PDHonline Course L154 (5 PDH)

Data in GIS

Instructor: Steve Ramroop, Ph.D.

2020

PDH Online | PDH Center

5272 Meadow Estates Drive
Fairfax, VA 22030-6658
Phone: 703-988-0088
www.PDHonline.com

An Approved Continuing Education Provider

Slide 1

Lecture 4 Content

■ **Geographic Information Systems (GIS)**

Data in GIS:

Common Data Formats in GIS

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

1

This lecture is a continuation of the data in GIS topics identified in the syllabus that is data in GIS – Common Data Formats in GIS.

Slide 2

Lecture content:
- **GIS Data format**
  - **Vector Data**
  - **Raster Data**
- **Data Examples:**
  - **Shapefiles and GeoMedia**
  - **Geodatabase and Warehouse**
  - **Coverage and E00 files**
  - **CAD: AutoCAD, Microstation**
  - **Lattice/Grid data**
  - **DLG**
  - **Image data**
  - **TIN**
  - **ASCII**
  - **Tabular database**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

2

This is the content of this lecture. The intended topics of discussion are identified. One of the main problems associated with the initial implementation of GIS is the use of data from other software that are of a different data format. The topics on this slide shows some of the common data formats used in GIS.

Slide 3

■ GIS Data Format

– **A GIS file format is a standard of <u>encoding</u> geographical information into a file.**

– **The GIS data formats are created mainly by government mapping agencies (such as the USGS) or by GIS software developers.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop   3

This slide identifies what is a GIS data format. It is generic explanation of GIS data format. GIS users are typically faced with the problem of how to convert data formats, or what is involved in the conversion with the loss of any component of the original data set. There are typically two types of data format creators: Government Mapping Agencies; or Software Developers.

Slide 4

GIS Data Format

– **Two general types of GIS Data Format exists**

  ▪ **Vector format**
  ▪ **Raster format**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

4

This slide identifies shows the two general GIS data formats used in GISs. They are the generic descriptions of GIS data formats: Vector and Raster formats.
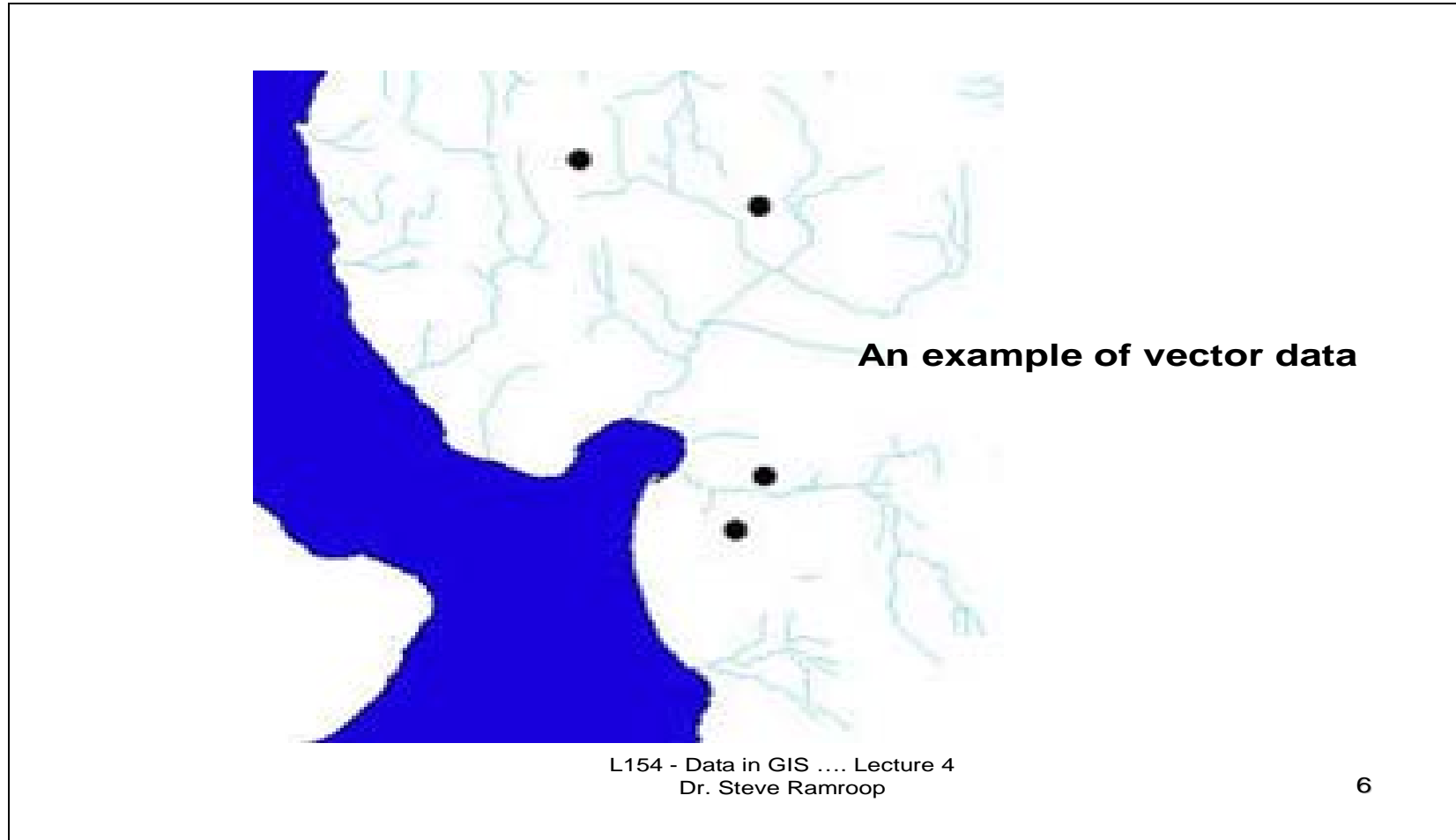
Slide 5

