Co-funded by the
Erasmus+ Programme
of the European Union

ECOL 572: ENVIRONMENTAL INFORMATICS AND MODLING

Course Teacher

Prof. S Jayakumar

Dept. of Ecology & Environmental Sciences

Pondicherry University

Puducherry, India

Pondicherry University, India

Co-funded by the
Erasmus+ Programme
of the European Union

Content

Pondicherry University, India

Co-funded by the
Erasmus+ Programme
of the European Union

1. General Information

Course Code	:	ECOL - 572
Course Title	:	Environmental Informatics and Modeling
Number of Credits	:	4.0 ECTS
Course duration	:	18 Weeks
Level	:	Postgraduate
Course Teacher	:	Prof. S. Jayakumar
Prerequisite	:	Basic understanding on Mathematics (school higher level), English language skill, computer operation (Windows/Mac).

Pondicherry University, India

Co-funded by the
Erasmus+ Programme
of the European Union

2. Course description

This course provides the fundamentals of environmental informatics, different types of database management system, spatial data structure, how to extract information from different data sources such as air borne data, space borne data, climate data, GPS, topographical maps and how to convert them into digital form. It also teaches the students how to integrate and analyse the trend and pattern. It also introduces the students the concept of weightage, how to assign class weightage and layer weightage. The fundamentals of spatial modelling and how to perform modelling to identify the suitability and vulnerability. It also introduces the components of an information system and of remote sensing. This course also demonstrates data collection using GPS, Map reading, DBMS, 2D analysis and 3D analysis

Pondicherry University, India

Co-funded by the
Erasmus+ Programme
of the European Union

3. Course goals

The main aim of the course is to provide students what is environmental informatics and modeling and how can it be accomplished. The objectives of the course are to provide the important aspects of DBMS, to explain how do extract information from various datasets, to provide a fundamental understanding on how to integrate data, perform analysis and interpret the outputs, to provide the students to know about the basic components of information system, to make them understand how to perform suitability and vulnerability modeling and to demonstrate 2D and 3D data analysis.

Co-funded by the
Erasmus+ Programme
of the European Union

4. Course outcome

By the end of the course, successful students will:

1. Know the significance of environmental informatics and its relevance to the natural resources management.
2. be familiar with various database management systems
3. be able to retrieve data from DBMS through query
4. know the different types of information extraction
5. be able to conceptualize the types of thematic maps required and the appropriate source of data to prepare the different types thematic maps in both analog and digital formats
6. be able to think spatially to manage the natural resources
7. be familiar with the concept and framework of information system.
8. know the concept of 2d and 3d data analysis.

5. Course structure

5.a. Course Content

Week - 1	Introduction to environmental informatics
	Environmental data, sampling, primary and secondary data
Week - 2	Data sources, data quality and standards
Week - 3	Introduction to database management system
Week - 4	Significance of DBMS and spatial data structures
Week - 5	Resource information extraction – air borne, space borne and data
	Resource information extraction from topomap and climate data
Week - 6	Digital database creation
	Introduction to data analysis and visualization
Week – 7	Data integration, trend analysis, pattern analysis
Week – 8	Understanding data layers and weightage
Week – 9	Spatial environment and spatial analysis
	Introduction to spatial modeling
Week – 10	Environmental suitability and vulnerability modeling
Week – 11	Structure and components of information system
Week – 12	Working with GPS and data handling
Week – 13	Map reading and information extraction
Week – 14	2d data analysis
Week – 15	3d data analysis

5. Course structure

Understanding the fundamentals of environmental informatics and modeling in natural resources management

5.b. Mode of delivery

In-Class Lectures

Google Classroom

On-line Lectures

Students will get enrolled in Google classroom

And online classes will be conducted if covid -19 conditions do not permit to conduct off-line classes

5. Course structure

5.c. In-class discussion

The main focus of discussion would be to understand the database management system, basic structure and function of information system

5.d. In-class assignments & field assignment

To understand information extraction from different data sources, basic framework of data collection, sampling, data quality standards

5.e. Reading and discussion of assigned papers for seminars

To understand the application of environmental informatics in various sectors, how do institutions make use of environmental informatics for better management of resources, how do people benefit of it, how to improve the information system more robust and user friendly.

5.f. Group project presentation

Ability to make presentation, effective communication, critical interpretation of data, response to audience

6. Course Assessment

Type of assessment	Percentage of Marks
In-class discussion	5
Assignment	5
Seminars	10
Group projects	10
Internal assessment test (MCQ types)	10
Final assessment	60
Total	**100**

7. References

1. Bungartz, HJ, Kranzlmuller D and Weinberg, V 2019. Advances and New Trends in Environmental Informatics: Managing Disruption, Big Data and Open Science Springer Publication. ISBN-13: 978-3030076191.

2. Avouris, NM., and Page, B. 2010. Environmental Informatics: Methodology and Applications of Environmental Information Processing, Springer Publication.

3. Agarwal, S.K., 2002. Eco-informatics, APH Publishing Corporation, 1535 pages, ISBN-13: 978-8176483247.

4. Coronel, C., Morris, S., Rob, P., 2009. Database Systems: Design, Implementation and Management, 9th Ed., Course Technology, 700 pages, ISBN-13: 978-0538748841.

5. Maguire, D., Batty, M., Goodchild, M., (Eds.) 2005, GIS, Spatial Analysis, and Modeling, Esri Press, 496 pages, ISBN-13: 978-1589481305.

6. Goodchild, M.F., Parks, B.O., Steyaert, L.T., (Eds.), 1993. Environmental Modeling with GIS (Spatial Information Systems) Oxford University Press, USA, 520 pages, ISBN-13: 978-0195080070.

7. Jorgensen, S. E., Chon, T-S., Recknage, F. A., 2009. Handbook of Ecological Modeling and Informatics, WIT Press, 448 pages, ISBN-13: 978-1845642075.

Co-funded by the
Erasmus+ Programme
of the European Union

1. What is Environmental Informatics?

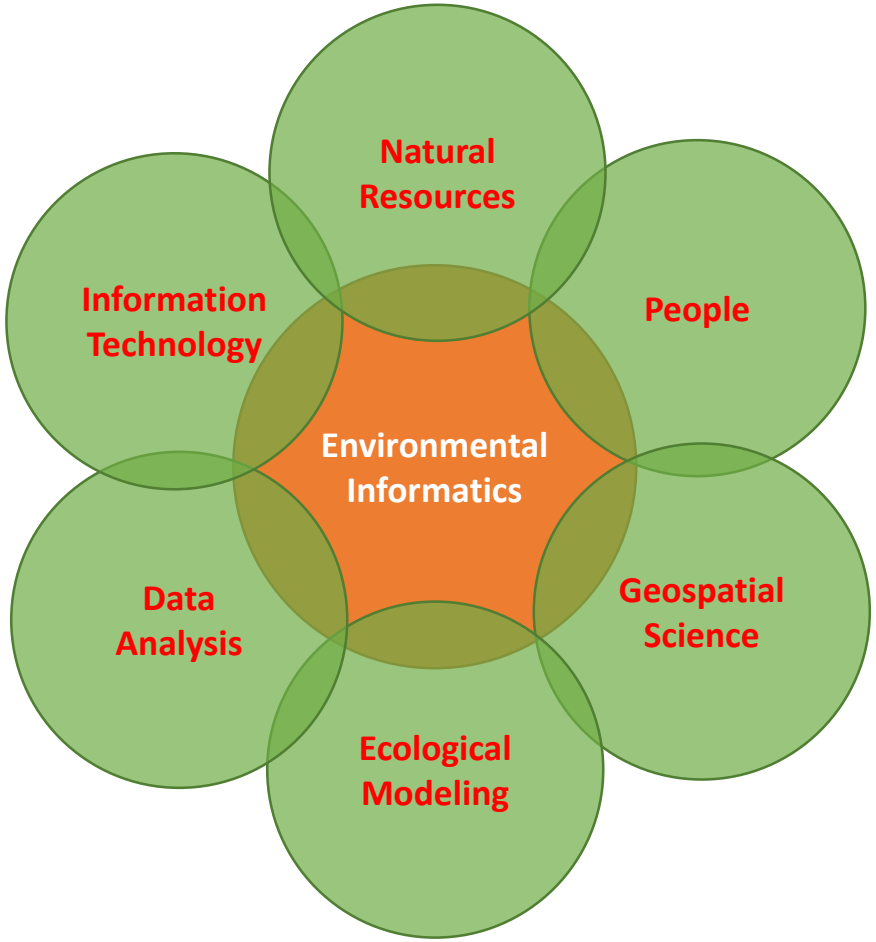
"Development of standard protocols/techniques for effective collection, storage, retrieval, sharing, integration and processing of complex environmental data to deliver comprehensive and meaningful information to address environmental problem."

EI –includes aspects of geographic information, mathematical and statistical modeling, remote sensing, database management, knowledge integration, and decision making

- Interdisciplinary field combining environment, information technology.
- Purpose is to translate environmental data into meaningful information.
- To deliver comprehensive and reliable information for environmental research and management
- International area of interest, used to address environmental problem worldwide.

Co-funded by the
Erasmus+ Programme
of the European Union

2. Components of Environmental Informatics



Natural resources – Spatial and Non-spatial data

Information Technology – Hardware and software database,

Data analysis – trend, decision, spatial

Ecological Modeling – Mathematical modeling/ spatial modeling

Geographical science - GIS

People – Experts, public, Govt. Depart. Planners decision makers, users

Co-funded by the
Erasmus+ Programme
of the European Union

2. Components of Environmental Informatics

Why is it so complex?

Environment = Resources = related to specific part of Earth.

Data → Spatial data and non-spatial data

Non-homogenous information sources

- text data (environmental laws)
- measurement data (monitoring network)
- geographically coded (specific to an area and time)
- multidimensional (complex geometric objects)

Visualization of environmental information

Co-funded by the
Erasmus+ Programme
of the European Union

3. Environmental data

- Data
  - Facts collected for calculation, analysis, reasoning, reference.
  - Primary data
  - Secondary data
  - Data collection

Co-funded by the
Erasmus+ Programme
of the European Union

3. Environmental data

- Data
- Spatial and Non-spatial data
  - Air
    - Ozone (O_3)
    - Particulate matter (PM_{10} and $PM_{2.5}$)
    - Carbon monoxide (CO)
    - Nitrogen dioxide (NO_2)
    - Sulfur dioxide (SO_2)
    - Lead (Pb)

