UNIT - I

(Part – 1)

A GENTLE INTRODUCTION TO GIS

1.1 THE NATURE OF GIS

Geographical Information System (GIS) is a computerized system for capturing, storing, querying, exploring, analyzing and displaying geospatial data.

- A town/urban planner might want to study the coverage by area, population growth. GIS helps a planner to explore and analyze this. Maharashtra state consists of about 44,000 villages spread over 35 districts in the State. MRSAC has completed the Geo-referencing of Village Maps Project (GVMP) in respect of all revenue village maps of the Maharashtra State. MahaRERA has mapped more than 4500 registered projects under GIS system enabling the town planner and citizens find minute details of property of their choice.
- A biologist might be interested in the study of impact of various natural phenomena on its surrounding geographic conditions and species.
- A natural hazard analyst might be interested in analyzing the risk areas of natural calamities by studying the rainfall pattern, climate change, earth observations etc.
- A Civil Engineer might be interested in planning new buildings and constructions according to the soil/rock information and geographical properties.
- A Mining Engineer might be interested in determining which prospective metal or mineral mines should be selected for future exploration, by considering the parameters like rock type extent, depth and quality of the ore body, etc.
- A geoinformatics engineer in a telecommunications company may want to determine the best sites for the company's relay stations, by considering various cost factors such as land prices, undulation of the terrain etc.
- A forest manager might want to optimize timber production using data on soil and current tree stand distributions, in the presence of a number of operational constraints, such as the need to preserve species diversity in the area;
- A hydrological engineer might want to study a number of water quality parameters of different sites in a freshwater lake to improve understanding of the current distribution of Typha reed beds to be laid for the Dam.

E-next
THE NEXT LEVEL OF EDUCATION

https://E-next.in
Scanned by CamScanner

In all of the above example GIS professionals work with Geospatial data, is a positional data used for referencing geographical data to store location based data and attributes like pattern, trends, conditions, forecasting, farm land, parcels, routes, forest land, vegetation, terrain and many more.

Spatial data refers to positional data relative to the Earth's surface.

1.2 SOME FUNDAMENTAL OBSERVATIONS

This world is dynamic. Many aspects of our daily lives and our environment are constantly changing, and not always for the better. Some of these changes appear to have natural causes (e.g. volcanic eruptions, meteorite impacts), while others are the result of human modification of the environment (e.g. land use changes or land reclamation from the sea, land filling used for e-waste). There are also a large number of global changes for which the cause remains un-clear: these include global warming, the El Nino/La Nina events, landslides and soil erosion. Changes to the Earth's geography can have natural or man-made causes, or both. If it is a mix of causes, we need a complete system to understand the changes.

To understand this, a study of the processes or phenomena that bring about geographic change is required. This can also help in decision making, to take the best course of action. For example, understood El Nino better, and can forecast that another event may take place in the year 2020; an action plan can be devised to reduce the expected losses in the fishing industry, to lower the risks of landslides caused by heavy rains or to build up water supplies in areas of expected droughts to minimize economic losses and death casualties.

The fundamental problem in many of GIS Applications is that of understanding phenomena that have a spatial or geographic dimension, along with temporal dimension. This means that GIS object under study has different characteristics for different locations (the geographic dimension) and also that these characteristics change over time (the temporal dimension). The El Nino event is a good example of such a phenomenon, because sea surface temperatures differ between locations, and sea surface temperatures change from one week to the next.

- **Defining GIS**

A GIS is a computer-based system that provides the following four sets of capabilities to handle geo-referenced data:

1. **Data capture and preparation**
2. **Data management,**
3. **Data manipulation and analysis**
4. **Data presentation**

