Lecture 1
Intro to GIS and GIS Vector and Raster Data Models

Tomislav Sapic
GIS Technologist
Faculty of Natural Resources Management
Lakehead University
GIS Around Us

Object
- GIS layer
- Use

Hydro Pole
- utilities
- utility maintenance

Street
- streets
- Maintenance
- Navigation
- Emergency response
- Traffic congestion analyses

Underground gas and water lines
- water_line, gas_line
- Maintenance
- Emergency response

People
- census
- Health, employment, social services planning

House
- buildings
- Urban planning
- Service maintenance
- Real estate database

Tree
- tree_inventory
- Maintenance
- Tree disease monitoring
- Urban forest environmental value

GIS Augmented Reality
Geographic Information Systems

Definition:
... a computer system capable of assembling, storing, manipulating, and displaying geographically referenced information. (United States Geological Survey (USGS 2005)).

Data – a collection of facts or figures that pertain to places, people, things, events and concepts. Data are usually represented by numerical values, alphanumeric characters, logical values, etc. In GIS, there are also locational data – x,y,z coordinates representing positions on the surface of the Earth.

Information – data structured, formatted, converted and modelled to become meaningful to a user.

System – many interrelated parts, which may themselves be operational systems, constructed together to achieve certain objectives or functions.

Source: Heywood et al. (2006)
• GIS is a special class of information systems.

• GIS refers to those spatial information systems that are used for processing and analyzing **geospatial data**.

Geospatial data:
- Reference to geographic space in an accepted geographical coordinate system.
- Are normally recorded at relatively small scales, meaning that they need to be generalized and symbolized.

• However, GIS is computer technology, which means that it is constantly changing due to continuously improved computer capabilities – indoor GIS is one of the newest developments.

Source: Heywood et al. (2006)
Geospatial vs. non-spatial information

- Non-spatial information
- Geospatial or non-spatial information?
- Geospatial information – GIS!
GIS

Data Storage and Management

Spatial, Attribute Analyses

Visualization

Data Capture and Input

Hardware and GIS Software

Something is missing
People!

Depending on our interests and needs we decide on:

• What type of geographic features to collect.

• How to store information, in what format, for what purpose.

• How to analyze data

• How to present data
GIS Components

- Data Capture and Input
- Hardware
- GIS Software
- Data Storage and Management
- Spatial and Attribute Analyses
- Visualization
- People
History of GIS

http://www.youtube.com/watch?v=eAFG6aQTwPk
## History of GIS

Canadian connection: The Canada Geographic Information System (CGIS), which was conceptualized in the early 1960s and became operational in 1972, as reported by Tomlinson et al. (1976), has been widely recognized as the first GIS ever produced.

<table>
<thead>
<tr>
<th>Stage of development</th>
<th>(a) The formative years</th>
<th>(b) Maturing technology</th>
<th>(c) GI infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technical environment</strong></td>
<td>• Mainframe computers • Proprietary software • Proprietary data structure • Mainly raster-based</td>
<td>• Mainframe and minicomputers • Georelational data structure • Graphical user interface • New data acquisition technologies (GPS, redefinition of datum, remote sensing)</td>
<td>• Workstations and PCs • Network/Internet • Open systems design • Multimedia • Data integration • Enterprise computing • Object-relational model and database • Geodatabase • Mobile computing</td>
</tr>
<tr>
<td><strong>Major users</strong></td>
<td>• Government • Universities • Military</td>
<td>• Government • Universities • Utilities • Business • Military</td>
<td>• Government • Universities and schools • Business • Utilities • Military • The general public</td>
</tr>
<tr>
<td><strong>Major application areas</strong></td>
<td>• Land and resource management • Census • Surveying and mapping</td>
<td>• Land and resource management • Census • Surveying and mapping • Facilities management • Market analysis</td>
<td>• Land and resource management • Census • Surveying and mapping • Facilities management • Market analysis • Utilities • Location-based services • Geographic data browsing</td>
</tr>
</tbody>
</table>

Source: Heywood et al. (2006)
The Canada Geographic Information System (CGIS)

- Developed as a digital technological alternative to paper storage of spatial (maps) and attribute data collected through Canada Land Inventory (CLI).

- CLI was designed to overcome inter-provincial differences and classify the land across Canada according to productive capability, be able to measure the areas and encourage the use of the collected data in planning, in the post-WW2 period of urban expansion.

- CGIS was fully operational in 1972 after 10 years of development.