Vector Format

– **Vector data use X and Y coordinates to define the locations of points, lines, and areas (polygons) that correspond to map features such as fire hydrants, trails, and parcels.**

– **Vector data tend to define centers and edges of features.**

– **Roadways depicted as lines, fire-stations as points and lakes and ponds shown as polygons (areas) are just some examples of vector data representing real-life features.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop   5

This slide identifies some of the basic characteristics of the Vector Format. The basic primitive is an X, Y coordinate. The repetition of coordinates defines lines; while features that starts and ends with the same coordinate defines polygon features.

Slide 6

An example of vector data

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

6

This is an example of an example of vector data. The diagram shows some points representing cities, lines show the drainage network, while the blue area is a polygon showing the water body (lake).
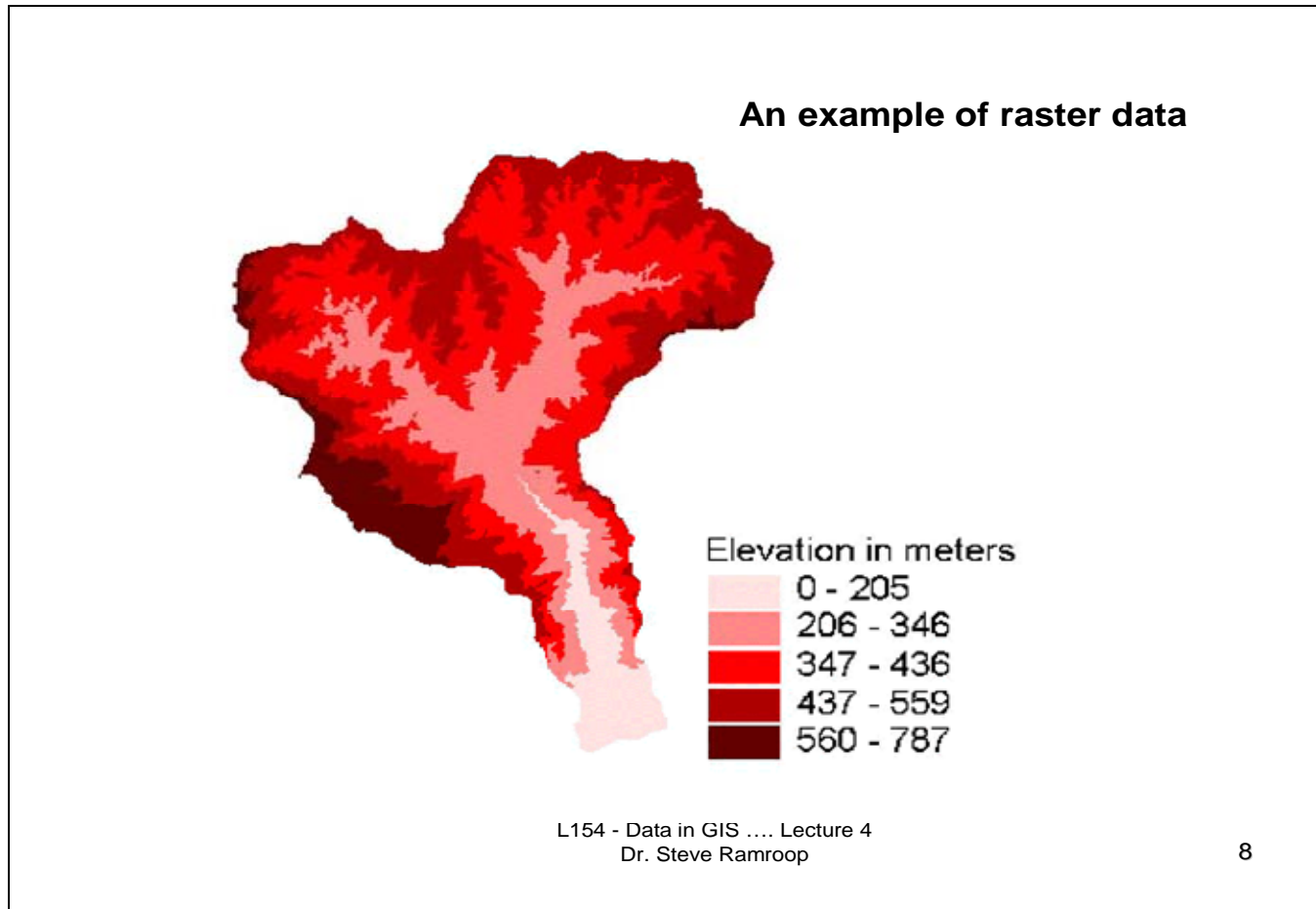
Slide 7

