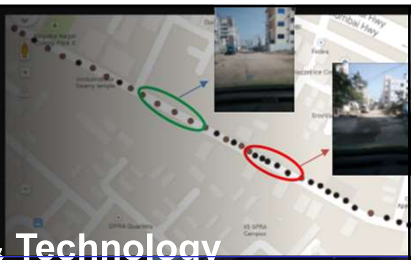


Intro to Spatial Science & Technology

Lecture 11b: Spatial Data Handling
- Data collection

Dr. K. S. Rajan
Lab for Spatial Informatics, IIIT Hyderabad
Feb 20th 2024

Spatial data

- Spatial data models and data structures
  - Two general ways to view reality: Object View (where is everything?), Field View (what occurs everywhere?)
  - Implemented naturally (yet not necessarily) as Vector / Raster
  - Vector: Point, Line, Polygon. Intuitive to human perception
    - Geometric / Topological data structures
  - Raster: Usually used for continuous natural objects or phenomena, e.g. elevation, rainfall, soil
    Simpler, faster for spatial analysis, large volumes
    Many compression methods developed
- Three dimensional data: Real 3D Vs. 2.5D
  - Representations: Raster (DN), TIN, Shading, Virtual GIS
- Temporal Dimension

GIS - Information stored in them

- Spatial (geometric/positional) info – locations of features, relative position to similar or other entities, shape/spread, distribution
- Coordinate System info – the exact location of the feature/s on the surface of the earth
- Attribute info – information pertaining to the spatial features or characteristics of the entities
- Symbology – how the spatial and attribute data are visualized or displayed
- Metadata – "Data about data"; information about the particular file containing the spatial data and its attributes

Approaches to Data Collection

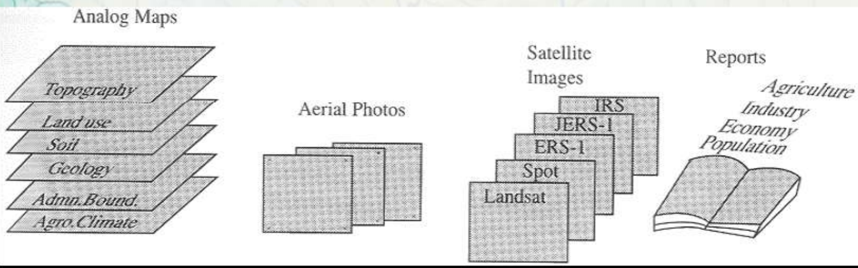
Depends on one or more of the following -

- Main Source of the data
  - Analog like Maps on paper or cloth, records;
  - Digital form with Geometry;
  - Only Attributes
- Data Format
  - Analog data – thematic or integrated maps; Scale or Not-to-Scale
  - Digital data – Data format – raster or vector
- Attribute Data with/without locational info

6

Main data sources for GIS

- Existing general-reference and thematic maps (digital or hardcopies);
- Ground Survey and Positioning;
- Remote Sensing Data Collection;
- Census and Sampling, Reports and Publications.

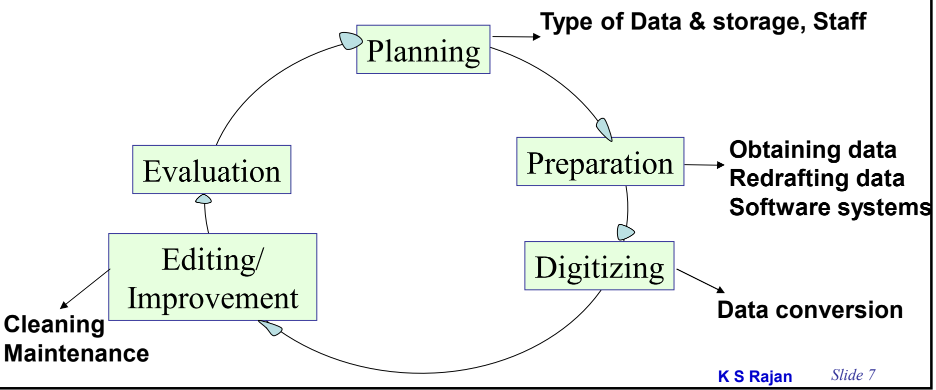


Data Format

- Either **digital** or **analog**.
- Analog data must always be digitized before being added to a geographic database.
- Data capture costs can account for up to **85%** of the costs of a GIS.
- When geographic data were very scarce, data collection was the main project task and it consumed the major of the available resources.
- Data collection still remains a time consuming, tedious, and expensive process. Usually it accounts for **15-50%** of the total cost of a GIS project.