Co-funded by the
Erasmus+ Programme
of the European Union

3. Environmental data

Air quality index

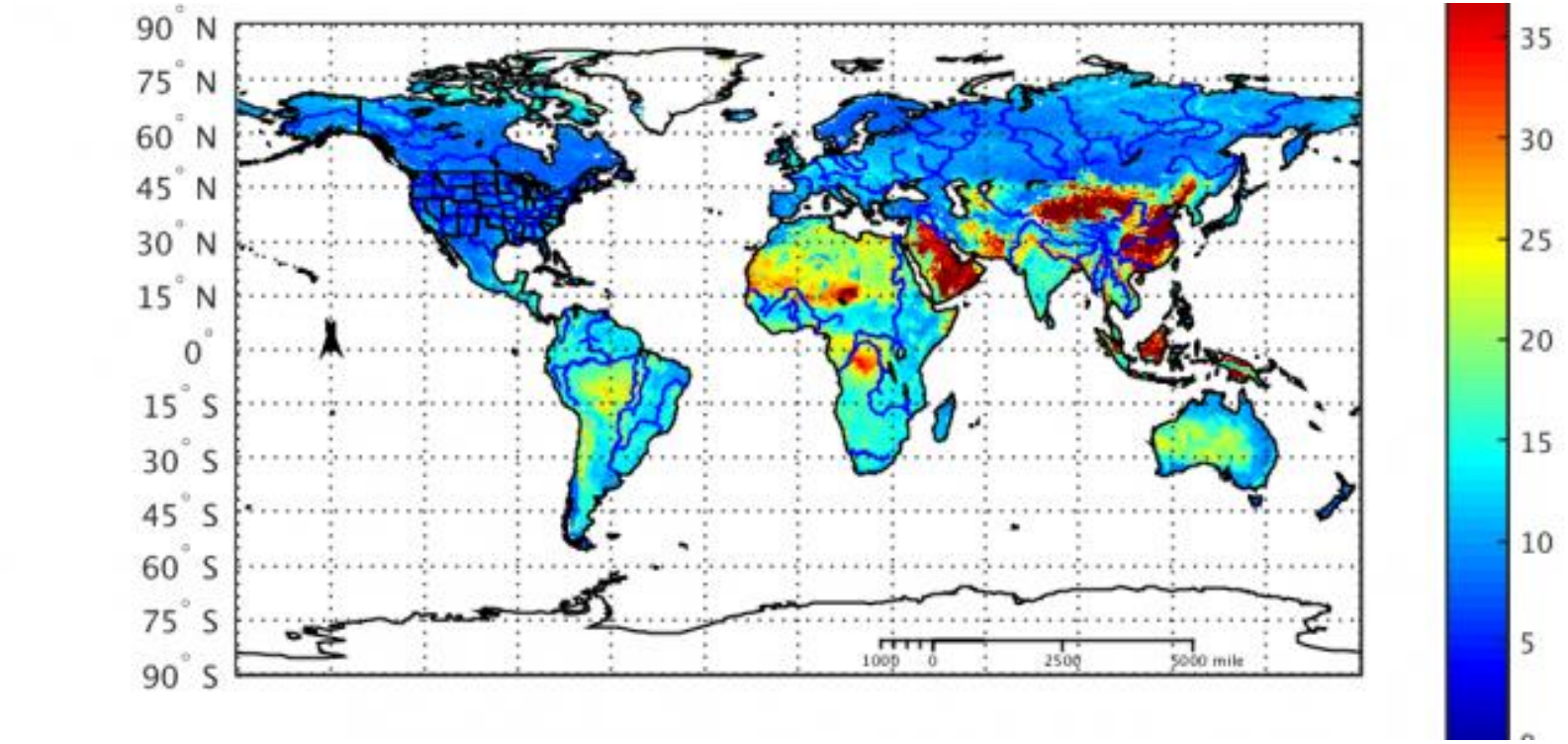
Air Quality Index (AQI) Values	Levels of Health Concern	Colors
When the AQI is in this range:	*..air quality conditions are:*	*...as symbolized by this color:*
0 to 50	Good	Green
51 to 100	Moderate	Yellow
101 to 150	Unhealthy for Sensitive Groups	Orange
151 to 200	Unhealthy	Red
201 to 300	Very Unhealthy	Purple
301 to 500	Hazardous	Maroon

Note: Values above 500 are considered Beyond the AQI. Follow recommendations for the Hazardous category. Additional information on reducing exposure to extremely high levels of particle pollution is available here.

Co-funded by the
Erasmus+ Programme
of the European Union

3. Environmental data

Air quality Map (PM 2.5 Micro gram /Meter cube)



Co-funded by the
Erasmus+ Programme
of the European Union

3. Environmental data

- Water
  - Acidity, Alkalinity, Ammonia, BOD, COD, Conductivity, Dissolved Oxygen, Fecal Coliform, Hardness, Metals, Nitrogen, Phosphorus, pH, Total Solids, Temperature, Turbidity,
- Soil
  - Physico-chemical properties
- Flora
  - Richness, diversity, abundance, distribution
- Fauna
  - Richness, diversity, abundance, distribution
- Weather
  - Temperature, Precipitation, Humidity

Co-funded by the
Erasmus+ Programme
of the European Union

4. Data variable and sampling

- Variable:
  - Variable = entity = data
  - Qualitative variable (Categorical variable)
  - Quantitative variable
    - Discrete variable
    - Continuous Variables
  - Univariate data
  - Bivariate data
  - Multivariate data

Co-funded by the
Erasmus+ Programme
of the European Union

4. Data variable and sampling

- Sampling
  - Sample – Observation drawn from Population
  - Types of sampling
    - Simple random
    - Stratified random
    - Cluster
    - Multistage
    - Systematic

Co-funded by the
Erasmus+ Programme
of the European Union

5. Data sources, quality and standards

- Data source
  - Maps
    - SOI Topo map
    - Other maps
  - Satellite data
    - Reflective
    - Thermal
    - Microwave
  - Aerial photo
  - Reports
  - Statistical data from Govt. departments/Institutes/Private

Co-funded by the
Erasmus+ Programme
of the European Union

5. Data sources, quality and standards

- Quality Check and Standards
  - Data quality
  - Scale
  - Pixel size – upscaling/downscaling
  - Unit
  - Analog to digital
  - Spatial reference
  - Data integration

Co-funded by the
Erasmus+ Programme
of the European Union

6. Data vs. information: What is the difference?

What is data?

- **Information science defines data as unprocessed information.**

What is information?

- **Information is data that have been organized/processed and communicated in a coherent and meaningful manner.**
- **Data is converted into information, and information is converted into knowledge.**

Co-funded by the
Erasmus+ Programme
of the European Union

7. Terms and Definitions

Data: An attribute

Field: one category of information, i.e., Name, Address, Semester Grade, Academic topic

Rec

Reg. No	Name	DOB	Blood group
1542689	Adrian	12/09/1985	B+
1542690	McGlove	05/11/1986	AB-

Data

File: A group or collection of similar records

Reg. No	Name	DOB	Blood group
1542689	Adrian	12/09/1985	B+
1542690	McGlove	05/11/1986	AB-

Co-funded by the
Erasmus+ Programme
of the European Union

8. Database management system

What is a database management system?

A collection of programs that enables you to store, modify, and extract information from a database.

There are many different types of DBMSs, ranging from small systems that run on personal computers to huge systems that run on mainframes.

The following are examples of database applications:

- computerized library systems
- automated teller machines
- flight reservation systems

Co-funded by the
Erasmus+ Programme
of the European Union

8. Database management system

Types of Databases

A database is any organized collection of data.

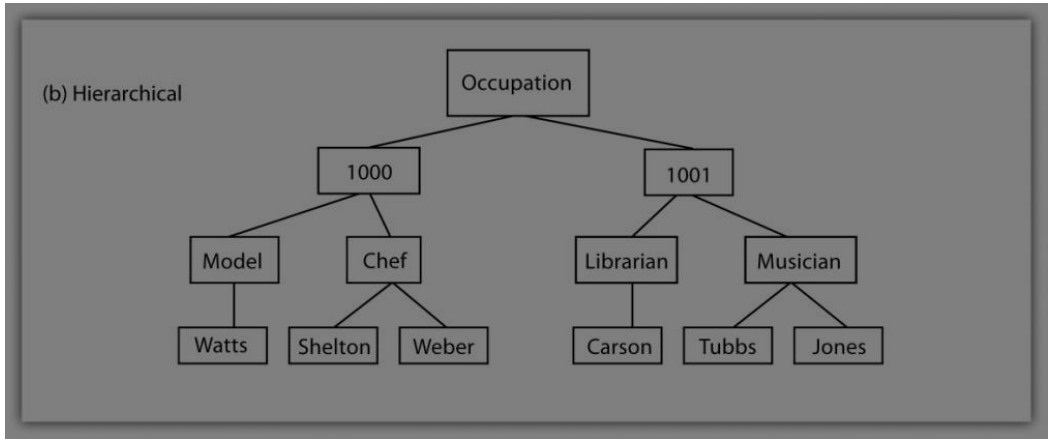
From a technical standpoint, DBMSs can differ widely.
Based upon their management of database structures.

- Hierarchical DBMS
- Network DBMS
- Relational DBMS
- Object oriented DBMS

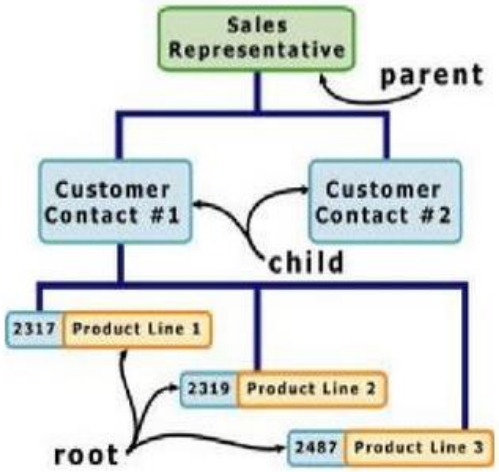
Co-funded by the
Erasmus+ Programme
of the European Union

Hierarchical DBMS

- Developed by IBM in 1968 and introduced in information management systems.
- Like a structure of a tree
- "parent-child" relationships
- Direct relationships exist between any two records that are stored consecutively
- one-to-many relationship
- No backward movement is possible/allowed



Co-funded by the
Erasmus+ Programme
of the European Union



- Parent record at the top of the pyramid is called the **root record**.
- A child record always has only one parent record to which it is linked, just like in a normal family tree.
- In contrast, a parent record may have more than one child record linked to it.
- Hierarchical databases work by moving from the top down.
- A record search is conducted by starting at the top of the pyramid and working down through the tree from parent to child until the appropriate child record is found.
- Furthermore, each child can also be a parent with children underneath it

Co-funded by the
Erasmus+ Programme
of the European Union

- The advantage - they can be accessed and updated rapidly because the tree-like structure and the relationships between records are defined in advance.
- The disadvantage - each child in the tree may have only one parent, and relationships
- Linkages between children are not permitted,
- Hierarchical databases are so rigid in their design that adding a new field or record requires that the entire database be redefined.