■ Raster Format

– **Raster data is defined by a matrix of square areas to define where features are located.**

– **The squares, also called pixels, cells, and grids, typically are of uniform size, and their size determines the detail that can be maintained in the dataset.**

– **Raster data represent areas that describe the interiors of features rather than boundaries as is the case with vector data.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

7

This slide identifies some of the basic characteristics of the Raster Format. The basic primitive of the raster format is the pixel. Raster data is space filling, in that every position on the data has a pixel value. Points are represented by a pixel; lines are represented by the set of connected pixels; while polygons are filled pixel areas.

Slide 8

An example of raster data

Elevation in meters
0 - 205
206 - 346
347 - 436
437 - 559
560 - 787

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop   8

This is an example of an example of raster data. The diagram shows some filled polygon areas for various elevation classes. Lightly colored red hue shows lower elevations when compared to the deeper red hues.
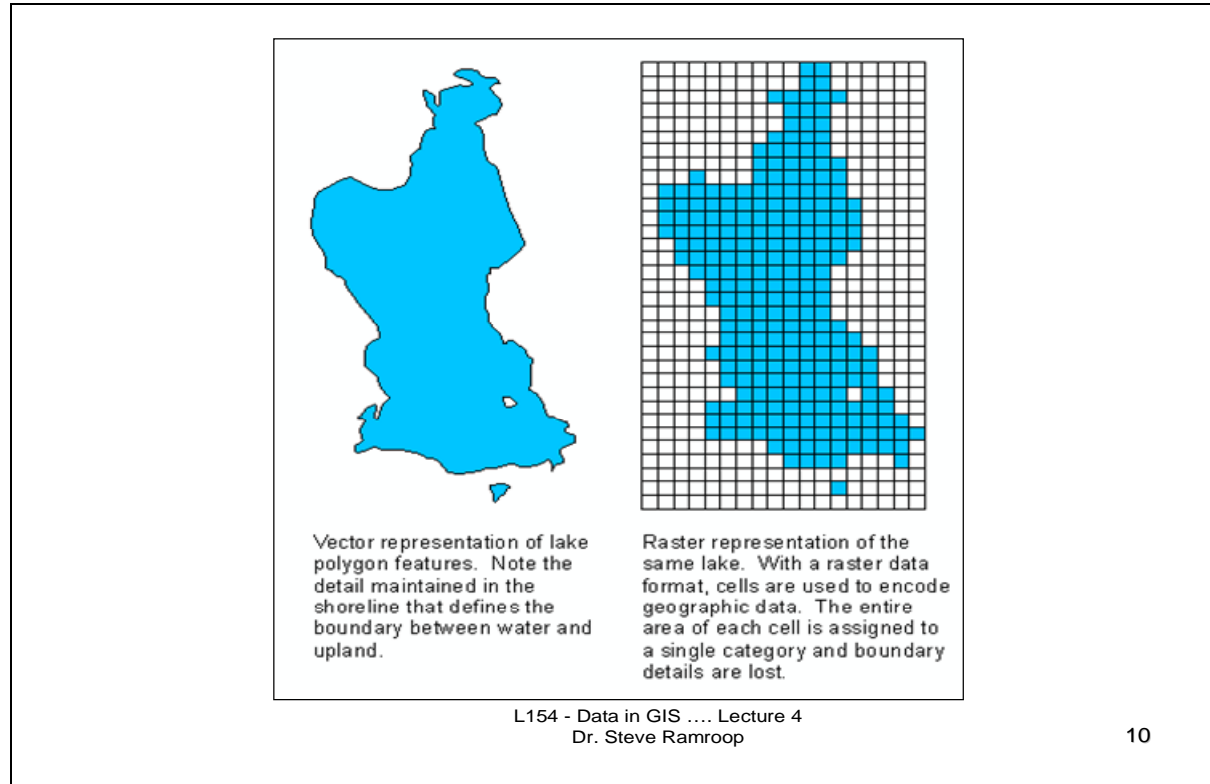
Slide 9