Data collection Workflow

- The workflow commences with **planning**, followed by **preparation**, **digitizing**, **editing and improvement** and, finally **evaluation**.
- Planning is obviously important to any project and data collection is no exception.



GIS Data Collection

Two broad types of data collection:

- **Data capture** (direct collection)
- **Data transfer** (importing data from other sources)

Two broad capture methods

- **primary**: direct measurement;
- **secondary**: indirect derivation from other sources (those reused from earlier studies)

Capturing attribute data

Managing data capture project

Data Capture Techniques

Data Capturing Method	RASTER	VECTOR
Primary	Digital Remote Sensing Images	GPS measurements
	Digital Aerial Photographs	Survey measurements
Secondary	Scanned maps of Photographs	Digitization of Published (like Topograpic) maps
	DEMs and other digital data sets	Toponymy (palce names) databases
	Rasterization	***Vectorization***

Data Conversion approaches too

Primary Geographic **Raster**
data capture

Where raster data come from ?

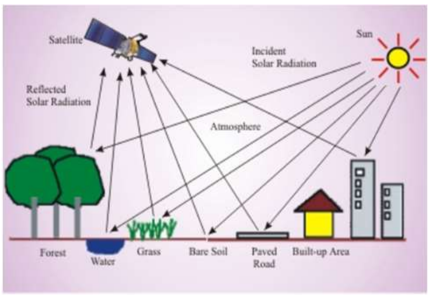
Satellite and air-photo imagery

Coded from existing maps or images – scanned maps, photographs etc.

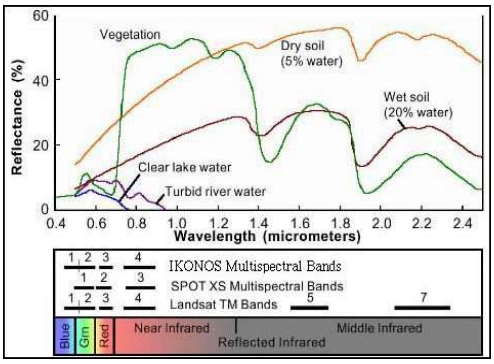
Derived from vector or other raster data. In GIS products, it is often called GRIDs that are results of Rasterization of vector data

Remote sensing

- most popular form of primary raster data capture is remote sensing
- a technique used to derive information about the physical, chemical and biological properties of objects without direct physical contact.



Imaging Spectroscopy



The consistency of data, systematic global coverage availability makes RS data useful for large area projects and for mapping inaccessible areas

Remote Sensing Data - Resolutions

Four key aspects of resolution are:

Spectral – wavelength ranges of the EMS that are measured

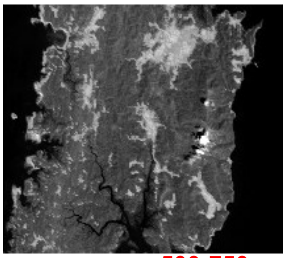
Radiometric – information quantization levels

Spatial- Size of the object that can be resolved

Temporal or repeat cycle or revisit time describes the frequency with which images are collected for the same area