Co-funded by the Erasmus+ Programme of the European Union

Network DBMS

- **Network databases** are similar to hierarchical databases by also having a hierarchical structure.
- Key differences - instead of looking like an upside-down tree, a network database looks more like a cobweb or interconnected network of records.
- In network databases, children are called **members** and parents are called **owners**.
- The most important difference is that each child or member can have more than one parent (or owner)

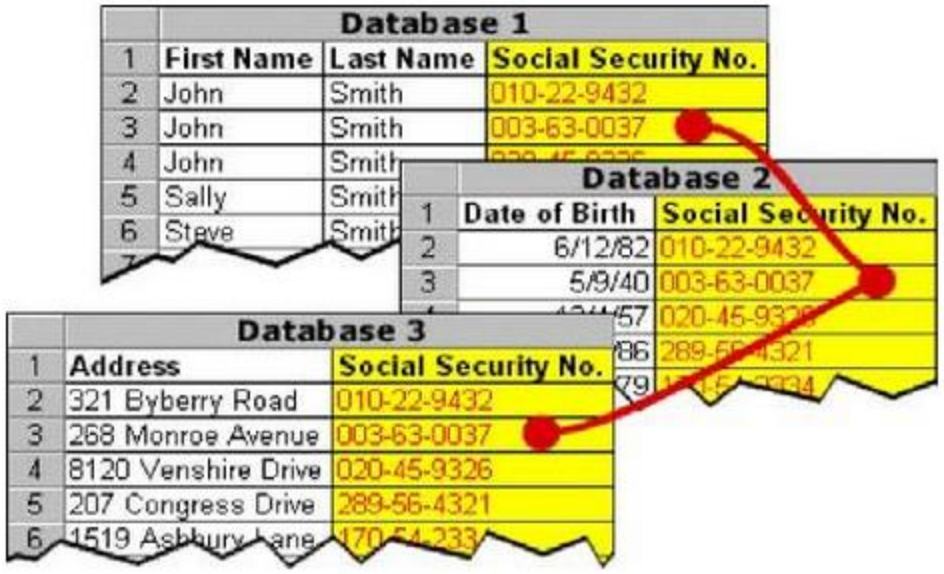
Co-funded by the
Erasmus+ Programme
of the European Union

- Network DBMS - relationships among data in the database are many-to-many
- Structure of a network database is extremely complicated because of these many-to-many relationships
- One record can be used as a key of the entire database
- Since more connections can be made between different types of data, network databases are considered more flexible.

- Two limitations –
  - Similar to hierarchical databases, network databases must be defined in advance.
  - There is also a limit to the number of connections that can be made between records.

Relational DBMS

- In **relational databases**, the relationship between data files is relational, not hierarchical.
- Hierarchical and network databases require the user to pass down through a hierarchy in order to access needed data.
- Relational databases connect data in different files by using common data elements or a key field.



Database 1

	First Name	Last Name	Social Security No.
2	John	Smith	010-22-9432
3	John	Smith	003-63-0037
4	John	Smith	
5	Sally	Smith	
6	Steve	Smith	

Database 2

	Date of Birth	Social Security No.
2	6/12/82	010-22-9432
3	5/9/40	003-63-0037

Database 3

	Address	Social Security No.
2	321 Byberry Road	010-22-9432
3	268 Monroe Avenue	003-63-0037
4	8120 Venshire Drive	020-45-9326
5	207 Congress Drive	289-56-4321
6	1519 Ashbury Lane	470-54-233

- Data is stored in different tables, each having a key field that uniquely identifies each row.
- Relational databases are more flexible than either the hierarchical or network database structures.
- In relational databases, tables or files filled with data are called **relations**, **tuples** designates a row or record, and columns are referred to as **attributes** or fields.

Co-funded by the
Erasmus+ Programme
of the European Union

- Relational databases work on the principle that each table has a key field that uniquely identifies each row, and that these key fields can be used to connect one table of data to another.
- Thus, one table might have a row consisting of a customer account number as the key field along with address and telephone number. The customer account number in this table could be linked to another table of data that also includes customer account number (a key field), but in this case, contains information about product returns, including an item number (another key field). This key field can be linked to another table that contains item numbers and other product information such as production location, color, quality control person, and other data. Therefore, using this database, customer information can be linked to specific product information.

- The relational database has become quite popular for two major reasons.
  - First, relational databases can be used with little or no training.
  - Second, database entries can be modified without redefining the entire structure.
- The downside of using a relational database is that searching for data can take more time.

Object-oriented **Databases (OODBMS)**

- Able to handle many new data types, including graphics, photographs, audio, and video
- **object-oriented databases** represent a significant advance over their other database cousins.
- Hierarchical and network databases are all designed to handle structured data; that is, data that fits nicely into fields, rows, and columns.
- They are useful for handling small snippets of information such as names, addresses, zip codes, product numbers, and any kind of statistic or number you can think of.
- On the other hand, an object-oriented database can be used to store data from a variety of media sources, such as photographs and text, and produce work, as output, in a multimedia format.

Co-funded by the Erasmus+ Programme of the European Union

Co-funded by the
Erasmus+ Programme
of the European Union

Disadvantages

- First, they are more costly to develop.
- Second, most organizations are reluctant to abandon or convert from those databases that they have already invested money in developing and implementing.

Advantages

- The ability to mix and match reusable objects provides incredible multimedia capability.
- Healthcare organizations, for example, can store, track, and recall CAT scans, X-rays, electrocardiograms and many other forms of crucial data.

Co-funded by the
Erasmus+ Programme
of the European Union

8. Database management system

Advantages of DBMS

It is important to design the database in such a way that:

A specific item can be reached easily

(maximum guarantee that the desired record will be reached)

- **The database can respond to the user's different questions/query easily**

  (necessary relationships are provided)
- **The database occupies minimum storage space**

  (choosing data types and how to express a certain concept is important)
- **The database contains no unnecessary data**

  (storing the gross salary is enough, the net salary can be calculated from the gross salary)
- **Data can be added and updated easily without causing mistakes**

  (no data redundancy)

Co-funded by the
Erasmus+ Programme
of the European Union

8. Database management system

Significance of DBMS

Data independence:

- Independent application programs
- Independent data representation and storage.

Efficient data access:

- Sophisticated techniques to store and retrieve data efficiently.

Data integrity and security:

- DBMS enforce integrity constraints on the data.
- Enforce access controls

Co-funded by the
Erasmus+ Programme
of the European Union

8. Database management system

Significance of DBMS

Data administration:

- Multiuser interface – internet/intranet
- Minimize redundancy
- Efficient retrieval

Concurrent access and crash recovery:

- Concurrent access to the data.
- DBMS protects users from the effects of system failures.

Reduced application development time:

- DBMS supports many important functions common to many applications
- Facilitates quick development of applications.

8. Database management system

Selecting a database system: Need Analysis

The needs analysis process will be specific to your organization/application but, at a minimum, should answer the following questions:

- How many records we will warehouse and for how long?
- Who will be using the database and what tasks will they perform?
- How often will the data be modified? Who will make these modifications?
- Who will be providing IT support for the database?
- What hardware is available? Is there a budget for purchasing additional hardware?
- Who will be responsible for maintaining the data?
- Will data access be offered over the Internet? If so, what level of access should be supported?

Co-funded by the Erasmus+ Programme of the European Union

Co-funded by the
Erasmus+ Programme
of the European Union

8. Database management system

Selecting a Database Management System

Database management systems (or DBMSs) can be divided into two categories -- **desktop databases** and **server databases**.

- Generally speaking, desktop databases are oriented toward single-user applications and reside on standard personal computers (hence the term desktop).

- Server databases contain mechanisms to ensure the reliability and consistency of data and are geared toward multi-user applications.

9. Datatype

Salary = Rs. 49,567.00

Dense evergreen forest

Co-funded by the Erasmus+ Programme of the European Union

Mark = 79

DOB: 02 Feb 2000

Name = Dr. Ramesh

Time = 11.30 am

pH value: 0.52

Temp = 29.5 deg C

Ans = YES/NO

Sl. No. 1- 1000

About Pondicherry University: Pondicherry University, established under an Act of Parliament in the year 1985, has grown from strength to strength in all possible ways all these years and has become a place on the educational hub of the country. 15 Schools, 37 Departments and 10 Centres offering 175 PG & Research programmes are within its fold and housed in the 800-acre sprawling Wi-Fi-enabled vibrant campus, which is just 12 kms. away from the Puducherry town. It has all the state-of-the-art facilities in all the Schools and Departments paving the way for the students to have a student-friendly, result-oriented academic environment with green ambience.

Co-funded by the
Erasmus+ Programme
of the European Union

Data types

- A data type in a programming language is a set of data with values having predefined characteristics.
- The *data type* of a programming element refers to what kind of data it can hold and how that data is stored.
- Data types apply to all values that can be stored in computer memory or participate in the evaluation of an expression.

Microsoft Access Data Types

Data type	Description
Text	Use for text or combinations of text and numbers. 255 characters maximum
Memo	Memo is used for larger amounts of text. Stores up to 65,536 characters. **Note:** You cannot sort a memo field. However, they are searchable
Byte	Allows whole numbers from 0 to 255
Integer	Allows whole numbers between -32,768 and 32,767
Long	Allows whole numbers between -2,147,483,648 and 2,147,483,647
Single	Single precision floating-point. Will handle most decimals
Double	Double precision floating-point. Will handle most decimals
Currency	Use for currency. Holds up to 15 digits of whole dollars, plus 4 decimal places. **Tip:** You can choose which country's currency to use
AutoNumber	AutoNumber fields automatically give each record its own number, usually starting at 1
Date/Time	Use for dates and times
Yes/No	A logical field can be displayed as Yes/No, True/False, or On/Off. In code, use the constants True and False (equivalent to -1 and 0). **Note:** Null values are not allowed in Yes/No fields
Ole Object	Can store pictures, audio, video, or other BLOBs (Binary Large OBjects)
Hyperlink	Contain links to other files, including web pages
Lookup Wizard	Let you type a list of options, which can then be chosen from a drop-down list

Co-funded by the Erasmus+ Programme of the European Union

Co-funded by the
Erasmus+ Programme
of the European Union

SQL Server data types

When you create a feature class or table in ArcGIS, there are 11 different data types available for each column. These types are mapped to SQL Server types in the table below.

ArcGIS data type	SQL Server data type
OBJECTID	INT(4)*
SHORT INTEGER	SMALLINT(2)
LONG INTEGER	INT(4)
FLOAT	REAL
DOUBLE	DOUBLE**
TEXT	VARCHAR(n)
DATE	DATETIME
BLOB	IMAGE
GUID	UNIQUEIDENTIFIER(16)
GEOMETRY	IMAGE
RASTER	IMAGE

*NOT NULL

**If scale is set to greater than 0 in ArcGIS, the SQL Server data type will be NUMERIC.

Co-funded by the
Erasmus+ Programme
of the European Union

Data types in Microsoft Access

The following list summarizes all the field data types available in Microsoft Access, their uses, and their storage sizes.

Text

- Use for text or combinations of text and numbers, such as addresses, or for numbers that do not require calculations, such as phone numbers, postal codes.
- Stores up to 255 characters. The FieldSize property controls the maximum number of characters that can be entered.

Memo

- Use for lengthy text and numbers, such as notes or descriptions.
- Stores up to 65,536 characters.

Number

- Use for data to be included in mathematical calculations, except calculations involving money (use Currency type).

Co-funded by the Erasmus+ Programme of the European Union

Data types in Microsoft Access contd.

Date/Time

- Use for dates and times.
- Stores 8 bytes.

Currency

- Use for currency values and to prevent rounding off during calculations.
- Stores 8 bytes.

AutoNumber

- Use for unique sequential (incrementing by 1) or random numbers that are automatically inserted when a record is added.
- Stores 4 bytes; stores 16 bytes for Replication ID (GUID).

Yes/No

- Use for data that can be only one of two possible values, such as Yes/No, True/False, On/Off. Null values are not allowed.
- Stores 1 bit.

Co-funded by the Erasmus+ Programme of the European Union

Data types in Microsoft Access contd.

Object Linking Embedding (OLE) Object

- Use for OLE objects (such as Microsoft Word documents, Microsoft Excel spreadsheets, pictures, sounds, or other binary data) that were created in other programs using the OLE protocol.
- Stores up to 1 gigabyte (limited by disk space).

Hyperlink

- Use for hyperlinks. A hyperlink can be a University Naming Conversion (UNC) path or a Uniform Resource Locator (URL)
- Stores up to 64,000 characters.

9. What is a database query and retrieval?

A database query is a piece of code (a query) that is sent to a database in order to get information back from the database. It is used as the way of retrieving the information from database.

The term 'query' means to **search**, to **question**, or to **find**. When you query a database, you are searching for information in the database. Different query languages exist for different type of databases. MS Access uses SQL, which stands for Structured Query Language. MS Access contains Tables, Forms, and Queries. The Forms are used to enter or display the data, the Tables where the data is saved, and the queries are used to search for specific data.

Co-funded by the
Erasmus+ Programme
of the European Union

What is a database query and retrieval? Contd..

An example of a query can be this:

Syntax

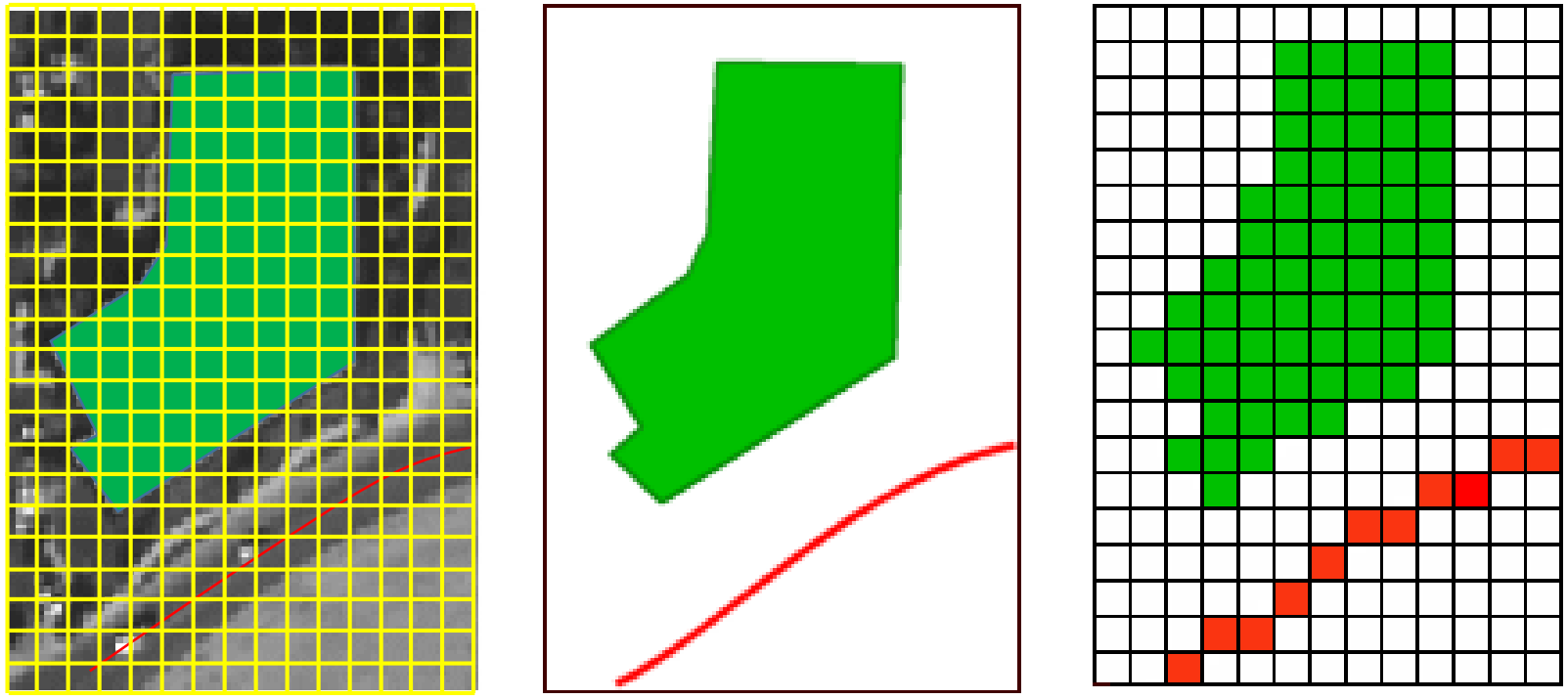
```
SELECT * FROM STUD_PU WHERE BG = ‘O+’
SELECT * FROM SOIL WHERE PH = 7
```

URGENT

Co-funded by the Erasmus+ Programme of the European Union

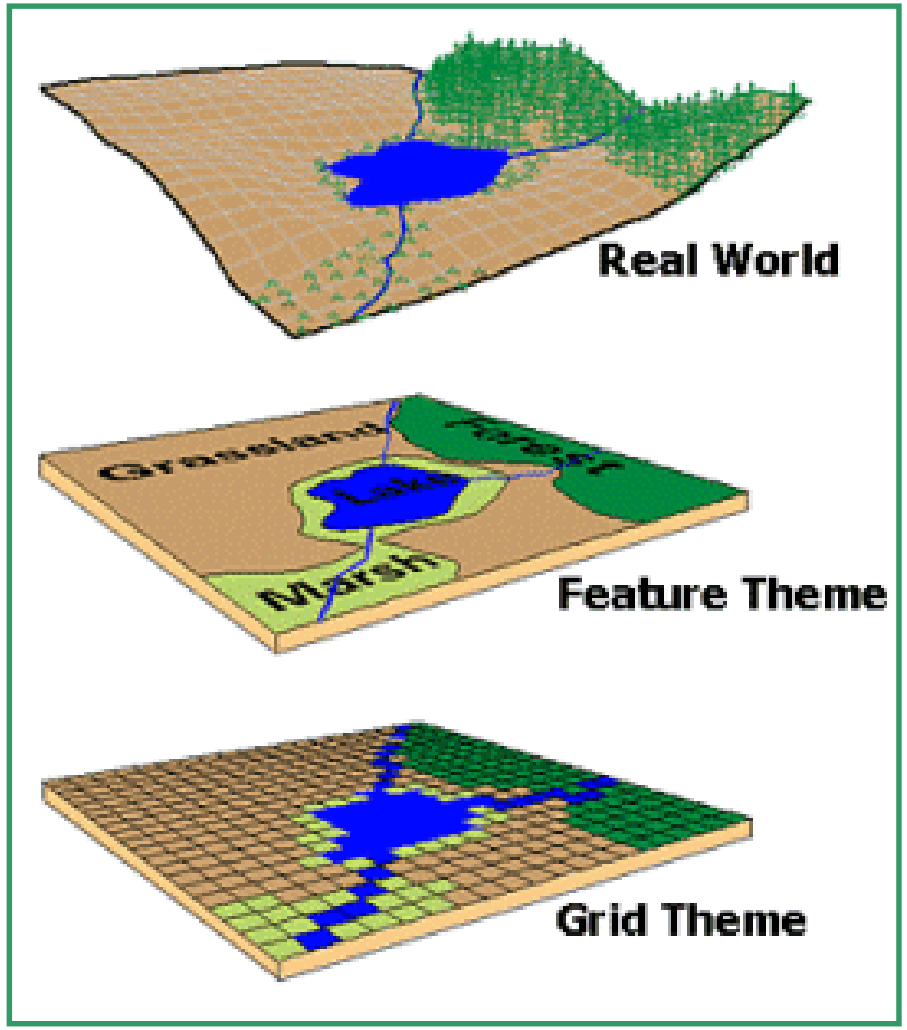
Co-funded by the
Erasmus+ Programme
of the European Union

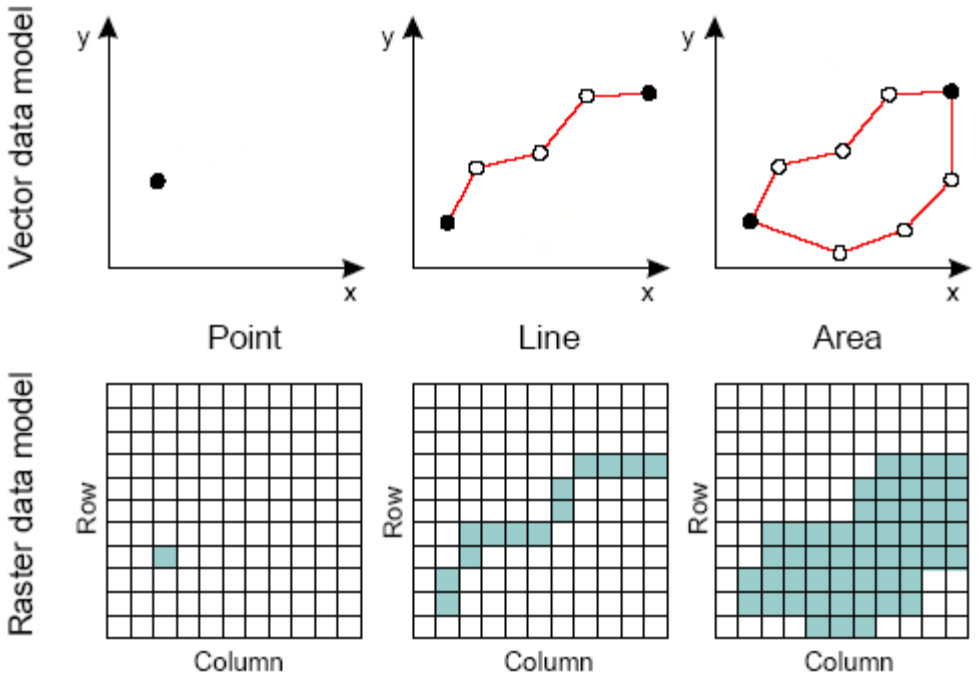
10. SPAITAL AND NON SPATIAL DATA



Co-funded by the
Erasmus+ Programme
of the European Union

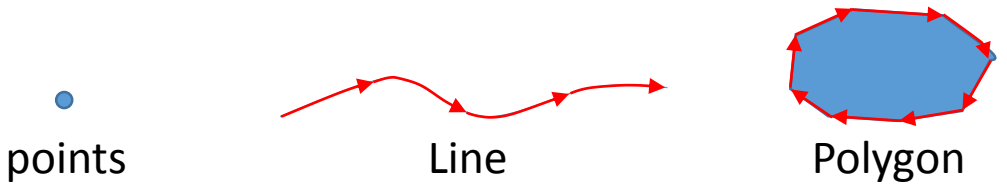
Raster vs Vector Data





Co-funded by the
Erasmus+ Programme
of the European Union

Vector Data



- Data are associated with points, lines, polygons or boundaries enclosing areas
- Points are located by coordinates
- Lines are described by a series of connecting vectors (line segments described by the coordinates of the start of the vector, its direction, and magnitude or length).
- Areas or polygons are described by a series of vectors enclosing the area.
- Any number of factors or attributes can be associated with a point, line, or polygon.
- Vector file stores
  - ❖ location information
  - ❖ Attribute information
  - ❖ information needed to link positional data with their attributes.

Co-funded by the
Erasmus+ Programme
of the European Union

Characteristics of vector data:

Points

- Zero-dimensional points are used for geographical features that can best be expressed by a single point reference — in other words, by simple location.

  Examples include wells, peaks, features of interest, and trailheads.

- Points convey the least amount of information.

- Points represent Areas - scale.

  For example, cities on a map of the world might be represented by points rather than polygons.