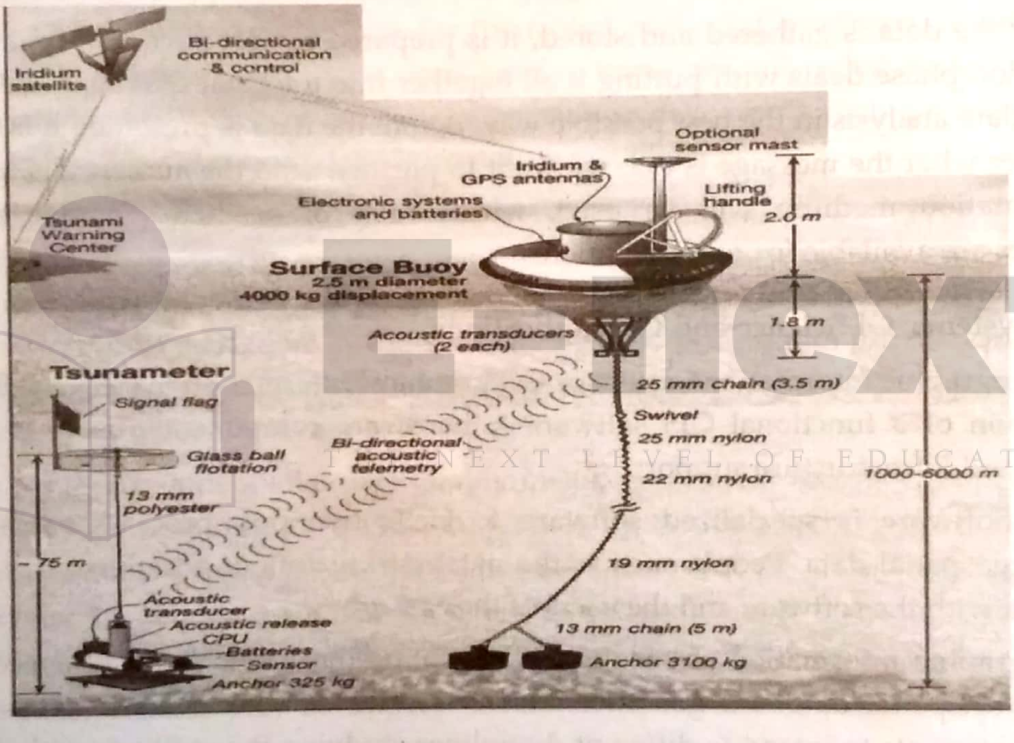
GIS is a complete system to enter (geo-referenced) data, to analyze it in various ways and to produce presentations (including maps and other types) from the data. Also support for various kinds of coordinate systems and transformations between them options for analysis of the geo-referenced data, and a large degree of freedom of choice in the way this information is presented (such as color scheme, symbol set, and medium used).

https://E-next.in
Scanned by CamScanner

1. **Data capture and preparation**

It is the most crucial and tedious task in GIS. Data capture and input is done using existing data or by creating new data. New data can be created from sensed images, GPS devices, field survey, user input, and text files etc. New data and map require editing and preparing it for presentation and further usage, this includes rectifying errors and geometric transformation.

In monitoring and forecasting when El Nino or La Nina is developing, data capture refers to the collection of sea water temperatures and wind speed measurements. This is achieved by placing buoys with measuring equipment at various places in the ocean. Each buoy measures a number of things: wind speed and direction; air temperature and humidity; and sea water temperature at the surface and at various depths down to 1000 to 6000 meters. All the data that a buoy obtains through its thermometers and other sensors, as well as the buoy's geographic position are transmitted by satellite communication daily.



2. **Data management**

Once the data is entered it must be verified and edited. Data is usually stored in tables in row and column format. Keys and their relationship establish connection between tables. Data manipulation includes data verification, attribute data management, insertion, updating, deleting and retrieval in different forms.

3. **Data manipulation and analysis**

Once the data has been collected and organized in a computer system, we can start analyzing it. Here, let us look at what processes were involved in the eventual production of the maps

The initial (buoy) data have been generalized from 70 point measurements (one for each buoy) to cover the complete study area. Clearly, for positions in the study area for which no data was available, some type of interpolation took place, probably using data of nearby buoys. This is a typical GIS function : de - riving an estimated value for a property

https://E-next.in
Scanned by CamScanner

for some location where we have not measured.

1. For each buoy, the average SST for each month was computed, using the daily SST measurements for that month. This is a simple computation.
2. For each buoy, the monthly average SST was taken together with the geo graphic location, to obtain a geo-referenced list of averages

Observation ID	Buoy ID	Position	December 2018 (Avg.) Sea Surface Temperature
OD3345	BO789	165º E 5 º N	26.32º C
OD3346	B7504	185º E 7 º N	27.86º C
OD3347	B1882	110º E 8 º 78′ N	26.96º C

4. **Data presentation**

After the data is gathered and stored, it is prepared for producing output. The data presentation phase deals with putting it all together into a format that communicates the result of data analysis in the best possible way. Before the data is presented it is necessary to consider what the message is that we want to portray, who the audience is, what kind of presentation medium will be used, which rules of aesthetics apply, and what techniques are available for representation.