Satellite Imagery

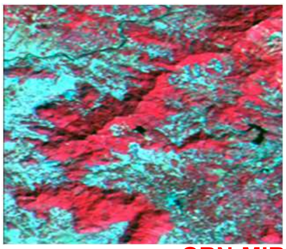
500-850nm
Cartosat-I



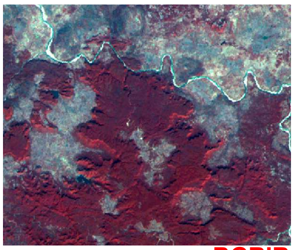
500-750nm
PAN

GRIR
LISS-IV

GRN-MIR
LISS III

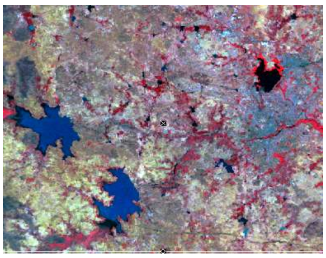


GRN-MIR
AWiFS

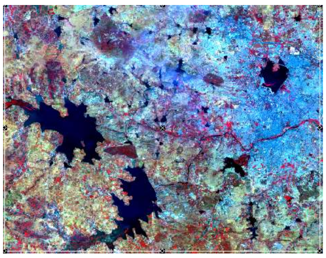


BGRIR
LISS I

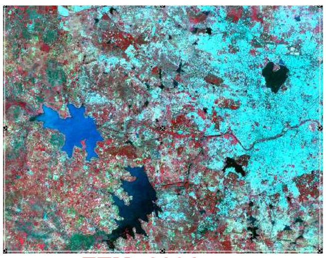
Part of Hyderabad city



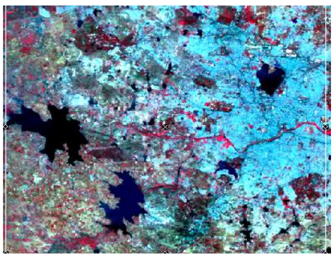
MSS -1976



TM -1989



ETM -2001



AWiFS -2001

Bits in RS data - Radiometric

- Data is stored in bits per pixel per channel
  - $\log_2 n$ (bits)
  - n is the quantization levels

sensor	satellite	level(bit)	descriptions
TM	Landsat	6	8bits data after radiometric correction
MSS	Landsat	8	
HRV(XS)	Spot	8	
HRV(PA)	Spot	6	
AVHRR	NOAA	10	both 10 and 16 bits data are available at distribution
SAR	JERS-1	3	real 3 bits, imaginary 3 bits
PAN	IRS-1C	6	
LISS-III, WiFS	IRS-1C	7	
PAN	Cartosat-1 (IRS-P5)	10 (1024 levels)	

Aerial Photographs

RS and AP are technically similar;
Difference in Capturing and Interpretation

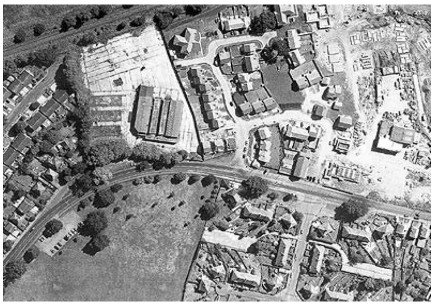
AP normally collected using analog optical camera, later rasterized by Scanning film negative

Cameras mounted in the nose or underbelly of an aircraft which flies at low altitudes (3000-9000m)

AP may be panchromatic or colour

AP are suitable for detailed surveying and mapping projects

A Part of UK



Both ***Satellite*** and ***Aerial Photographs*** are subjected to the process Of ***Georeferencing*** before being utilized for application purpose.

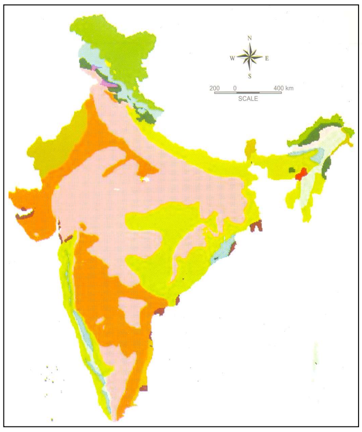
Toposheets

GCPs collected by GPS / DGPS

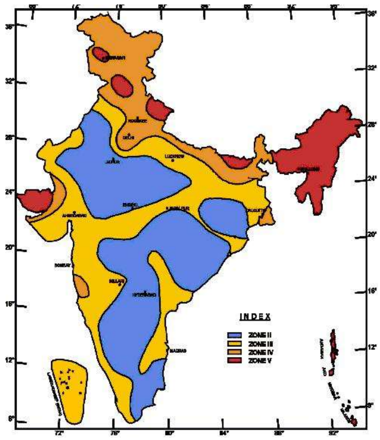
Secondary Raster data capture :

Using scanner
- Scanned maps and documents are used extensively in GIS as background maps and data stores.