- No measurements are possible with point features.

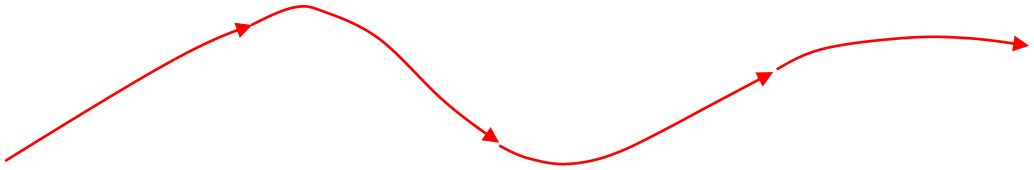
Co-funded by the
Erasmus+ Programme
of the European Union

Lines

One-dimensional lines or polylines are used for linear features such as **rivers, roads, railroads, trails**, and **topographic lines**.

Again, as with point features, linear features displayed at a small scale will be represented as linear features rather than as a polygon.

Line features can measure distance.



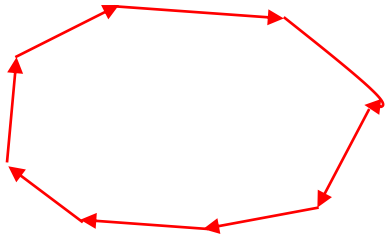
Co-funded by the
Erasmus+ Programme
of the European Union

Polygons

Two-dimensional polygons are used for geographical features that cover a particular area of the earth's surface. Such features may include **lakes, park boundaries, buildings, city boundaries**, or **land uses**.

Polygons convey the most amount of information of the file types.

Polygon features can measure **perimeter** and **area**.



Co-funded by the
Erasmus+ Programme
of the European Union

Vector data sets

Each of these geometries is linked to a row in a database that describes their **attributes**.

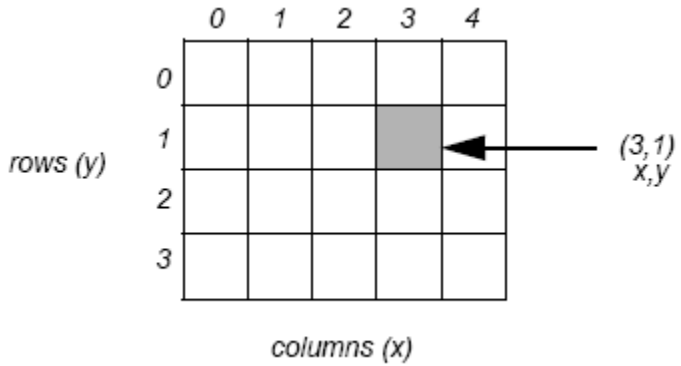
For example, a database that describes a water body may contain a water body's depth, water quality, pollution level.

Vector data can also be used to represent continuously varying phenomena.

Contour lines and triangulated irregular networks (TIN) are used to represent elevation or other continuously changing values.

Co-funded by the
Erasmus+ Programme
of the European Union

Spatial Data Structure



File coordinate system

Raster Data

- Data are divided into cells, pixels, or elements
- Cells are organized in arrays
- Each cell has a single value
- Row and Column Numbers are used to identify the location of the cell within the array.
- Perhaps the most common example of raster data is a digital image.

Co-funded by the
Erasmus+ Programme
of the European Union

Characteristics of raster data:

- A raster data type is, in essence, any type of digital image represented by **reducible** and **enlargeable grids**.
- Raster data type consists of rows and columns of cells, with each cell storing a **single value**.
- Raster data can be images (raster images) with each pixel (or cell) containing a **color value**.
- Additional values recorded for each cell may be a **discrete value**, such as land use, a **continuous value**, such as temperature, or a **null value** if no data is available.
- Raster data is stored in various **formats**; from a standard file-based structure of TIF, JPEG, etc. to binary large object (BLOB) data stored directly in a relational database management system (RDBMS) similar to other vector-based feature classes

Co-funded by the
Erasmus+ Programme
of the European Union

Advantages and disadvantages - Vector

- The main advantage of the vector data format is that it allows precise representation of points, boundaries, and linear features.

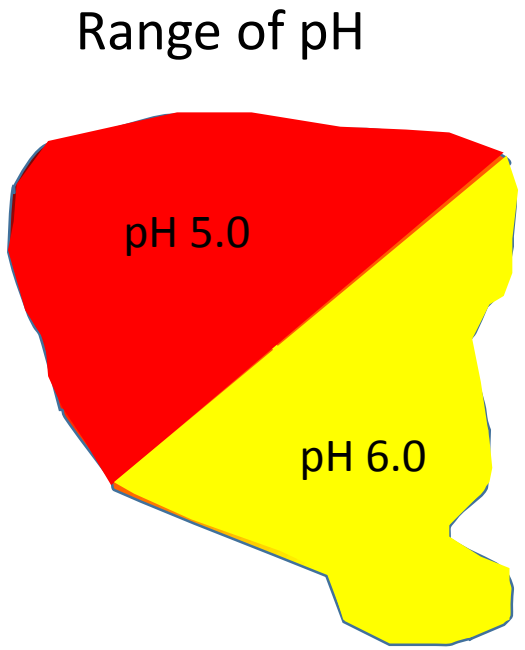
- This makes it particularly useful for analysis tasks that require accurate positioning,
  for example in engineering or cadastral boundary databases.

- It is also possible in a vector-based GIS to define the spatial relationship (i.e., the connectivity and adjacency) between coverage features. This aspect of GIS is known as topology, and is important for such purposes as network analysis
  for example to find an optimal path between two nodes in a complex transport network.

Co-funded by the
Erasmus+ Programme
of the European Union

Advantages and disadvantages – Vector contd..

- The main disadvantage of vector data is that the boundaries of the resulting map polygons are discrete (enclosed by well-defined boundary lines), whereas in reality the map polygons may represent continuous gradation or gradual change, as in soil maps.



Co-funded by the
Erasmus+ Programme
of the European Union

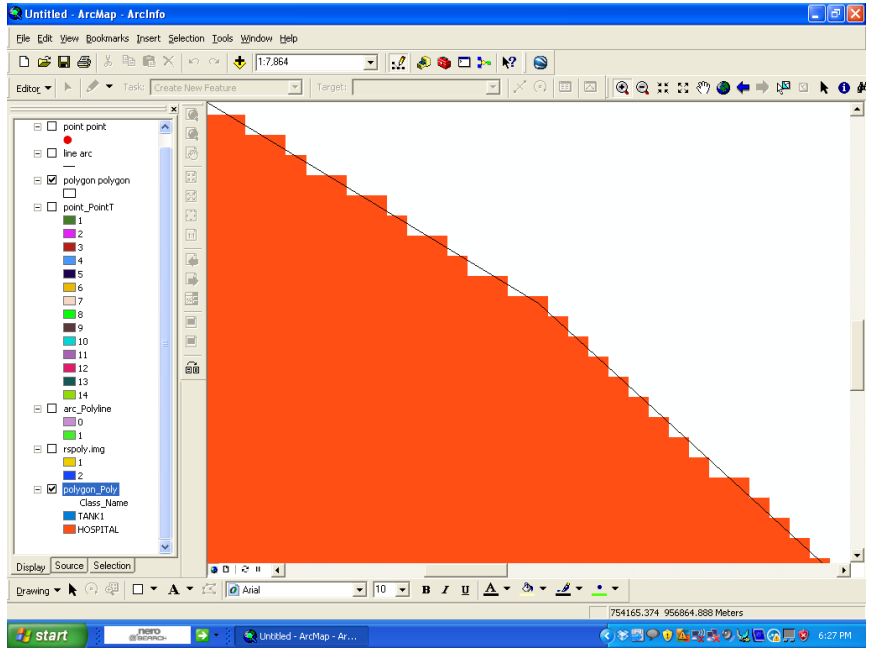
Advantages and disadvantages – Raster

- Grid Data Models use a raster matrix (a grid of image cells) to represent information. Grid data are also known as raster data.

- The resolution (visual definition) of the raster depends on its pixel (cell) size. In other words, pixel resolution represents the size of the ground area covered by each pixel in the image.

- The smaller the cell size, the higher the resolution.

- The raster data model is, therefore, good for representing indistinct boundaries, such as thematic information on soil types, soil moisture, vegetation, ground temperatures, and so on.



Co-funded by the
Erasmus+ Programme
of the European Union

Advantages and disadvantages – Raster contd..

- Furthermore, as reconnaissance satellites and aerial surveys use raster-based scanners, the information (i.e., scanned images) can be directly incorporated into GIS programmes capable of working with raster data.

- However, the higher the grid resolution, the larger the data file is going to be. This is the main limitation of raster based GIS.

Co-funded by the
Erasmus+ Programme
of the European Union

11. Which data model is suitable?

- The question of which data model to use in GIS depends on the nature and objective of the GIS project.

- Primarily the model type will depend on the nature of the data. Issues of concern are
  - the volume of the data generated,
  - ease of analysis and accuracy.

- Generally, vector data sets are economical in terms of file size, and have a high level of positional precision, but are relatively difficult to use in mathematical computations.

- On the other hand, grid data sets tend to take up more file space and have a coarser resolution, but are easier to work with mathematically.

Co-funded by the
Erasmus+ Programme
of the European Union

12. Attribute data

- Attribute data are descriptions, measurements, and/or classifications of the geographic features.
- Information attached to each object in a layer
- Attribute (tabular) data is the descriptive data that GIS links to map features.
- Attributes are often stored in database files kept separately from the graphic portion of the map.
- GIS software packages maintain internal links tying each graphical map entity to its attribute information.

Co-funded by the
Erasmus+ Programme
of the European Union

L3. Hardware and Software requirements

- The effectiveness of Information System projects depend as much on the futuristic planning of hardware and software requirements, as on technicalities of its implementation.

- It is necessity to analyze the requirements keeping in view the following:
  1. Organization's goals and objectives
  2. Analysis of current practices and problems encountered and redundancies
  3. Organizational issues - management structure, resources and personnel
  4. Training
  5. Applications
  6. Base map requirements
  7. Conceptual data base design
  8. System integration
  9. Cost of GIS implementation – applications development, data conversion, hardware, software, resources, training, update and maintenance
  10. The Strategic Implementation Plan over a period of 3-5 years

Co-funded by the
Erasmus+ Programme
of the European Union

Hardware requirements

It is obvious that hardware and software requirements vary considerably depending on the tasks undertaken. The following minimum configuration allows installation of most modern Information system applications for work with small components. Recommended configurations are noted in parentheses for work with anything other than small drawings.

- 2.8 Ghz i7 true PC compatible
- 12 GB RAM
- 1200 x 1800 SVGA (Super Video Graphic Array) Display (1280 x 1024 or greater recommended).
- 1 TB hard disk free space (gigabytes of free space recommended).
- Windows 10, with most recent service pack, in standard 64-bit versions

Co-funded by the
Erasmus+ Programme
of the European Union

Hardware requirements contd..

- Recent Internet Explorer version plus most recent service pack.
- Microsoft's .NET Framework 2.0 or more recent.
- Internet Information Service (IIS) 5.1 or greater to operate IMS.

Memory Requirements for Large Projects

- For best performance it is recommended to install lots of RAM and having plenty of free disk space for temporary files.

Co-funded by the
Erasmus+ Programme
of the European Union

Software requirements

- The choice of software shall depend on the needs of the organization, functionality desired and the money available, and the period for which the planning is being done.
- One may need to make a comparison of costs and benefits (both of which keep changing rapidly) before making a final decision.

  - **Hydro GeoAnalyst** – Environmental data management and visualization software by Schlumberger Water Services.
  - **Autodesk** – Products include **MapGuide** and other products that interface with its flagship **AutoCAD** software package.
  - **Cadcorp** – Developers of GIS software and OpenGIS standard (e.g. Read/Write Open Source PostGIS database).
  - **Intergraph** – Products include **GeoMedia**, **GeoMedia Profesional**, **GeoMedia WebMap**, and add-on products for industry sectors, as well as photogrammetry.

Co-funded by the
Erasmus+ Programme
of the European Union

Software requirements contd..

- **ERDAS -Intergraph** – A proprietary GIS, Remote Sensing, and Photogrammetry software developed by ERDAS, Inc.
- **ESRI** – Products include **ArcView 3.x**, **ArcGIS**, **ArcSDE**, **ArcIMS**, and **ArcWeb** services.
- **IDRISI** – Proprietary GIS product developed by Clark Labs.
- **MapInfo** – Products include **MapInfo Professional** and **MapXtreme**. integrates GIS software, data and services.
- **MapPoint** – Proprietary GIS product developed by Microsoft.
- **GISNet – A** web-based GIS system developed by MRF Geosystems Corporation.
- **Caliper** – Products include **Maptitude**, **TransCAD** and **TransModeler**. Develops GIS and the only GIS for transportation.
- **Pictometry** – Proprietary software which allows oblique images to be draped with shapefiles. Many gov't applications (fire, 911, police, assessor) as well as commercial.
- **Black Coral Inc** — a leading edge product company developing geospatial collaboration capabilities that enable better outcomes for personnel and tactical teams operating in emergency response and military environments.

Co-funded by the
Erasmus+ Programme
of the European Union

14. Resource information extraction

What is the need for resource information?

- better management of natural resources
- present and past status
- conservation
- sustainable utilization
- solution to resource problems

Co-funded by the
Erasmus+ Programme
of the European Union

14. Resource information extraction

Information extraction from Airborne and Space borne data

- Airborne data – taking photograph of the ground from an elevated position – air photo (1-2 km from ground)

- Space borne data – sensing the ground from far away distance (700 – 800 km from ground) – Satellite Remote Sensing

15. Airborne data

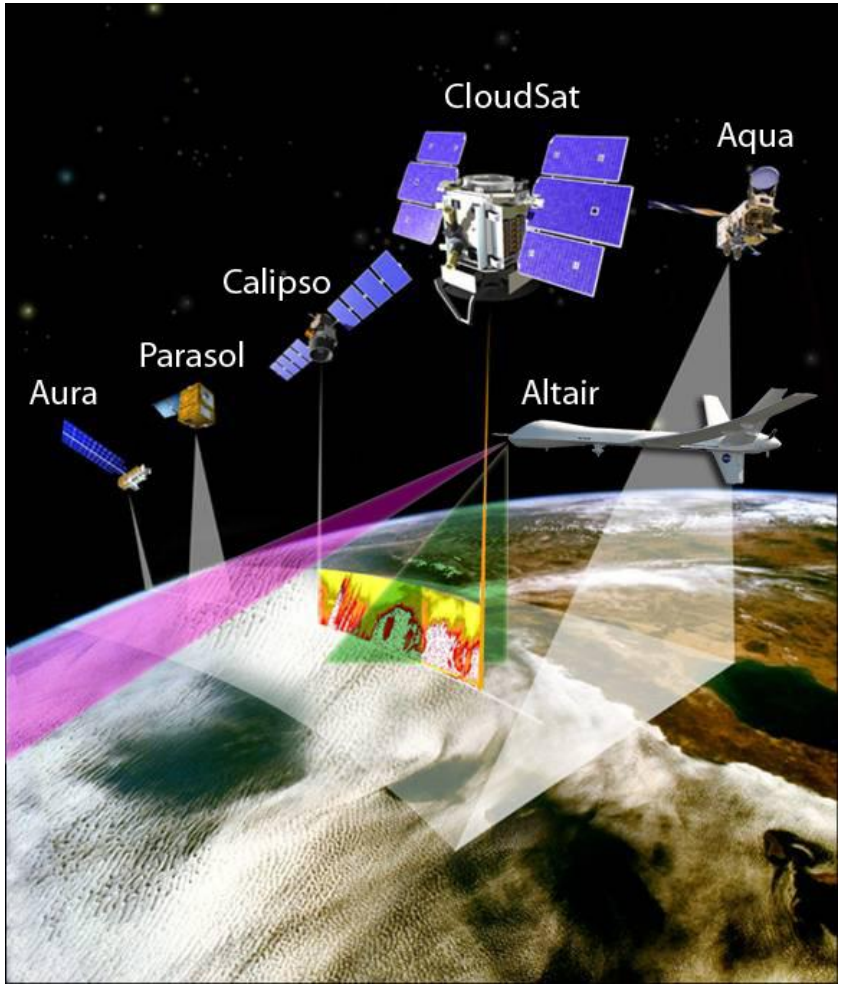
Airborne data – Aerial/air photos

Co-funded by the
Erasmus+ Programme
of the European Union

Co-funded by the
Erasmus+ Programme
of the European Union

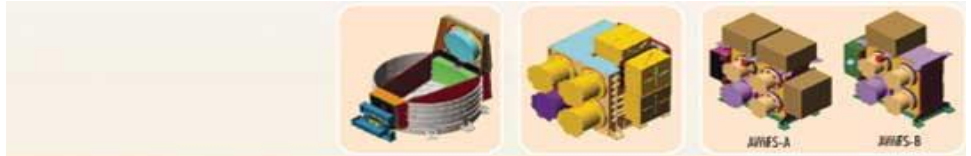
16. Space borne data

Space borne data – Remote Sensing images



- Satellites take picture of the earth surface from 700-800 km - picture from space
- Continuously take picture
- Sun-synchronous
- Pixel size varies from 0.6m to 1 km
- Micro level to global cover
- Information extraction
  - Image interpretation
  - Classification
    - Visual
    - Digital
      - Unsupervised
      - Supervised
      - Expert
      - Object oriented
      - Head-up

16. Space borne data



PAYLOADS	LISS-4	LISS-3	AWiFS
Spatial Resolution (m)	5.8	23.5	56
Swath (km)	23.9 (MX mode) 70.3 (PAN mode)	141	740
Spectral Bands (micron)	0.52-0.59 0.62-0.68 0.77-0.86	0.52-0.59 0.62-0.68 0.77-0.86 1.55-1.70	0.52-0.59 0.62-0.68 0.77-0.86 1.55-1.70
Quantisation (bits)	7	7	10
Square Wave Response (at Nyquist)	>0.20	B2>0.40 B3>0.40 B4>0.35 B5>0.20	B2>0.40 B3>0.40 B4>0.35 B5>0.20
Power (W)	216	70	114
Weight (kg)	169.5	106.1	103.6
Data Rate (MBPS)	105	52.5	52.5

IRS P6 (RESOURCESAT- 1)

Co-funded by the
Erasmus+ Programme
of the European Union

17. Classification

Supervised training

- It is controlled by the analyst
- Analysts selects pixels that represent patterns or land cover features
- Knowledge of data and the classes desired, is required before classification
- By identifying patterns, you can instruct the computer system to identify pixels with similar characteristics

Unsupervised training

- It is more computer automated
- Requires very few parameters from the user/analyst
- Based on the parameters, computer makes cluster of pixels with similar spectral characteristics
- These clusters do not necessarily correspond to directly meaningful classes
- This method is used when less is known about the data before classification

Co-funded by the
Erasmus+ Programme
of the European Union

18. Positioning System

Global navigation satellite system (GNSS)

- NAVSTAR Global Positioning System (GPS) -USA 32 medium earth orbit satellites operations since 1978
- GLONASS - Russian since 1995
- Galilio - European Union - 30 MEO satelites some are operations since 2014 , expected to be in full service till 2020
- Beidou-2 - China, 30 MEO satellites, expanding current regional service into global by 2020

Regional satellite navigation system

- Beidou-1 - China 16 satellites , covering Asia pacific since DEC 2012
- IRNSS - India 7 satellites covering India , 7 satellites launched , soon it will be operational.
- QZSS - Japan 3 satellites system for Japan, first satellite launched in 2010

Co-funded by the
Erasmus+ Programme
of the European Union

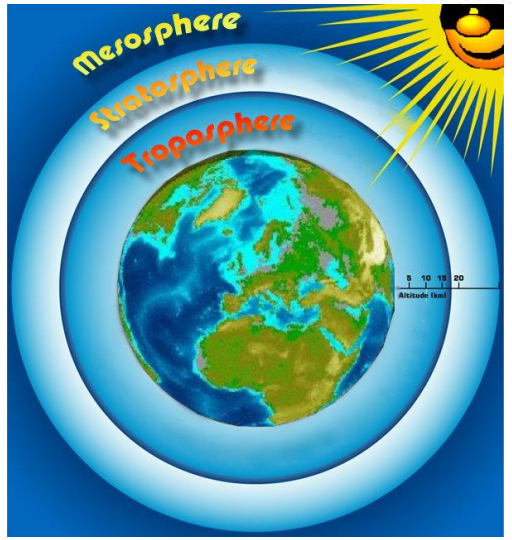
19. Topographical maps – Information extraction

- Multi-thematic maps
- Provides land cover and land use
- Information extraction is by means of manual drawing and digitization
- Digital topographic maps are very handy - easy to extract thematic information
- Details available,
  - Transport network (All types of roads and path)
  - Drainage networks (Rivers, various drainage orders)
  - Settlements
  - Post office, Police station, PWD offices, Telegraph office,
  - Power lines, Telephone lines,
  - Forest area – reserved forest, reserved lands, types of forest
  - Wells, bore wells, tanks, reservoirs
  - Agriculture – crop type, and so on

20. Information extraction from climatic data

Weather - BACKGROUND

Co-funded by the Erasmus+ Programme of the European Union

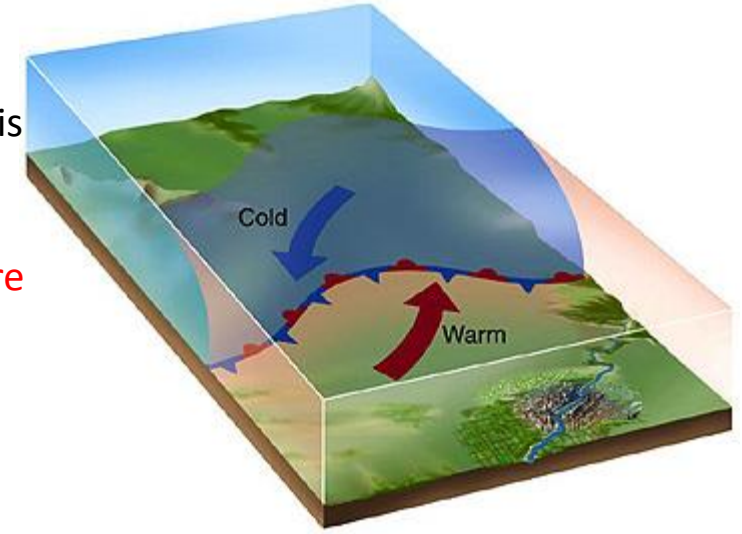


When you talk about weather, you are really talking about the air.

How hot or cold is the air?
How much moisture is in it?
How fast is the air moving?
How heavily does it press on the earth?

The Oxford Dictionary defines weather as *the condition of atmosphere at a certain place and time, with reference to the presence or absence of sunshine, rain, wind, etc*.

The Sun is basically the 'engine' that fuels the world's weather. The equator is the area on Earth which faces closest to the Sun is heated the most while the two poles receive the least heat.



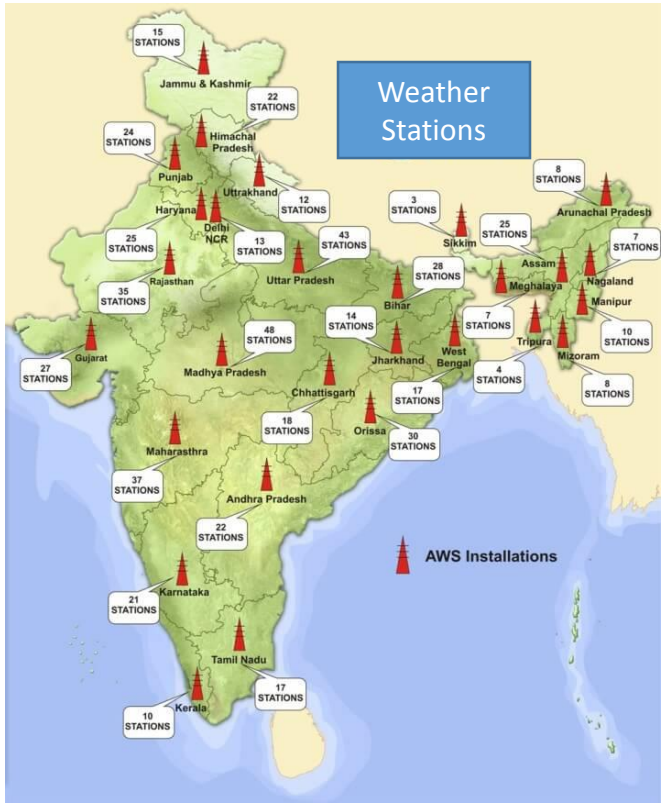
The atmosphere is constantly trying to even out the temperature irregularities by carrying warm air to places which are cold and cold air to places which are warm.

In fact, weather plays a big part in our lives and affects many of the things that we do and also affects all things on Earth.

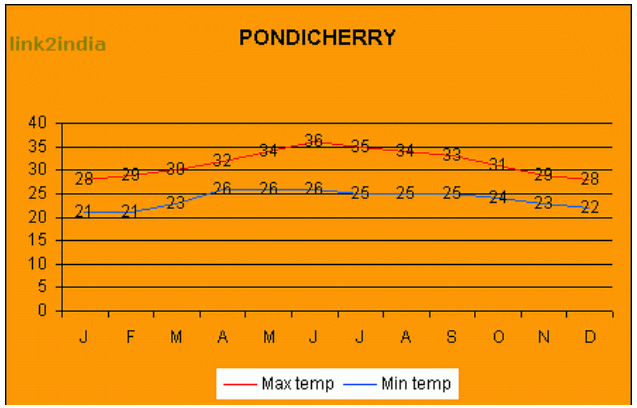
20. Information extraction from climatic data

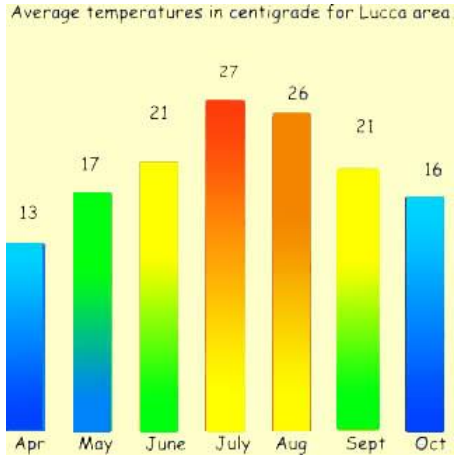
Information extraction from rainfall and temperature data