- **GI Systems, GI Science and GI Applications**

A Geographic information System or Geographical Information System (GI System) is combination of a functional GIS software & hardware components, users to work on software and infrastructure support.

GIS Software is specialized software to facilitate input, process, transform and analyze geospatial data. People such as the database creators or administrators, analysts who work with the software and the users of the end product.

Geographic information science is the discipline that deals with all aspects of the handling of spatial data and geo-information. Geo-Information Science is the scientific field that attempts to integrate different disciplines studying the methods and techniques of handling spatial information.

Geoinformatics, geomatics, and spatial information science are related terms, these are all similar terms which have much the same meaning, although each approach has slight differences in the way it deals with problems, some emphasizing engineering approaches, others computational solutions, and so on.

The difference between a geographic information system and a GIS application is that, GIS software can generically be applied to many different applications. Eg. If ESRI ArcGIS is used in mapping crime information it is one application of ArcGIS and if it is used for Geo-referencing of Village Maps Project it is another application of the same GIS system.

- **Spatial data & Geo-information**

Date represents facts and figures in raw form, while information is processed data. Spatial data represents positional values. Spatial data is also known and geo-referenced

THE NEXT LEVEL OF EDUCATION

Scanned by CamScanner

data. Processing of this geo referenced data into using knowledge tools and its application specific interpretation is known as geo-information.

As this information is intended to reduce uncertainty in decision-making, any errors and uncertainties in spatial information products may have practical, financial and even legal implications for the user. For these reasons, it is important that those involved in the acquisition and processing of spatial data are able to assess the quality of the base data and the derived information products. The International Standards Organization (ISO) considers quality to be "the totality of characteristics of a product that bear on its ability to satisfy a stated and implied need"

Increasing availability and decreasing cost of data capture equipment has resulted in many users collecting their own data. However, the collection and maintenance of 'base' data remain the responsibility of the various governmental agencies, such as National Mapping Agencies (NMAs), which are responsible for collecting topographic data for the entire country following pre-set standards. Other agencies such as geological survey companies, energy supply companies, local government departments, and many others, all collect and maintain spatial data for their own particular purposes. If data is to be shared among different users, these users need to know not only what data exists, where and in what format it is held, but also whether the data meets their particular quality requirements. This 'data about data' is known as metadata.

Since the real power of GIS lies in their ability to combine and analyze geo-referenced data from a range of sources, we must pay attention to the issues of data quality and error, as data from different sources are also likely to contain different kinds of error. This may include mistakes or variation in the measurement of position and/or elevation, in the quantitative measurement of attributes or in the labeling or classification of features. Some degree of error is present in every spatial data set.

Key components of spatial data quality include positional accuracy (both horizontal and vertical), temporal accuracy (that the data is up to date), attribute accuracy (e.g. in labeling of features or of classifications), lineage (history of the data including sources), completeness (if the data set represents all related features of reality), and logical consistency (that the data is logically structured).

These components play an important role in assessment of data quality for several reasons :

1. Even when source data, such as official topographic maps, have been subject to stringent quality control, errors are introduced when these data are input to GIS.
2. Unlike a conventional map, which is essentially a single product, a GIS database normally contains data from different sources of varying quality.
3. Unlike topographic or cadastral databases, natural resource databases contain data that are inherently uncertain and therefore not suited to conventional quality control procedures.
4. Most GIS analysis operations will themselves introduce errors.

1.3 THE REAL WORLD AND REPRESENTATIONS OF IT

- **Models and modeling**

A model is a representation of whole or some part of the real world having certain

https://E-next.in
Scanned by CamScanner

characteristics in common with the real world. It is used to study and operate on the model itself instead of the real world in order to test what happens under various conditions, and help us answer to 'what if' analysis, by changing the data or alter the parameters of the model, and analyze the effects of the changes.

Models are of different type like Mathematical model, simulation model, data model, clay model, architectural model etc.

Map is the most commonly used model in GIS. It is a miniature representation of some part of the real world. A 'real world model' is a representation of a number of phenomena that we can observe in reality, usually to enable some type of study, administration, computation and/or simulation. Most maps and databases can be considered static models. At any point in time, they represent a single state of affairs. New developments or changes in the real world are not easily recognized in these models. Dynamic models or - process models are used for such systems, to emphasize changes that have taken place, are taking place or may take place sometime in the future. Dynamic models are more complicated than static models, and usually require much more computation. Simulation models are an important class of dynamic models that allow the simulation of real world processes.

