in statistics the ‘nominal scale’ 4λ Some entities may need to be categorized – e.g., roadways as a class: with categories for national highways, urban roads, private roads λ Each entity type must be unique (no ambiguity) – e.g., detached house classified under house; not industrial building λ Assumes identical occurrences can be classified λEntity: Type Classiﬁcation
5. [5.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-5-638.jpg?cb=1446831726)Entity: Attributes • Each entity type may have one or more attributes – e.g., buildings may have attributes characterizing material (frame or masonry), as well number of stories • Attributes may describe quantitative data ranked in three levels of accuracy Ordinal (Ranks) – Good – Better – Best Interval (numeric) – Age – Income Ratio (scale) – Length – Area 5
6. [6.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-6-638.jpg?cb=1446831726)Spatial Data Types • continuous: elevation, rainfall, ocean salinity • areas: – unbounded: landuse, market areas, soils, rock type – bounded: city/county/state boundaries, ownership parcels, zoning – moving: air masses, animal herds, schools of fish • networks: roads, transmission lines, streams • points: – ﬁxed: wells, street lamps, addresses – moving: cars, fish, deer 6
7. [7.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-7-638.jpg?cb=1446831726)Attribute data types Categorical (name): – nominal • no inherent ordering เช่น land use types, county names – ordinal • inherent order เช่น road class; stream class • บางครั้งก็แสดงเป็นตัวเลขที่ถือเป็นเพียงแค่ ชื่อเฉพาะ ไม่นำไปใช้ในการคำนวณ Numerical Known difference between values – interval • No natural zero • can’t say ‘twice as much’ • temperature (Celsius or Fahrenheit) – ratio • natural zero • ratios make sense (e.g. twice as much) • income, age, rainfall • ค่าตัวเลขเป็นเลขเต็ม integer [whole number] or หรือจุดทศนิยม ﬂoating point [decimal fraction] 7
8. [8.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-8-638.jpg?cb=1446831726)Real World > Data Modeling Source: Bernhardsen, Tor. (1999). 2nd Ed. Geographic Information Systems: An Introduction. p 38. 8
9. [9.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-9-638.jpg?cb=1446831726)Real World > Modeling Process Source: Bernhardsen, Tor. (1999). 2nd Ed. Geographic Information Systems: An Introduction. p 39. Fig 3.2. 9
10. [10.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-10-638.jpg?cb=1446831726)Modeling: Geometric & Attribute Data Source: Bernhardsen, Tor. (1999). 2nd Ed. Geographic Information Systems: An Introduction. p. 40. 10
11. [11.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-11-638.jpg?cb=1446831726)Modeling: Attribute Data Source: Bernhardsen, Tor. (1999). 2nd Ed. Geographic Information Systems: An Introduction. pp 40. 11
12. [12.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-12-638.jpg?cb=1446831726)Modeling: Entity Relations Source: Bernhardsen, Tor. (1999). 2nd Ed. Geographic Information Systems: An Introduction. pp 40. 12
13. [13.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-13-638.jpg?cb=1446831726)Data Model > Real-world entities correspond to database objects – carrier of information = entityλEntities as Objects > object(s) Image: Bernhardsen, Tor. (1999). 2nd Ed. Geographic Information Systems: An Introduction. p 42. 13
14. [14.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-14-638.jpg?cb=1446831726)Objects Characterized by: • Type (unique ID, type code/object class) • Attributes (qualitative/quantitative data) • Relations (calculable vs. attributable) • Geometry (point, line, area/polygon) • Quality (accuracy, resolution, coverage extent, representation, etc.) 14
15. [15.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-15-638.jpg?cb=1446831726)Data Base Management Systems (DBMS) Contain Tables or feature classes in which: – rows: entities, records, observations, features: • ‘all’ information about one occurrence of a feature – columns: attributes, fields, data elements, variables, items (ArcInfo) • one type of information for all features The key ﬁeld is an attribute whose values uniquely identify each row entity AttributeKey field 15
16. [16.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-16-638.jpg?cb=1446831726)Flat File Database Record Value Value Value Attribute Attribute Attribute Record Value Value Value Record Value Value Value 16