■ Raster and Vector Comparison

– **Vector data are excellent for capturing and storing spatial details using a set of coordinated points to define points, lines, and areas**

– **Raster data are well suited for capturing, storing, and analyzing data such as elevation, temperature, soil pH, etc. that vary continuously from location to location.**

– **Raster data formats also are used to store aerial and satellite imagery.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

9

This slide identifies a comparison between the raster and vector data formats. The choice of data formats is variable and all depends upon the intended GIS application. In some projects they are typically using vector formats (e.g. road network analysis); while in other projects raster formats are used (e.g. small scale site suitability selection); and other projects where a combination of the two formats are used (e.g. 3-D GIS applications). Therefore neither is the preferred choice.

Slide 10

Vector representation of lake polygon features. Note the detail maintained in the shoreline that defines the boundary between water and upland.

Raster representation of the same lake. With a raster data format, cells are used to encode geographic data. The entire area of each cell is assigned to a single category and boundary details are lost.

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop
10

This is an example of an example of a vector and raster representation of the same area of study. The vector diagram shows a lake bounded by a closed polygon then colored in blue. Notice the island polygon within the lake area. In the raster format the entire lake area is divided into a set of equal area squares. The blue squares represent the lake area while the white squares represent the non-lake areas. Notice in the raster format, the island polygon in the lake is represented by a single square.

Slide 11

Raster-to-Vector & Vector-to-Raster

- Moving from vector to raster is not that difficult.
- A line or polygon is simply given a pixel value.
- The opposite is not true though.
- The problem is that one line might be several pixels wide, therefore one has to skeletonize the line, often leaving it very jagged.
- This is a time consuming and complicated procedure.
- Sometimes it is impossible to exchange, and one cannot move between the formats. If this is the case, the map has to be re-digitized.
- In other instances, there is just a poor conversion, and data is lost in the exchange.

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

11

This slide show the two common methods used to convert between raster and vector data formats. To convert from raster to vector it is called "vectorization", while the reverse, vector to raster is called "rasterization". This slide shows some of the common characteristics of both conversions.

Slide 12

Raster and Vector formats

Raster	Vector
❑ National Geospatial-Intelligence Agency (NGA)'s ARC Digitized Raster Graphics ❑ Band Interleaved by Line (image format linked with satellite derived imagery) ❑ National Geospatial-Intelligence Agency (NGA)'s Compressed ARC Digitized Raster Graphics ❑ Digital raster graphic (DRG) - digital scan of a paper USGS topographic map ❑ ECW - Enhanced Compressed Wavelet (from ERMapper) ❑ ESRI grid - proprietary binary and metadata less ASCII raster formats used by ESRI ❑ GeoTIFF - TIFF variant enriched with GIS relevant metadata ❑ IMG - ERDAS IMAGINE image file format ❑ JPEG2000 - Open-source raster format. ❑ MrSID - Multi-Resolution Seamless Image Database	❑ Geography Markup Language (GML) - XML based open standard (by OpenGIS) for GIS data exchange ❑ AutoCAD DXF - Contour elevation plots in AutoCAD DXF format ❑ Shapefile - ESRI's open, hybrid vector data format using SHP, SHX and DBF files ❑ Simple Features - Open Geospatial Consortium specification for vector data ❑ MapInfo TAB format - MapInfo's vector data format using TAB, DAT, ID and MAP files ❑ National Transfer Format (NTF) - National Transfer Format (mostly used in the UK) ❑ TIGER - Topologically Integrated Geographic Encoding and Referencing ❑ Cartesian coordinate system (XYZ) - Simple point cloud ❑ Vector Product Format - National Geospatial-Intelligence Agency (NGA)'s format of vectored data for large geographic databases. ❑ GeoMedia - Intergraph's Microsoft Access based format for spatial vector storage ❑ ESRI's closed, integrated vector data storage strategy using Microsoft's Access MDB format ❑ ESRI's file-based geodatabase format, stored as folders in a file system ❑ Coverage - ESRI's closed, hybrid vector data storage strategy. Legacy / ArcInfo format with reduced support in ArcGIS Desktop lineup ❑ Spatial Data File - Autodesk's high-performance geodatabase format, native to MapGuide