Quality of outputs depend on
- quality of source
- data quality
- scanning device and
- type of preparation



Forest types of India

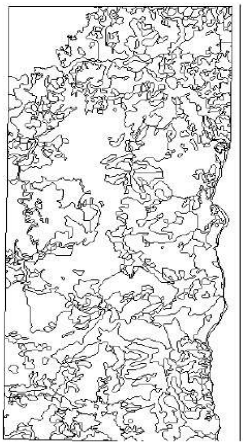


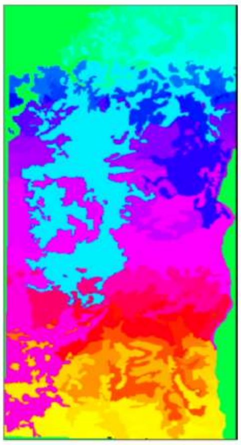
Seismic zones of India

Rasterization

Rasterization refers to conversion from vector to raster data.
Raster format is more convenient to produce color coded polygon maps such as color coded land use map
Rasterization is also useful to integrate GIS with remote sensing because remote sensing images are in raster format.

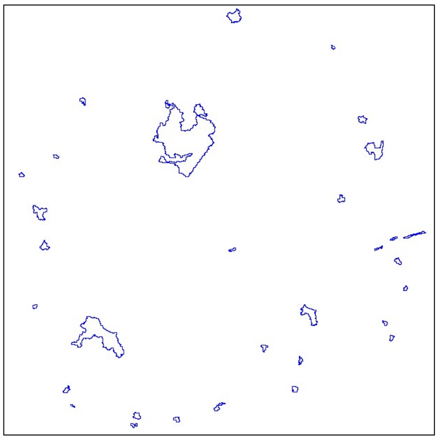
Rasterization

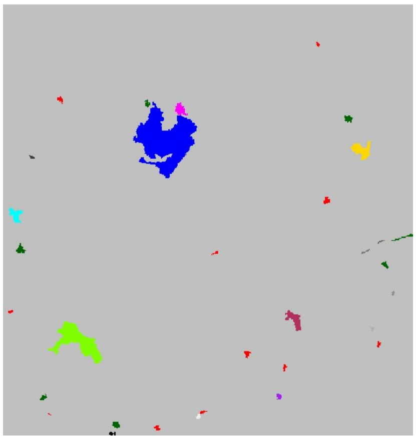




Forest type map of North Andaman

Hyderabad water bodies





Data Transfer

Interchange between different file formats

RAW (bsq, bil, bip), TIFF, JPEG, IMG, GEOTIFF, GRID, etc

Vector data capture

Primary data collection – direct measurement.

Two main branches of vector data capture

Ground Surveying (traditional)
GPS (Global Positioning System)

NEW Tech-Mobile platforms, LiDAR, IoT, etc

Ground Surveying:

- based on the principle that the 3D location of any point can be determined by measuring angles and distances from other known points.
- Surveys begin from a benchmark point, if the coordinate system of this point is known, all the subsequent points can be collected in this coordinate system.
- uses measurements to determine the locations of objects.
- Traditional survey equipment like theodolites (angles) tapes and chains (distance) are replaced by electro-optical devices called total stations, which can measure both angles and distances to an accuracy of 1mm.
- very time consuming and expensive, but still the best way to obtain highly accurate point locational data.