- Roger Tomlinson was one of the main actors in the development; IBM was brought to the table to provide computer programming and data management.
GIS Then and Now

1970’s

Present

OpenStreetMap
The Free Wiki World Map

Search
more examples...
Where am I?

OpenStreetMap is a free worldwide map, created by people like you.
The data is free to download and use under its open license. Create a user account to improve the map.
### History of ESRI GIS Software

<table>
<thead>
<tr>
<th>Time frame</th>
<th>(a) The formative years</th>
<th>(b) Maturing technology</th>
<th>(c) GI infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960–1980</td>
<td>Mainframe computers</td>
<td>Mainframe and minicomputers</td>
<td>Workstations and PCs</td>
</tr>
<tr>
<td></td>
<td>Proprietary software</td>
<td>Georelational data structure</td>
<td>Network/Internet</td>
</tr>
<tr>
<td></td>
<td>Proprietary data structure</td>
<td>Graphical user interface</td>
<td>Open systems design</td>
</tr>
<tr>
<td></td>
<td>Mostly raster-based</td>
<td>New data acquisition technologies (GPS, redefinition of datum, remote sensing)</td>
<td>Multimedia</td>
</tr>
<tr>
<td>1980–mid-1990s</td>
<td></td>
<td></td>
<td>Data integration</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Enterprise computing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Object-relational model and database</td>
</tr>
<tr>
<td>Mid-1990s–Now</td>
<td></td>
<td></td>
<td>Geodatabase</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mobile computing</td>
</tr>
<tr>
<td></td>
<td>Universities</td>
<td>Universities</td>
<td>Universities and schools</td>
</tr>
<tr>
<td></td>
<td>Military</td>
<td>Military</td>
<td>Business</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Utilities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Military</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The general public</td>
</tr>
<tr>
<td>2002–2010</td>
<td>Land and resource management</td>
<td>Land and resource management</td>
<td>Land and resource management</td>
</tr>
<tr>
<td></td>
<td>Census</td>
<td>Census</td>
<td>Census</td>
</tr>
<tr>
<td></td>
<td>Surveying and mapping</td>
<td>Surveying and mapping</td>
<td>Surveying and mapping</td>
</tr>
<tr>
<td></td>
<td>Facilities management</td>
<td>Facilities management</td>
<td>Facilities management</td>
</tr>
<tr>
<td></td>
<td>Market analysis</td>
<td>Market analysis</td>
<td>Market analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Utilities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Utilities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The general public</td>
</tr>
<tr>
<td>2010–Now</td>
<td></td>
<td></td>
<td>Location-based services</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Geographic data browsing</td>
</tr>
</tbody>
</table>

- **1982**
  - ArcINFO released.
  - Command prompt based
  - Datasets: coverages, grids.
  - Discontinued in 1999.

- **1992**
  - ArcView 3.x released.
  - Graphical User Interface.
  - New dataset: *shapefile*.
  - Discontinued in 2002.

- **1999**
  - ArcGIS 8.0 released.
  - Combination between ArcINFO and ArcView 3.x
  - New dataset: *geodatabase*.

- **2004**
  - ArcGIS 9.3 released.

- **2010**
  - ArcGIS 10 released.

- There are other, including free and open source, GIS software available to GIS users.

- Practical GIS work for the course has been delivered mostly on ESRI GIS software.
<table>
<thead>
<tr>
<th>Sectors</th>
<th>Application Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
<td>Research in humanities, science, and engineering</td>
</tr>
<tr>
<td></td>
<td>Primary and secondary schools—school district delineation, facilities management, bus routing</td>
</tr>
<tr>
<td></td>
<td>Spatial digital libraries</td>
</tr>
<tr>
<td>Business</td>
<td>Banking and insurance</td>
</tr>
<tr>
<td></td>
<td>Real estate—development project planning and management, sales and renting services, building management</td>
</tr>
<tr>
<td></td>
<td>Retail and market analysis</td>
</tr>
<tr>
<td></td>
<td>Delivery of goods and services</td>
</tr>
<tr>
<td>Government</td>
<td>Federal government—national topographic mapping, resource and environmental management, weather services, public land management, population census, election, and voting</td>
</tr>
<tr>
<td></td>
<td>State/provincial government—surveying and mapping, land and resource management, highway planning and management</td>
</tr>
<tr>
<td></td>
<td>Local/municipal government—social and community development, land registration and property assessment, water and wastewater services</td>
</tr>
<tr>
<td></td>
<td>Homeland security and law enforcement—crime analysis, deployment of human resources, community policing, emergency planning and management</td>
</tr>
<tr>
<td></td>
<td>Health care—pattern and spread of diseases, distribution and delivery of health services</td>
</tr>
<tr>
<td></td>
<td>International development and humanitarian relief</td>
</tr>
<tr>
<td>Industry</td>
<td>Engineering—surveying and mapping, site and landscape development, pavement management</td>
</tr>
<tr>
<td></td>
<td>Transportation—route selection for goods delivery, public transit, vehicle tracking</td>
</tr>
<tr>
<td></td>
<td>Utilities and communications—electricity and gas distribution, pipelines, telecommunications networks</td>
</tr>
<tr>
<td></td>
<td>Forestry—forest resource inventory, harvest planning, wildlife management and conservation</td>
</tr>
<tr>
<td></td>
<td>Mining and mineral exploration</td>
</tr>
<tr>
<td></td>
<td>Systems consulting and integration</td>
</tr>
<tr>
<td>Military</td>
<td>Training</td>
</tr>
<tr>
<td></td>
<td>Command and control</td>
</tr>
<tr>
<td></td>
<td>Intelligence gathering</td>
</tr>
</tbody>
</table>