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop   12

This slide identifies some of the common raster and vector data formats used in GIS. It is good practice to get familiar with a few of such formats.

Slide 13

Shapefile

– **A shapefile is a vector data format for storing the location, shape, and attributes of geographic features.**

– **Format developed by the Environment Systems Research Institute (ESRI)**

– **A shapefile is stored in a set of related files and contains one feature class.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

13

This slide show one of the common data formats used in GIS. Shapefile is the proprietary file format developed by ESRI.

It is typically a vector data format that has graphic and attribute component that are related.

Slide 14

■ Geomedia and Data Warehouses

– **A .GWS file is known as a GeoMedia GeoWorkspace File.**

– **Developed by Intergraph Cooperation**

– **GeoMedia enables the user to bring data from disparate databases into a single GIS environment for viewing, analysis, and presentation.**

– **No translation of data is required.**

– **Users avoid problems with redundant and out-of-date data because everyone is getting their information from the source.**

– **GeoMedia's data server technology supports open standards, providing direct access to all major geospatial/CAD data formats and to industry-standard relational databases.**

– ***GeoMedia* has no proprietary data format – simultaneous access to "*data warehouses*" in a multitude of formats is achieved in a standard and seamless manner.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

14

This slide shows some of the characteristics of a GeoMedia GeoWorkspace File by the Intergraph Cooperation. There is no proprietary data format of Geomedia. The strategy adopted by the Intergraph Cooperation is to access all types of data formats where there is no translation required.

Slide 15

Arc-Info Coverage

– **This is a data model for storing geographic features using ArcInfo software (originally written by ESRI)**
– **A coverage stores a set of thematically associated data considered to be a unit.**
– **A coverage usually represents a single layer, such as soils, streams, roads, or land use.**
– **In a coverage, features are stored as both primary features (points, arcs, polygons) and secondary features (tics, links, annotation).**
– **Feature attributes are described and stored independently in feature attribute tables.**
– **Coverages cannot be edited in ArcGIS.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop   15

This slide gives some characteristics of what is an Arc-Info coverage data format. Arc-Info is the original ESRI GIS software that has all of the processing capabilities. The data format used in Arc-Info is called coverage. It is a directory under which all of the related files associated with a data set are stored. For example, a road coverage data set will have

the projection files; registration files; attribute files; point/arc/polygons files; extent file; and such like. This slide gives some characteristics of the coverage data set.

Slide 16

Arc-Info Interchange File (.E00)

– **An ArcInfo interchange file**
– **Also known as an export file**
– **Is a file format used to enable a coverage, grid or TIN and an associated INFO table to be transferred between different machines.**
– **ArcInfo interchange files have a .E00 extension, which increments to .E01, .E02, and so on, if the interchange file is composed of several separate files.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop   16

This slide show the format coverage data sets is shared and exported to other computer applications. The .E00 file format is the file format which is typically shared by organizations that has coverage data. It is also called ArcInfo **interchange file**.

Slide 17

■ GeoDatabase

— **A geo-database is an object-oriented data model that represents geographic features and attributes as objects and the relationships between objects but is hosted inside a relational database management system.**

— **Developed by ESRI**

— **A geodatabase can store objects, such as feature classes, feature data sets, nonspatial tables, and relationship classes.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop   17

This slide shows a new format which is used by ESRI in their ArcGIS software that is built upon the ArcInfo concept. It represents a repository format that stores all of the shapefiles and other image data files for a given project. Such a repository has a file extension of .mdb. Geodatabases are shared easily among organizations and via the Internet.

Slide 18

Lattice/Grid

– **Lattice and Grid, as well as the term Raster, can be used synonymously to describe a data format that stores positional (horizontal) location information in a row-column (Cartesian) structure**

– **The individual row-column intersects are called cells, grid cells, or in the case of imagery, pixels (see next slide).**

– **Besides these functional advantages, grids or rasters are designed to store attributes.**

– **Though some grids may store multiple attributes just like vector data, grids usually store only a single numerical value which can represent a range of real life values such as biological sample data, rainfall amounts, or a gray or color-scale value representing a picture element.**

– **One of the most common applications of grids is where this numerical value, called the 'Z' value, stores a number representing elevation.**

– **This makes grid data very useful for 3-D analysis and display when you have the appropriate software.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop   18

This slide shows another format which is typical of Arc/Info. As shown are the characteristics of what is a lattice or grid data format. This is a format that is used to develop 3-D models.

Slide 19

Example of a Lattice/Grid

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop   19
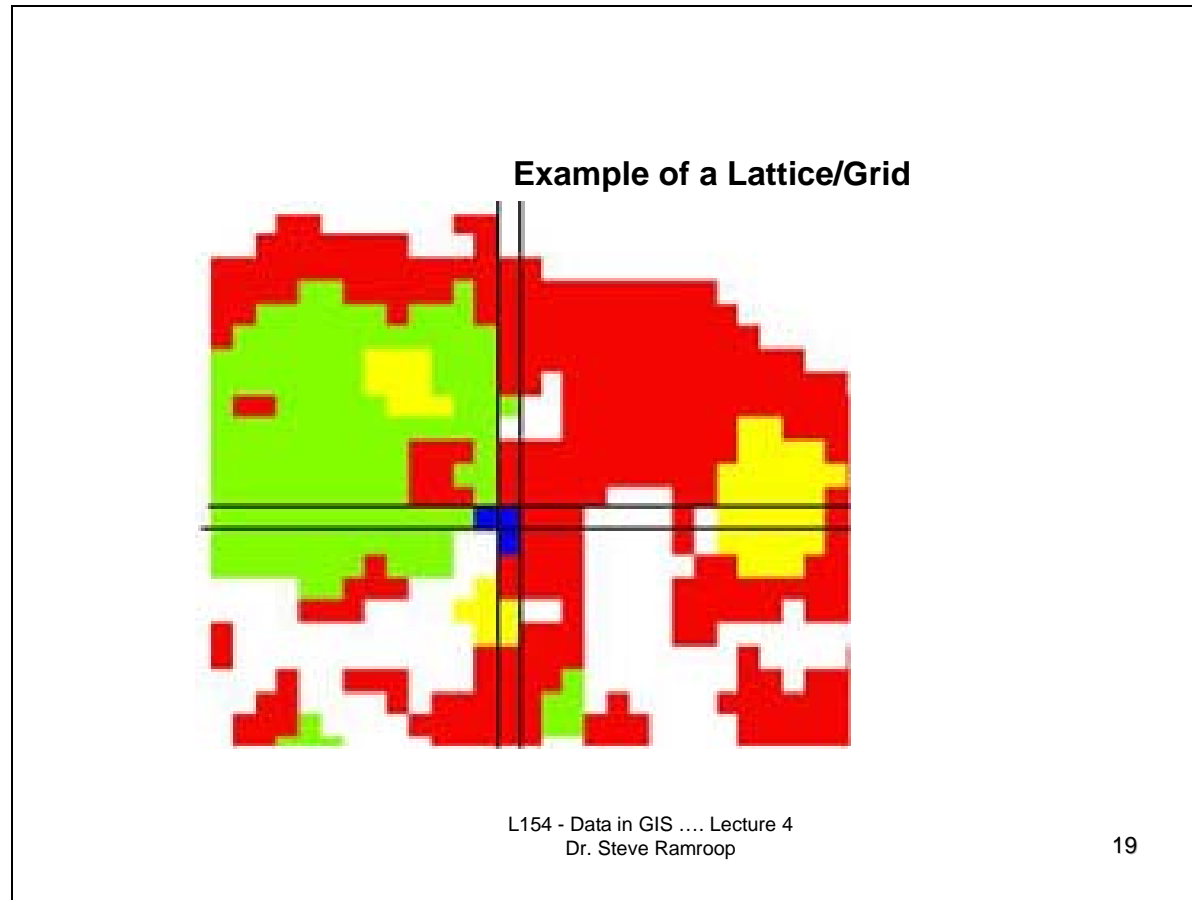
This is an example of a lattice or grid data format. It is a raster data set. The diagram shows pixel data where each color represents a specific attribute of the GIS application being developed.   The diagram has five distinct colors which represents five distinct attributes.

Slide 20

Digital Line Graph (DLG) :

– **DLGs are distributed by the government, and are available at 1:100,000 and 1:24,000 scales.**

– **Features are in separate files that most GIS packages will import, although extra data manipulation is often necessary.**

– **DLGs consist of line work with the contours removed, therefore elevation is not available.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

20

This slide shows another data format called digital line graph. It is typically line data which shows elevation data. GIS software would typically import and export into DLG formats.

Slide 21

■ Image

- – Images are similar to a grid or raster. Even though there are multiple image types, a common one is orthophotography (i.e., aerial or high-resolution satellite imagery where the relief and tilt errors are removed and a coordinates system is defined).

- – Images store their positional, that is x, y, location information in a pixel by pixel pattern just like grids, but in this case the 'Z' value is a number which is interpreted by software as a shade of gray, as in a panchromatic image, or a Red-Blue-Green color pattern as in color photography.

- – The 'Z' value is just a number so it can be manipulated as in a grid, allowing image analysis to be performed or imagery color or display characteristics to be modified.

- – Imagery provides a key cartographic role such as orthoimagery serving as an up-to-date background to other vector datasets.

- – Because of the common usage of imagery in GIS, most software supports a range of image file types such as TIF, IMG, etc., with installed or no-cost extensions.

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

21

This slide shows some additional data formats that can be used in a GIS. Image data is raster data. The format for image data are the common .tif, .img formats. To use such image data they need to be georeferenced for it to be included into a GIS. The gereferencing helps to add realistic coordinates to the pixels of the image.

Slide 23

■ TIN

– **Triangulated Irregular Network refers to a TIN.**

– **It is another format for storing 3 dimensional data that has an x, y and z value.**

– **In the case of a TIN the Z value is stored with links to adjacent nodes so that the data more closely represents a continuous surface than a grid or lattice representation. Like in a grid or lattice, the Z value can represent any quantifiable value, but TINs are often associated with storing and displaying elevation data.**

– **They are somewhat specialized in that they require 3-D analysis and display software such as ArcGIS 3-D analyst.**

– **TINs can be converted into grids and also in vector equivalents. However this changes the way the data is modeled and usually involves some interpolation of the data thus reducing the functionality of the TIN format.**

– **Even though TINs generally store only a single Z value as an attribute, the TIN format creates very large files as they store the relationship between all the features within the data.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

23

This slide show the TIN format that uses a set of irregularly shaped triangles to represent the phenomena being mapped. TINs are commonly used in representing elevation data. This slide gives some details into TINs.
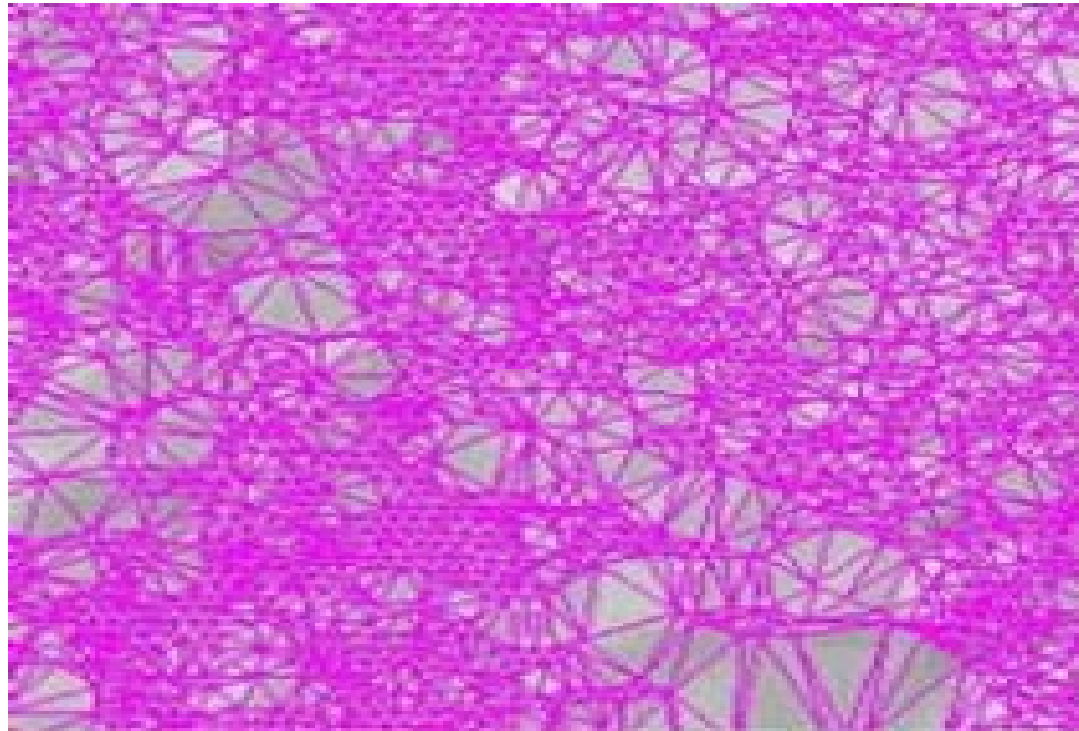
Slide 24

This slide gives an example of a TIN. Dense areas of triangles show more changes in the phenomena being mapped while the sparse triangles represent areas where the data is not changing a large amount.
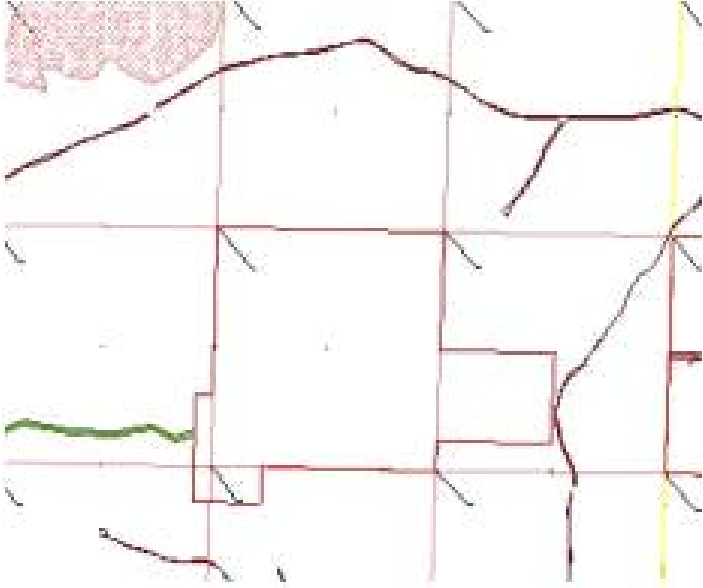
Slide 25

ASCII

– **Another type of vector data developed for and used extensively in engineering CAD (Computer Aided Drawing) software.**

– **ASCII refers to American Standard Code for Information Interchange**

– **Drawing files (DWG) and the ASCII export version (DXF) are broadly used to store planimetric linework such as roads, water/sewer infrastructure, and legal description information by public work agencies, survey departments and utility companies.**

– **For GIS users this data is often converted to GIS-type formats such as vector shapefiles, but DXF and DWG can also be read directly by most GIS software.**

– **These CAD data types provide a key bridge between GIS and engineering applications. For example the LiDAR-derived elevation contours are provided in both vector shapefile and vector DWG format.**

– **Though CAD formats provide accurate and detailed location information they do not store attribute information in the same way as GIS vector data does but rather provide more limited descriptive information in the LAYER and other DWG entity values.**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

25

This slide shows a common data exchange format used in CAD software. It is called ASCII. This slide gives some details into the ASCII format.

Slide 27

Example of an ASCII file

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

27

This slide shows an example of an ASCII file. It is a file which can be readily opened by most CAD and GIS software.

Slide 28

Tabular data set

– **Data in the form of tables**
– **In GIS they are referred to as attribute tables that are linked to the spatial database (which is the map)**
– **The link is through a unique identifier which can be a number or a string**

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop   28

This slide shows the characteristics of a common data set used in GIS – Tabular data set. The tables can be created using any database software such as MS-Access, MS-Excel, Dbase, Oracle, and such like.

Slide 29

Example of a tabular data set

Chscinam	Crit	Class	Sppno	Phyld
Acrocheilus alutaceus		F	224	F
Accipiter cooperii		B	1920	Y
Accipiter gentilis	B	B	1960	Y
Actitis macularia		B	3320	Y
Acipenser medirostris		F	32	F
Accipiter striatus		B	1880	Y
Acipenser transmontanus		F	40	F
Adelpha bredowii californica	IO	I	199	A
Aegolius acadicus		B	5340	Y
Aechmophorus clarkii	B	B	95	Y
Aegolius funereus	B	B	5330	Y
Aechmophorus occidentalis	B	B	90	Y

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

29

This slide shows an example of a tabular data set. The tables used have a set of rows and columns. Unique IDs can be used from the columns to link to other tables or geographic data set.

Slide 30

… The End …

L154 - Data in GIS …. Lecture 4
Dr. Steve Ramroop

30