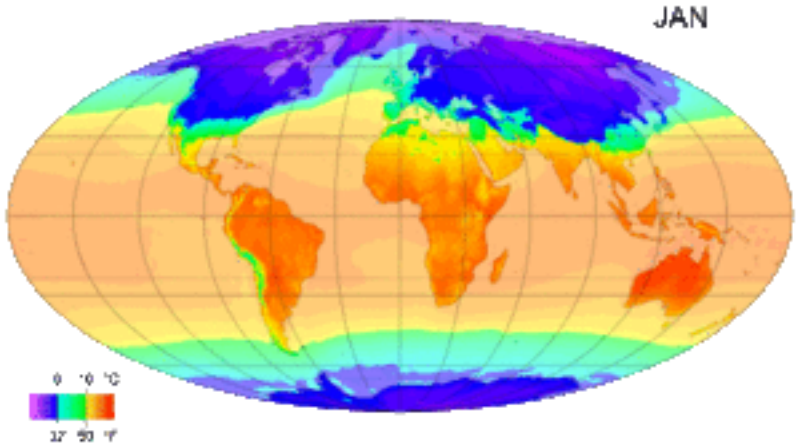
Air Temperature



Location of weather stations

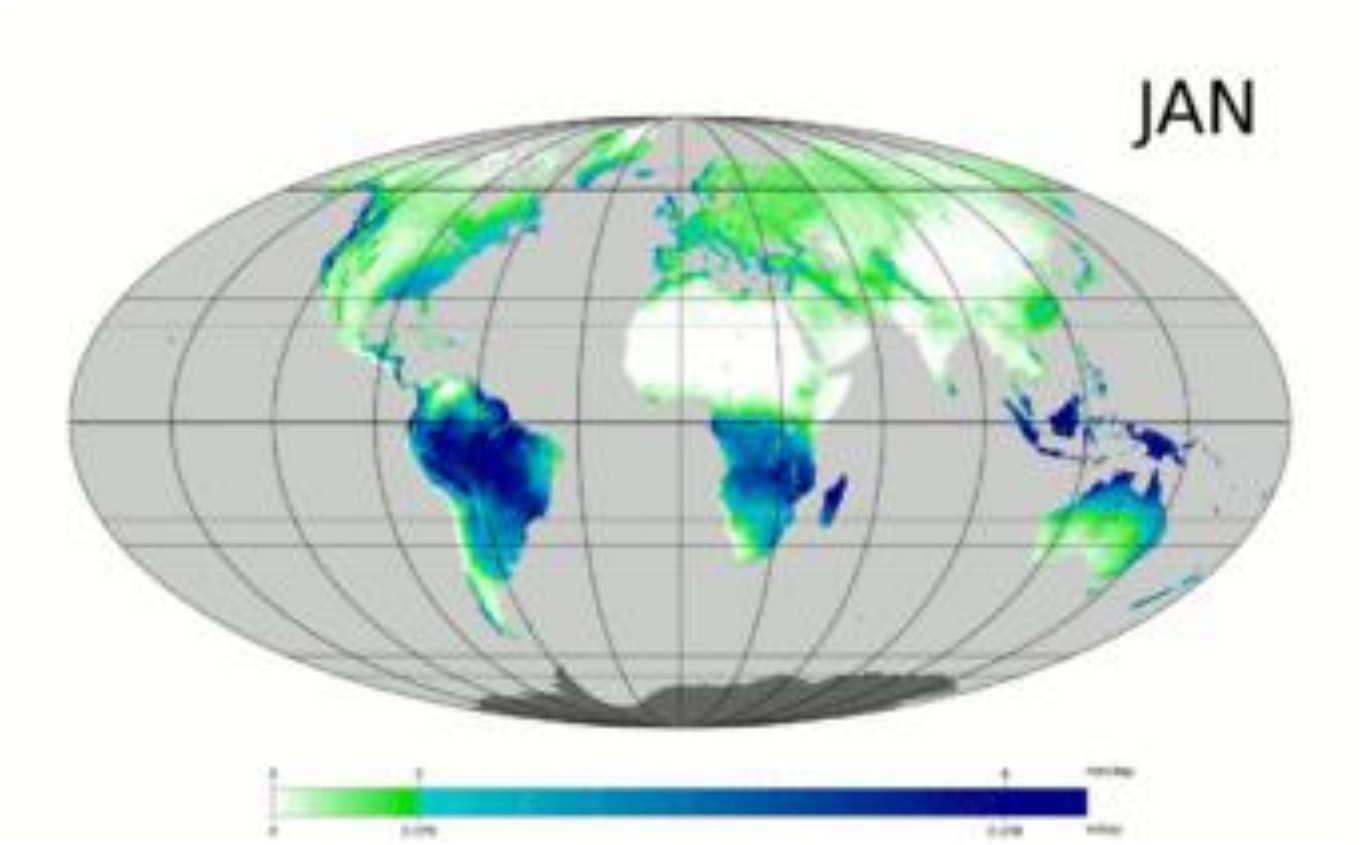






A map of global long term monthly average surface air temperatures

21. Precipitation Information extraction



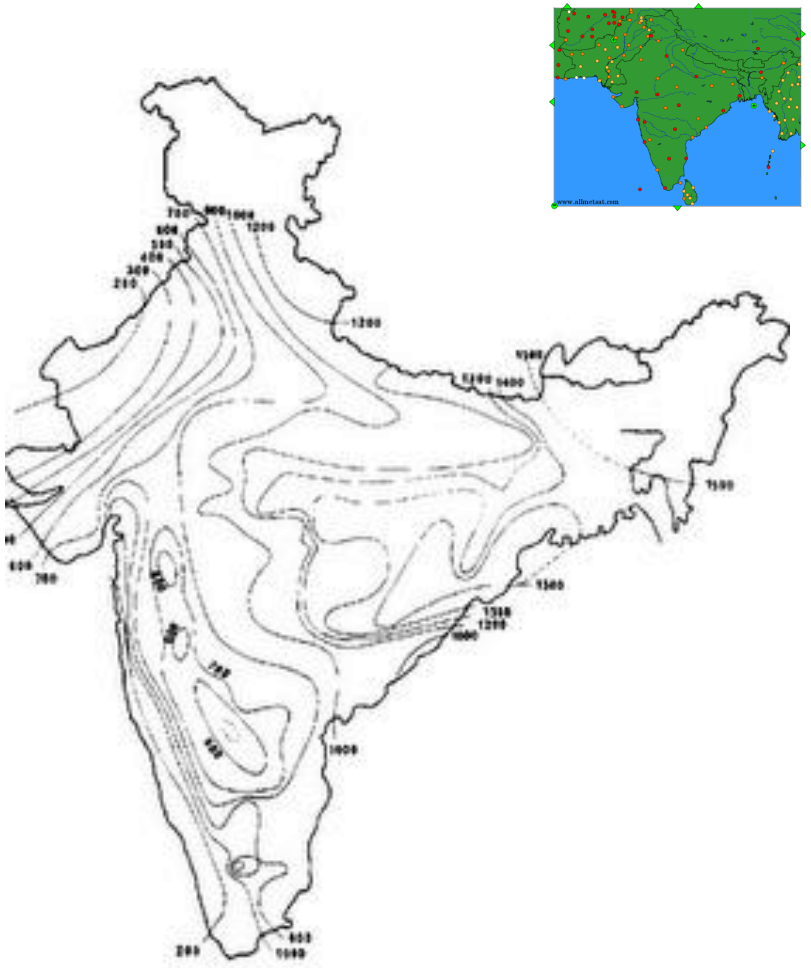
Long-term mean precipitation by month

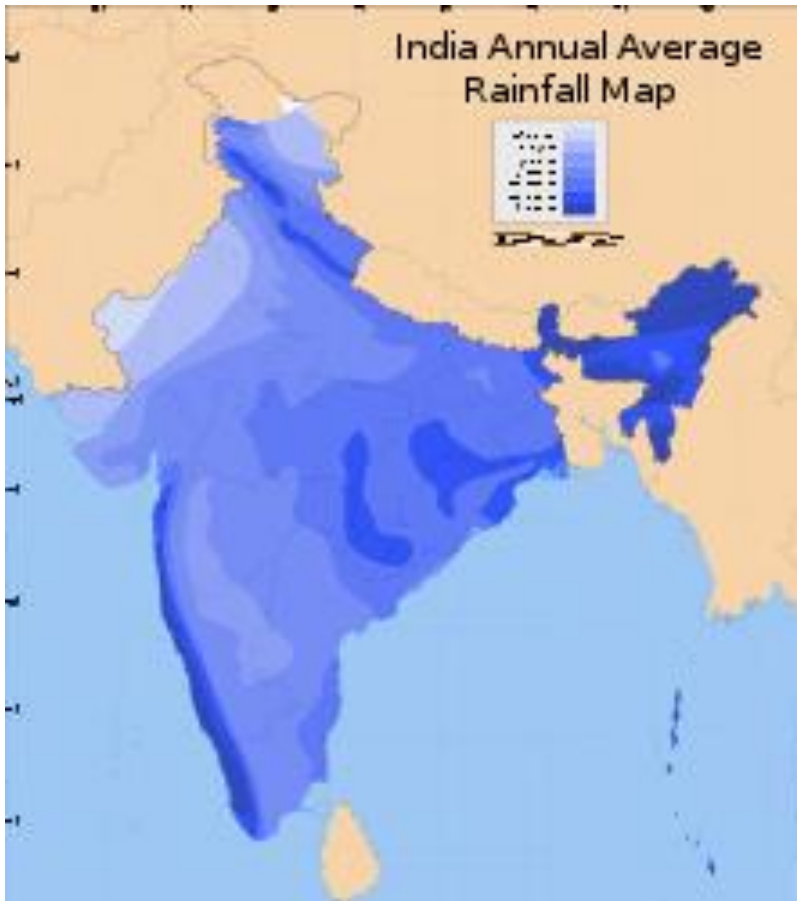
Co-funded by the Erasmus+ Programme of the European Union

Co-funded by the
Erasmus+ Programme
of the European Union

21. Information extraction from Rain guage stations

Rainfall Isohyet: A line drawn through geographic points recording equal amounts of precipitation for a specified period or for a particular storm

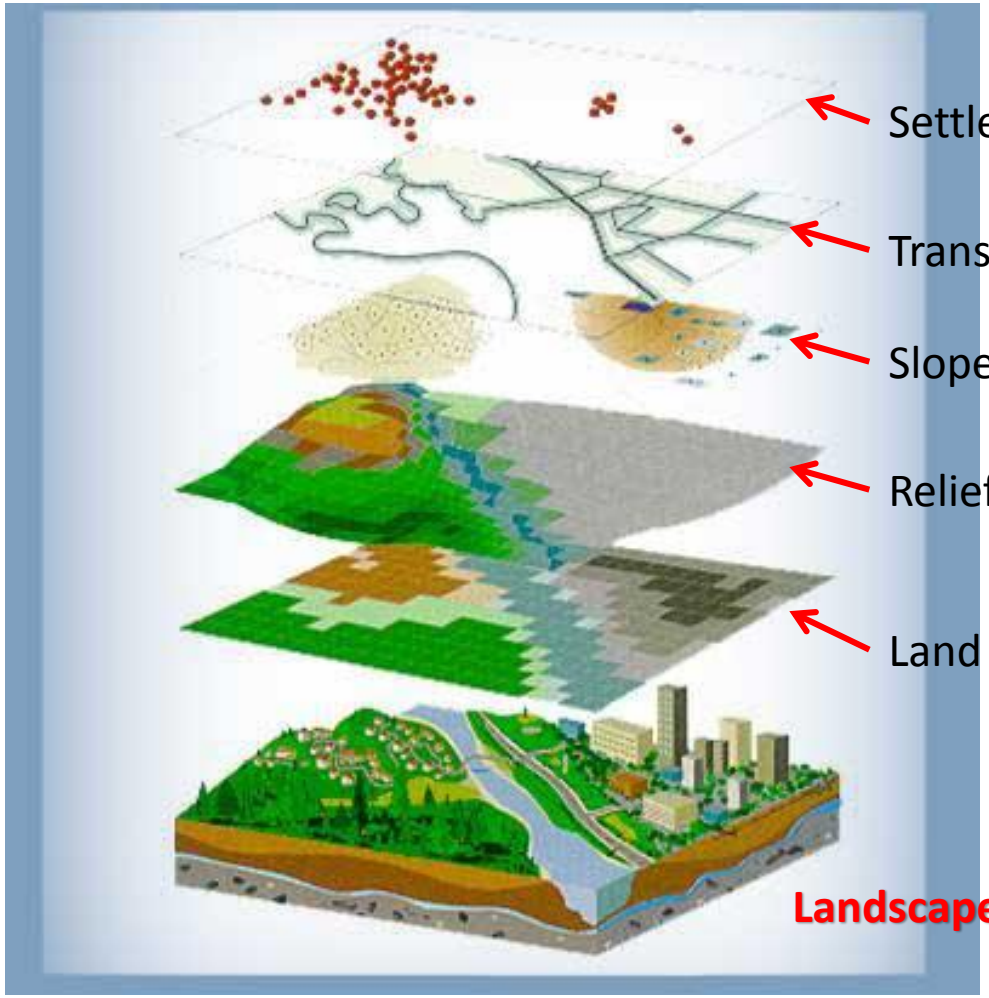




Co-funded by the
Erasmus+ Programme
of the European Union

22. Spatial database creation

Spatial database creation: Thematic maps



THEMATIC MAPS

Thematic map is a also a map prepared based on single theme

Co-funded by the Erasmus+ Programme of the European Union

Thematic map

- A **thematic map** is a type of map especially designed to show a particular theme connected with a specific geographic area.

- These maps can portray physical, social, political, cultural, economic, sociological, agricultural, or any other aspects of a city, state, region, nation, or continent

- Thematic maps are sometimes referred to as **graphic essays** that portray spatial variations and interrelationships of geographical distributions. Location, of course, is important to provide a reference base.

Co-funded by the
Erasmus+ Programme
of the European Union

23. Cartography

- Cartography is the art, science, and technology of making maps
- A Map is a geographic representation of a portion of the Earth's surface drawn to scale as seen from above. It's uses colors, symbols and labels to represent features found on the ground.
- A spatial model of the real world, but differentiated from it by: abstraction, focus, simplification, symbolization scale, projection, and purpose

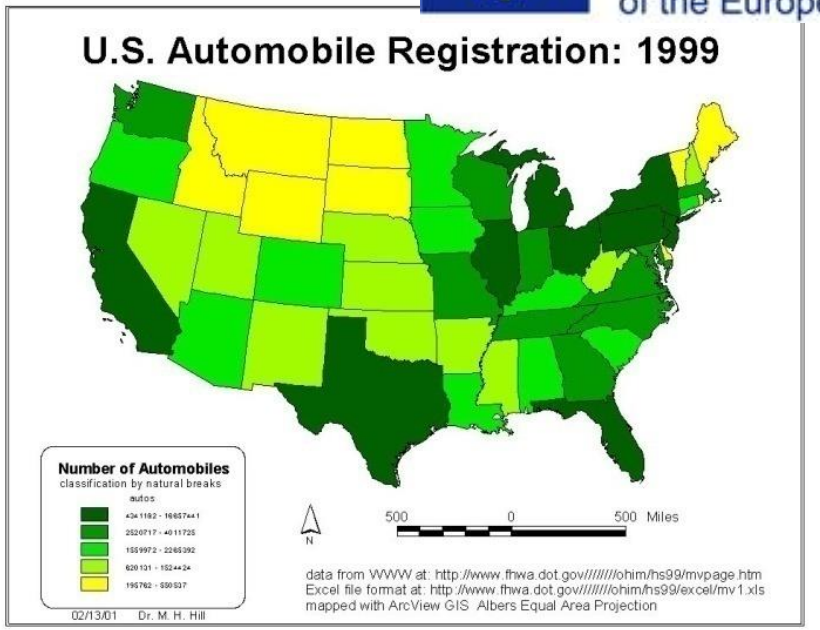
The Nature of Maps

- Maps are an abstracted representation
- Maps simplify the truth that is being represented
- The model of the Earth being represented in a map can rarely be complete
- Maps are used as means of communication
- Every map has the goal to convey a message

Co-funded by the
Erasmus+ Programme
of the European Union

24. Elements/Components of a Map

- Title (What, Where, When)
- North Arrow
- Legend (distinct categories and symbols, gives details about symbols and describes what values are mapped and how they are classified)
- Scale bar (given as graphic scale because reproduction will change in the scale in the same order as it changes the map)
- Source and credits
- Insets
- Date (when were data obtained and what period do they represent)
- Spatial Referencing Systems (projection and geodetic parameter)



25. Accuracy Assessment

Co-funded by the Erasmus+ Programme of the European Union

Confusion matrix

		Ground Truth			No. Classified pixels
		Water	Agri.	Forest	
Classified in map as:	Water	**46**	2	2	**50**
	Agriculture	10	**37**	3	**50**
	Forest	5	1	**44**	**50**
No. ground truth pixel		**61**	**40**	**49**	**150**

User accuracy

Producer accuracy

Overall accuracy = 46+37+44/150 = 84.7%

Producer's accuracy : the probability that a certain land-cover of an area on the ground is classified as such

User's accuracy : the probability that a pixel labeled as a certain land-cover class in a map is really this class

As a producer, I only classified 75.4% of all the water pixels as such.

As a user of a map, I can expect that roughly 92% of all the pixels classified as water are indeed water on the ground.

Producer accuracy

Water = 46/61 = 75.4%
Agriculture = 37/40 = 92.5%
Forest = 44/49 = 89.7%

User accuracy

Water = 46/50 = 92%
Agriculture = 37/50 = 74%
Forest = 44/50 = 88%

Co-funded by the
Erasmus+ Programme
of the European Union

26. Digital database creation

Manual Digitizing

This is traditionally the most common way to convert paper-based sources of spatial information (e.g. maps) to digital data.

The paper map is attached by tape to a digitizing table (or tablet as the smaller digitizers are known). Usually between 4-6 initial points of which the coordinates are known are logged.

Optimally these points are such locations as the intersections of graticule lines. In the absence of an overlying grid system, points are taken from identifiable locations such as street intersections or landmarks.

The data is then digitized by tracing the features of interest with a mouse like hand held device called a puck.