Source: Heywood et al. (2006)
Vector and Raster Data Models
• Spatial data can represent real-world features that have discrete boundaries (such as roads, buildings, lakes, rivers, administrative boundaries) as well as real-world phenomena/features that have non-discrete boundaries (such as precipitation and nutrient levels, terrain).

• In general, the vector data model is used to represent the former and the raster data model to represent the latter.

• However, the lines of separation are sometimes crossed.

• As well, a typical GIS process involves moving back-and-forth between the two models.

Source: Lo et al. (2007)
Raster vs. Vector Data Model Representations of the Real World

- Although discrete objects such as buildings, ski lifts, roads, are usually not represented by a raster data model, they are handy for illustrating a direct comparison of representation between a raster and a vector data model.

- Elevation, as a continuous land (field) characteristic, is usually stored in a vector file for easier data management, but is more suitable to be represented by a raster data model.

Source: Heywood et al. (2006)
Vector Data Model

- Spatial features are usually stored in separate theme GIS files.

Roads are typically stored in a separate line vector GIS file, called, e.g., roads.

Rivers would typically be stored in a separate line vector GIS files, called, e.g., rivers or streams.

Lakes would typically be stored in a polygon (poly) vector GIS file, called, e.g., lakes.

A small pond might be stored as part of a wetlands GIS file.

A forest area (stand) is usually delineated and stored in a separate poly GIS file, often called fri.
Vector Data Model

The three main geometric shapes used in the vector data model, to represent real world features are:

- Point
- Line
- Polygon

Points, lines and polygons are stored in separate, point, line, and polygon, GIS files.

Each GIS file along with spatial features also contains a table in which each row (record) represents one of the spatial features.

(An exception to the above are so called multipart features, where one row represents multiple features.)

Columns (fields) in the table are used to store data (attributes) describing each feature.
**Point GIS File**

- The shapefile represented by this layer contains point features representing municipalities.

- A feature is represented by a row (record) in the file’s table.

- The Shape column (field) contains the spatial information, i.e., x and y coordinates for each point.
Line features are made of two or more vertices (sing. vertex).
A vertex is a point georeferenced by an \(x,y\) pair.
Each record (row) in the table represents a whole line (all vertices composing the line), which is often just a segment of a longer line.

End vertices are called nodes.

In ArcGIS, vertices are not usually visible but can be made visible during the editing of the line.
Polygon features are made of one or more lines that enclose(s) an area. Since lines are made up of vertices that means that polygons are, in the end, made up of vertices (points), i.e. they are represented by a group of x,y coordinates. Each polygon is represented by one record in the file’s table.
• Circles, as well, are actually a series of vertices, i.e. a series of straight edges (not consisting of an infinite number of points!).

• This is changing as well, with GIS software now giving an option to connect vertices with parameterized segments, such as circular arcs, elliptical arcs, and Bézier curve.
x,y Pairs as Spatial Representation

- Ultimately, GIS vector files, point, line or polygon files, have their spatial representation stored in the computer memory through x,y coordinate pairs.
- There are no ‘lines’ stored – the GIS software recognizes the type of the file (a point, line, or a polygon GIS file) and displays the features as points only or connects the points (vertices) with lines and displays them as lines or fills the line-enclosed areas with a colour and displays the features as polygons.

Source: Lo et al. (2007)
Singleparts vs. Multipart

Singlepart
- One feature is represented by one record.