- **Maps**

A map is a graphic representation of real world at a certain level of detail, which is determined by the scale, having physical boundaries, and features spanning two map sheets have to be cut into pieces. Cartography is the art and science of map making, functions as an interpreter, translating real world phenomena (primary data) into correct, clear and understandable representations for our use. Maps also become a data source for other applications, including the development of other maps.

Maps have been used for thousands of years to represent information about the real world, and today also it continued to be extremely useful for many applications in different domains. A disadvantage of the traditional paper map is that it is generally restricted to two dimensional static representations, and that it is always displayed in a fixed scale. The map scale determines the spatial resolution of the graphic feature representation. The smaller the scale, the less detail a map can show. The accuracy of the base data, on the other hand, puts limits to the scale in which a map can be sensibly drawn. Therefore, the selection of a proper map scale is one of the first and most important steps in map design.

Digital mapping also called digital cartography is the process by which a collection of data is compiled and formatted into a virtual image. The primary function of this technology is to produce maps that give accurate representations of a particular area, detailing major road arteries and other points of interest.

- **Databases**

A database is a repository for storing large amounts of data. It comes with a number of useful functions :

1. A database can be used by multiple users at the same time—i.e. it allows concurrent access.
2. A database offers a number of techniques for storing data and allows the use of the most efficient one—i.e. it supports storage optimization,

THE NEXT LEVEL OF EDUCATION
https://E-next.in
Scanned by CamScanner

3. A database allows the imposition of rules on the stored data; rules that will be automatically checked after each update to the data—i.e. it supports data integrity,
4. A database offers an easy to use data manipulation language, which allows the execution of all sorts of data extraction and data updates—i.e. it has a query facility,
5. A database will try to execute each query in the data manipulation language in the most efficient way—i.e. it offers query optimization.

Databases can store almost any kind of data. A database may have many such tables, each of which stores data of a certain kind. It is not uncommon for a table to have many thousands of data rows, sometimes even hundreds of thousands. For the El Niño project, one may assume that the buoys report their measurements on a daily basis and that these measurements are stored in a single, large table.

Buoy	Date	Sea Surface Temperature	Wind Speed	Humidity (%)	Temp	....
B0799	2018/12/03	26.02º C	NNW4.5	72.09	22.31º C	....
B9214	2018/12/03	25.86º C	NW4.6	45.35	23.12º C	....
B3686	2018/12/03	26.24º C	NNW3.8	42.36	23.02º C	....
B5908	2018/12/03	27.12º C	N1.6	79.35	22.45º C	....
B2871	2018/12/03	28.35º C	W3.5	85.25	21.96º C	....
B4205	2018/12/03	27.45º C	NW 3.1	45.65	21.47º C	....
B1234	2018/12/03	27.86º C	SSW4.5	79.25	22.19º C	....
B1380	2018/12/03	26.79º C	W1.3	76.12	22.09º C	....
B6543	2018/12/03	25.96º C	S4.1	68.25	21.21º C	....
....	....	....	....	....	....	....
....	....	....	....	....	....	....

- **Spatial databases and spatial analysis**

GIS applications uses spatial for spatial analysis, and uses a spatial database for storage. Spatial database also known as geo-database can store representations of real world geographic phenomena for use in a GIS. It stores about spatial reference systems, and supports all kinds of analyses that are inherently geographic in nature, such as distance and area computations and spatial interpolation. The phenomena for which we want to store representations in a spatial database may have point, line, and area or image characteristics. Different storage techniques exist for each of these kinds of spatial data. These geographic phenomena have various relationships with each other and possess spatial (geometric), thematic and temporal attributes (they exist in space and time). For data management purposes, phenomena are classified into thematic data layers. The purpose of the database is usually described by a description such as cadastral, topographic, land use, or soil database.

Spatial analysis is the generic term for all manipulations of spatial data carried out to improve one's understanding of the geographic phenomena that the data represents. It

https://E-next.in
Scanned by CamScanner

involves questions about how the data in various layers might relate to each other, and how it varies over space.

Geospatial analysis is used in application areas like the environmental and life sciences, in particular ecology, geology and epidemiology. It has extended to almost all industries including defense, intelligence, utilities, Natural Resources (i.e. Oil and Gas, Forestry ... etc.), social sciences, medicine and Public Safety (i.e. emergency management and criminology), disaster risk reduction and management, and climate change adaptation.

E-next THE NEXT LEVEL OF EDUCATION

Scanned by CamScanner