17. [17.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-17-638.jpg?cb=1446831726)Relational DBMS: Goal: produce map of values by district/ neighborhood Problem: no district code available in Parcel Table Solution: join Parcel Table, containing values, with Geograpahy Table, containing location codings, using Block as key field Tables are related, or joined, using a common record identiﬁer (column variable), present in both tables, called a secondary (or foreign) key, which may or may not be the same as the key field. Secondary or foreign key 17
18. [18.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-18-638.jpg?cb=1446831726)18
19. [19.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-19-638.jpg?cb=1446831726)Arc/node map data structure with files Arc/Node Map Data Structure with Files. 1 1,2,3,4,5,6,7 Arcs File POLYGON “A” A: 1,2, Area, Attributes File of Arcs by Polygon 1 2 3 4 5 6 7 8 9 10 11 12 13 1 x y 2 x y 3 x y 4 x y 5 x y 6 x y 7 x y 8 x y 9 x y 10 x y 11 x y 12 x y 13 x y PointsFile 1 2 2 1,8,9,10,11,12,13,7 19
20. [20.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-20-638.jpg?cb=1446831726)GIS Data Models: Raster v. Vector “raster is faster but vector is corrector” Joseph Berry • Raster data model – location is referenced by a grid cell in a rectangular array (matrix) – attribute is represented as a single value for that cell – much data comes in this form • images from remote sensing (LANDSAT, SPOT) • scanned maps • elevation data from USGS – best for continuous features: • elevation • temperature • soil type • land use • Vector data model – location referenced by x,y coordinates, which can be linked to form lines and polygons – attributes referenced through unique ID number to tables – much data comes in this form • DIME and TIGER files from US Census • DLG from USGS for streams, roads, etc • census data (tabular) – best for features with discrete boundaries • property lines • political boundaries • transportation 20
21. [21.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-21-638.jpg?cb=1446831726)Real World Vector RepresentationRaster Representation Concept of Vector and Raster line polygon point 21
22. [22.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-22-638.jpg?cb=1446831726)Representing Data using Raster Model • area is covered by grid with (usually) equal-sized cells • location of each cell calculated from origin of grid: – “two down, three over” • cells often called pixels (picture elements); raster data often called image data • attributes are recorded by assigning each cell a single value based on the majority feature (attribute) in the cell, such as land use type. • easy to do overlays/analyses, just by ‘combining’ corresponding cell values: “yield= rainfall + fertilizer” (why raster is faster, at least for some things) • simple data structure: – directly store each layer as a single table (basically, each is analagous to a “spreadsheet”) – computer data base management system not required (although many raster GIS systems incorporate them) corn wheat fruit clover fruit oats 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 1 1 1 1 1 4 4 5 5 5 1 1 1 1 1 4 4 5 5 5 1 1 1 1 1 4 4 5 5 5 1 1 1 1 1 4 4 5 5 5 1 1 1 1 1 4 4 5 5 5 2 2 2 2 2 2 2 3 3 3 2 2 2 2 2 2 2 3 3 3 2 2 2 2 2 2 2 3 3 3 2 2 4 4 2 2 2 3 3 3 2 2 4 4 2 2 2 3 3 3 22
23. [23.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-23-638.jpg?cb=1446831726)• grid often has its origin in the upper left but note: – State Plane and UTM, lower left – lat/long & cartesian, center • single values associated with each cell – typically 8 bits assigned to values therefore 256 possible values (0-255) • rules needed to assign value to cell if object does not cover entire cell – majority of the area (for continuous coverage feature) – value at cell center – ‘touches’ cell (for linear feature such as road) – weighting to ensure rare features represented • choose raster cell size 1/2 the length (1/4 the area) of smallest feature to map (smallest feature called minimum mapping unit or resel--resolution element) • raster orientation: angle between true north and direction defined by raster columns • class: set of cells with same value (e.g. type=sandy soil) • zone: set of contiguous cells with same value • neighborhood: set of cells adjacent to a target cell in some systematic manner Raster Data Structures: Concepts 23
24. [24.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-24-638.jpg?cb=1446831726)Full Matrix--162 bytes 111111122222222223 111111122222222233 111111122222222333 111111222222223333 111113333333333333 111113333333333333 111113333333333333 111333333333333333 111333333333333333 1,7,2,17,3,18 1,7,2,16,3,18 1,7,2,15,3,18 1,6,2,14,3,18 1,5,3,18 1,5,3,18 1,5,3,18 1,3,3,18 1,3,3,18 Raster Data Structures Runlength Compression (for single layer) Run Length (row)--44 bytes “Value thru column” coding. 1st number is value, 2nd is last column with that value. Now, GIS packages generally rely on commercial compression routines. Pkzip is the most common, general purpose routine. MrSid (from Lizard Technology)and ECW (from ER Mapper) are used for images. All these essentially use the same concept. Occasionally, data is still delivered to you in run-length compression, especially in remote sensing applications. This is a “lossless” compression, as opposed to “lossy,” since the original data can be exactly reproduced. 24
25. [25.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-25-638.jpg?cb=1446831726)Raster Data Structures Quad Tree Representation (for single layer) • sides of square grid divided evenly on a recursive basis – length decreases by half – # of areas increases fourfold – area decreases by one fourth • Resample by combining (e.g. average) the four cell values – although storage increases if save all samples, can save processing costs if some operations don’t need high resolution • for nominal or binary data can save storage by using maximum block representation – all blocks with same value at any one level in tree can be stored as single value store this quadrant as single 1 store this quadrant as single zero 1 1 1 1 1 1 1 1 I 1,0,1,1 II 1 III 0,0,0,1 IV 0 Essentially involves compression applied to both row and column. 2 2 1 2 3 4 4 4 4 54 4 4 3 4 2 3 4 2.5 3.5 3.25 25
26. [26.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-26-638.jpg?cb=1446831726)Raster Data Structures: Raster Array Representations for multiple layers • raster data comprises rows and columns, by one or more characteristics or arrays – elevation, rainfall, & temperature; or multiple spectral channels (bands) for remote sensed data – how organise into a one dimensional data stream for computer storage & processing? • Band Sequential (BSQ) – each characteristic in a separate file – elevation file, temperature file, etc. – good for compression – good if focus on one characteristic – bad if focus on one area • Band Interleaved by Pixel (BIP) – all measurements for a pixel grouped together – good if focus on multiple characteristics of geographical area – bad if want to remove or add a layer • Band Interleaved by Line (BIL) – rows follow each other for each characteristic A B B B III IV I II 150 160 120 140 Elevation Soil Veg File 1: Veg A,B,B,B File 2: Soil I,II,III,IV File 3: El. 120,140,150,160 A,I,120, B,II,140 B,III,150 B,IV,160 A,B,I,II,120,140 B,B,III,IV,150,160 Note that we start in lower left. Upper left is alternative. 26
27. [27.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-27-638.jpg?cb=1446831726)Vector Data Model Representing Data using the Vector Model: formal application • point (node): 0-dimension – single x,y coordinate pair – zero area – tree, oil well, label location • line (arc): 1-dimension – two (or more) connected x,y coordinates – road, stream • polygon : 2-dimensions – four or more ordered and connected x,y coordinates – first and last x,y pairs are the same – encloses an area – census tracts, county, lake Point: 7,2 Line: 7,2 8,1 Polygon: 7,2 8,1 7,1 7,2 8 . x=7 y=2 1 2 7 1 2 7 1 2 7 8 8 27
28. [28.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-28-638.jpg?cb=1446831726)Vector Data Structures: Whole Polygon Whole Polygon (boundary structure): polygons described by listing coordinates of points in order as you ‘walk around’ the outside boundary of the polygon. – all data stored in one file • could also store--inefficiently--attribute data for polygon in same file – coordinates/borders for adjacent polygons stored twice; • may not be same, resulting in slivers (gaps), or overlap • how assure that both updated? – all lines are ‘double’ (except for those on the outside periphery) – no topological information about polygons • which are adjacent and have common boundary? • how relate different geographies? e.g. zip codes and tracts? – used by the first computer mapping program, SYMAP, in late ‘60s – adopted by SAS/GRAPH and many business thematic mapping programs. Topology --knowledge about relative spatial positioning --managing data cognizant of shared geometry Topography! --the form of the land surface, in particular, its elevation 28
29. [29.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-29-638.jpg?cb=1446831726)Whole Polygon: illustration A 3 4 A 4 4 A 4 2 A 3 2 A 3 4 B 4 4 B 5 4 B 5 2 B 4 2 B 4 4 C 3 2 C 4 2 C 4 0 E A B C D 1 2 3 4 5 0 1 2 3 4 5 C 3 0 C 3 2 D 4 2 D 5 2 D 5 0 D 4 0 D 4 2 E 1 5 E 5 5 E 5 4 E 3 4 E 3 0 E 1 0 E 1 5 Data File 29
30. [30.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-30-638.jpg?cb=1446831726)Vector Data Structures: Points & Polygons Points and Polygons: polygons described by listing ID numbers of points in order as you ‘walk around the outside boundary’; a second file lists all points and their coordinates. – solves the duplicate coordinate/double border problem – lines can be handled similar to polygons (list of IDs) , but how handle networks? – still no topological information – first used by CALFORM, the second generation mapping package, from the Laboratory for Computer Graphics and Spatial Analysis at Harvard in early ‘70s 30
31. [31.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-31-638.jpg?cb=1446831726)Points and Polygons: Illustration 1 3 4 2 4 4 3 4 2 4 3 2 5 5 4 6 5 2 7 5 0 8 4 0 9 3 0 10 1 0 11 1 5 12 5 5 E A B C D 1 2 3 4 5 0 1 2 3 4 5 A 1, 2, 3, 4, 1 B 2, 5, 6, 3, 2 C 4, 3, 8, 9, 4 D 3, 6, 7, 8, 3 E 11, 12, 5, 1, 9, 10, 11 Points File 1 2 34 5 6 78910 11 12 Polygons File 31
32. [32.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-32-638.jpg?cb=1446831726)Comprises 3 topological components which permit relationships between all spatial elements to be defined (note: does not imply inclusion of attribute data) • ARC-node topology: – defines relations between points, by specifying which are connected to form arcs – defines relationships between arcs (lines), by specifying which arcs are connected to form routes and networks • Polygon-Arc Topology – defines polygons (areas) by specifying which arcs comprise their boundary • Left-Right Topology – defines relationships between polygons (and thus all areas) by • defining from-nodes and to-nodes, which permit • left polygon and right polygon to be specified • ( also left side and right side arc characteristics) Vector Data Structure: Node/Arc/Polygon Topology Left Right from to 32
33. [33.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-33-638.jpg?cb=1446831726)Node/Arc/ Polygon and Attribute Data Relational Representation: DBMS required! Spatial Data Attribute Data Birch Cherry I II III IV 1 4 3 A35 Smith Estate A34 2 33
34. [34.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-34-638.jpg?cb=1446831726)1 2 3 4 5 X Representing Point Data using the Vector Model: data implementation Y • Features in the theme (coverage) have unique identifiers--point ID, polygon ID, arc ID, etc • common identifiers provide link to: – coordinates table (for ‘where) – attributes table (for what) • Again, concepts are those of a relational data base, which is really a prerequisite for the vector model 34
35. [35.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-35-638.jpg?cb=1446831726)Variety of Vector Models • Spaghetti model • Topological model (most common) • Triangulated irregular network (TIN) • Dime files and TIGER files • Network model • Digital Line Graph (DLG) • Shapefile (ArcView/ArcGIS; ESRI) • Others: HPGL, PostScript/ASCII, CAD/.dxf 35
36. [36.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-36-638.jpg?cb=1446831726)Vector Model: Spaghetti Source: Lakhan, V. Chris. (1996). Introductory Geographical Information Systems. p. 54. 36
37. [37.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-37-638.jpg?cb=1446831726)Vector Model: Topological Bernhardsen, Tor. (1999). 2nd Ed. Geographic Information Systems: An Introduction. p. 62. fig. 4.12. 37
38. [38.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-38-638.jpg?cb=1446831726)Why Topology Matters • Connections & relationships between objects are independent of their coordinates • Overcomes major weakness of spaghetti model – allowing for GIS analysis (Overlaying, Network, Contiguity, Connectivity) • Requires all lines be connected, polygons closed, loose ends removed. 38
39. [39.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-39-638.jpg?cb=1446831726)Vector Model: Network Source: Heywood, Ian and Sarah Cornelius and Steve Carver. An Introduction to Geographical Information Systems. p. 60. fig. 3.14. 39
40. [40.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-40-638.jpg?cb=1446831726)Vector Model: TIN: Triangulated Irregular Network Surface A B CD 6 1 2 3 4 5 E F G H Elevation points (nodes) chosen based on relief complexity, and then their 3- D location (x,y,z) determined. Points Elevation points connected to form a set of triangular polygons; these then represented in a vector structure. Polygons Attribute Info. Database Attribute data associated via relational DBMS (e.g. slope, aspect, soils, etc.) Advantages over raster: •fewer points •captures discontinuities (e.g ridges) •slope and aspect easily recorded Disadvantages: Relating to other polygons for map overlay is compute intensive (many polygons) 40
41. [41.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-41-638.jpg?cb=1446831726)Overview: Representing Surfaces • Surfaces involve a third elevation value (z) in addition to the x,y horizontal values • Surfaces are complex to represent since there are an infinite number of potential points to model • Three (or four) alternative digital terrain model approaches available – Raster-based digital elevation model • Regular spaced set of elevation points (z-values) – Vector based triangulated irregular networks • Irregular triangles with elevations at the three corners – Vector-based contour lines • Lines joining points of equal elevation, at a specified interval – Massed points and breaklines • The raw data from which one of the other three is derived • Massed points: Any set of regular or irregularly spaced point elevations • Breaklines: point elevations along a line of significant change in slope (valley floor, ridge crest) x y z 41
42. [42.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-42-638.jpg?cb=1446831726)Digital Elevation Model • a sampled array of elevations (z) that are at regularly spaced intervals in the x and y directions. • two approaches for determining the surface z value of a location between sample points. – In a lattice, each mesh point represents a value on the surface only at the center of the grid cell. The z-value is approximated by interpolation between adjacent sample points; it does not imply an area of constant value. – A surface grid considers each sample as a square cell with a constant surface value. Advantages • Simple conceptual model • Data cheap to obtain • Easy to relate to other raster data • Irregularly spaced set of points can be converted to regular spacing by interpolation Disadvantages • Does not conform to variability of the terrain • Linear features not well represented 42
43. [43.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-43-638.jpg?cb=1446831726)Triangulated Irregular Network • Advantages – Can capture significant slope features (ridges, etc) – Efficient since require few triangles in flat areas – Easy for certain analyses: slope, aspect, volume • Disadvantages – Analysis involving comparison with other layers difficult a set of adjacent, non-overlapping triangles computed from irregularly spaced points, with x, y horizontal coordinates and z vertical elevations. 43
44. [44.](https://image.slidesharecdn.com/2-151106173605-lva1-app6891/95/gis-data-structure-44-638.jpg?cb=1446831726)Contour (isolines) Lines Advantages • Familiar to many people • Easy to obtain mental picture of surface – Close lines = steep slope – Uphill V = stream – Downhill V or bulge = ridge – Circle = hill top or basin Disadvantages • Poor for computer representation: no formal digital model • Must convert to raster or TIN for analysis • Contour generation from point data requires sophisticated interpolation routines, often with specialized software such as Surfer from Golden Software, Inc., or ArcGIS Spatial Analyst extensionridge valley hilltop Contour lines, or isolines, of constant elevation at a specified interval, 44