Multipart
- Multiple features are represented by one record.
- Applications for multipart features are for example, multiple polygons representing protective buffers around the same type of features (e.g. eagle nests).
- A downside of a multipart: only one attribute is assigned to all multipart parts, which, e.g. for areas, can be deceiving when looking at individual parts.
Vector Data Model Attributes

- In GIS vector files attributes are, in a simple form, stored in tables (databases).

- A table consists of records (rows) representing individual features, fields (columns) representing a particular theme describing the feature, and attributes – an intersection between a record and a field.

- In ArcGIS, FID and Shape fields, although shown in the Table of Attributes, are not actually part of the attributes, but rather represent the spatial and index information (e.g. the .shp and .shx content in shapefiles). Because of it FID and Shape fields cannot be deleted from the table, unlike any other pure attribute field.
## Typical Storages of Vector Data

- All here presented happen to be created by ESRI.

<table>
<thead>
<tr>
<th>Vector File Format</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Coverage**       | - first vector file format developed by ESRI  
                     - it was the main vector file format of, at the time, major ESRI software, ArcINFO  
                     - composite  
                     - has inherent topology  
                     - discontinued but can still be found in older GIS databases |
| **Shapefile**      | - developed as a simpler file format than coverages, at first meant only for converting coverages into it and viewing it in a more user friendly GIS software called ArcView  
                     - composite  
                     - become wide spread and adopted by all newly developed GIS software, including open source  
                     - often less stable than coverages, doesn`t have inherent topology |
| **Feature Class**  | - successor to shapefile  
                     - composite but residing in a geodatabase, in the form of bit tables  
                     - can only exist within a geodatabase  
                     - as part of a geodatabase can have, e.g., table joins and topology defined. |

*Feature Class* (not a true file but a geodatabase component)
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 entire shapefile is shown as one item.
- All components of a shapefile need to be present together (important when transferring files!), otherwise the shapefile can be defunct or incomplete.
- The projection file is a beneficial addition to a shapefile, although not a necessary one—shapefiles can be used without a projection file but this is to be avoided.
- All elements have the same filename (e.g., roads).
- Shapefiles can be either a point or a line (arc) or a polygon file -- they cannot contain more than one shape type!

Possible composite file extensions:

- **.dbf** – dBase table (database) file, containing attributes.
- **.shp** – the file that stores the feature geometry, i.e., x,y coordinates.
- **.shx** – the 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** and **.aih** – attribute index file.
- **.atx** – new, ArcGIS attribute index file.
- **.lxs, .mxs** – geocoding index files.
- **.cpg** – specifies character set code page.
Raster Data Model

- Raster data model is represented by square cells of same size organized in horizontal rows and vertical columns (rows and columns are never slanted).

A raster’s position (geographic location) is defined by assigning x,y coordinates to one or two corners -- positions of the cells themselves flow from this reference point and the overall Cartesian coordinate system structure.

Spatial resolution of a raster is defined by the cell size (1 m, ... 20 m, ... 100 m, ...).

Each cell contains a value. Values are always numeric and can be either integer or decimal numbers (or nodata in grids).

Nominal or ordinal values can be used in the raster when something is not measured but classified into categories (themes).
An ArcMap layer representing a raster that contains elevation values. The raster itself doesn’t contain colours – colours are assigned to raster cells by software for interpretation purposes and are based on cell values.

The raster converted into a simple text raster format, ASCII.

• A raster dataset representing terrain.
Raster’s Spatial Resolution

- Expresses the size of the earth surface represented by one (square) cell.
- A 10 m raster is a raster whose each cell represents 10 x 10 meters of the earth surface.
- Because of the nature of rasters and their easy computational resampling, raster spatial resolution does not necessarily represent the spatial variability on the ground.
- The measure of represented spatial variability is expressed with **Ground Sampling Distance**.
- For example, the 80 m raster in the example on the right can be taken and resampled into a 40 m raster, but in that case all four new 40 m cells within each original 80 m cell area would have the same cell value – the spatial resolution would become 40 m, but the Ground Sampling Distance would stay 80 m.

- Chosen coarseness of raster resolution has implications on the feature representation.
Raster Types Based on Cell Values

- In terms of the cell values, generally, rasters can be divided into four broad groups: image rasters, interpolated rasters, thematic rasters, and derivatives of each of the group.

- **Image rasters** have values that represent measurements of energy (reflections) captured by a sensor (camera).

- Image raster cell values are usually whole numbers (integers).
• Image raster files are often multiband

- Values representing different electromagnetic spectrum segments (i.e., primary colours, NIR, etc.), are stored in separate bands (channels) – spectral resolution!