Theodolite

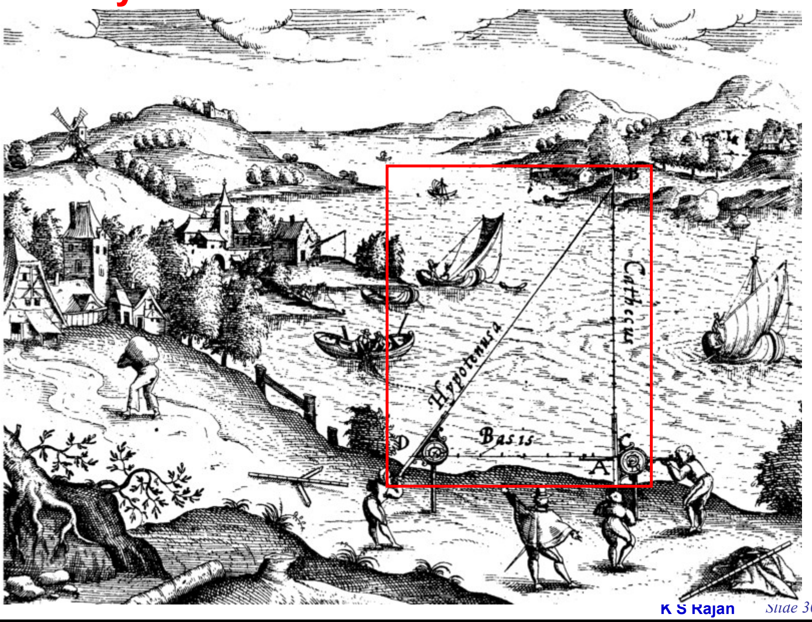
'08 Spatial Informa

Total stations automatically log data and the most sophisticated can create vector point, line and polygon objects in the field, thus providing direct validation.

Trimble 3600: designed to optimize workflow and field productivity. 3600 series delivers huge productivity gains with features such as DR (Reflectorless Mode), QuickDrive, Clamp Free Endless Slow Motion and Tracklight. ACU is an on-board, color, Windows CE device equipped with a graphical touch-screen that can also be used with Trimble 5800 RTK Rovers and 5600 Total Stations.

Trimble 5600: uses the best, most-productive measuring method available—ideal for a wide range of applications. For specialized high-precision applications, the 5600 IR Total Station is the ideal solution, providing with the capability to measure distances to an accuracy of +/- 0.8 mm +1 ppm.

A survey model



Geodata from Surveying

Data collected by field surveying

- Points coordinates
- Distances
- Elevations
- Attribute data

Data entry
- Import, if in digital form
- Manual entry

Geodata from GPS (Global Positioning System)

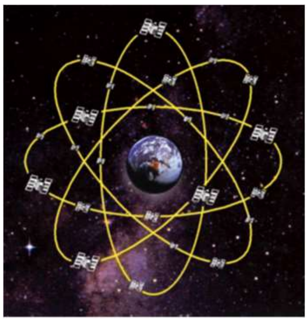
GPS data

- Handheld devices which can record location in the Field
- Used for locating features and entering attributes in the field
- Creates file of x,y,z coordinates to build point, line or polygon layers
- Data can be imported directly in to GIS for 'automatic' mapping or georeferencing

What is GPS?

GPS stands for ***Global Positioning System*** **(Constellation of 24+ satellites)**

Satellites broadcast precise time information. Using this information, one can easily calculate exact location on the Earth.



Feasibility studies begun in 1960's.
Pentagon appropriates funding in 1973.
First satellite launched in 1978.
System declared fully operational in April, 1995

GNSS (Global Navigation Satellite Systems) is a common acronym encompassing all existing and planned navigation systems.

Operational -
- *Navigation System with Timing and Ranging: Global Positioning System* or ***NAVSTAR*** *GPS*
- *GLONASS (Russian) – Global Navigation Satellite System*

Partially Operational
- European Galileo
- Indian NAVIC (IRNSS)
- Chinese Compass → now, BeiDou

See https://www.onesdr.com/list-of-gnss-frequency-and-accuracy/ for a list of GNSS Frequencies and Accuracies

Secondary Vector data capturing

- Methods are:
  - **Manual digitizing:** involves digitizing vector objects from maps and other geographic data sources.
  - Manual digitizing is still the simplest, easiest and cheapest method of capturing vector data from existing maps.
  - **Heads-up digitizing and Vectorization**:
  - **Photogrammetry**

Digitizing using digitizing table

•The process of converting continuous lines into discrete points so that they can be stored in a computer is called **digitizing**
Special table used for tracing features from a map in to a GIS
• Map is fixed to the table and a special mouse is used to trace over features
• Table is connected to a computer running GIS software, which records the movement of the mouse and creates points, lines and polygons according to the users controls
Tedious and time consuming, but fairly accurate

Tools for Manual digitizing

Digitizing board

Major problems in map digitizing

•The map will stretch or shrink day by day which makes the newly digitized points slightly off from the previous points;
•The map itself has errors;
•Discrepancies across neighboring map sheets will produce disconnectivity;
•Operators will make a lot of errors and mistakes while digitizing.

Attribute data

Import from GIS databases (ArcGIS, MapInfo, OSM, etc)

Import from general databases (MS Access, Oracle etc)

Manual entry

Derive new attribute data from existing data (classification, computation)

Import from field observations

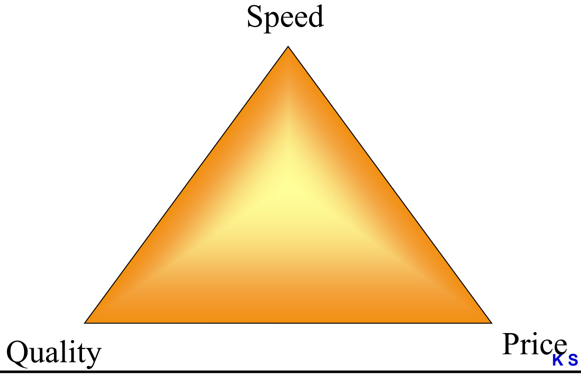
Socio-economic & environmental data

Socio-economic data is widely available, often from national and local government, and is usually the product of population surveys and censuses

This data combines with other datasets to produce neighbourhood Profiles to classify area for marketing purposes - **Geodemographics**

Managing a Data Capture Project

- In any data capture project, there is a fundamental tradeoff between **quality, speed and price**.
- Capturing high quality data quickly is possible, but it is very expensive. If price is a key consideration then lower quality data can be captured over a longer period.



Geographic data formats

- One of the biggest problems of external sources is many different formats
- As no single format is appropriate for all tasks and applications, many different formats have evolved in response to diverse user requirements.
- Data can be transferred between systems by direct read into memory or via an intermediate file format.
- More than 25 organizations are involved in the standardization of various aspects of geographic data and geoproccessing. Several of these are country and domain specific.
- At the global level, ISO (the International Standard Organization) is responsible for coordinating efforts through the work of technical committees TC211 and 287.

Open GeoSpatial Consortium (OGC) is the Standards' agency in GIS

Popular Geographic Data Formats

Vector	Raster (Image)
Automated Mapping System (AMS)	Arc Digitized Raster Graphics (ADRG)
ESRI Coverage	Band Interleaved by Line (BIL)
Computer Graphics Metafile (CGM)	Band Interleaved by Pixel (BIP)
Digital Feature Analysis Data (DFAD)	Band SeQuential (BSQ)
Encapsulated Postscript (EPS)	Windows Bitmap (BMP)
Microstation drawing file format (DGN)	Device-Independent Bitmap (DIB)
Dual Independent Map Encoding (DIME)	Compressed Arc Digitized Raster Graphics (CADRG)
Digital Line Graph (DLG)	Controlled Image Base (CIB)
AutoCAD Drawing Exchange Format (DXF)	Digital Terrain Elevation Data (DTED)
AutoCAD Drawing (DWG)	ERMapper
MapBase file (ETAK)	Graphics Interchange Format (GIF)
ESRI Geodatabase	ERDAS IMAGINE (IMG)
Land Use and Land Cover Data (GIRAS)	ERDAS 7.5 (GIS)
Interactive Graphic Design Software (IGDS)	ESRI GRID file (GRID)
Initial Graphics Exchange Standard (IGES)	JPEG File Interchange Format (JFIF)
Map Information Assembly Display System (MIADS)	Multi-resolution Seamless Image Database (MrSID)
MOSS Export File (MOSS)	Tag Image File Format (TIFF; GeoTIFF tags are supported)
TIGER/Line file: Topologically Integrated Geographic Encoding and Referencing (TIGER)	Portable Network Graphics (PNG)
Spatial Data Transfer Standard/Topological Vector Profile (SDTS/TVP)	
ESRI ArcView GIS (Shapefile)	
Vector Product Format (VPF)	
UK National Transfer Format (NTF)	