- Red band: ~ 610 – 660 nm
- Green band: ~ 530 – 590 nm
- Blue band: ~ 430 – 490 nm

One byte (0-255) per pixel, per channel, if an 8 bit image; three bytes per pixel (0-16,777,215) per channel, if a 24 bit image, etc.

**Thematic rasters** have cell values that represent a theme, category, class. Usually these are whole, integer numbers.

- Cell values in **interpolated rasters** represent interpolated measurements such as terrain elevations, precipitation levels, soil pH levels, etc. Usually the cell values are decimal numbers (floats), although they can be integers as well.

- Thematic and interpolated rasters can be multi-band but are predominantly one-band.
Raster Types Based on Cell Values

• Examples of derivative rasters:


  • A zonal statistics raster can be created from a thematic and an interpolated raster [link](http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/h-how-zonal-statistics-works.htm).
Raster Types Based on Cell Values

- **Interpolated raster.**
- **Derivative raster.**
- **Image raster.**
- **Thematic raster.**

E.g., slope raster

Classification into classes

- Real world
- Object-based model
- Field-based model
- Regular tessellation
- User view level
- Exact object
- Inexact object
- Irregular sampling points
- Data model level
- Vector model
- Raster model
- Vector-raster conversion
- Interpolation/Gridding
- Object recognition
- Spatial database
Regular pattern of sizes and distances in the raster data model means that topology is inherent to it, allowing for a range of spatial, topological analyses.

Relation Analyses Through Overlapping Neighbourhood Analyses

Cost-Path Analyses
Raster Dataset Properties

- **Format**: the type of file used to store the raster (for example, .jpg, .tiff).

- **Number of Bands**: the number of layers in the raster file, representing the same area but through different values.

- **Data Type**: the type of data assigned to raster cells; integer (whole number), float (decimal number), etc. The data type is assigned to the entire raster, all cells have the same data type – there can’t be one cell having an integer value and the next cell a float number value.

- **Data Depth**: also known as pixel or bit depth – the per-cell binary range assigned to the raster; e.g., an 8-bit integer raster can have integers (whole numbers) from 0 – 255 (8 bits = 255) assigned to a cell, only.

- **Statistics**: include minimum and maximum value, mean, and standard deviation of all values assigned to raster cells.

- **Extents**: left, right, top and bottom coordinates of the raster dataset.

- **Projection**: raster’s coordinate system (map projection).

- **Size of the Raster**: the number of rows and columns in the raster.
Raster Data Model

• In the raster data model, attributes are limited to the numeric values of the cells themselves, and while it is possible to link additional attributes to the groups of cells having same values, this is rarely done in practice for the reasons of low utilization value and cumbersome data management.

• Raster data models often take more memory space than the spatial component of the vector data models (attributes attached to the vector data models can tip the balance the other way).

• Rasters can also have NoData values, which indicates absence of data (not the same as zero!).

Typical Raster Data Formats in ArcGIS

GRID – a raster data format used by ESRI platforms; can contain both integers and floats; usually one-band but can be stacked into so called grid stacks.

.img – a raster file format created for ERDAS, a remote sensing software.; can contain both integers (up to 32 bit) and floats; can be both one-band or multiband (including > 3 bands).

.tif - can contain both integers (up to 32 bit) and floats; can be both one-band or multiband (including > 3 bands).

.jpg - cannot contain > 8-bit (unsigned) values and cannot have > 3 bands.

ASCII – uncompressed, simple raster data format, used to transfer information.
## Vector and Raster Data Models and Their Advantages

<table>
<thead>
<tr>
<th>Vector Data Model</th>
<th>Raster Data Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suitable for:</strong></td>
<td><strong>Suitable for:</strong></td>
</tr>
<tr>
<td>• Recording discrete features with definable boundaries.</td>
<td>• Representing continuous spatial features.</td>
</tr>
<tr>
<td>• Attaching both textual and numerical attributes to spatial features, and managing and analyzing these attributes and through them the features themselves.</td>
<td>• Containing images (aerial, satellite).</td>
</tr>
<tr>
<td>• Editing, update, and management of the spatial features.</td>
<td>• Analyzing relations between continuous phenomena taking place on the same geography.</td>
</tr>
<tr>
<td>• Representing natural shapes.</td>
<td>• Analyzing spatial relations between features (cost-path, density, interspersion, edge contact, etc.).</td>
</tr>
<tr>
<td></td>
<td>• Neighbourhood analyses.</td>
</tr>
</tbody>
</table>

In many respects, vector and raster data models complement each other and get switched between along the processes of data collection, observation, analyses, and presentation.
References: