



GIS-503/DGIS-503/CGIS-503

M.A. /M.Sc. Geoinformatics/ DGIS/ CGIS

FUNDAMENTALS OF GIS

**DEPARTMENT OF REMOTE SENSING AND GIS
SCHOOL OF EARTH AND ENVIRONMENT SCIENCE
UTTARAKHAND OPEN UNIVERSITY
HALDWANI (NAINITAL)**



M.A. /M.Sc. Geo-informatics/DGIS/CGIS

UTTARAKHAND OPEN UNIVERSITY

FUNDAMENTALS OF GIS

FUNDAMENTALS OF GIS



**DEPARTMENT OF REMOTE SENSING AND GIS
SCHOOL OF EARTH AND ENVIRONMENTAL SCIENCE
UTTARAKHAND OPEN UNIVERSITY**

Phone No. 05946-261122, 261123

Toll free No. 18001804025

Fax No. 05946-264232,

E-mail info@uou.ac.in

Website <http://uou.ac.in>

FUNDAMENTALS OF GIS

BOS Members

Chairman

Vice Chancellor
Uttarakhand Open University, Haldwani

Professor R.K.Pande
Dean Arts, DSB Campus
Kumaun University
Nainital

Professor Pradeep Goswami
Department of Geology
DSB Campus, Nainital

Dr. Ranju Joshi Pandey
Department of Geography & NRM
Department of Remote Sensing and GIS
School of Earth and Environmental Science
Uttarakhand Open University, Haldwani

Convener

Professor P.D. Pant
School of Earth and Environment Science
Uttarakhand Open University, Haldwani

Professor D.D.Chauniyal
Retd. Professor
Garhwal University
Srinagar

Dr. Suneet Naithani
Associate Professor,
Department of Environmental Science
Doon,University, Dehradun

Programme Coordinator

Dr. Ranju J. Pandey
Department of Geography & NRM
Department of Remote Sensing and GIS
School of Earth and Environment Science
Uttarakhand Open University, Haldwani

Units Written By	Unit No.
Dr. D.N.Pant Ex- Scientist & Group Head IIRS, Dehradun	1, 2, 3, 4
Dr. Manish Kumar Assistant Professor Department of Geography, School of Basic Sciences, Central University of Haryana & Amit Ray	5
Dr. Sharad Goyal (Project Scientist) Forest Geoinformatics Division Forest Survey of India MOEF & CC, Dehradun	6,7&8

Course Editor

Dr. Ranju Joshi Pandey
Department of Geography & NRM
Department of Remote Sensing and GIS
School of Earth and Environment Science
Uttarakhand Open University, Haldwani

Title:	Fundamentals of GIS
ISBN No.:	978-93-90845-49-1
Copyright:	Uttarakhand Open University
Edition:	First (2021)

Published By: Uttarakhand Open University, Haldwani, Nainital-263139

Printed By:

CONTENTS

BLOCK 1: INTRODUCTION OF GIS

UNIT 1: Definition and History of GIS	1-20
UNIT 2: Data Models	21-37
UNIT 3: Data Query	38-55
UNIT 4: Data output and Visualization	56-77

BLOCK 2: OVERVIEW ON RS & GIS SOFTWARE

UNIT 5: Introductory idea about RS software	78-90
UNIT 6: Overview of open source software	91-105
UNIT 7: Overview of Quantum GIS software	106-127
UNIT 8: Overview of GIS portals and open data archives	128-147

BLOCK 1 - INTRODUCTION OF GIS

UNIT 1 - DEFINITION AND HISTORY OF GIS

1.1 OBJECTIVES

1.2 INTRODUCTION

1.3 DEFINITION AND HISTORY OF GIS

1.4 SUMMARY

1.5 GLOSSARY

1.6 ANSWER TO CHECK YOUR PROGRESS

1.7 REFERENCES

1.8 TERMINAL QUESTIONS

1.1 OBJECTIVES

After reading this unit you will be able to understand that:

- Concept and Definition of GIS
- Need, Scope and Importance of GIS
- History of GIS
- Components and types of GIS

1.2 INTRODUCTION

In the previous book you have studied about the basics of remote sensing, electromagnetic radiation, platforms, sensors, resolution types and analysis, Aerial Photography. Remote sensing techniques provide you very quick, timely and reliable information about the land surface features based on your requirement and objectives. That information facilitates various kinds of planning, management and developmental activities for socio-economic development and prudent/judicious use of natural/cultural resources. The temporal remote sensing data highlights the changes and dynamism of earth surface cover types and land uses which is utmost essential for future planning and taking precautionary measures about the danger/alarming/crucial situations if those are showing negative trends but if positive the planning will be followed accordingly. But in most of the cases, the collection or obtaining information/data itself does not prove the complete solution until and unless we integrate, analyze it and make modeling for the required purposes. For example you have the information about the availability of land or kind of resources but you do not know about their suitability or fitness for a particular kind of use or management and planning towards their development. Therefore the GIS (Geographical Information System), which is the most powerful computerized tool for collection, storage, analysis, manipulation, modeling and retrieval of information, is equally essential like remote sensing, for successful planning, management and optimization of land uses/resource utilization. In fact remote sensing based data/information and its analysis and modeling through GIS tools and techniques are complementary to each other for all kinds of developmental planning.

GIS is a multi-disciplinary subject that includes a variety of technologies and concepts. In order to understand the capabilities of GIS, it is important to learn cartography, mapping, spatial and statistical analysis, different types of database, database management, and programming. All of them will be discussed in the next units except GIS objectives, concept, definitions, need, scope, importance, and history, components of GIS and GIS types which are being described/ explained in this unit.

1.3 DEFINITION AND HISTORY OF GIS

1.3.1 GIS objectives, concept and definition

1.3.1.1 GIS Objectives

- 1) To improve the efficiency of decision making processes and planning.
- 2) Provide efficient means for data distribution and handling.
- 3) Eradication of the duplicated data,
- 4) Integration of information from many sources.
- 5) Analysis of queries involving geographical reference data for generation of new information, Update data quickly and at the minimum cost.

1.3.1.2 Concepts and Definition

People around the world may like to know the geographic, social, economic, political, and environmental information in digital form for their practical knowledge, academics, services, responsibilities etc. To explore geo-referenced digital information, electronic tools designed for acquiring, presenting, and interacting with information that links location with measured values are needed. One such tool is called a geographic information system, better known as GIS.. GIS technology powers solutions for telecommunications, utilities, agriculture, defense, oil, health care, transportation, mining, environmental management, petroleum, water/wastewater, and many other industries as well as local, state, and federal government agencies. The following points highlight the concepts and definitions of GIS:

- A geographic information system (GIS), also known as a geospatial information system, is a system for capturing, storing, analyzing and managing data and associated attributes which are spatially referenced to the Earth.
- GIS provides Geographic/geospatial information about the places on earth's surface, knowledge about "what, where, when, how, how far etc within the spatial domain and the Geographic/ geospatial synonym.
- It is an Information System capable of integrating, storing, editing, analyzing, sharing, and displaying Geographically Referenced information.
- Generally, GIS is a tool that allows users to create interactive queries, data editing, analyze the spatial information, creation of maps, and presentation of results.
- GIS and its location based intelligence applications becomes the base for many of the other location-enabled services that rely on analysis and visualization.
- GIS makes relationships with unrelated information by using location as the key index variable.
- All Earth-based spatial-temporal location and extent references should be relatable to one another and ultimately to a "real" physical location or extent.

Generally, GIS is based on an integrating technology consisting of:

- Remote Sensing
- Cartography and Mapping
- GPS
- Computers

- RDMS
- Information Technology
- Communication technology
- Survey and field data collection

GIS is a type of software which consists of a computer system that allows us to handle information about the location of features or phenomena on the Earth's surface. It has all the functionality of a conventional DBMS plus much of the functionality of a computer mapping system. GIS as a DBMS that allows us to explicitly handle the spatial data. Common examples are QGIS, Arc GIS and MapInfo.

1.3.1.3 Need, scope and importance of GIS

Geographic information systems are among the most exciting and powerful geomatics decision-making tools in the world. A GIS uses computer technology to integrate, manipulate and display a wide range of information to create a picture of an area's geography, environment and socio-economic characteristics. Beginning with a computerized topographic map as its base, a GIS *overlays and integrates* graphic and textual information from separate databases. The end result is a customized and reliable tool that can support decision making and problem solving and provide almost instantaneous answers to complex questions (Geomatics Canada Web Site, 2000). The following lines highlight this sub-title:

- Geographic information system technology can be used for scientific investigations, resource management, asset management, Environmental Impact Assessment, Urban planning, etc. For example, GIS might allow emergency planners to easily calculate emergency response times in the event of a natural disaster.
- GIS is used to find wetlands that need protection from pollution, or GIS can be used by a company to site a new business to take advantage of a previously underserved market.
- In the modern concept of Information technology, GIS is termed as Geo-Information technology. The technology is involved with the integration of Surveying and Mapping techniques, Remote sensing and Satellite Imagery, Photogrammetry, Geography, Geology, Geomorphology, Cartography and Global Positioning Systems (GPS).
- The scope of using GIS techniques is a well known aspect in the sectors of forestry and forest management, land suitability classification, agricultural and horticultural land optimization, irrigation, rural development, industrial land uses and industries, minerals, energy, transport, communications, Science, Technology and Environment. GIS can be used for scientific investigations, resource management, and development planning.
- GIS used in Computer Science, Civil Engineering, Mathematics, Statistics and Operations Research. With GIS, we can create new approaches that help us understand the relationship between man and the environment. This calls for more integrated tools that build a holistic and comprehensive approach to resolving planning problems.

- GIS includes the merging of cartographical data, statistical analysis, and database technology. As GIS can be thought of as a system, it digitally creates and "manipulates" spatial areas that may be jurisdictional, purpose or application oriented for which a specific GIS is developed. Hence, a GIS developed for an application, jurisdiction, enterprise, or purpose may not be necessarily interoperable or compatible with a GIS that has been developed for some other application, jurisdiction, enterprise, or purpose.
- Any organization, government private is in some way or another strongly linked to the geography in which it operates. A GIS that has been designed in a proper manner has the capability of providing quick and easy access to large volumes of data of these Geographical features.

GIS has a great role in the organizational benefits and in almost every industry. There is a growing interest and awareness of the economic and strategic value of GIS because of more standards-based technology demonstrated by its users. The number of GIS enterprise solutions and its strategies that include GIS are growing rapidly. The importance of GIS generally falls into the following five basic categories:

i)Efficiency based cost saving

This is associated with automating or improving a workflow and improvements in the mission itself. GIS functions as a repository of information associated with fixed assets. Municipal GIS captures information about the city's assets – typically land use information (such as parcel number, parcel owner, permits and zoning) and physical items (such as street lights, traffic signals, utility meters, water pipes, sewer infrastructure, street trees, pavement, curbs and sidewalks). What makes GIS different from traditional databases is that it associates geographic placement, in latitude and longitude, with each asset. Consequently it's possible to make decisions based not only on a type of asset but also on the asset's location.

According to a study carried out in California, by using GIS to automate the notification process, the city reduced by 90 percent the amount of staff time needed to do the job. Here the GIS-based process had been simple and concerned to the following points:

- a) Identify the parcel containing the planned development;
- b) Instruct the GIS system to identify all parcels whose boundaries cross or are inside a specific radius; and
- c) Generate mailing labels to the parcel owners.

The cost and time savings associated with even this simple GIS operation paid for the system deployment.

ii) Prompt and the Best Decision Making

This typically has to do with making better decisions about location. Common examples include real estate site selection, route/corridor selection, zoning, planning, conservation, natural resource extraction, etc. People are beginning to realize that making the correct decision about a location is strategic to the success of an organization.

iii) Improved communication

GIS-based maps and visualizations greatly assist in understanding situations and storytelling. They are a new language that improves communication between different teams, departments, disciplines, professional fields, organizations, and the public.

iv) Better geographic information recordkeeping

Many organizations have a primary responsibility of maintaining authoritative records about the status and change of geography (geographic accounting). Cultural geography examples are zoning, population census, land ownership, and administrative boundaries. Physical geography examples include forest inventories, biological inventories, environmental measurements, water flows. GIS provides a strong framework for managing these types of systems with full transaction support and reporting tools. These systems are conceptually similar to other information systems in that they deal with data management and transactions, as well as standardized reporting (e.g., maps) of changing information. However, they are fundamentally different because of the unique data models and hundreds of specialized tools used in supporting GIS applications and workflows.

v) Managing geographically

In government and many large corporations, GIS is becoming essential to understand what is going on. Senior administrators and executives at the highest levels of government use GIS information products to communicate. These products provide a visual framework for conceptualizing, understanding, and prescribing action. Examples include briefings about various geographic patterns and relationships including land use, crime, the environment, and defence/security situations.

GIS is increasingly being implemented as enterprise information systems. This goes far beyond simply spatially enabling business tables in a DBMS. Geography is emerging as a new way to organize and manage organizations. Just like enterprise-wide financial systems transformed the way organizations were managed in the _60s, _70s, and _80s, GIS is transforming the way that organizations manage their assets, serve their customers/citizens.

Any organization, government private is in some way or another strongly linked to the geography in which it operates. A GIS that has been designed in a proper manner has the capability of providing quick and easy access to large volumes of data of these geographical features. The user can access & select information by area or by theme to merge one data set with another, to analyze spatial characteristics of data, to search for particular features, to update quickly and cheaply and asses alternatives.

In simpler terms, GIS allows the user to understand geographic information in an easy manner without having to go through large volumes of confusing data that is in tabular form. Visualizing the geography of a particular location is no doubt easier that trying to analyze raw data. The use of modern GIS offers many advantages over paper maps. The important

characteristics of GIS are having its tool-kit which highlights the following important functional characteristics:

- Use either SI or CGS as primary units. (SI units)
- **Manipulate spatially:**
 - Calculate distances and adjacencies
 - Change projections and scales
 - Integrate disparate sources
- **Analyse spatially:**
 - Quantitative analysis
 - Exploratory spatial data analysis
 - Qualitative analysis
- **Visualise data:**
 - Maps!
 - Tables, graphs, etc.
 - Animations
 - Virtual landscapes
- More robust and resistant to damage
- Faster and more efficient
- Requires less person time and money

The importance of GIS is also based on the facts that this system collects and stores information about the world as a collection of thematic layers that can be linked together by geography. This simple but extremely powerful and versatile concept has proven invaluable for solving many real-world problems from tracking delivery vehicles, to recording details of planning applications, to modeling global atmospheric circulation.

1.3.2 Components of GIS

A GIS system in working order integrates the following five key components:

- Data
- Hardware
- Software
- Users/People
- Methods

1.3.2.1 Data

Data is a set of values of subjects with respect to qualitative or quantitative variables. Data and information or knowledge is often used interchangeably. You may also define data as a collection of facts, such as numbers, words, measurements, observations or even just descriptions of things.

Data may be classified as i) qualitative and ii) quantitative. Qualitative data is descriptive information whereas Quantitative data is numerical information. For example slope of the hill is very steep is a qualitative information whereas slope of the hill is in between the slope range of 30 - 40% is quantitative information.

Geographic data refers to information about the earth's surface and the objects found on it. This information comes in three basic forms: spatial data, tabular data, and image data. Spatial data contains the locations and shapes of map features. Tabular data is collected and compiled for specific areas and is the descriptive data that GIS links to map features. Image data includes such diverse elements as satellite images, aerial photographs, and scanned data--data that's been converted from printed to digital format. Data can be created or bought. For example, a GPS receiver can be used to identify sites in an agricultural field where weed data is collected. A table can be created in the GIS showing location as well as species and number of weeds present in the measured area. Alternatively, data can be purchased. In most cases, images are bought from satellite or aircraft companies that used cameras to collect images of the Earth's surface.

1.3.2.1a Map Data

Map is an essential tool for execution of various kinds of activities. Maps or GPS based ground coordinates act as input coordinate system in GIS spatial data domain. Geo-referenced remote sensing data is also equally helpful in creating spatial information. It becomes more realistic approach if you take the help of all three data types to prepare the spatial database. Map data provides information about location (Figure 1.1). Figure 1.1 represents the latitude and longitude points (graticules) on the polygon graphics formed by digitizing the feature from map.

1.3.2.1b Attribute Data

It provides you the information about what can be found at a particular location. The attributes of a river, for example, might include its name, length, and average depth, rate of flow, water quality, how many dams are on it, and how many bridges cross it. Figure 1.2 represent both map and attribute data.

1.3.2.1c Hardware

Hardware is a Computer on which GIS software runs. There are a different range of computers like Desktop or server based. Arc GIS, Arc Info and Arc view GIS software servers are server based computers where GIS software runs on network computer. The good computer hardware components must have high capacity. Examples of hardware components are: server, digitizer, PC, Printer, plotter, Hard driver, processor, graphics card, etc. These all component function together to run GIS software smoothly. A mini hardware system of GIS is shown in figure 1.3.

1.3.2.1d Software

Software is a GIS component, a technology for storing and analyzing location and attributes data. GIS Software) is designed to store, retrieve, manage, display, and analyze all types of geographic

and spatial data. Figure 1.4 represent the spatial and attribute data within the computer software. Examples of GIS software are Arc GIS and Arc View for desktop, Web GIS, Arc Info as server GIS etc.

1.3.2.1e GIS User/People/Personnel

GIS technology may have many limitations without the involvement of people who manage the system and develop plans for applying it to real world problems. GIS users range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work. GIS user/people/personnel's specifically include the following categories of working environment:

- Project coordinators
- Data analysts
- Programmers
- Data and knowledge managers
- Librarians Types of GIS

Figure 1. 1 Map data of GIS showing the Graticules after digitization of a particular feature from hard copy of a map

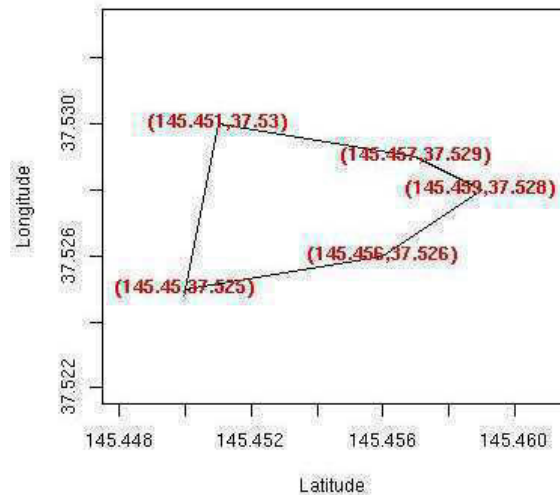


Figure 1.2 GIS Map showing Map and Attribute Data

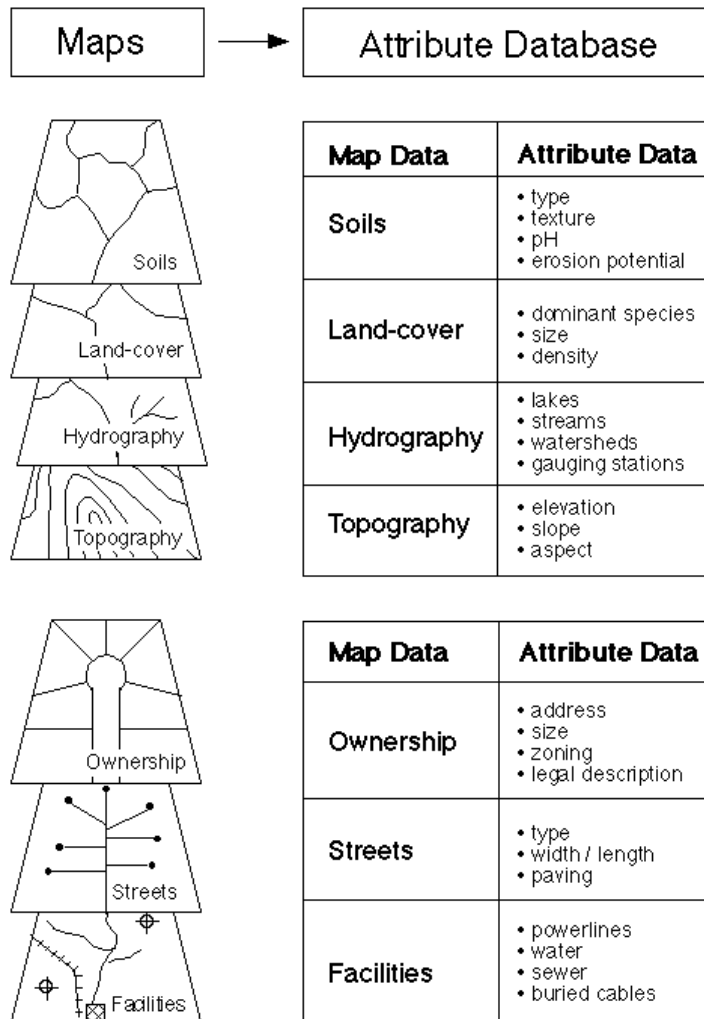
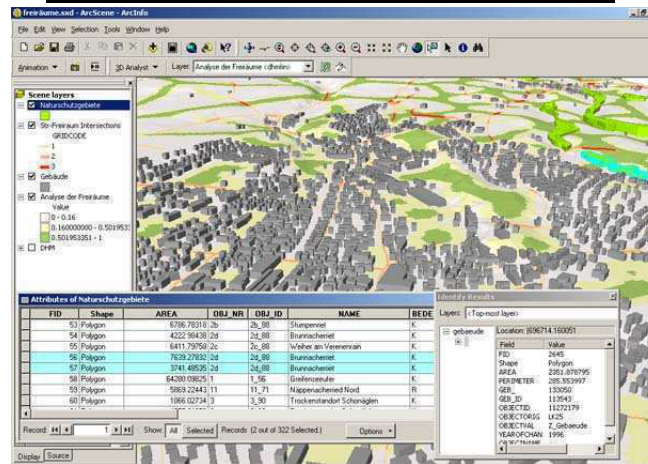


Figure 1.3 GIS Hardware consists of CPU, Monitor, Plotter, and Printer



Figure 1.4 GIS Software on the computer monitor

1.3.2.1d Methods

A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

1.3.3 Types of GIS

In recent time, there are nine types of GIS in practice. These are categorised as below:

- Desktop GIS
- Professional GIS
- Enterprise GIS
- Web GIS
- Internet GIS
- Mobile GIS
- Distributed GIS
- 4 D GIS
- Cloud GIS

1.3.3.1 Desktop GIS

Desktop GIS is an immensely powerful computer mapping system. It is a tool for managing information of any kind according to where it is located. For example, businesses can track customer locations, optimize delivery routes, or decide where to site future businesses using GIS; scientists use GIS to manage sensitive wildlife habitats or track animal movements in an ecosystem; and health care specialists track the spread of infectious disease with GIS

Desktop GIS represents the real geographical features on a computer similar to the way maps represent the same on paper. Both GIS and paper maps convey information about places. However, desktop GIS have power and flexibility to change, if any, that paper maps lack. The scale of the map influences the size of what appears on it. With GIS, however, you can

store and link huge amounts of information about the objects represented on maps. These objects are called features. Each map feature has a location, a representative shape, and a symbol that represents one or more of its characteristics. Because features on maps are organized according to relative location or position, maps are good for showing the relationships among feature locations. These relationships, called spatial relationships, are important because understanding them helps us solve problems.

While preparing the maps we use three basic shapes-- points, lines and areas to represent real-world objects location on the basis of geographical coordinates. Points represent objects that have discrete locations and are too small to be depicted as areas. Lines represent objects that have length but are too narrow to be depicted as areas. Areas represent objects too large to be depicted as points or lines. Shapes alone do not give you enough information, so maps use graphic symbols to help identify features and provide information about them.

Most features can be represented as more than one shape. The scale of a map tells how the size of the map features compares with the size of the geographic objects they represent. Map scales vary from small-scale to large-scale. For example, on a small-scale map a city may be represented as a point. That same city would be represented as an area on a large-scale map.

With desktop GIS, you are not limited to the amount of information you can get about what you see on the map. Desktop GIS stores all the information about map features in a GIS database and links the features on the map to the information about them. This means that you can access all the information about a feature by simply clicking on it.

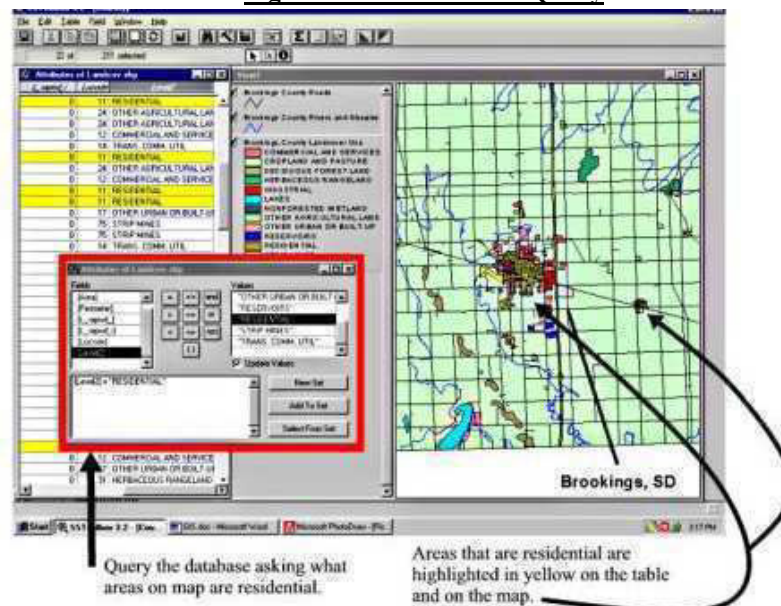
The information that a desktop GIS stores about map features is referred to as attribute information, or attributes. Desktop GIS formats attributes in rows and columns, and stores them as tables. Each column stores a different attribute and each row relates to a single feature. The link between map features and their attributes is the basic principle behind how a desktop GIS works, and is the source of its power. Once the map features and attributes are linked, you can access the attributes for any map feature or locate any feature from its attributes in a table. GIS can also display features based on any attribute in the table. Because the link between features and attributes is a two-way relationship, changing an attribute in the table automatically results in a change on a map. Desktop GIS links sets of features and their attributes and manages them together in units called themes. A theme contains a set of related features, such as roads, streams, parcels, or wildlife habitat areas, along with the attributes for those features.

All the themes for a geographic area taken together make up a GIS database. The design of a GIS database is strong because it's flexible. You can add new themes to a GIS database or delete old ones; you can separate themes to create more themes, or combine themes if they have common characteristics. What you want to do with a GIS database, and what information you need, will determine the best design for you.

The GIS database can be 'queried'. This means a user can ask questions and get answers about the database. For example, in Figure 1.5, the user queries the database about location of the residential area in and around Brookings, SD. The results of the query are highlighted in yellow in both the table and the map.

Information can be presented as maps, charts, and tables, along with graphics you import from other programs or even graphics you draw yourself. The presentations you create can be output to a printer to produce hard copy, or displayed on your computer's screen. You decide what information to present and how much--how much detail, which colors and symbols, and how the final pieces will be arranged. And if your audience or your objective should change, it's easy to make your presentation reflect those changes, without having to start over.

Figure 1.5 GIS Database Query



1.3.3.2 Professional GIS

Professional GIS is important in the field of GIS as it helps to establish GIS as a true profession, ensures those who produce geographic information have competency in a core set of knowledge and skills, encourages long-term professional development, and promotes ethical behavior by members of the profession. For individuals, successful certification may improve advancement opportunities; result in higher compensation and greater career mobility, and greater recognition from employers and colleagues.

1.3.3.3 Enterprise GIS

Enterprise GIS is a geographic information system that is integrated through an entire organization so that a large number of users can manage, share, and use spatial data and related information to address a variety of needs, including data creation, modification, visualization, analysis, and dissemination.

1.3.3.4 Web GIS

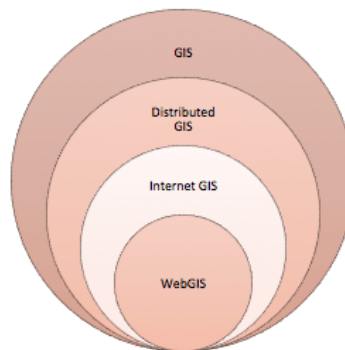
Web GIS is a GIS system that uses web technologies. It often uses web technologies to communicate among different components of the system. Web GIS originates from a

combination of web technology and the Geographical Information System, which is a recognized technology that is mainly composed of data handling tools for storage, recovery, management and analysis of spatial data. Web GIS is a kind of distributed information system. The simplest architecture of a Web GIS must have at least one client and one server that client is a desktop application or web browser application that allows users to communicate with server, and the server is a web server application.

1.3.3.5 Internet GIS

Web GIS is a close term to Internet GIS. These two words are always used as synonymous with each other. There is a slight difference between these two words. The Internet supports many services with the Web being one of these services. So we can call a system as Internet GIS if it uses many of services of Internet not only Web service and if it uses only Web we should name

Figure 1.6 Internet GIS and web GIS Comparison



it Web GIS. This definition makes Internet GIS boarder than Web GIS. In real world Web is the most attractive service of Internet and it is why Web GIS is more common than Internet GIS.

The geospatial Web, or GeoWeb, is another term that uses to refer Web GIS, However the definition of Web GIS is not identical to Web GIS. GeoWeb can be defined by merging geospatial information with none geospatial information such as news, photos, stories and so on.

1.3.3.6 Mobile GIS

Geographic Information Systems (GIS) being used out of the office and into the field is called Mobile GIS. A mobile GIS is used in the field to capture, store, update, manipulate, analyze, and display geospatial data and information. Mobile GIS integrates one or more of the following technologies:

- Mobile devices (such as a PDA, tablet, or laptop computer, and in some countries mobile phones)
- Geographic Information System (GIS) software.
- The Global Positioning System (GPS).
- Wireless communications for Internet-based GIS access.

For most applications, it is an extension of desktop GIS, although increasingly users are taking mobile GIS data and uploading it directly to powerful visualization tools online such as Google

Earth. Mobile GIS can allow for edits and changes to be made in the field, increasing accuracy and saving time. Many mobile GIS systems are relatively inexpensive.

1.3.3.7 Distributed GIS

It refers to GI Systems that do not have all of the system components in the same physical location. This could be the processing, the database, the rendering or the user interface. Examples of distributed systems are web-based GIS and Mobile GIS.

1.3.3.8 4D GIS

4D GIS has developed into an essential component of GIS applications, incorporating 2D and 3D data with time. 4D GIS integrates, manages, and analyzes spatial and non spatial information, providing quality visualization, simulation, and communication. While also facilitating better decision making by providing a geographic representation of the full scope of a project.

1.3.3.9 Cloud GIS

This refers to hosted services on the internet meant for users of GIS technology or members of public who want to access maps. Services include map service, data storage and access, powerful analysis with applications, to manage assets and information.

1.3.4 GIS Data Types

There are two types of data stored for each item in the database.

i) Attribute data:

Attribute data says *what* a feature is, eg. Statistics, text, images, sound, etc.

ii) Spatial data:

Spatial data reflects where the features are existing, eg. All features, objects and related Phenomenon. It has the following characteristics:

- a) It is co-ordinate based
- b) It may be a **Vector** data showing the following discrete features:
 - i) Points
 - ii) Lines
 - iii) Polygons (zones or areas)
- a. It may be a **Raster** data which represent a continuous surface.

Details of raster and vector data will be discussed in the next unit.

1.3.5 History of GIS

While considering about the history of GIS, we must think over the conventional methods of map preparation as map is always at the heart of GIS. Mapping has revolutionized how we think about location. Maps are important decision-making tools. They help us get to places. And they are becoming more immersed in our everyday lives.

Conventional methods of map preparation dates back 1854. With the passage of time, there had

been the tremendous changes in the field of cartography and mapping. In fact GIS has no meaning and no importance if we do not consider its strong powerful computerized tool and the varieties of software. GIS was developed by Geographers just because to overcome the conventional method of map making and the integration of different GIS layers, its analysis and modeling. The development of GIS has saved money, man power and the time. But the important point is saving paper containing map which we were bound to torn after a frequent stipulated time when the environment and surface features were changing and the maps had no meaning /no importance.

Advancements in GIS were the result of several technologies. Databases, computer mapping, remote sensing, programming, geography, mathematics, computer aided design, and computer science all played a key role in the development of GIS.

Today, we'll uncover some of the key moments in the history of GIS that has shaped it what it has become today:

1.3.6 Paper Mapping Analysis with Cholera Clusters

The history of GIS all started in 1854 when British physician John Snow began mapping outbreak locations, roads, property boundaries and water lines of Cholera hit city of London. When he added these features to a map, something interesting happened: He saw that Cholera cases were commonly found along the water line.

John Snow's Cholera map was a major event connecting geography and public health safety. Not only was this the beginning of spatial analysis, it also marked the start of a whole field of study: Epidemiology – the study of the spread of disease. To this date, John Snow is known as the father of epidemiology. The work of John Snow demonstrated that GIS is a problem-solving tool. He put geographic layers on a paper map and made a life-saving discovery.

During 20th century, the concepts of printing different theme maps and to separate layers from a map were developed. But this concept did not represent a full GIS as there was no option to analyze mapped data. With the passage of time, many other concepts were developed but that did not full fill the whole concept of GIS. The actual concept of GIS was first introduced in the early 1960s, by Dr. Roger Tomlinson and it was subsequently researched and developed as a new discipline. The GIS history views Roger Tomlinson as a pioneer of the concept, where the first iteration was designed to store, collate, and analyze data about land usage in Canada. Thus Dr. Roger Tomlinson is generally recognized as the "father of GIS." He is the visionary geographer who conceived and developed the first GIS for use by the Canada Land Inventory in the early 1960s.

During 1970s, the second phase of development in GIS history was continued and by the 1980s the concept progressed in a big way and likely to be adopted by different agencies. By the late 1980s, there was a focus on improving the usability of technology and making facilitates more user-specific.

Throughout the 1990s, software company Esri released ArcView, a desktop solution for mapping systems. The influx of the Internet saw widespread adoption of GIS heading into the millennium,

and the technology reached governmental authorities. Many companies adopted the technology to provide services to cities, municipalities and private organizations to manage assets in the field, gather business intelligence, and easily send data to the company headquarters to analyze. In recent past, the focus of GIS shifted to sharing data across multiple platforms, and if its history is anything to go by, the industry will continue to debate how to resolve problems arising from data ownership.

4D mapping is a logical step for GIS, and this will help specifically for urban planning. GIS is not only essential for developing an urban area, but it is also necessary for today's fast-moving, technological landscape.

1.4 SUMMARY

GIS is the most powerful computerized tool for collection, storage, analysis, manipulation, modeling and retrieval of information. GIS is equally important like remote sensing for successful planning, management and optimization of land uses/resource utilization. In fact remote sensing based data/information and its analysis and modeling through GIS tools and techniques are complementary to each other for all kinds of developmental planning.

In this unit, the basics of GIS include GIS objectives, concept, definition need, scope, importance, history, components of GIS and GIS types. Generally, GIS is based on an integrating technology consisting of remote Sensing, cartography and mapping, GPS, Computers, RDMS, information technology, communication technology, survey and field data collection.

Geographic information systems are among the most exciting and powerful geomatics decision-making tools in the world. A GIS uses computer technology to integrate, manipulate and display a wide range of information to create a picture of an area's geography, environment and socio-economic characteristics. Beginning with a computerized topographic map as its base, a GIS *overlays and integrates* graphic and textual information from separate databases. The end result is a customized and reliable tool that can support decision making and problem solving and provide almost instantaneous answers to complex questions. The important characteristics of GIS is having its tool-kit which highlights to manipulate and analyse the data spatially and visualise it in the form of maps, tables, graphs, animations and virtual landscapes.

Geographic data refers to information about the earth's surface and the objects found on it. This information comes in three basic forms: spatial data, tabular data, and image data. Spatial data contains the locations and shapes of map features. Tabular data is collected and compiled for specific areas and is the descriptive data that GIS links to map features. Image data includes such diverse elements as satellite images, aerial photographs, and scanned data--data that's been converted from printed to digital format. In addition to data GIS components include hardware, software, users and methods.

GIS types are categorised as desktop GIS, professional GIS, enterprise GIS, web GIS, internet GIS, mobile GIS, distributed GIS, 4 d GIS and cloud GIS. GIS data types consists of attribute data and spatial data. Examples of attribute data are statistics, text, images and sound of a feature. Similarly spatial data indicate the existance of feature, objects and related phenomenon. The nature of spatial data is either vector or raster.

Historically, the concept of GIS started a long back during 1854 when British physician John Snow began mapping of the probable sites for Cholera outbreak disease locations and found the real locations along the water lines of London city. During the period of 1854 -1960, there were some concepts developed related to this subject but none of them had been satisfactory.

The actual concept of GIS was first introduced in the early 1960s, by Dr. Roger Tomlinson, Geographer who is generally recognized as the "father of GIS." Since that time, there had been tremendous developments in this field. Today the focus of GIS shifted to sharing data across multiple platforms, multiple numbers of Government and private organizations and even the critical problems are being solved through GIS and GIS software's like ArcView, *ERDAS Imagine etc.*

1.5 GLOSSARY

Altitude: Altitude is specified relative to either mean sea level (MSL) or an ellipsoid. In simplest term altitude is the vertical distance from mean sea level.

Analysis: Analysis is the process of identifying a question or issue to be addressed, modeling the issue, investigating model results, interpreting the results, and possibly making a recommendation.

ArcGIS: A comprehensive desktop GIS software package developed by ESRI.

ArcMap: Editing and map making module of ArcGIS.

ARC/INFO: Private domain complete GIS software package from ESRI, Inc. that has very powerful modeling, analysis and output capabilities.

ArcView: Private domain GIS software from ESRI, Inc. that allows users to organize, maintain, visualize, and disseminate maps and spatial information. This GIS software does not have the analysis and modeling capabilities like ARC/INFO.

Attribute: i) A characteristic of a feature in a Geographic Information System (GIS). Each identifiable feature has attributes. One common attribute of all geographic features is its position. other attributes depend on the type of feature. Example: a road may have a name or designation number, pavement type, width, number of lanes, etc. Each attribute has a range of possible values called its domain.

ii) Attribute is also called a column in a database table.

Attribute table: A tabular file containing rows and columns. In a GIS, attribute tables are associated with a class of geographic features, such as wells or roads. Each row

represents a geographic feature. Each column represents one attribute of a feature, with the same column representing the same attribute in each row.

Data type: The characteristic of columns and variables that defines what types of data values they can store. Examples include character, floating point and integer.

DBMS: Database management system. It is a set of computer programs for organizing the information in database. A DBMS supports the structuring of the database in a standard format and provides tools for data input, verification, storage, retrieval, query, and manipulation.

Geographic data: The locations and descriptions of geographic features. The composite of spatial data and descriptive data.

Geographic database: A collection of spatial data and related descriptive data organized for efficient storage and retrieval by many users.

Geographic feature: A user-defined geographic phenomenon that can be modeled or represented using geographic data sets in a GIS. Examples of geographic features include streets, sewer lines, manhole covers, accidents, lot lines, and parcels.

1.6 ANSWER TO CHECK PROGRESS

Q1- Write the definitions of GIS?

Q2- Write the objectives of GIS?

Q3- Write the Concept of GIS?

Q4- Write the need of GIS?

Q5- When was GIS started?

Q6- Write the Components of GIS?

Q7- Write the types of GIS?

1.7 REFERENCES

1. P. Fu, J Sun, 2011, Web GIS principles and applications, ESRI Press.

2. V Pessina , F Meroni , 2004, A web GIS tool for seismic hazard scenarios and risk analysis Geomatics Canada Web Site, 2000.

3. Pandey, Dharendra et al (2013). Int. Conference on Advances in Engineering & Technology - AET - 2013 - ISBN- 978 - 81 -927082-1-7

4. <https://www3.northern.edu/natsource/soils/GIS.htm>

1.8 TERMINAL QUESTIONS

- 1-What is GIS and what does it do?
- 2-How does Desktop GIS work?
- 3-What do you need to know about the data?
- 4-Explain the importance of GIS in the field of Efficiency based cost saving, Prompt and the best decision making and improved communication.
- 5-List the GIS types and differentiate between mobile and Internet GIS.
- 6-What are the different GIS components? Explain GIS software
- 7-Write the historical overview of GIS.
- 8-Both remote sensing and GIS are complementary to each other for efficient planning and decision making processes. Elaborate this statement in context with the role of information technology.

UNIT 2 - DATA MODELS

2.1 OBJECTIVES

2.2 INTRODUCTION

2.3 DATA MODELS

2.4 SUMMARY

2.5 GLOSSARY

2.6 ANSWER TO CHECK YOUR PROGRESS

2.7 REFERENCES

2.8 TERMINAL QUESTIONS

2.1 OBJECTIVES

After reading this unit you will be able to understand that:

- Concepts of data models.
- Implementation of raster data models in GIS applications.
- Implementation of vector data models in GIS applications.
- Advantages and limitations of raster and vector data.

2.2 INTRODUCTION

In the previous unit you have learnt about the basics of GIS which included GIS objectives, concept, definition, need, scope, importance, history; components of GIS and GIS types. GIS allows the user to understand geographic information in an easy manner without having to go through large volumes of confusing data that is in tabular form. Visualizing the geography of a particular location is no doubt easier than trying to analyze raw data. The use of modern GIS offers many advantages over paper maps. The importance of GIS is based on the facts that this system collects and stores information about the world as a collection of thematic layers that can be linked together by geography. This simple but extremely powerful and versatile concept has proven invaluable for solving many real-world problems. One of the most important components of GIS is the geospatial Data. In this unit we will discuss how the geospatial data collected from various means and from various sectors of earth's surface is represented in the computer.

Visualization of manmade or natural features/objects//phenomenon need to determine the best presentation of geographic space. Data models are a set of rules used to describe or represent aspects of real world landscape in a computer. You may also say a data model is a way of defining and representing real world surfaces and characteristics in GIS. The basic units of spatial data/information are points, lines and polygons. Data models belong to represent those geospatially basic units. A GIS data model enables a computer to represent the real geographical elements as graphical elements. The data models are classified into the categories of vector and raster data models. We may consider the following different aspects of data models:

A data model is also called an abstract model that organizes elements of data and standardizes how they relate to one another and to properties of the real world entities. For instance, a data model may specify that the data element representing a car be composed of a number of other elements which, in turn, represent the color and size of the car and define its owner.

The term data model is used in two distinct but closely related senses. Sometimes it refers to an abstract formalization of the objects and relationships found in a particular application domain, for example the customers, products, and orders found in a manufacturing organization. At other times it refers to a set of concepts used in defining such formalizations: for example concepts such as entities, attributes, relations, or tables. So the "data model" of a banking application may be defined using the entity-relationship "data model". This article uses the term in both senses.

The main function of any information system is to manage the large quantities of structured and unstructured data. Data models describe the structure, manipulation and integrity aspects of the data stored in data management systems. They typically do not describe unstructured data, such as word processing documents, email messages, pictures, digital audio, and video.

2.3 DATA MODELS

2.3.1 Data Model

- A data model in GIS is the technique of representing geographic objects or features as data.
- The data model in GIS is a mathematical construct for representing geographic objects or surfaces as data.
- Data model is based on the data itself, its relationship, semantic and constraint. Data model provides the details of information to be stored, and is of primary use when the final product is the generation of computer software code for an application or the preparation of a functional specification to aid a computer software make-or-buy decision.
- A data model specifically determines the structure of data. Data models are specified in a data modeling notation, which is often graphical in form.
- A data model can sometimes be referred to as a data structure, especially in the context of programming languages. Data models are often complemented by function models, especially in the context of .
- A geodata model is an abstract, artificially created data model. GIS data model should not only describe the content, properties and data structures, but also rules and relations between the entities of a data model. A proper data model should facilitate data updation possibility.

2.3.1a Raster Data Model

- The raster data model represent the geographical feature as cell matrices that store numeric values.
- A raster consists of a matrix of cells (or pixels) organized into rows and columns (or a grid). Each cell contains a value representing information.
- The method using dots, in which each pixel (picture element) is bitmap imaging, and is known as raster graphics.

2.3.1b Vector Data Model

- The vector data model represents geographical object/feature/points as collections of points, lines, and polygons.
- The vector data model represents geography as collections of points, lines, and polygons..The method using geometrical formulas for data model is known as vector or vectorised graphics. Vector storage implies the use of vectors (directional lines with their x,y coordinates) to represent a geographic feature.

2.3.2 Data model characteristics

A data model is a picture or description which depicts how data is to be arranged to serve a specific purpose. The data model depicts what data items are required, and how that data must look. There are various kinds of data models for various uses.

Models are used in a vast array of GIS applications, from simple evaluation to the prediction of future landscapes. In the past it has often been necessary to couple GIS with special software designed for high performance in dynamic modeling. But with the increasing power of GIS hardware and software, it is now possible to reconsider this relationship. Modeling in GIS raises a number of important issues, including the question of validation, the roles of scale and accuracy, and the design of infrastructure to facilitate sharing of models.

A data model is defined in to the following two ways –

i) A set of data into which the data needed for a particular application can be fitted. For example, a table is a very simple example of a data model, and in the way in which tables are often used in GIS, the rows of the table correspond to a group or class of real-world features, such as counties, lakes, or trees, and the columns correspond to the various characteristics of the features, in other words, the attributes. This table template turns out to be very useful because it provides a good fit to the nature of data in many GIS applications. In essence, GIS data models allow the user to create a representation of how the world looks.

ii) A model is a representation of one or more processes that are believed to occur in the real world—in other words, of how the world works.

A model is a computer program that takes a digital representation of one or more aspects of the real world and transforms them to create a new representation. Models can be static, if the input and the output both correspond to the same point in time, or dynamic, if the output represents a later point in time than the input. The common element in all of these models is the operation of the GIS in multiple stages, whether they are used to create complex indicators from input layers or to represent time steps in the operation of a dynamic process.

Static models often take the form of indicators, combining various inputs to create a useful output. For example, the Universal Soil Loss Equation (USLE) combines layers of mapped information about slope, soil quality, agricultural practices, and other properties to estimate the amount of soil that will be lost to erosion from a unit area in a unit time (Wischmeier and Smith 1978). Dynamic models, on the other hand, represent a process that modifies or transforms some aspect of the Earth's surface through time. Contemporary weather forecasts are based on dynamic models of the atmosphere; dynamic models of stream flow are used to predict flooding from storms; and dynamic models of human behavior are used to predict traffic congestion.

2.3.3 Data modeling

Data modeling is simply an act of exploring data-oriented structures. Data modeling can also be defined as the process of documenting a complex software system design as an easily understood diagram, using text and symbols to represent the way data needs to flow. The diagram can be used to ensure efficient use of data, as a blueprint for the construction of new software.

Figure 2.1 Data modelling –step-1



Figure 2.2 Data modelling –step-2

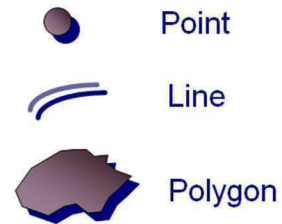
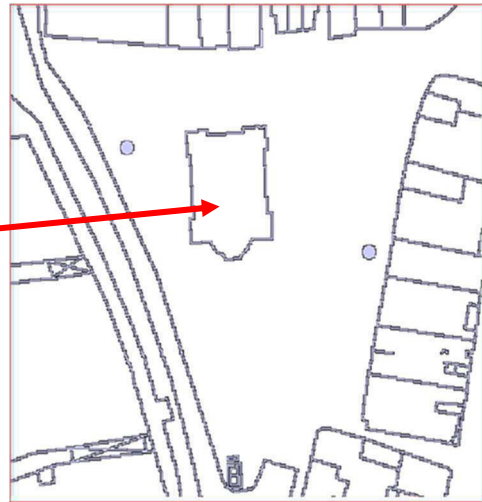


Figure 2.3 Data modelling –step-3



Feature: Building
 Object: Polygon
 Entity: Tourist Information Bureau



Data models are used for a variety of purposes, from high-level conceptual models to physical data models. From the point of view of an object-oriented developer data modeling is conceptually similar to class modeling. With data modeling you identify entity types whereas with class modeling you identify classes. Data attributes are assigned to entity types just as you would assign attributes and operations to classes. There are associations between entities, similar to the associations between classes – relationships, inheritance, composition, and aggregation are all applicable concepts in data modeling.

Traditional data modeling is different from class modeling because it focuses solely on data – class models allow you to explore both the behavior and data aspects of your domain, with a data model you can only explore data issues. Because of this focus data modelers have a tendency to be much better at getting the data “right” than object modelers. However, some people will model database methods (stored procedures, stored functions, and triggers) when they are physical data modeling. Steps for data modeling are shown in figures 2.1, 2.2 and 2.3.

2.3.4 Practical use of Data Models

We need to discuss how data models can be used in practice to better understand them. Following are three basic styles of data model:

- **Conceptual data models.** These models, sometimes called domain models, are typically used to explore domain concepts with project stakeholders.
- **Logical data models (LDMs).** LDMs are used to explore the domain concepts, and their relationships, of your problem domain. This could be done for the scope of a single project or for the entire enterprise. LDMs depict the logical entity types, typically referred to simply as entity types, the data attributes describing those entities, and the relationships between the entities.
- **Physical data models (PDMs).** PDMs are used to design the internal schema of a database, depicting the data tables, the data columns of those tables, and the relationships between the tables. PDMs often prove to be useful on both Agile and traditional projects and as a result the focus of this article is on physical modeling.

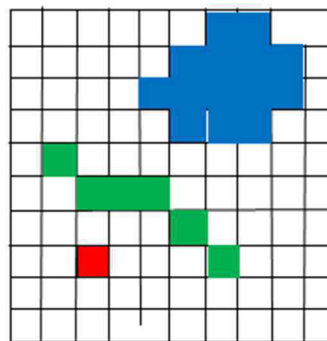
Data models are a set of rules to describe and represent aspects of the real world in a computer. Two primary data models are available to complete this task: raster data models and vector data models.

2.3.5 Raster data model

The raster data model is widely used in applications ranging far beyond geographic information systems (GISs). Most likely, you are already very familiar with this data model. Digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps contain raster data. The ubiquitous JPEG, BMP, and TIFF file formats (among others) are based on the raster data model. A digital image of remote sensing data itself a raster data model. If you zoom this digital image and see it deeply, you will notice that it is composed of an array of tiny square pixels (or picture elements). Each of these uniquely colored pixels, when viewed as a whole, combines to form a coherent image.

The raster data model consists of rows and columns of equally sized pixels interconnected to form a planar surface (Figure 2.4). These pixels are used as building blocks for creating points, lines, areas, networks, and surfaces illustrates how a land parcel can be converted to a raster representation. Although pixels may be triangles, hexagons, or even octagons, square pixels represent the simplest geometric form with which to work. Accordingly, the vast majority of available raster.

Figure 2.4 Raster Model - Image files composed of grid cells (pixels)



GIS data are built on the square pixel. These squares are typically reformed into rectangles of various dimensions if the data model is transformed from one projection to another.

Because of the reliance on a uniform series of square pixels, the raster data model is referred to as a grid-based system. Typically, a single data value will be assigned to each grid locale. Each cell in a raster carries a single value, which represents the characteristic of the spatial phenomenon at a location denoted by its row and column. The data type for that cell value can be either integer or floating-point alternatively, the raster graphic can reference a database management system wherein open-ended attribute tables can be used to associate multiple data values to each pixel. The advance of computer technology has made this second methodology increasingly feasible as large datasets are no longer constrained by computer storage issues as they were previously.

The raster model averages all values within a given pixel to yield a single value. Therefore, the more area covered per pixel, the less accurate the associated data values. The area covered by each pixel determines the spatial resolution of the raster model from which it is derived. Specifically, resolution is determined by measuring one side of the square pixel. A raster model with pixels representing 10 m by 10 m (or 100 square meters) in the real world would be said to have a spatial resolution of 10 m; a raster model with pixels measuring 1 km by 1 km (1 square kilometer) in the real world would be said to have a spatial resolution of 1 km; and so forth.

Care must be taken when determining the resolution of a raster because using an overly coarse pixel resolution will cause a loss of information, whereas using overly fine pixel resolution will result in significant increases in file size and computer processing requirements during display and/or analysis. An effective pixel resolution will take both the map scale and the minimum mapping unit of the other GIS data into consideration. In the case of raster graphics with coarse spatial resolution, the data values associated with specific locations are not necessarily explicit in the raster data model. For example, if the location of telephone poles were mapped on a coarse raster graphic, it would be clear that the entire cell would not be filled by the pole. Rather, the pole would be assumed to be located somewhere within that cell (typically at the center).

Imagery employing the raster data model must exhibit several properties. First, each pixel must hold at least one value, even if that data value is zero. Furthermore, if no data are present for a given pixel, a data value placeholder must be assigned to this grid cell. Often, an arbitrary, readily identifiable value (e.g., -9999) will be assigned to pixels for which there is no data value. Second, a cell can hold any alphanumeric index that represents an attribute. In the case of quantitative datasets, attribute assignment is fairly straightforward. For example, if a raster image denotes elevation, the data values for each pixel would be some indication of elevation, usually in feet or meters. In the case of qualitative datasets, data values are indices that necessarily refer to some predetermined translational rule.

The cell size is an important concept in raster data model. Within a GIS, the uses of raster data fall under four main categories:

2.3.5.1 Encoding Raster Data

Several methods exist for encoding raster data from scratch. Three of these models are as follows:

i) Cell-by-cell raster encoding:

This method encodes a raster by creating records for each cell value by row and column. This method could be thought of as a large spreadsheet wherein each cell of the spreadsheet represents a pixel in the raster image. This method is also referred to as “exhaustive enumeration.”

ii) Run-length raster encoding:

This method encodes cell values in runs of similarly valued pixels and can result in a highly compressed image file. The run-length encoding method is useful in situations where large groups of neighboring pixels have similar values. It is less useful where neighboring pixel values vary widely.

iii) Quad-tree raster encoding:

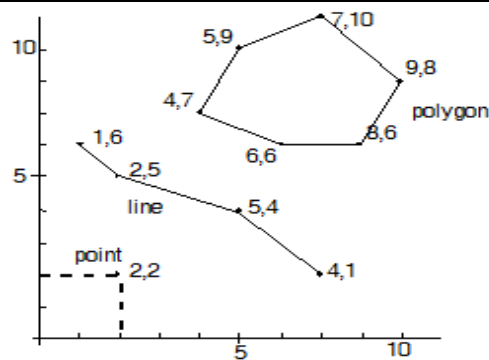
This method divides a raster into a hierarchy of quadrants that are subdivided based on similarly valued pixels. The division of the raster stops when a quadrant is made entirely from cells of the same value. A quadrant that cannot be subdivided is called a “leaf node.”

Following are some specific characteristics of raster model:

- **Rasters as base maps**
- Three main sources of raster base maps are orthophotos from aerial photography, satellite imagery, and scanned maps.
- **Rasters as surface maps**
Rasters are well suited for representing data that changes continuously across a landscape (surface). They provide an effective method of storing the continuity as a surface. They also provide a regularly spaced representation of surfaces.
- **Rasters as thematic maps**
Rasters representing thematic data can be derived from analyzing other data.
- **Rasters as attributes of a feature**
Rasters used as attributes of a feature may be digital photographs, scanned documents, or scanned drawings related to a geographic object or location.

2.3.6 Vector data model characteristics

Vector data is characterized by the use of sequential points called vertices. Vertex is a linear segment (distance and direction) - may be line or arc. Each vertex consists of an X coordinate and a Y coordinate. In vector representation, the storage of the vertices for each feature and connectivity between features, e.g. the sharing of common vertices are important. Points are represented by a single x, y pair. Points are used to model single, discrete features such as buildings, wells, power poles, sample locations, and similar items.

Figure 2.5 Vector model-Geometric objects (Points, Lines and Polygons)

In general, vector system is composed of two components: i) that manages spatial data and ii) manages thematic data. This is called as hybrid organization system. For all vector datasets, there is a scale and type dependency of spatial data. For example, depending on the scale and choice of presentation/map designing; an airport is represented as a point or a polygon.

2.3.6.1 Vector Data Structures

Vector data models can be structured in different ways. The simplest vector data structure is called the spaghetti data model. In the spaghetti model, each point, line and polygon (Figure 2.5) is represented as a string of X, Y coordinate pairs. In spaghetti data model, though the location associated with each line, spatial relationships are not clearly encoded. This reflects a lack of topological information, which is problematic if the user attempts to make measurements or analysis. In contrast to the spaghetti data model, the topological data model is characterized by the inclusion of topological information within the dataset, as the name implies.

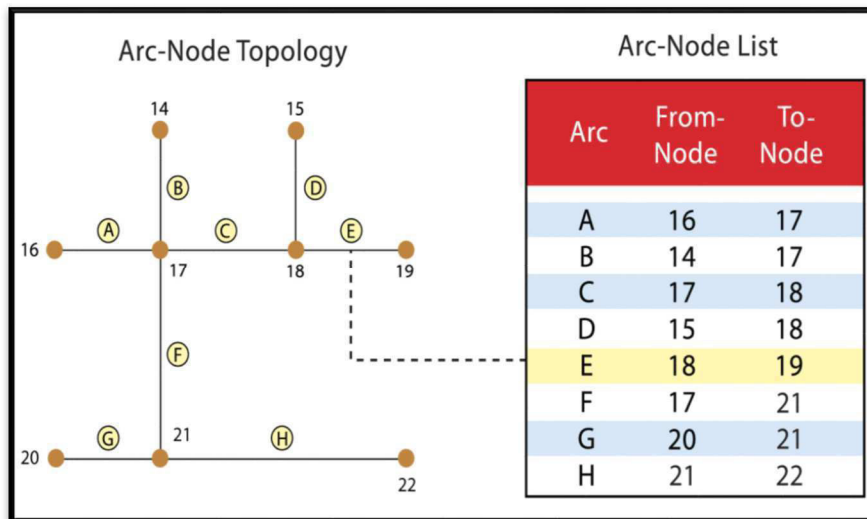
2.3.6.2 Concepts of Topology

Topology is a set of rules that model the relationships between neighboring points, lines, and polygons and determines how they share geometry. For example, consider two adjacent polygons. In the spaghetti model, the shared boundary of two neighboring polygons is defined as two separate, identical lines. The inclusion of topology into the data model allows for a single line to represent this shared boundary with an explicit reference to denote which side of the line belongs with which polygon. Topology is also concerned with preserving spatial properties when the forms are bent, stretched, or placed under similar geometric transformations, which allows for more efficient projection and reprojection of map files. Following are three topological precepts to understand the topological model.

i) Connectivity

It describes the arc-node topology for the feature dataset. In the topological data model, nodes are the intersection points where two or more arcs meet. In the case of arc-node topology, arcs have both a from-node (i.e., starting node) indicating where the arc begins and a to-node (i.e., ending node) indicating where the arc ends (Figure 2. 6). In addition, between each node pair is a line segment, sometimes called a link, which has its own identification number and references both its from-node to node.

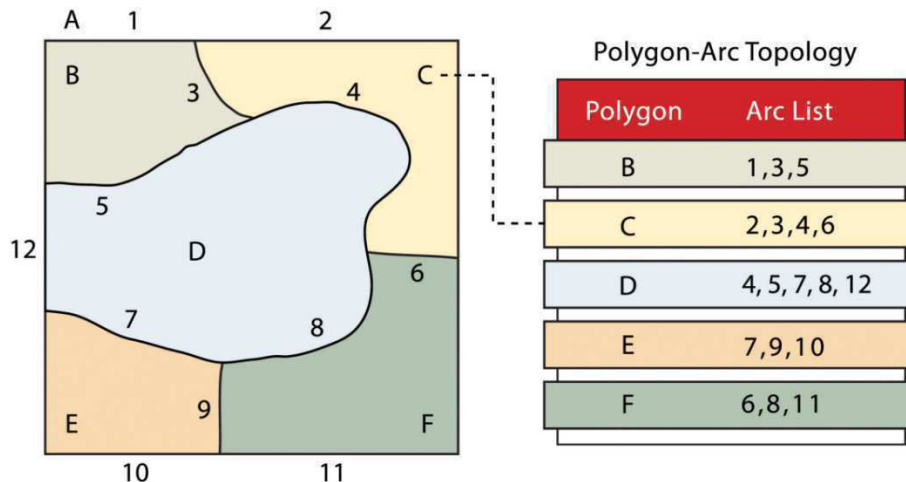
Figure 2. 6: Arc-Node Topology



ii) Area Definition

The second basic topological precept is area definition. Area definition states that an arc that connects to surrounding area defines a polygon, also called polygon-arc topology. In the case of polygon-arc topology, arcs are used to construct polygons, and each arc is stored only once (Figure 2.7). The figure shows that the polygon F is made up of arcs 8, 9, and 10.

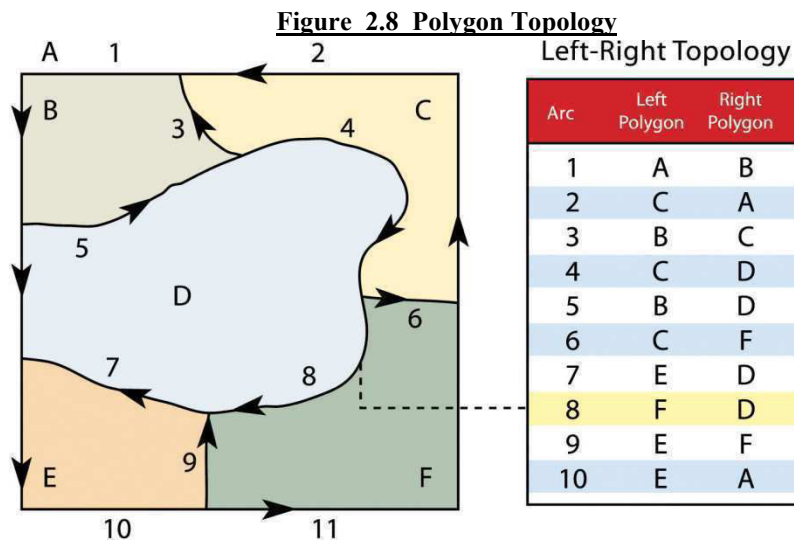
Figure 2.7: Polygon-arc Topology



iii) Contiguity

The concept of contiguity is based on the polygons that share a adjacent boundary. In a polygon topology, all arcs have a direction (node to-node), which allows adjacency information to be determined (Figure 2. 8). Polygons that share an arc are considered adjacent, or contiguous, and therefore the “left” and “right” side of each arc can be defined. This left and right polygon information is stored within the attribute information of the topological data model. The

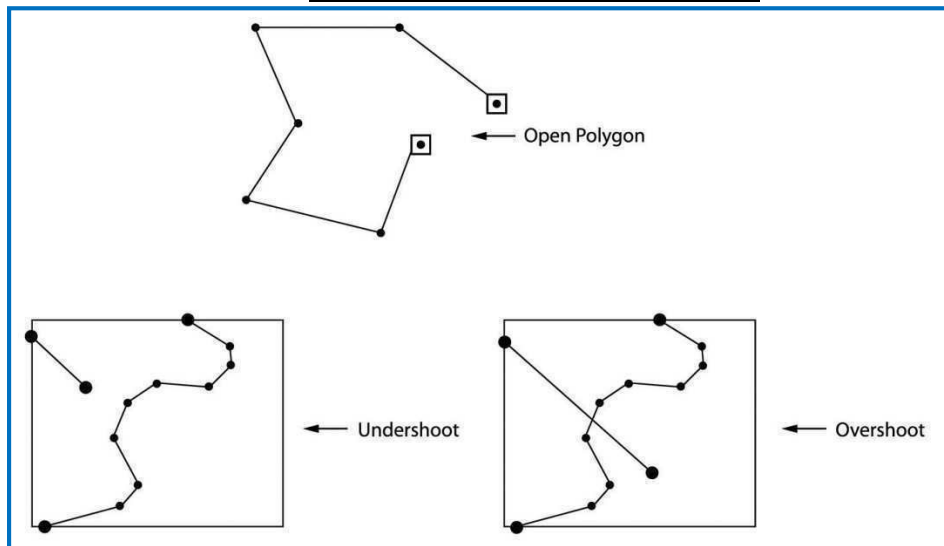
“universe polygon” is an essential component of polygon topology that represents the external area located outside of the study area. The figure shows that arc 6 is bound on the left by polygon B and to the right by polygon C. The universe polygon A is to the left of arcs 1, 2, and 3.



2.3.6.3 Topological Errors in GIS

Topology allows the computer (with GIS software) to rapidly determine and analyze the spatial relationships of all its included features. In addition, topological information is important because it allows for efficient error detection within a vector dataset. In the case of polygon features, open or unclosed polygons, which occur when an arc does not completely loop back upon itself, and unlabeled polygons, which occur when an area does not contain any attribute information, violate polygon-arc topology rules. Another topological error found with polygon features is the sliver. Slivers occur when the shared boundary of two polygons do not meet exactly (Figure 2.9).

In the case of line features, topological errors occur when two lines do not meet perfectly at a node. This error is called an “undershoot” when the lines do not extend far enough to meet each other and an “overshoot” when the line extends beyond the feature. The result of overshoots and undershoots is a “dangling node” at the end of the line. Dangling nodes aren’t always an error, however, as they occur in the case of dead-end streets on a road map.

Figure 2.9 Common Topological Errors

Many types of spatial analysis require the degree of organization offered by topologically explicit data models. In particular, network analysis (e.g., finding the best route from one location to another) and measurement (e.g., finding the length of a river segment) relies heavily on the concept of to- and from-nodes and uses this information, along with attribute information, to calculate distances, shortest routes, quickest routes, and so forth. Topology also allows for sophisticated neighborhood analysis such as determining adjacency, clustering, nearest neighbors, and so forth.

The basic concepts of topology explained in the above paragraphs provide full understanding of topological data model. In this model, the node acts as more than just a simple point along a line or polygon. The node represents the point of intersection for two or more arcs. Arcs may or may not be looped into polygons. Regardless, all nodes, arcs, and polygons are individually numbered. This numbering allows for quick and easy reference within the data model.

2.3.7. Advantages and disadvantages of raster and vector models

2.3.7.1 Advantages of Raster Model

The use of a raster data model confers many advantages like -

- i) The technology required to create raster graphics is inexpensive and ubiquitous. Nearly everyone currently owns some sort of raster image generator, namely a digital camera.
- ii) Raster graphics are having the relative simplicity of the underlying data structure. Each grid location represented in the raster image correlates to a single value (or series of values if attributes tables are included).
- iii) Because of simplicity of structure of raster data, it is exceptionally useful for a wide range of applications.
- iii) The simple data structure may also help explain why it is relatively easy to perform overlay analyses on raster data.

iv) Simplicity also lends itself to easy interpretation and maintenance of the graphics, relative to its vector counterpart.

v) A powerful format for advanced spatial and statistical analysis.

vi) The ability to represent continuous surfaces and perform surface analysis.

vii) The ability to uniformly store points, lines, polygons, and surfaces.

viii) The ability to perform fast overlays with complex datasets.

Despite the advantages, there are also the following disadvantages to use the raster data model.

2.3.7.2 Disadvantages of Raster Model

i) Raster files are typically very large.

ii) Raster images built from the cell-by-cell encoding methodology, the sheer number of values stored for a given dataset result in potentially enormous files.

iii) Any raster file that covers a large area and has somewhat finely resolved pixels will quickly reach hundreds of megabytes in size or more.

iii) These large files are only getting larger as the quantity and quality of raster datasets continues to keep pace with quantity and quality of computer resources and raster data collector (e.g., digital cameras, satellites).

iv) The output images are less “pretty” than their vector counterparts. This is particularly noticeable when the raster images are enlarged or zoomed. Depending on how far one zooms into a raster image, the details and coherence of that image will quickly be lost amid a pixilated sea of seemingly randomly colored grid cells.

v) The geometric transformations that arise during map re-projection efforts can cause problems for raster graphics to use the raster data model as the change in map projections will alter the size and shape of the original input layer and frequently result in the loss or addition of pixels.

vi) These alterations will result in the perfect square pixels of the input layer taking on some alternate rhomboidal dimensions. However, the problem is larger than a simple reformation of the square pixel. Indeed, the re-projection of a raster image dataset from one projection to another brings change to pixel values that may, in turn, significantly alter the output information.

vii) It is not suitable for some types of spatial analyses. For example, difficulties arise when attempting to overlay and analyze multiple raster graphics produced at differing scales and pixel resolutions. Combining information from a raster image with 10 m spatial resolution with a raster image with 1 km spatial resolution will most likely produce nonsensical output information as the scales of analysis are far too disparate to result in meaningful and/or interpretable conclusions.

viii) In addition, some network and spatial analyses (i.e., determining directionality or geo-coding) can be problematic to perform on raster data.

2.3.7.3 Advantages of Vector Data Structure/Model

- Vector data provide good representation of features/objects/phenomenon
- It is a Compact data structure
- Topology can be completely described
- Highlight accurate graphics
- Retrieval, updating and generalization of graphics and attributes are possible.
- Vector data are also in the form of images.
- Vectorised images describe every aspect of their shape in terms of a mathematical formula. We can consider this advantage of vector image over raster one on the basis of the following lines:
 - On a vector image, a simple mathematical formula would describe the radius of a circle and the processor could calculate the rest.
 - In contrast of the above, a raster image of the said circle will have to store the large number of pixels determined (on the basis of spatial resolution) within that circle.
- The above fact indicate that a vector image is manipulated much more easily – grown or shrunk, twisted and bent – without any distortion or loss of quality.
- Vector data also provides an increased ability to alter the scale of observation and analysis.
- Zooming deep into a vector image does not change the view of a vector graphic in the way that it does a raster graphic.
- Vector data tend to be more compact in data structure, so file sizes are typically much smaller than their raster counterparts.
- Topology is inherent in the vector model and this topological information helps in various simplified spatial analysis (e.g., error detection, network analysis, proximity analysis, and spatial transformation) when using a vector model.

2.3.7.4 Disadvantages of Vector Data Model

- The vector data structure tends to be much more complex than the simple raster data model.
- As the location of each vertex must be stored explicitly in the model, there are no shortcuts for storing data like raster models.
- The implementation of spatial analysis can also be relatively complicated due to minor differences in accuracy and precision between the input datasets. Similarly, the algorithms for manipulating and analyzing vector data are complex and can lead to intensive processing requirements, particularly when dealing with large datasets.

2.4 SUMMARY

Geography and GIS are based on spatial science. It is important to represent geographical features with respect to their geographic space or geospatial distribution. Data model is a way of defining and representing such features/ the real world surfaces and their characteristics

in GIS. The basic units of spatial data/information are points, lines and polygons. Data models belong to represent those geospatially basic units. A GIS data model enables a computer to represent the real geographical elements as graphical elements. It is evident that an information system manages the large quantities of structured and unstructured data. Data models describe the structure, manipulation and integrity aspects of the data stored in data management systems. They typically do not describe unstructured data, such as word processing documents, email messages, pictures, digital audio, and video.

In fact a model is a computer program that takes a digital representation of one or more aspects of the real world and transforms them to create a new representation. There is a difference between data model and data modeling. Where data models represent those geospatially basic units, data modeling is a process of documenting a complex software system design.

The data models are classified into the categories of raster and vector data models. The raster model represent the geographical feature as cell matrices that store numeric values. The vector data model represents geographical object/feature/points as collections of points, lines, and polygons. Both the Models are used in a vast array of GIS applications, from simple evaluation to the prediction of future landscapes. In the past it has often been necessary to couple GIS with special software designed for high performance in dynamic modeling. But with the increasing power of GIS hardware and software, it is now possible to reconsider this relationship. Modeling in GIS raises a number of important issues, including the question of validation, the roles of scale and accuracy, and the design of infrastructure to facilitate sharing of models.

Raster GIS data models are built on the square pixel. These squares are typically reformed into rectangles of various dimensions if the data model is transformed from one projection to another. The raster model averages all values within a given pixel to yield a single value. Therefore, the more area covered per pixel, the less accurate the associated data values. The area covered by each pixel determines the spatial resolution of the raster model from which it is derived. Specifically, resolution is determined by measuring one side of the square pixel.

Vector data is characterized by the use of sequential points called vertices. In vector representation, the storage of the vertices for each feature and connectivity between features, e.g. the sharing of common vertices are important. Points are represented by a single x,y pair. Points are used to model single, discrete features such as buildings, wells, power poles, sample locations, and similar items. The simplest vector data structure is called the spaghetti data model. In the spaghetti model, each point, line and polygon is represented as a string of X, Y coordinate pairs. In contrast to the spaghetti data model, the topological data model is characterized by the inclusion of topological information within the dataset, as the name implies. Topology is a set of rules that model the relationships between neighboring points, lines, and polygons and determines how they share geometry.

The main advantage of a raster data model is based on its powerful format for advanced spatial and statistical analysis and has the ability to perform fast overlays with complex datasets. But the disadvantage or limitation is also equally prominent keeping in view of its large files. These large files are only getting larger and the output images are less “pretty” than their vector

counterparts. Vector data model is a Compact data structure which provides good representation of features. However this data structure tends to be much more complex than the simple raster data model. And another main limitation highlights that the implementation of spatial analysis can be relatively complicated due to minor differences in accuracy and precision between the input datasets.

2.5 GLOSSARY

Arc: An arc consists of a string of vertices terminated by a node.

Line: A line is a collection of related points.

Point: Points are zero-dimensional objects.

Polygon: A polygon is a collection of related lines.

USLE: Universal Soil Loss Equation.

2.6 ANSWER TO CHECK PROGRESS

Q1- Write the definition of data models?

Q2- Write the concepts of data models?

Q3- Write the advantage of raster data?

Q4- Write the limitations of raster data?

Q5- Write the advantage of vector data?

Q6- Write the limitations of vector data?

2.7 REFERENCES

1. American National Standards Institute. 1975. *ANSI/X3/SPARC Study Group on Data Base Management Systems; Interim Report*. FDT (Bulletin of ACM SIGMOD) 7:2.
2. Dangermond and Dangermod 1982. "A Classification of Software Components Commonly used in Geographic Information Systems." In *Proceedings of the U.S.-Australia Workshop on the Design and Implementation of Computer-Based Geographic Information Systems*, 70–91. Honolulu, HI.
3. White, D. 2006. "Display of Pixel Loss and Replication in Reprojecting Raster Data from the Sinusoidal Projection." *Geocarto International* 21 (2): 19–22.
4. Seong, J. C. 2003. "Modeling the Accuracy of Image Data Reprojection." *International Journal of Remote Sensing* 24 (11): 2309–21.
5. Wlschmeler. W.H.. and Smith, D.D. 1978. Predicting rainfall-erosion losses-a guide to conservation planning. USDA Agr. Handbk. No. 537, 58p.
6. <https://www.slideshare.net/bala1957/four-data-models-in-gis>

2.8 *TERMINAL QUESTIONS*

- 1-What is data model in GIS and what is its role?
- 2-What do you need to know about the data modeling?
- 3-Explain the practical use of data modeling.
- 4-Explain raster data model and its encoding systems.
- 5-Define topology and elaborate its role in vector data modeling.
- 6-Describe the advantages and limitations of raster data model.
- 7-Explain which data model is better for overlaying and analysis under GIS domain.
- 8-What are the various topological Errors in GIS? Explain each of them diagrammatically.
- 9-Vector Data model is a compact data structure and provide good representation of features/ objects/ phenomenon. Elaborate this statement in comparison to raster data model.
- 10-Explain the specific advantages and limitations of vector model.

UNIT 3 - DATA QUERY

3.1 OBJECTIVES

3.2 INTRODUCTION

3.3 DATA QUERY

3.4 SUMMARY

3.5 GLOSSARY

3.6 ANSWER TO CHECK YOUR PROGRESS

3.7 REFERENCES

3.8 TERMINAL QUESTIONS

3.1 OBJECTIVES

The objectives of this unit are to acquaint, explain, discuss and describe the following:

- Database concept and importance
- Spatial data analysis and data query
- Data queries in detail
- Salient points of network analysis for querying

3.2 INTRODUCTION

In the previous unit you have learnt basics of GIS and GIS data, data model and modeling, types of GIS data, their nature and characteristics, role of topology in GIS data model, advantages and limitations of GIS data types. After studying this topics/sub a topics you must focus your attention for GIS data query as it is one of the most important function /operation within the GIS data base and GIS spatial and a spatial data analysis.

It is a well known fact that the GIS are not just a display tool. The geographic data connected to map which shows geospatial locations of various features, but this kind of analogue information does not have full computational /statistical values so as to extract any specific inferences. Features displayed in a GIS have two key components consisting of information describing their spatial location such as bounding latitudes and longitudes and the attribute data describing their non-spatial qualities. The non-spatial data in a GIS are presented in tables that make up a database linked to the map. The geographic features in the table are presented in horizontal rows, where each row represents a single record. The attributes of the features in the table are *listed in* vertical columns, with field names at the top. For example, in the Attribute Table of **Rivers** shown to the right, individual rivers are the geographic features that are listed in horizontal rows. The attributes of rivers include their length, name, and system and are listed in the vertical columns. In the image, the second record displaying the attributes of the Arkansas River is selected.

The power of geographic analysis is the ability to ask and answer questions about geographic features and their attributes and the relationship between them. This is what is known as a *Query* or selection. A query chooses a subset of records from the database. Queries may be simple, only searching for one feature, or more complex, selecting a group of features. Basic query operations may be combined to select complex sets of information. **Query or data query is one of the operations used in GIS to analyze data. While analyzing the spatial data we have to use the operations like attribute query, spatial query and alteration of original data. It is not necessary that we are bound to use all these operations. It depends on the type and objectives of study.** In this Unit you need to know database, database management, spatial data analysis and network analysis for understanding and learning the topic of data query under GIS domain. All these sub-topics are the vital parts of GIS. Spatial data analysis can be done in two ways. One is vector based and another one in raster based. You have already learnt about basic characteristics of raster and vector models. The analysis part is being dealt in this unit. Keeping in view of contents to be explained/discussed in this unit the following objectives have been outlined:

3.3 DATA QUERY

3.3.1 Data base

Geographical information, derived from thematic maps, field investigation or any other source commonly contains more than one attribute associated with its spatial location. The additional attributes like soil properties, tree species characteristics are called spatial or non-spatial database.

A spatial database describes a collection of entities, which may have permanent location on some global, dimensional space. Normally there is a mixture of spatial and non-spatial entity types. Spatial entity types have the basic topographical properties of location, dimension and shape. A spatial or non-spatial entity types do not have location. Thus, before the analysis can be done, the additional data need to be specified and incorporated in the geographical database.

3.3.1.1 GIS as a Database

In the past, attempts were made to build GIS from scratch using very limited tools like operating systems and compilers for computer programming language. More recently, GIS have been built around existing database management systems (DBMS). This is illustrated by the following:

- Purchase or lease of the DBMS becomes a major part of the system's software cost and
- The DBMS handles many functions which would otherwise have to be programmed into the GIS.

Any DBMS makes assumptions about the data which is handled. To make effective use of a DBMS, it is necessary to fit those assumptions. Certain types of DBMS are more suitable for GIS than others because their assumptions fit spatial data better than others.

3.3.2 Definition

A database is a collection of non-redundant data which can be shared by different application systems.

A database implies separation of physical storage from use of the data by an application program to achieve program/data independence. Using a database system, the user or programmer or application specialist need not know the details of how the data are stored and such details are "transparent to the user". Changes or updating can be made to data without affecting other components of the system. These changes include, for example, change of data format or file structure or relocation from one device to another.

3.3.3 Database Query

A Query, which makes the importance of database, is used to retrieve data from database. Generally, a query is defined as the action of retrieving data from database after its analysis/semi-analysis/manipulation. Sometimes, data retrieval is based on the selective procedure. Only that data is selected which fits a certain criteria based on the objectives already framed.

Database queries use set algebra and Boolean algebra. Set algebra uses the comparison operations less than, greater than, equal to and not equal to. An example of set algebra is, "Show all the records that are greater than or equal to X". Boolean algebra uses the words And, Or, Like and Not. Both of these types of queries focus on the features and attributes that are in the dataset. These selection criteria may be applied individually or in combination.

One may query a database either using i) the computer program or ii) via a user interface. While using computer programming, one may retrieve data from the database with programming language like SQL and use the SELECT statement. Using the SELECT statement, one can retrieve all records or just some of the records.

i) Using the computer program

Examples:

- a) *SELECT * FROM Albums; -- (for all the records)*
- b) *SELECT * FROM Albums --- (for some of the records)*
Where ArtistId = 1;

ii) User Interface

This type of query only returns records where the value in the ArtistId column equals 1. So if there are say, three albums belonging to artist 1, then three records would be returned.

SQL is a powerful language and the above statement is very simple. One can use SQL to choose a particular column to display. One may add further criteria, and can even query multiple tables at the same time. SQL is explained under the head-data query language.

Database management systems usually offer a "design view" for the queries. Design view enables to pick and choose which columns you want to display and what criteria you'd like to use to filter the data.

3.3.4 Advantages

There are several advantages of Database management system over file system. Few of them are as follows:

- **No redundant data:** Redundancy removed by data normalization. No data duplication saves storage and improves access time.
- **Data Consistency and Integrity:** As we discussed earlier the root cause of data inconsistency is data redundancy, since data normalization takes care of the data redundancy, data inconsistency also been taken care of as part of it
- **Data Security:** It is easier to apply access constraints in database systems so that only authorized user is able to access the data. Each user has a different set of access thus data is secured from the issues such as identity theft, data leaks and misuse of data.
- **Privacy:** Limited access means privacy of data.
- **Easy access to data** – Database systems manages data in such a way so that the data is easily accessible with fast response times.
- **Easy recovery:** Since database systems keep the backup of data, it is easier to do a full recovery of data in case of a failure.
- **Flexible:** Database systems are more flexible than file processing systems.

3.3.5 Spatial data analysis and data query

In addition to the basic functions related to automatic cartography and database management system, the most important uses of GIS are spatial analysis capabilities. As spatial information is organized in a GIS domain, it should be able to answer complex questions regarding space. Spatial analysis in GIS involves three types of operation:

- i) Attribute query or non-spatial query
- ii) Spatial query and
- iii) Generation of new data sets from the original database

The scope of spatial analysis ranges from a simple query about the spatial phenomenon to complicated combinations of attribute queries, spatial queries and alteration of original data.

3.3.5.1 Attribute Query

It requires the processing of attribute data exclusive of spatial information. In other words, it is a process of selecting information by asking logical questions.

Example: From a database of a city parcel map (Figure 3.1) where every parcel is listed with a land use code, a simple attribute query require the identification of all parcels for a specific land use type. Such a query can be handled through the table without referencing the parcel map. Because no spatial information is required to answer this question, the query is considered an attribute query. In this example, the entries in the attribute table (Table 3.1) that has land use codes identical to the specified type are identified.

Attribute queries are complex. Query expressions, like the one above, can be strict together to form long equations that could include any of the operators listed above and any number of existing attribute fields. Once the desired features are selected, you can perform a number of analytical processes on just the selected features, or, alternatively, you could save the highlighted features to a new layer. Attribute query is a vector process, but reclassification is a similar raster-based process.

3.3.5.2 Spatial Query

Spatial query involves selecting features based on location or spatial relationships, which requires processing of spatial information. For instance a question may be raised about parcels within one mile of the free way and each parcel. In this case, the answer can be obtained either from a hard copy map or by using a GIS with the required geographic information (Figure 3.1).

Figure: 3.1

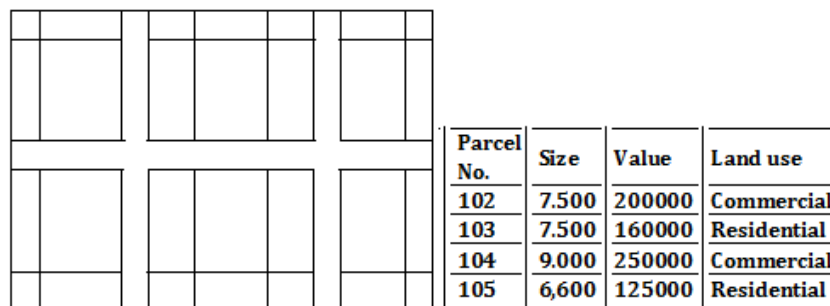


Figure 1: A sample parcel map

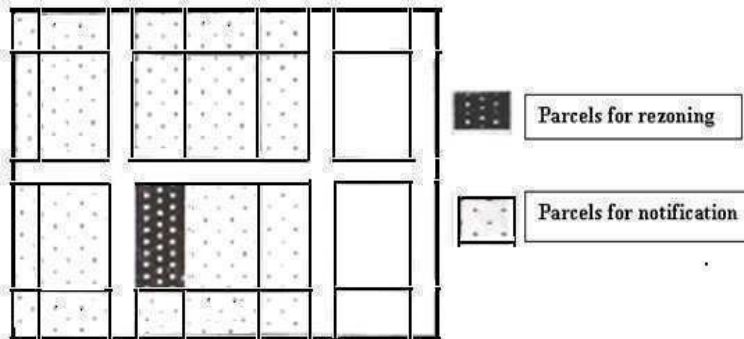
Table 1: Attributes of parcel map

Note: 1) Listing of parcel No. and value with land use = 'commercial' is an attribute query.

2) *Identification of all parcels within 100 m distance is a spatial query.*

Let us take one spatial query example where a request is submitted for rezoning, all owners whose land is within a certain distance of all parcels that may be rezoned must be notified for public hearing. A spatial query is required to identify all parcels within the specified distance. This process cannot be accomplished without spatial information. In other words, the attribute table of database alone does not provide sufficient information for solving problems that involve location.

Figure 3.2 Land owners within a specified distance from the parcels to be Rezoned, identified through spatial query



3.3.5.3 Generation of new data sets from the original database

While basic spatial analysis involves some attribute queries and spatial queries, complicated analysis typically requires a series of GIS operations including multiple attribute and spatial queries, alteration of original data and generation of new data sets. The methods for structuring and organizing such operations are major concern in spatial analysis. An effective spatial analysis is one in which the best available method are appropriately employed for different types of attribute queries, spatial queries and data alteration. The design of the analysis depends on the purpose of study.

3.3.5.4 GIS Uses in Spatial Analysis

GIS can interrogate geographic features and retrieve associated attribute information, called identification. It can generate new set of maps by query and analysis. Following are GIS operational procedure and analytical tasks that are particularly useful for spatial analysis:

- Single layer operations
- Multilayer operations/Topological overlay
- Spatial Modeling
- Geometric modeling which includes calculating the distance between geographic features; calculating area, length and perimeter and the geometric buffers.
- Point pattern analysis
- Network analysis
- Surface analysis
- Raster/Grid analysis
- Fuzzy spatial analysis
- Geostatistical tools for spatial analysis

All the above operations, analysis and modeling are described in the concerned book of GIS.

3.3.6 Data query definition, concepts, and languages

You have got the idea regarding the role of data query in the above paragraphs related to database, database management and spatial data analysis. Under this heading you will have the detailed study about data query. Access to robust search and query tools is essential to examine the general trends of a dataset. Queries are essentially questions posed to a database. You may also define a query is a request for data or information from a database table or combination of tables. This data may be generated as results returned by Structured Query Language (SQL) or as pictorials, graphs or complex results, e.g., trend analyses from data-mining tools. One of several different query languages may be used to perform a range of simple to complex database queries. SQL, the most well-known and widely-used query language, is familiar to most database administrators (DBAs).

3.3.6.1 Attribute Query

Attribute query is way to search for and retrieve records of features in set of data based on its attribute values. **Example:** In a country's map, name of states which contains "PRADESH"

3.3.6.2 Spatial Query

It involves selecting features based on location or spatial relationships, which require processing of spatial information. **Example:** Like areas within some distance
Spatial queries use spatial logic. For example, "Show all the X within a relationship to space Y". It is more like Venn diagram logic. These queries focus more on the physical relationships between datasets such as finding all the rivers that are within a given state or all the states that are adjacent to a river.

Following are three broad categories of spatial query

- Direction
- Distance
- Topology

3.3.6.3 Query Layer

A query layer is a layer or stand-alone table that is defined by a SQL query. Query layers allow both spatial and non-spatial information stored in a DBMS to be easily integrated into GIS projects within ArcMap. Since query layers are using SQL to directly query database tables and views, spatial information used by a query layer is not required to be in a geo-database.

When working in ArcMap, you create query layers by defining a SQL query. The query is then run against the tables and views in a database, and the result set is added to ArcMap as a layer or stand-alone table (depending on the query).

The query is executed every time the layer is displayed or used in ArcMap. This allows the latest information to be visible without making a copy or snapshot of the data and is

especially useful when working with dynamic information that is frequently changing. The query layers functionality works with any of the DBMSs supported by ArcGIS.

Query layers allow ArcMap to integrate data from geo-databases as well as from DBMSs. Therefore, query layers can quickly integrate spatial and non-spatial information into GIS projects independently of where and how that information is stored.

3.3.6.4 Query Language

Query languages or data query languages (DQLs) are computer languages used to make queries in databases and information systems. Query languages can be classified according to whether they are database query languages or information retrieval query languages. The difference is that a database query language attempts to give factual answers to factual questions, while an information retrieval query language attempts to find documents containing information that is relevant to an area of inquiry.

Specifically, SQL (Structured Query Language) is a commonly used computer language developed to query attribute data within a relational database management system. Created by IBM in the 1970s, SQL allows for the retrieval of a subset of attribute information based on specific, user-defined criteria via the implementation of particular language elements. More recently, the use of SQL has been extended for use in a GIS (Shekhar and Chawla 2003). One important note related to the use of SQL is that the exact expression used to query a dataset depends on the GIS file format being examined. For example, ANSI SQL is a particular version used to query ArcSDE geo-databases, while Jet SQL is used to access personal geo-databases. Similarly, shape files, coverages, and dBASE tables use a restricted version of SQL that doesn't support all the features of ANSI SQL or Jet SQL.

Query languages generate different data types according to function. For example, SQL returns data in neat rows and columns and is very similar to Microsoft Excel in appearance.

Other query languages generate data as graphs or other complex data manipulations, e.g., data mining, which is the deep analysis of information that uncovers previously-unknown trends and relationships between distinct or divergent data. For example, a SQL manufacturing company query may reveal that monthly sales peak in June and July, or that female sales representatives continually outperform male counterparts during holiday months.

3.3.6.5 Methods for Searching and Querying Attribute Data

The selective display and retrieval of information based on these queries are essential components of any geographic information system (GIS). There are three basic methods for searching and querying attribute data:

- i. **Selection,**
- ii. **Query by Attribute, and**
- iii. **Query by Geography.**

i) Selection

Selection represents the easiest way to search and query spatial data in a GIS. Selecting features highlight those attributes of interest, both on-screen and in the attribute table, for subsequent display or analysis. To accomplish this, one selects points, lines, and polygons

simply by using the cursor to “point-and-click” the feature of interest or by using the cursor to drag a box around those features. Alternatively, one can select features by using a graphic object, such as a circle, line, or polygon, to highlight all of those features that fall within the object. Advanced options for selecting subsets of data from the larger dataset include creating a new selection, selecting from the currently selected features, adding to the current selection, and removing from the current selection.

ii) Query by Attribute

An attribute query is the process of searching and retrieving records of features in a database based on desired attribute values. Such queries are a fundamental part of managing and analyzing GIS data. Typically, this is performed using a criteria-based query language, most commonly SQL.

In most cases, the desired information can be given as a set of criteria based on the available attributes, such as "having a code of 'CURRENT'." These criteria are formatted in the appropriate query language as a Boolean expression, which can be validated as either true or false for each record in the database. Individual criteria can be constructed and combined using logical operators such as comparisons (>, <, =, >=, <=), Boolean algebra (and, or, not), and functions (sin, cos, sqrt, etc.). An alternative to using attribute query is to perform a spatial selection.

Map features and their associated data can be retrieved via the query of attribute information within the data tables. For example, search and query tools allow a user to show all the census tracts that have a population density of 500 or greater, to show all counties that are less than or equal to 100 square kilometers, or to show all convenience stores within 1 mile of an interstate highway.

In Geospatial Database Management (GDBM), all attribute tables in a relational database management system (RDBMS) used for an SQL query must contain primary and/or foreign keys for proper use. In addition to these keys, SQL implements clauses to structure database queries. A clause is a language element that includes the SELECT, FROM, WHERE, ORDER BY, and HAVING query statements.

- SELECT denotes what attribute table fields you wish to view.
- FROM denotes the attribute table in which the information resides.
- WHERE denotes the user-defined criteria for the attribute information that must be met in order for it to be included in the output set.
- ORDER BY denotes the sequence in which the output set will be displayed.
- HAVING denotes the predicate used to filter output from the ORDER BY clause.

While the SELECT and FROM clauses are both mandatory statements in an SQL query, the WHERE is an optional clause used to limit the output set. The ORDER BY and HAVING are optional clauses used to present the information in an interpretable manner.

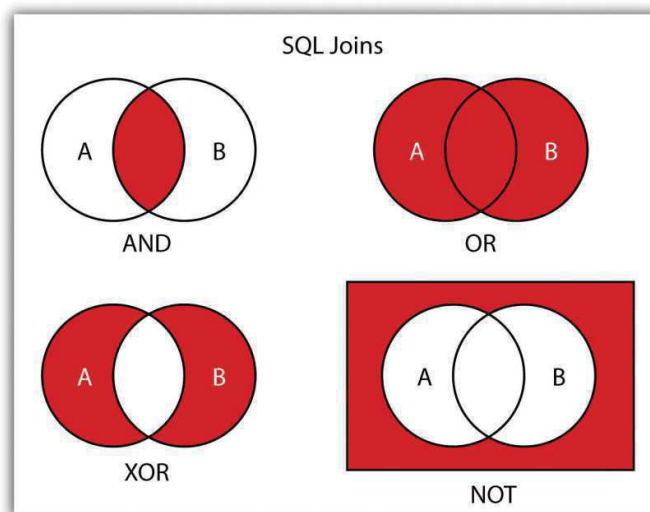
In addition to clauses, SQL allows for the inclusion of specific operators to further delimit the result of query. These operators can be relational, arithmetic, or Boolean and will typically appear inside of conditional statements in the WHERE clause. A relational operator employs the statements equal to (=), less than (<), less than or equal to (<=), greater than (>), or greater than or equal to (>=). Arithmetic operators are those mathematical functions that include addition (+), subtraction (-), multiplication (*), and division (/). Boolean

operators (also called Boolean connectors) include the statements AND, OR, XOR, and NOT. The AND connector is used to select records from the attribute table that satisfies both expressions. The OR connector selects records that satisfy either one or both expressions. The XOR connector selects records that satisfy one and only one of the expressions (the functional opposite of the AND connector). Lastly, the NOT connector is used to negate (or unselect) an expression that would otherwise be true. Put into the language of probability, the AND connector is used to represent an intersection, OR represents a union, and NOT represents a complement. Figure 3 "Venn Diagram of SQL Operators" illustrates the logic of these connectors, where circles A and B represent two sets of intersecting data. Keep in mind that SQL is a very exacting language and minor inconsistencies in the statement, such as additional spaces, can result in a failed query.

iii) Query by geography.

Query by geography, also known as a “spatial query,” allows one to highlight particular features by examining their position relative to other features. For example, a GIS provides robust tools that allow for the determination of the number of schools within 10 miles of a home. Several spatial query options are available, as outlined here. Throughout this discussion, the “target layer” refers to the feature dataset whose attributes are selected, while the “source layer” refers to the feature dataset on which the spatial query is applied. For example, if we were to use a state boundary polygon feature dataset to select highways from a line feature dataset (e.g., select all the highways that run through the state of Uttarakhand), the state layer is the source, while the highway layer is the target.

Figure 3.3: "Venn Diagram of SQL Operators"



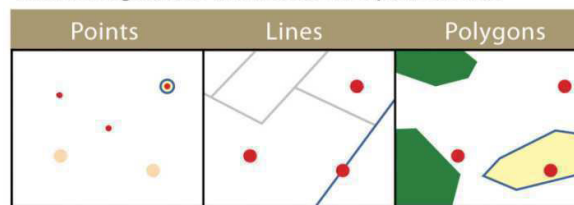
- **INTERSECT.** This oft-used spatial query technique selects all features in the target layer that share a common locale with the source layer. The “intersect” query allows points, lines, or polygon layers to be used as both the source and target layers (Figure 3.4).
- **ARE WITHIN A DISTANCE OF.** This technique requires the user to specify some distance value, which is then used to buffer the source layer. All features that intersect this

buffer are highlighted in the target layer. The “are within a distance of” query allows points, lines, or polygon layers to be used for both the source and target layers (Figure 3.5).

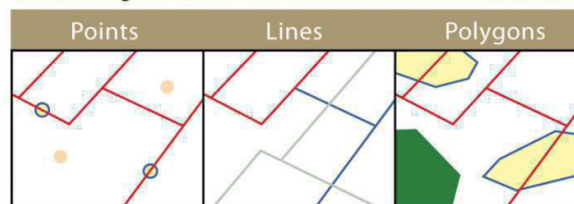
- **COMPLETELY CONTAIN.** This spatial query technique returns those features that are entirely within the source layer. Features with coincident boundaries are not selected by this query type. The “completely contain” query allows for points, lines, or polygons as the source layer, but only polygons can be used as a target layer (Figure 3.6).
- **ARE COMPLETELY WITHIN.** This query selects those features in the target layer whose entire spatial extent occurs within the geometry of the source layer. The “are completely within” query allows for points, lines, or polygons as the target layer, but only polygons can be used as a source layer (Figure 3.7).

Figure 3.4 Query by Intersect method

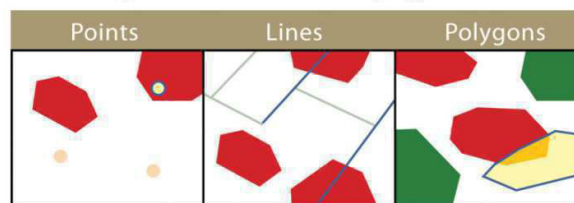
When finding features that intersect with point features



When finding features that intersect with line features



When finding features that intersect with polygon features

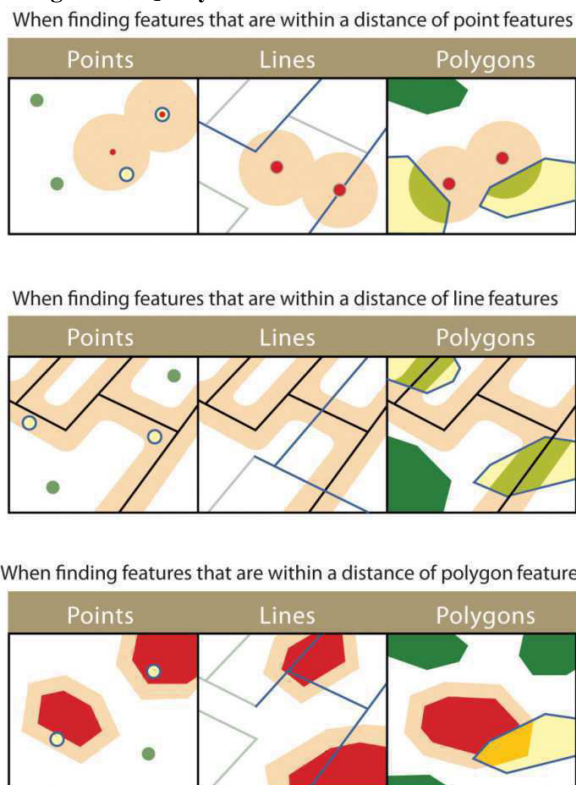


Note: The highlighted blue and yellow features are selected because they intersect the red features.

- **HAVE THEIR CENTER IN.** This technique selects target features whose center, or centroid, is located within the boundary of the source feature dataset. The “has their center in” query allows points, lines, or polygon layers to be used as both the source and target layers (Figure 3.8).
- **SHARE A LINE SEGMENT.** This spatial query selects target features whose boundary geometries share a minimum of two adjacent vertices with the source layer. The “share a line segment” query allows for line or polygon layers to be used for either of the source and target layers.

- **TOUCH THE BOUNDARY OF.** This methodology is similar to the INTERSECT spatial query; however, it selects line and polygon features that share a common boundary with target layer. The “touch the boundary of” query allows for line or polygon layers to be used as both the source and target layers.
- **ARE IDENTICAL TO.** This spatial query returns features that have the exact same geographic location. The “are identical to” query can be used on points, lines, or polygons, but the target layer type must be the same as the source layer type.
- **ARE CROSSED BY THE OUTLINE OF.** This selection criteria returns features that share a single vertex but not an entire line segment. The “are crossed by the outline of” query allows for line or polygon layers to be used as both source and target layers.
- **CONTAIN.** This method is similar to the COMPLETELY CONTAIN spatial query; however, features in the target layer will be selected even if the boundaries overlap. The “contain” query allows for point, line, or polygon features in the target layer when points are used as a source; when line and polygon target layers with a line source; and when only polygon target layers with a polygon source.
- **ARE CONTAINED BY.** This method is similar to the ARE COMPLETELY WITHIN spatial query; however, features in the target layer will be selected even if the boundaries overlap. The “are contained by” query allows for point, line, or polygon features in the target layer when polygons are used as a source; when point and line target layers with a line source; and when only point target layers with a point source.

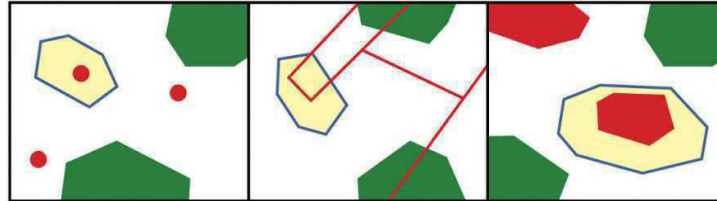
Figure 3.5 Query based on “are within a distance of”



Note: The highlighted blue and yellow features are selected because they are within the selected distance of the red features; tan areas represent buffers around the various features.

Figure 3.6 Query based on “completely contain”

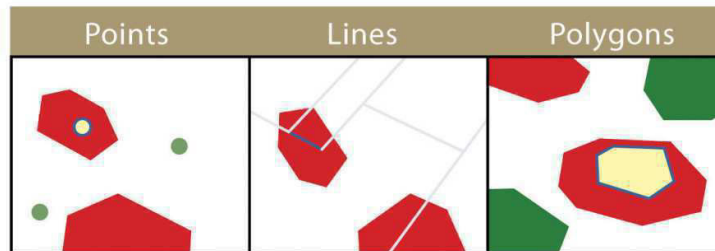
When finding point, line, and polygon features completely contained by polygon features



Note: The highlighted blue and yellow features are selected because they completely contain the red features.

Figure 3.7 Query based on “are completely within”

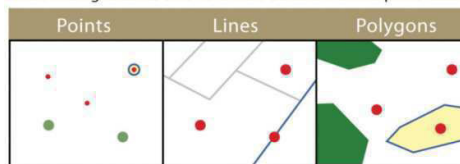
When finding features that are completely within polygon features



Note: The highlighted blue and yellow features are selected because they are completely within the red features.

Figure 3.8 Query based on “have their center in”

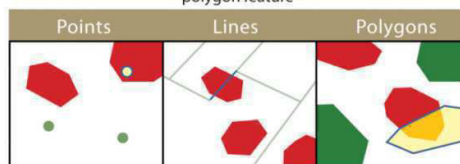
When finding features that have their centers within a point feature



When finding features that have their centers within a line feature



When finding features that have their centers within a polygon feature



Note: The highlighted blue and yellow features are selected because they have their centers in the red features.

3.3.7 Query to Retrieve Data

The ability to query and retrieve data based on some user defined criteria is a necessary feature of the data storage and retrieval subsystem. Data retrieval involves the capability to easily select data for graphic or attribute editing, updating, querying, analysis and/or display. The ability to retrieve data is based on the unique structure of the DBMS and command interfaces are commonly provided with the software. Most GIS software also provides a programming subroutine library, or macro language, so the user can write their own specific data retrieval routines if required.

Querying is the capability to retrieve data, usually a data subset, based on some user defined formula. These data subsets are often referred to as *logical views*. Often the querying is closely linked to the data manipulation and analysis subsystem. Many GIS software offerings have attempted to standardize their querying capability by use of a *Standard Query Language (SQL)*. This is especially true with systems that make use of an external relational DBMS. Through the use of SQL, GIS software can interface to a variety of different DBMS packages. This approach provides the user with the flexibility to select their own DBMS. This has direct implications if the organization has an existing DBMS that is being used for to satisfy other business requirements. Often it is desirable for the same DBMS to be utilized in the GIS applications. This notion of integrating the GIS software to utilize an existing DBMS through standards is referred to as corporate or enterprise GIS. With the migration of GIS technology from being a research tool to being a decision support tool there is a requirement for it to be totally integrated with existing corporate activities, including accounting, reporting, and business functions.

There is a definite trend in the GIS marketplace towards a generic interface with external relational DBMS's. The use of an external DBMS, linked via a SQL interface, is becoming the norm. Flexibility as such is a strong selling point for any GIS. SQL is quickly becoming a standard in the GIS software marketplace.

3.3.8 Salient points of network analysis for querying

Network analysis enables to solve problems, such as finding the most efficient travel route, generating travel directions, finding the closest facility, defining service areas based on travel time. **Examples:** Transportation network, Utility network.

Transportation network: Rivers, Roadways, Rail route etc

Utility network: Telecommunication line, sewage,

The points of network analysis have briefly been outlined below:

3.3.8.1 Network classification based on route system

i) Directed

Utility network is directed network. It is not allowed to take different path than selected path

ii) Undirected

Transportation network is undirected network where user is free to take decision

3.3.8.2 Modeling a Network

Data requirement for network analysis

- Attribute data
- Spatial data

3.3.8.3 Network characteristics

- Length
- Direction
- Connectivity
- Pattern

3.3.8.4 Network Analysis Operations

- Shortest path analysis
- Best route
- Closest facility
- OD-cost matrix
- Network partitioning

3.3.8.5 Network Analysis Best Routes

Best route can be the quickest, shortest, or most scenic route. It depends on the Impedances chosen like Time, Distance.

3.3.8.6 Applications of Network Analysis

- Routing
- Travelling salesperson problem
- Resource allocation
- Isochrones

3.4 SUMMARY

This unit introduces the foundations of data analysis in the context of query. It explains how data are queried and extracted within Geographic Information Systems (GIS). Attribute and topological query operations are also introduced. Spatial query refers to the process of retrieving a data subset from a map layer by working directly with the map features. In a spatial database, data are stored in attribute tables and feature/spatial tables. Structured query language (SQL) is a database “query” language designed for extracting information from relational databases. Although both raster and vector data use Boolean connectors, raster data query uses these connectors to isolate raster cell values from other cell values within one raster map and link the selected cell values to cell values of other raster layers. Spatial queries can only be performed on a single vector file as they can store multiple attributes simultaneously. The unit outlines, describes and explain the following lines:

- Concepts of Database and data query and language and the inputs of data query within those terms.

- The three basic methods for searching and querying attribute data are selection, query by attribute, and query by geography.
- SQL is a commonly used computer language developed to query by attribute data within a relational database management system.
- Queries by geography allow a user to highlight desired features by examining their position relative to other features. The eleven different query-by-geography options listed here are available in most GIS software packages.

The entire unit is based on the following sub-topics-

database concept and importance, spatial data analysis and data query, data queries in detail and salient points of network analysis for querying.

3.5 GLOSSARY

SQL - Structured Query Language

NoSQL- 1) A **NoSQL** database provides a mechanism for storage and retrieval of data that is modeled in means other than the tabular relations used in relational databases. 2) The **NoSQL** database, or Not Only SQL database, is an alternative to established SQL approaches and is used in web and cloud applications.

ANSI SQL or Jet SQL – 1) A specification, not a particular product. It's a document, describing the official features of the SQL language.2) ANSI (American Standard Institute) developed a common standard for SQL which runs in all modern vendor databases.

ArcGIS - A comprehensive desktop GIS software package developed by ESRI.

ArcMap - Editing and map making module of Arc GIS.

ARC/INFO -Private domain complete GIS software package from ESRI, Inc. that has very powerful modeling, analysis and output capabilities.

Attribute- 1. A characteristic of a feature in a Geographic Information System (GIS). Each identifiable feature has attributes. One common attribute of all geographic features is its position. Other attributes depend on the type of feature. Example: a road may have a name or designation number, pavement type, width, number of lanes, etc. Each attribute has a range of possible values called its domain. 2. A column in a database table.

Attribute Table -A tabular file containing rows and columns. In a GIS, attribute tables are associated with a class of geographic features, such as wells or roads. Each row represents a geographic feature. Each column represents one attribute of a feature, with the same column representing the same attribute in each row.

Boolean expression - A type of expression that reduces to a true or false (logical) condition. A Boolean expression contains logical expressions (e.g., DEPTH > 100) and Boolean operators. A Boolean operator is a keyword that specifies how to combine simple logical

expressions into complex expressions. Boolean operators negate a predicate (NOT), specify a combination of predicates (AND), or specify a list of alternative predicates (OR). For example, DEPTH > 100 AND DIAMETER > 20.

Column -The vertical dimension of a table. A column has a name and a data type applied to all values in the column.

Connectivity -The topological identification of connected arcs by recording the from- and to-node for each arc. Arcs that share a common node are connected.

Contiguity - The topological identification of adjacent polygons by recording the left and right polygons of each arc.

Coverage- 1. A digital version of a map forming the basic unit of vector data storage in ARC/INFO. Coverage stores geographic features as primary features (such as arcs, nodes, polygons, and label points) and secondary features (such as tics, map extent, links, and annotation). Associated feature attribute tables describe and store attributes of the geographic features. 2. A set of thematically associated data considered as a unit. Coverage usually represents a single theme such as soils, streams, roads, or land use.

Database- A logical collection of interrelated information managed and stored as a unit, usually on some form of mass- storage system such as magnetic tape or disk.

GIS database- includes data about the spatial location and shape of geographic features recorded as points, lines, areas, pixels, grid cells, or tins, as well as their attributes.

Database design - The formal process of analyzing facts about the real world into a structured database model. Database design is characterized by the following phases: requirement analysis, logical design and physical design.

Data model -1. The result of the conceptual design process. A generalized, user-defined view of the data related to applications. 2. A formal method of describing the behavior of the real-world entities. A fully developed data model specifies entity classes, relationships between entities, integrity rules and operations on the entities.

Database management system (DBMS) -A set of computer programs for organizing the information in a database. A DBMS supports the structuring of the database in a standard format and provides tools for data input, verification, storage, retrieval, query, and manipulation. Examples of database management systems include Access, SQLserver, Oracle, Informix, and DBASE.

Domain- The domain of a GIS attribute determines the set of possible values for that attribute. A domain can be set of numbers, characters, or strings. For example the domain for the attribute 'Pavement type' may be the set of strings ('Gravel','Bitumen', and Concrete).

Feature- In a GIS, a physical object or location of an event. Features can be points (a tree or a traffic accident), lines (a road or river), or areas (a forest or a parking lot).

Geodatabase- An object-based GIS data model developed by ESRI for ArcGIS that stores each feature as rows in a table. Personal geodatabases store data in a Microsoft Access .mdb file. Corporate geodatabases store data in a DBMS such as SQLserver or Oracle. This data structure supports rules-based topology and allows the user to assign behavior to data.

Geographic data - The locations and descriptions of geographic features. The composite of spatial data and descriptive data.

Geographic database - A collection of spatial data and related descriptive data organized for efficient storage and retrieval by many users.

Geographic feature- A user-defined geographic phenomenon that can be modeled or represented using geographic data sets in a GIS. Examples of geographic features include streets, sewer lines, manhole covers, accidents, lot lines, and parcels.

3.6 ANSWER TO CHECK PROGRESS

- 1- Explain definition, concepts and importance of database and database management?
- 2- Explain spatial data analysis types and methods.
- 3- Explain data queries, concepts and languages?
- 4- Explain Methods of Geographical data querying techniques.
- 5- Network analysis-salient points for querying.

3.7 REFERENCES

1. Exploring spatial analysis in Geographical Information System by Yue Hong Chou.
2. Fundamentals of Spatial Analysis by Robert Laurini and Derek Thompson.
3. Shashi Shekhar & Sanjay Chawla-2003, Spatial Databases: A Tour. Pub. Pearson, Apr18, 2019.
4. http://planet.botany.uwc.ac.za/nisl/GIS/GIS_primer/page_31.htm

3.8 TERMINAL QUESTIONS

- 1-Define database, database query and list out its advantages.
- 2-What is query language and structured query language (SQL). Explain the method of querying by attribute.
- 3- When querying by geography, what is the difference between a source layer and a target layer?
- 4-What is the difference between the CONTAIN, COMPLETELY CONTAIN and ARE CONTAINED BY queries?
- 5- List out the salient points of network analysis for finding the problems under GIS domain.

UNIT 4 - DATA OUTPUT AND VISUALISATION

4.1 OBJECTIVES

4.2 INTRODUCTION

4.3 DATA OUTPUT AND VISUALIZATION

4.4 SUMMARY

4.5 GLOSSARY

4.6 ANSWER TO CHECK YOUR PROGRESS

4.7 REFERENCES

4.8 TERMINAL QUESTIONS

4.1 OBJECTIVES

The objectives of this unit is to understand the role of GIS database management, analysis of GIS data and network analysis for data output and visualization. The topic is based on the following contents to be discussed:

- Definitions and concepts
- Database Management System
- GIS Analysis through the following functions/operations
 - Selection and Measurement
 - Overlay Analysis
 - Neighborhood Operation
 - Connectivity Analysis
- Network Analysis
- GIS data Visualization.
- Mapping Contents

4.2 INTRODUCTION

You have learnt about GIS basics, data models, data query and data input. These topics and related sub-topics have provided you full understanding, knowledge and expertise about the subject GIS. The last topic under this stream is GIS data output and visualization which will highlight the actual role and utility of GIS and geospatial data.

Data output and data visualization are the most important steps to consider during any GIS related themes. The way in which the data is visualized will directly impact audience's interpretation of the final product. Data output and Visualization elements include the results of data modeling, colorful maps, coloring charts, diagrams, map extent, labels, boundaries, interactivity, 3D models, varieties of charts, tables and more. Data output provides the geospatial information based on the GIS data analysis methods.

Geospatial data is any data that contains location information. Geographic information can come in a variety of forms & dash geographic coordinates such as latitude and longitude, street addresses, and others – and can be captured by a variety of methods – GPS, IP addresses, LIDAR, historical records, digitization, and others. In short, if data can be placed on a map, it can be characterized as geospatial data. Data output and visualization is the outcome of GIS spatial analysis and modeling. Spatial analysis encompasses everything we do with our geospatial data, from framing our research question to presenting our final results. The spatial analysis process involves using analytical techniques to examine our geospatial data and answer questions by highlighting or creating new information.

Due to the flexibility of GIS, spatial analysis can constitute one simple task or a series of complex tasks. A simple spatial analysis process might consist solely of visualizing data on a map for users to interpret, and a complex spatial analysis process can incorporate multiple datasets, spatial statistics, and Python scripts. The graphic on the right represents the spatial analysis process in general terms.

When working with geospatial data, it is important to have a fairly specific idea of what you want your final product to look like. Understanding your desired result makes it easier to

think about how you should manipulate your data and what spatial analyses you should conduct.

A GIS operation used to manipulate GIS data. A typical geo-processing operation takes an input dataset, creates database, performs an operation on that database and returns the result of the operation as an output dataset. Common geo-processing operations include geographic feature overlay, feature selection and analysis, topology processing, raster processing, and data conversion. Geo-processing allows for definition, management, and analysis of information used to form decisions.

GIS data output visualization/ mapping and data management. Database management services provide consolidation of information and improved understanding of each data set and how it relates to other data within a spatial and temporal context.

Geo-processing provides the tools and framework to manage and manipulate geospatial data. The main difference between geo-processing and spatial analysis is that geo-processing includes the **tools** for the analysis of data, and spatial analysis provides a detailed process of acquiring, analyzing, and presenting geospatial data. For example, we can use the "table join" geo-processing tool in Arc GIS to combine two tables and produce a new joined table. This new table is one of the many geo-processing tasks undertaken in a broader spatial analysis, which incorporates acquisition and visualization of that data.

4.3 DATA OUTPUT AND VISUALIZATION

4.3.1 Database Management System (DBMS)

- A database management system (DBMS) is system software for creating and managing database.
- The DBMS provides users and programmers with a systematic way to create, retrieve, update and manage data.
- A DBMS makes it possible for end users to create, read, update and delete data in a database.
- The DBMS essentially serves as an interface between the database and end users or application programs, ensuring that data is consistently organized and remains easily accessible.

4.3.2 Data base management system characteristics

The DBMS manages three important things: the data, the database engine that allows data to be accessed, locked and modified -- and the database schema, which defines the database's logical structure. These three foundational elements help provide concurrency, security, data integrity and uniform administration procedures. Typical database administration tasks supported by the DBMS include change management, performance monitoring/tuning and backup and recovery. Many database management systems are also responsible for automated rollbacks, restarts and recovery as well as the logging and auditing of activity.

The DBMS is perhaps most useful for providing a centralized view of data that can be accessed by multiple users, from multiple locations, in a controlled manner. A DBMS can limit what data the end user sees, as well as how that end user can view the data, providing many views of a single database schema. End users and software programs are free from

having to understand where the data is physically located or on what type of storage media it resides because the DBMS handles all requests.

The DBMS can offer both logical and physical data independence. That means it can protect users and applications from needing to know where data is stored or having to be concerned about changes to the physical structure of data (storage and hardware). As long as programs use the application programming interface (API) for the database that is provided by the DBMS, developers won't have to modify programs just because changes have been made to the database.

With relational DBMSs (RDBMSs), this API is SQL, a standard programming language for defining, protecting and accessing data in a RDBMS.

4.3.3 DBMS Types

The main DBMS types include the following:

i) **Relational database management system (RDMS)**

RDMS is adaptable to most of the cases, but its products are quite expensive.

ii) **NoSQL DBMS**

It is well-suited for loosely defined data structures that may evolve over time.

iii) **In-memory database management system (IMDBMS)** - provides faster response times and better performance.

iv) **Columnar database management system (CDBMS)**

It is well-suited for data warehouses that have a large number of similar data items.

v) **Cloud-based data management system (CBDMS)**

It is the cloud service provider and is responsible for providing and maintaining the DBMS.

4.3.4 Advantages of a DBMS

i) One of the biggest advantages of using a DBMS is that it lets end users and application programmers access and use the same data while managing data integrity.

ii) Data is better protected and maintained when it can be shared using a DBMS instead of creating new iterations of the same data stored in new files for every new application.

iii) The DBMS provides a central store of data that can be accessed by multiple users in a controlled manner. Central storage and management of data within the DBMS provides the following:

- Data abstraction and independence
- Data security
- A locking mechanism for concurrent access
- An efficient handler to balance the needs of multiple applications using the same data
- The ability to swiftly recover from crashes and errors, including restart ability and recoverability
- Robust data integrity capabilities

- Logging and auditing of activity
 - Simple access using a standard application programming interface (API)
 - Uniform administration procedures for data.
- iv) DBMS can be used to impose a logical, structured organization on the data.
- v) It delivers economy of scale for processing large amounts of data because it is optimized for such operations.
- vi) DBMS can also provide many views of a single database schema. A view defines what data the user sees and how that user sees the data.
- v) The DBMS provides a level of abstraction between the conceptual schema that defines the logical structure of the database and the physical schema that describes the files, indexes and other physical mechanisms used by the database.
- vi) When a DBMS is used, systems can be modified much more easily when business requirements change.
- vii) New categories of data can be added to the database without disrupting the existing system and applications can be insulated from how data is structured and stored.
- In fact, a DBMS perform additional work to provide these advantages. A DBMS uses more memory and CPU than a simple file storage system. Different types of DBMSes require different types and levels of system resources.

4.3.5 GIS data analysis

GIS data analysis process starts when the data input process is completed and the GIS layers are preprocessed. Analysis of geographic data requires thoughts, reasoning and the knowledge of the subject matter for which you are going to analyze data and bringing out the desired results. Critical thinking and reasoning. One should investigate and search for patterns, associations, inter-connections, interactions, and evidences of changes during the period and the over space. GIS helps to analyze the data sets, test for spatial relationships and evaluation of features for the particular judicious/optimum uses, but it does not replace the necessity, if you need it and think spatially without any reasoning. Before starting analysis and bringing outputs one must conceive of the possible spatial patterns and relationships.

Once the GIS layers are integrated you may make up your mind to ask questions based on the objectives and specific requirements.. Following is the sample examples:

“What is at...?”

“Where is it...?”

“What has changed since?”

“What spatial patterns exist?”,

and “What if...?” (The scenario question).

For what is it suitable.....? (Based on the parameters involved and the criterion already given).

The first two of these questions inventory features and minimally examine feature location and relationships. The last three questions are more complex. To answer these questions, you must use or string together some of the analytical functions that you will learn about in this chapter. The particular analytical functions—and their order—are up to you. The GIS

analysis for data output under this sub-topics focuses on the following GIS functions that assist to evaluate, estimate, predict, interpret, and understand spatial data.

- Selection and Measurement
- Overlay Analysis
- Neighborhood Operation
- Connectivity Analysis

4.3.5.1 Selection & Measurement

Selection is not an analysis function, but it is an important first step for many analysis functions. It is included because of its heavy use in the analytical phase. Following are two selection processes under this category:

- **Attribute Query (Boolean Selection)**

Attribute query selects features based on their attribute values. It involves picking features based on query expressions, which use Boolean algebra (and, or, not), set algebra ($>$, $<$, $=$, $>=$, $<=$), arithmetic operators ($=$, $-$, $*$, $/$), and user-defined values. Simply put, the GIS compares the values in an attribute field with a query expression that you define. For example, if you want to select every restaurant whose price is considered inexpensive, you would use a query expression like “PRICE = “INR” (where “PRICE” is the attribute field under investigation, “=” is the set algebra operator, and “INR” is the value). Your software then looks for a value equal to INR in the price field of each record, and selects only those records that satisfy the equation.

- **Spatial Selection (Spatial Searches)**

GIS provides very effective tools for generating maps and statistical reports from a database. However, GIS functionality far exceeds the purposes of mapping and report compilation.

While attribute queries select features by sorting through records in a data file, spatial selection chooses features from the map interface. In most cases, it selects features from one layer that fall within or touch an edge of polygon features in a second layer (or an interactively drawn graphic polygon).

There are many types of spatial selection. Point in polygon, perhaps the most used, selects the points of one layer if they are contained within a selected polygon (or polygons) of a different layer (or graphic). Line in Polygon, a similar operation, selects line features that are wholly or partially contained within a different layer’s polygon. Polygon in Polygon is another variation that selects polygon features within (or overlapping) selected polygons from a second layer. Another type of spatial selection is point distance (which has line and polygon versions too), which identifies all the points in one layer that are within a specified distance of a selected point(s) in a different layer. Like any type of selection, one can perform analytical processes on those highlighted features or save them to a new layer.

One can mix spatial selections and attribute queries. Here spatial features are selected first and then from the feature’s attribute file. It will reduce (or alternatively increase) the selected records through attribute query expressions.

Spatial selection is a vector process, but when one combines reclassification and overlay, the combination produces a raster version of spatial selection.

- **Measuring Distance**

There are many ways to measure distance. Most GIS programs, both raster and vector, have a ruler button that allows you to measure distances across a map. After clicking the button, you point on the map where you want to begin your distance measurement and then click at the ending point (or intervening points that define the path you want to measure).

Many vector-based systems measure distances along existing vector line networks, like streets, sewers, and railroads. This type of distance measurement relies on topological network relationships, which are discussed later (see Connectivity Analysis). In addition, some vector systems automatically generate length measurements for line features as you enter them. They store the length result in an attribute field within the layer's data file. Those systems that do not have this automatic function usually provide a way for you to calculate line feature length and store the result in an attribute field that you define. Once calculated and stored, you can sum the length of multiple line features by selecting them and calculating their sum.

Raster-based systems allow generating distance measurements in all directions away from a selected pixel or group of pixels. These distances are placed in a new layer where each cell's value represents the distance from that cell to the nearest selected pixel. These "distance" layers are often used for spread functions.

- **Measuring Area/Perimeter**

Many vector systems automatically generate area and perimeter measurements for polygon features and store these values in prescribed attribute fields. Those systems that do not have this automatic function do provide a way for you to generate area and perimeter and store the results in user-defined fields. Once calculated and stored, you can select multiple polygon features and sum their area and perimeter.

Calculating areas and perimeters are done differently in raster systems. Instead of measuring and storing each polygon's area and perimeter in the feature's pixels, raster systems already know the size—the area covered—by a single pixel. To calculate area, it simply adds up the number of pixels with a specified attribute and multiplies the count by the area contained in a single pixel. It is easy math. For example, a layer might have 100 polygons that possess one of twelve land cover categories. The routine finds each occurrence of the twelve categories (even if they are not contiguous) and sums the category's area and perimeter. Perimeter is usually equally easy if the pixels are square, and in the vast majority of cases they are. These measurements are provided either in standard tables or in new layers where the pixels exhibit the sums of the area and perimeter of the category to which it originally belonged.

- **Calculating Descriptive Statistics**

Descriptive statistics summarize attribute data. They reduce the complexities of numerous individual values into a few meaningful numbers that describe the individual features collectively. Descriptive statistics are organized into the following two groups:

- Measurement of central tendency and
- Measurement of dispersion.

Central tendency describes the center of the attribute data's distribution. The mean, median, and mode are its three common measures, but which measure one should use depends largely

on the attribute's level of measurement. Table 4.1 shows the three central tendency measures for the attribute values of a single field.

Table 4.1 Central tendency measures. These attribute values attained from a vector layer's attribute field or from selected pixels in a raster layer

Attribute values:	24,25,30,39, 40,45,45,45,45,48,50, 50,55,58, 60,61,65,65,65,70,72, 75,200,205
Mean	63.28
Median	50
Mode	45

- The most used measure is the mean, which is calculated by adding together each feature's attribute value and dividing the sum by the number of features. Like all measures of central tendency, the mean acts as a representative to be used to describe all the values within a single attribute field. This measure requires interval or ratio data.
 - If the attribute values are in ascending or descending order, the median is the middle score in the distribution (this works for an odd number of cases). In other words, half of the attribute values are above and half are below this value. In an even numbered distribution, the median is the average of the two middle scores. Median is used for ordinal and derived (aggregated) data.
 - Mode is the most frequent score in a distribution. Of course, some distributions do not have a mode if there are no repeated values. At times, the only repeated value might be at the low or high end of the distribution, making this measure a bit unreliable and certainly un-central. The measure, however, is helpful in describing leading categories (for instance the different political parties). It is the only measure for describing the central tendency of nominal data.
- Dispersion, the second group of descriptive statistics, looks at the attribute data's spread. Its measures (including range, variance, and standard deviation) describe how much the attribute values vary around the distribution's center (its central tendency measures). Are the values clustered tightly or are they spread out? These measures help you judge how well the central tendency characterizes all of the values in the attribute field. If the measure of dispersion is small, the values are clustered and the central tendency measure describes the distribution well. There are several types of measures of dispersion (Table 10.2).
- Counts and frequencies are not measures of dispersion, but they are basic ways to summarize data. Counting simply denotes quantity. Frequency is the number of times an attribute field has a particular value. A frequency distribution, usually in the form of a histogram, describes the shape (or structure) of the attribute data by tabulating the frequencies of each value (or range of values).
 - Range is the distance between the minimum and maximum attribute values. To derive it, simply subtract the minimum value from the maximum value. It is the simplest measure of dispersion, but it is vulnerable to outliers (rogue values that are significantly different from the rest of the attribute values). If you think outliers affect the range, use the interquartile range instead. It divides the distribution, arranged from low to high, into four parts each containing 25 percent of the attribute values, and it is the difference between the 25th & 75th percentiles.

•The variance looks at the difference between the distribution’s values and its central tendency measure (in this case the mean). It is more complex than computing the average difference that each attribute value falls from the mean. Such a score does not provide enough numeric emphasis to the attribute values on the low and high end of the distribution. The variance adjusts for this by squaring the difference, summing the squares, and dividing by the count.

•Standard deviation is the square root of the variance. Like the variance, it describes the dispersion around the mean and allows you to evaluate how closely the numbers in the dataset are packed around the mean (in other words, how well the mean describes or summarizes the set of numbers). Similarly, the smaller the number the tighter the values are clustered around the central tendency measure. Unlike the variance’s higher values, however, standard deviation uses numbers that are similar to the original dataset. Still, the two are essentially the same thing.

In vector systems, descriptive statistics are usually generated within the attribute file interface. In raster layers, menu commands process the descriptive statistics. Each attribute field can be summarized in its entirety or confined to selected records or pixels. Which descriptive statistics are calculated depends on the attribute data’s level of measurement (Table 4.3).

Table 4.2 Measures of Dispersion

Attribute values: 24,25,30,39, 40,45,45,45,45,48,50, 50,55,58, 60,61,65,65,65,70,72, 75,200,205	
Count	25
Range	205-24=181
Variance	1866.4416
Std. Dev.	43.2023

Table 4.3 Depicts which descriptive statistics can be used with different data levels

Types of Descriptive Statistics	Level of Measurement			
	Nominal	Ordinal	Interval	Ratio
Mode, Counts, Frequencies	yes	yes	yes	yes
Median, Minimum, Maximum, Range	No	yes	yes	yes
Mean, Variance, Std. Deviation	No	No	yes	yes

4.3.5.2 Overlay Analysis

Overlay is one of the most common and powerful GIS functions. It investigates the spatial association of features by “vertically stacking” feature layers to investigate geographic patterns and determine locations that meet specific criteria.

Today, there are many types of GIS overlay. Vector and raster models both perform overlay, but their overlay functions differ considerably and thus will be discussed separately.

- **Vector (Logical) Overlay**

Vector overlay predominantly overlays polygons in one layer over polygons in another layer, but it can also be used to overlay point or line features over polygon layers. Sometimes referred to as topological or logical overlay, it is conceptually and mathematically more demanding than raster overlay. There are three types of vector overlay operations:

Polygon on polygon is where one polygon layer is superimposed over another polygon layer to create a new output polygon layer. The resultant polygons may contain some or all of the attributes from the polygons in which they were created. Several types of polygon on polygon overlay exist, including intersection (A and B), union (A or B), and clip (A not B). These Boolean operators work both on the attribute table and the geography.

- Intersection computes the geometric intersection of all of the polygons in the input layers. Only those features that share a common geography are preserved in the output layer. Any polygon or portion of a polygon that falls outside of the common area is discarded from the output layer. The new polygon layer can possess the attribute data of the features in the input layers.

- Union combines the features of input polygon layers. All polygons from the input layers are included in the output polygon layer. It can also possess the combined attribute data of the input polygon layers.

- Clip removes those features (or portions of features) from an input polygon layer that overlay with features from a clip polygon layer. The clip layer acts as a cookie cutter to remove features (and portions of features) that fall inside the clip layer.

Point in polygon is where a layer of point features is superimposed over a layer of polygon features. The two layers produce a point layer that includes attributes from the surrounding input layer polygons. Alternatively, one can tally the number of point features falling within each polygon and store the sum as a new attribute in the polygon layer. Other point attributes can be aggregated (summed, averaged, etc.) and included as attributes in the polygon's data file. The transferring of attributes based on their geographic position is called a spatial join.

Line on polygon is similar to point in polygon, but lines are superimposed on polygons. This type of spatial join either appends polygon attributes to line features falling within them or counts and aggregates line attribute data to the polygon layer's data file.

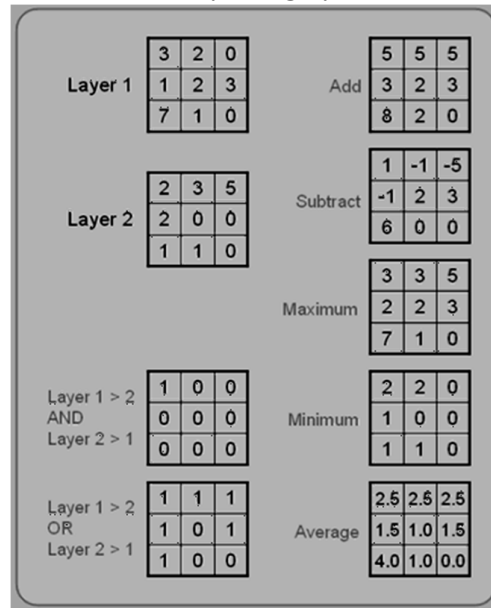
- **Raster (Arithmetic) Overlay**

Raster overlay superimposes at least two input raster layers to produce an output layer. Each cell in the output layer is calculated from the corresponding pixels in the input layers. To do this, the layers must line up perfectly; they must have the same pixel resolution and spatial extent. If they do not align, they can be adjusted to fit by the preprocessing functions. Once preprocessed, raster overlay is flexible, efficient, quick, and offers more overlay possibilities than vector overlay.

Raster overlay, frequently called map algebra, is based on calculations which include arithmetic expressions and set and Boolean algebraic operators to process the input layers to create an output layer. The most common operators are addition, subtraction, multiplication, and division, but other popular operators include maximum, minimum, average, AND, OR, and NOT. In short, raster overlay simply uses arithmetic operators to compute the corresponding cells of two or more input layers together, uses Boolean algebra like AND or

OR to find the pixels that fit a particular query statement, or executes statistical tests like correlation and regression on the input layers (Figure 4.1).

Figure 4.1 Raster Overlay. Using layers 1 and 2, all sorts of overlay are possible

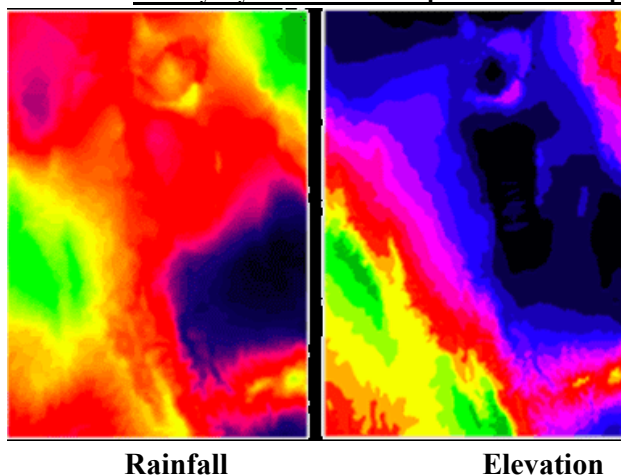


4.3.5.3 Correlation and Regression

Correlation and Regression are two ways to compute the degree of association between two (or sometimes more) layers. With correlation, one cannot assume a causal relationship. In other words, one layer is not affecting the spatial pattern of the other layer. The patterns may be similar, but no cause and effect is implied.

Regression is different in which any one can make the assumption that one layer (and its variable) influences the other. You specify an independent variable layer (sometimes more than one) that affects the dependent variable layer. Figure 4.2 shows a precipitation (dependent) and elevation (independent) as the layers.

Figure 4.2: Spatial relationship between these two layers? Correlation and regression tests allow to overlay layers to test their spatial relationship



4.3.5.4 Neighborhood Operations

Neighborhood operations, also called proximity analyses, consider the characteristics of neighboring areas around a specific location. These functions either modify existing features or create new feature layers, which are influenced, to some degree, by the distance from existing features. All GIS programs provide some neighborhood analyses, which include buffering, interpolation, The issen polygons, and various topographic functions.

- **Buffering**

Buffering creates physical zones around features. These “buffers” are usually based on specific straight-line distances from selected features (Figure 4.3). Buffers, common to both raster and vector systems, are created around point, line, or polygon features. The resulting buffers are placed in an output polygon feature layer. Once complete, buffer layers are used to determine which features (in other layers) occur either within or outside the buffers (spatial queries), to perform overlay, or to measure the area of the buffer zone. They are the most used neighborhood operation.

Figure 4.3 Buffering around a selected line feature

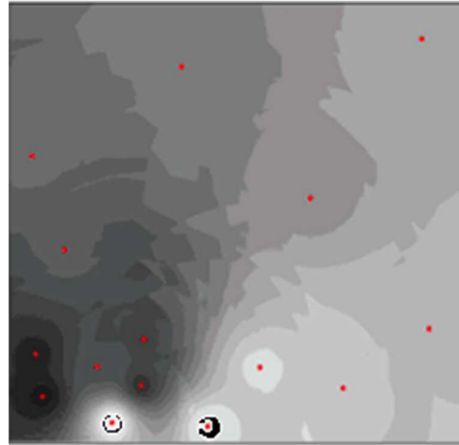


Most buffers use constant widths to generate zones, but some buffers have variable widths that depend on feature attribute values. As an example, the figure to the right depicts a variable distance buffer based on decibels from an airport.

- **Interpolation**

Interpolation is a method of predicting or estimating pixel values at unsampled locations based on the known values of neighboring pixels (Figure 4.4). Since it is impractical for you to take measurements at all locations across your study area due to money, time, legal, and physical constraints, you interpolate between known pixel values (sampled locations). With interpolation, one creates a continuous surface like elevation, temperature, and soil characteristics that occur everywhere. Because of its continuous nature, interpolation is only available within raster-based systems.

Figure 4.4: Interpolating between point features. The red dots are the points where values are known. The gray cells are the estimated data based on the known values.

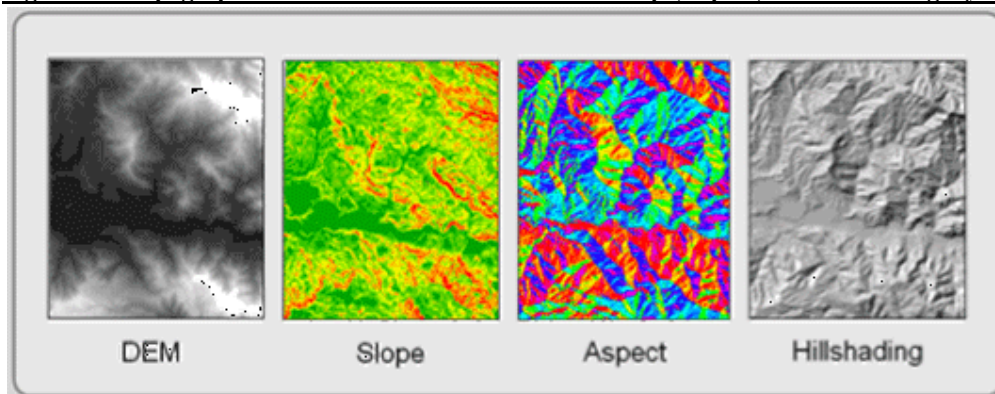


- **Topographic Functions**

Topographic functions use Digital Elevation Models (DEMs) to illustrate the lay of the land. DEMs are raster layers containing elevation data in each pixel. From these values, you produce output layers to portray slope (inclination), aspect (direction), and hill shading (Figure 4.5). These topographic functions are typical neighborhood processes; each pixel in the resultant layer is a product of its own elevation value as well as those of its surrounding neighbors.

- Slope layers exhibit the inclination or steepness of the land. It is the change in elevation over a defined distance.
- Aspect is the compass direction in which a slope faces. From north, it is usually expressed clockwise from 0 to 360 degrees.

Figure 4.5: Topographic Functions. The DEM creates the slope, aspect, and hill shading layers



- Hill shading, which is cartographically called shaded relief, is a lighting effect which mimics the sun to highlight hills and valleys. Some areas appear to be illuminated while others lie in shadows.

While these functions are raster processes, most can be mimicked in a vector environment by Triangulated Irregular Networks (TIN). In addition, topographic functions can derive vector isolines (contours).

- **Connectivity Analyses**

Connectivity analyses use functions that accumulate values over an area traveled. Most often, these include the analysis of surfaces and networks. Connectivity analyses include network analysis, spread functions, and visibility analysis. This group of analytical functions is the least developed in commercial GIS software, but this situation is changing as commercial demand for these functions is increasing.

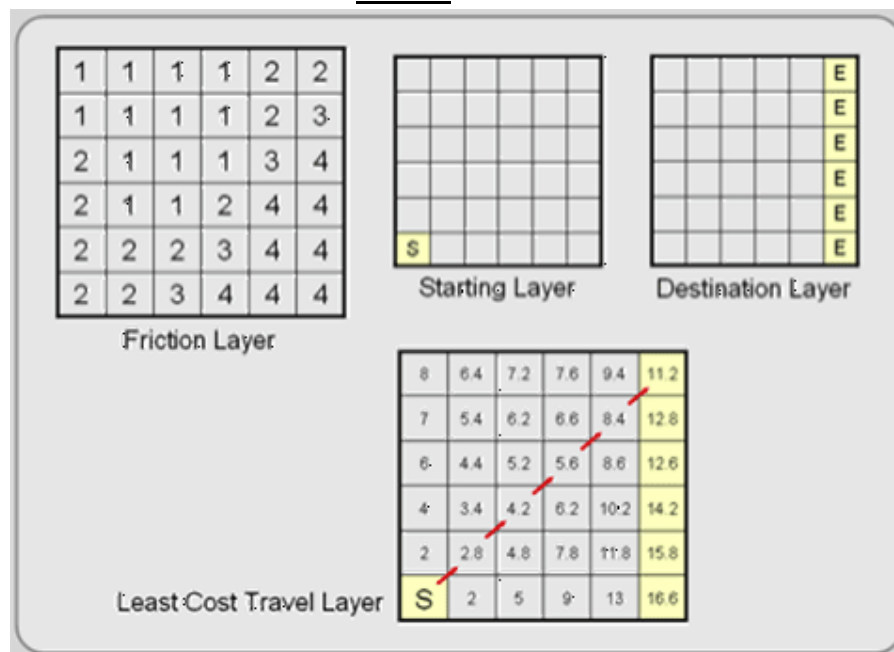
Vector-based systems generally focus on network analysis capabilities. Raster-based systems provide visibility analysis and sophisticated spread function capabilities.

- **Spread Functions (Surface Analysis)**

Spread functions are raster analysis techniques that determine paths through space by considering how phenomena (including features) spread over an area in all directions but with different resistances. You begin with an origin or starting layer (a point where the path begins) and a friction layer, which represents how difficult—how much resistance—it is for the phenomenon to pass through each cell. From these two layers, a new layer is formed that indicates how much resistance the phenomenon encounters as it spreads in all directions (Fig. 4.6).

Add a destination layer, and you can determine the “least cost” path between the origin and the destination. “Least cost” can be a monetary cost, but it can also represent the time it takes to go from one point to another, the environmental cost of using a route, or even the amount of effort (calories) that is spent.

Figure 4.6 Spread Functions. This example shows that the shortest distance is not always the least cost distance.



- **View shed Modeling (Inter-visibility Analysis)**

View shed modeling uses elevation layers to indicate areas on the map that can and cannot be viewed from a specific vantage point. The non-obscured area is the view shed. View sheds are developed from DEMs in raster-based systems and from TINs in vector systems. The ability to determine view shed (and how they can be altered) is particularly useful to national and state park planners and landscape architect

4.3.6 Network analysis

Network analyses involve analyzing the flow of networks—a connected set of lines and point nodes (sometimes called centers or hubs). These linear networks most often represent features such as rivers, transportation corridors (roads, railroads, and even flight paths), and utilities (electric, telephone, television, sewer, water, gas). Point nodes usually represent pickup or destination sites, clients, transformers, valves, and intersections. People, water, consumer packages, kilowatts, and many other resources flow to and from nodes along linear features.

Each linear feature affects the resource flow. For example, a street segment might only provide flow in one direction (a one-way street) and at a certain speed. Nodes can also affect flow. A stuck valve might allow too much of a resource to stream out and away from its intended destination. Network analysis tools help you analyze the “cost” of moving through the network. Like spread functions, “cost” can represent money, time, distance, or effort. Network analyses are vector-based applications, but there are similarities with raster-based spread functions.

Following are the three major types of network analyses which include route selection (shortest path), resource allocation, and network modeling.

i) Route Selection

Route selection attempts to identify the least “cost” route. As described above, cost can be defined a number of ways. You might want to find the shortest path between your home and a weekend destination or the least costly route that delivers UPS packages to their recipients. In any route selection routine, two or more nodes, including an origin and a destination point, must be identified and be able to be visited on the network. Sometimes there are a large number of possible routes. It is the job of the network analysis algorithm to determine the least cost route. Multiple paths are tested until the least cost path connects the starting and destination points.

ii) Resource Allocation

Resource allocation is the second major type of network analysis, involves the apportionment of a network to nodes. To do this, it defines one or more allocation nodes on the network. Territories of linear features, like streets, are defined around each of these allocation nodes. The linear features are usually assigned to the nearest node, where distance is measured in time, length, money, or effort

iii) Network Modeling

Network modeling uses interconnected linear features and point nodes to analyze how resources travel through networks. The linear features, like streets or river channels, have attributes that might define travel speed, number of lanes, and volume of flow. Nodes also have attributes that might identify vehicular turns and the time or cost required for each turn. Resources like water or traffic are placed in the network and their movement modeled. This way, problems with the network load can be identified.

4.3.7 GIS data visualization

GIS mapping services allow visualizing and understanding data, by identifying relationships, patterns and trends by providing a digital platform for viewing and processing layers of spatial information. It reflects that GIS clearly supports mapping and visualization. Maps of any discipline attributes or derivative attributes can be readily created in a GIS. Most GIS software packages support mapping symbols. The map overlay functionality of GIS, which allows maps to be stacked on top of each other, clearly enhances visualization of spatial geo-information derived from different GIS projects.

In GIS data output visualization, maps and related diagrams, charts, graphs etc are the most important part of the process of spatial data handling. They are used to visualize spatial data, to reveal and understand spatial distributions and relations. Recent developments such as scientific visualization and exploratory data analysis have had a great impact.

GIS have introduced the integration of spatial data from different kinds of sources, such as digital maps, documents, remote sensing, statistical databases, and recycled paper maps. Their functionality offers the ability to manipulate, analyze, and visualize the combined data. Their users can link application-based models to them and try to find answers to (spatial) questions. The purpose of most GIS is to function as decision support systems in the specific environment of an organization. Maps are important tools in this process. They are used to visualize spatial data, to reveal and understand spatial distributions and relations. However, maps are no longer the final products they used to be. Maps are now an integral part of the process of spatial data handling. The growth of GIS has changed their use, and as such has changed the world of those involved in cartography and those working with spatial data in general. This is caused by many factors which can be grouped in to the following three main categories.

i) Technological developments

In the fields of databases, computer graphics, multimedia and virtual reality has boosted interest in graphics and simulated sophisticated (spatial) data presentation. From this perspective it appears that there are almost no barriers left.

ii) User oriented developments

Often as an explicit reaction to technological developments, have stimulated scientific visualization and exploratory data analysis. Also, the cartographic discipline has reacted to these changes. New concepts such as dynamic variables, digital landscape models, and digital

cartographic models have been introduced. Map-based multimedia and cartographic animation, as well as the visualization of quality aspects of spatial data, is core topics in contemporary cartographic research. Tomorrow's users of GIS will require a direct and interactive interface to the geographical and other (multimedia) data. This will allow them to search spatial patterns, steered by the knowledge of the phenomena and processes being represented by the interface. One of the reasons for this is the switch from a data-poor to a data-rich environment, but it is also because of the intensified link between GIS and application-based models. As a result, an increase in the demand for more advanced and sophisticated visualization techniques can be seen.

According to Taylor (1994) the trend of visualization should be seen as an independent development that will have a major influence on cartography. In his view the basic aspects of cognition (analysis and applications), communication (new display techniques), and formalism (new computer technologies) are linked by interactive visualization. Following are three roles for data visualization:

i) Visualization is used to present spatial information. The results of spatial analysis operations can be displayed in well-designed maps easily understood by a wide audience. Questions such as 'what is?', or 'where is?', and 'what belongs together?' can be answered. The cartographic discipline offers design rules to help answer such questions through functions which create proper well-designed maps (Kraak and Ormeling 1996; Mac Eachren 1994a; Robinson et al 1994).

ii) Visualization is used to analyze and manipulate the known data. In a planning environment the nature of two separate datasets can be fully understood, but not their relationship. A spatial analysis operation, such as (visual) overlay, combines both datasets to determine their possible spatial relationship. Questions like 'what is the best site?' or 'what is the shortest route?' can be answered. What is required are functions to access individual map components to extract information and functions to process, manipulate, or summaries that information (Bonham-Carter 1994).

iii) Visualization is used to explore and to play with unknown data. Questions like 'what is the nature of the dataset?', or 'which of those datasets reveal patterns related to the current problem studied?', and 'what if . . .?' have to be answered before the data can actually be used in a spatial analysis operation. Functions are required to allow the user to explore the spatial data visually (Mac Eachren 1995; Peterson 1995).

4.3.7.1 Presenting Spatial Distributions

Maps are the unique powerful tools for the transfer of spatial information. Using a map one can locate geographical objects, while the shapes and colours of its signs and symbols inform us about the characteristics of the objects represented. Maps reveal spatial relations and patterns, and offer the user insight into the distribution of particular phenomena. Board (1993) defines the map as 'a representation or abstraction of geographical reality' and 'a tool for presenting geographical information in a way that is visual, digital or tactile'. Traditionally cartographers have concentrated most of their research efforts on enhancing the transfer of spatial data. This knowledge is very valuable, although some additional new concepts need to be introduced as illustrated in Figure 4.7. The traditional paper map functioned not only as an analogue database but also as an information transfer medium.

Today a clear distinction is made between the database and presentation functions of the map, known respectively as the Digital Landscape Model (DLM) and Digital Cartographic Model (DCM). A DLM can be considered as a model of reality, based on a selection process. Depending on the purpose of the database, particular geographical objects have been selected from reality, and are represented in the database by a data structure.

To visualize data in the form of a paper map requires a different approach to an onscreen visualization, and a road map for a vehicle navigation system will look different from a map designed for a casual tourist. Both, however, can be derived from the same DLM.

4.3.7.2 Map Contents

Next in importance to its contents, the usefulness of a map depends on its scale. For certain GIS applications one needs very detailed large-scale maps, while others require small-scale maps. Figure 4.8 shows a small-scale map (on the left) and a large-scale map. Traditionally maps have been divided into topographic and thematic types. Topographic maps portray the Earth's surface as accurately as possible subject to the limitations of the map scale. Topographic maps may include such features as houses, roads, vegetation, relief, geographical names, and a reference grid. Thematic maps represent the distribution of a particular phenomenon. Figure 4.9 shows the topography of the peak of Mount Kilimanjaro in Africa and the figure 4.10 shows the geology of the same area.

Figure 4.7 Spatial data characteristics: from reality to the map via a digital landscape model and a digital cartographic model

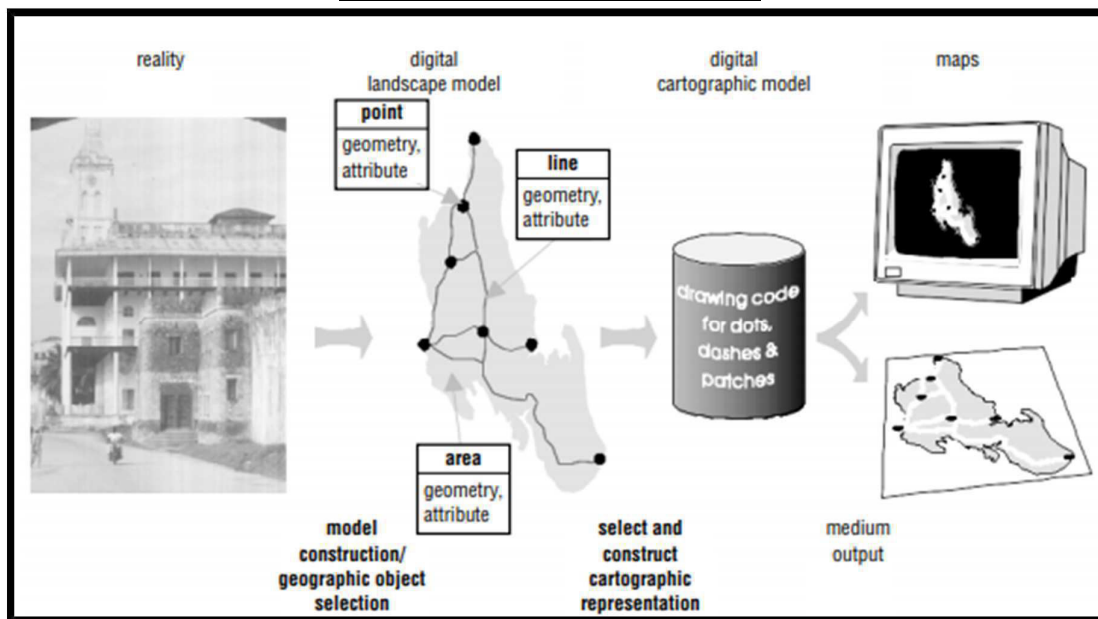


Figure 4.8 A small-scale map of East Africa, and a large-scale map of Stone Town (Zanzibar).

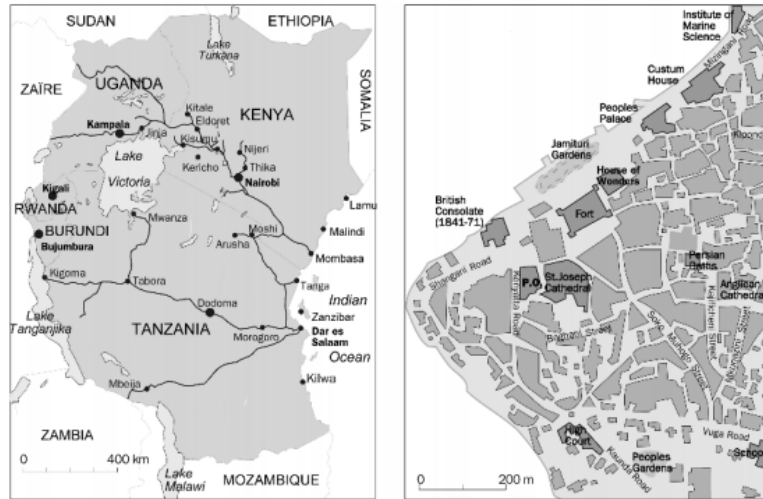


Figure 4.9 Topography of the peak of Mount Kilimanjaro in Africa

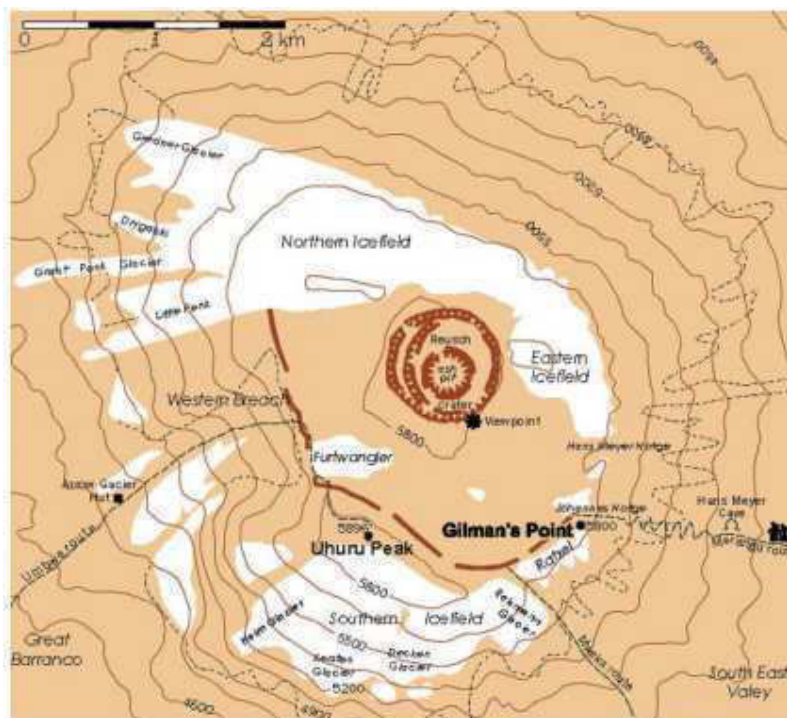
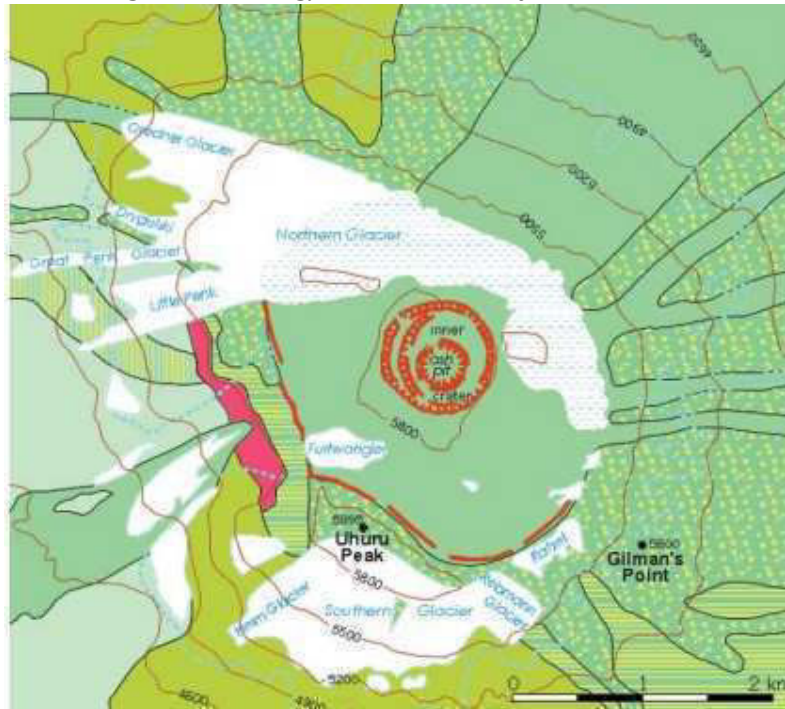


Figure 4.10 Geology of Mount Kilimanjaro in Africa

4.3.8 Common map elements and design attributes under GIS data visualization

i) Map Elements:

- Title
- Map Symbols
- Legend
- Map Scale
- North Arrow
- Map Area
- Labels
- Borders and Neatlines
- Graticule or Coordinate Grid

ii) Map Design:

- Colors and Shading
- Typography/Fonts
- Focus of attention

ii) Attributes

- Attribute tables
- Non-spatial information

4.4 SUMMARY

The topic entitled “GIS Data Output Analysis and Visualization” need the study of GIS data base , database management, GIS analysis, data query and network analysis.

Data output and visualization is the outcome of GIS spatial analysis and modeling. Spatial analysis encompasses everything we do with our geospatial data, from framing our research question to presenting our final results. The spatial analysis process involves using analytical techniques to examine our geospatial data and answer questions by highlighting or creating new information.

The Database Management System (DBMS) provides users and programmers with a systematic way to create, retrieve, update and manage data. I makes possible for end users to create, read, update and delete data in a database. It essentially serves as an interface between the database and end users or application programs, ensuring that data is consistently organized and remains easily accessible. In view of data output analysis and visualization, it acts as catalyst for providing such services as and when required.

After the completion of the processes i) GIS data input and storage ii) overlaying and integration of data layers and iii) database creation and management,; the GIS and network analysis and modeling are started for data output and visualization.. Analysis of geographic data requires thoughts, critical thinking and reasoning and the knowledge of the subject under consideration. The particular analytical functions and their order are up to the user`s choice. The analysis for data output under this sub-topics focuses on the GIS functions- i) Selection and Measurement ii) Overlay Analysis iii) Neighborhood Operation and iv) Connectivity Analysis. These functions assist to evaluate, estimate, predict, interpret, and understand spatial data.

Network analyses involve analyzing the flow of networks.. These linear networks most often represent features such as rivers, transportation corridors (roads, railroads, and even flight paths), and utilities (electric, telephone, television, sewer, water, gas). Point nodes usually represent pickup or destination sites, clients, transformers, valves, and intersections. People, water, consumer packages, kilowatts, and many other resources flow to and from nodes along linear features. There are three major types of network analyses which include route selection (shortest path), resource allocation, and network modeling. These are -i) Route Selection ii) Resource Allocation and iii) Resource allocation. Network modeling uses interconnected linear features and point nodes to analyze how resources travel through networks.

GIS data output includes all the models of spatial and attribute data and many of the other outputs based on the objectives of projects and themes. GIS data visualization, particularly includes GIS mapping services which allow to visualize and understand data, by identifying relationships, patterns and trends by providing a digital platform for viewing and processing layers of spatial information.

4.5 GLOSSARY

View shed - A view shed is the geographical area that is visible from a location.

Python scripts - Python script is an interpreter, high-level, general-purpose programming language.

IP address - An IP (Internet Protocol) Address is a numerical identifier for a computer or device that is participating in a computer network.

API - Application Programming Interface

4.6 ANSWER TO CHECK PROGRESS

Q1- What do you understand by database Management System?

Q2- Define Network analysis?

Q3- What do you mean by GIS data Visualization?

Q4- What do you understand by Map elements?

4.7 REFERENCES

1. Bonham-Carter G F 1994 Geographical information systems for geo-scientists: modeling with GIS. New York, Pergamon Press.
 2. Journal of Child Psychology, ed. Eric Talor, (1994) (<https://doi.org/10.1111/j.1469-7610.1994.tb01228.x>.)
 3. Kraak M-J, Ormeling F J 1996. Cartography, visualization of spatial data. Harrow, Longman.
 4. Mac Eachren A M 1994a. Some truth with maps: a primer on design and symbolization. Washington DC, Association of American Geographers.
 4. Mac Eachren A M 1995. How maps work. New York, Guilford Press.
 5. Peterson M P 1995. Interactive and animated cartography. Englewood Cliffs, Prentice-Hall.
 6. Robinson et al 1994. Robinson A H, Morrison J L, Muehrcke P C, Kimerling A J, Guptill S C 1994. Elements of cartography, 6th edition. New York, John Wiley & Sons Inc.
 7. <http://www.esri.com/software/arcgis/arcgisonline/features/visualization>.
-

4.8 TERMINAL QUESTIONS

- 1) Define database management and its role in data output and visualization.
- 2) List out the advantages of database management.
- 3) Explain Attribute Query (Boolean Selection).
- 4) List out the GIS analysis functions. Describe an overlay analysis.
- 5) How would you calculate Descriptive Statistics under GIS data analysis techniques?
- 6) Describe Vector Logical Overlay.
- 7) Explain Neighborhood Operations.
- 8) The growth of GIS has changed in the modern period of technological development. Elaborate this sentence in the context of recent GIS developments.

BLOCK 2: OVERVIEW ON RS & GIS SOFTWARE

UNIT 5 - INTRODUCTORY IDEA ABOUT RS SOFTWARE

5.1 OBJECTIVES

5.2 INTRODUCTION

5.3 INTRODUCTORY IDEA ABOUT RS SOFTWARE

5.4 SUMMARY

5.5 GLOSSARY

5.6 ANSWER TO CHECK YOUR PROGRESS

5.7 REFERENCES

5.8 TERMINAL QUESTIONS

5.1 OBJECTIVES

Over the years, there has been a significant trend toward open, collaborative software development. Open source software is software that is made available to the general communities. Any functions are performed by remote sensing applications, including: Detecting changes in photographs obtained at various times of the same location is, Orthorectification, Spectral Analysis -for example, determining the health of a forest using non-visible regions of the electromagnetic spectrum, Image categorization entails categorizing pixels into distinct land cover groups based on reflectance (e.g. Supervised classification, Unsupervised classification and Object Oriented classification). The wealth and security that advanced nations desire is built on a foundation of earth information. Urban planning, disaster preparedness, precision farming, climate treaty verification, efficient airline and ship routing, energy usage optimization, and natural resource access all need accurate and thorough information on the Earth's dynamic environment. This information, aside from fundamental weather forecasts, has long been the realm of academic and professional professionals. Raw observations were difficult to come by, tools for converting data into knowledge were difficult to use, and methods for disseminating outcomes were inefficient. Due to technology breakthroughs in sensors, processing ability to integrate and analyses sensor data, and communications capabilities for disseminating this knowledge, all of this has altered in the previous few years. As a result, mass-market consumer applications are emerging that will fundamentally alter how we utilize earth data during the next decade. Everything from schooling to planning is made possible by internet maps and virtual worlds such as Google Earth and Microsoft Virtual Earth, and all these developments occurs and in fact become possible due availability of various tools of remote sensing software's which clearly marked its objectives.

5.2 INTRODUCTION

The technique of identifying and monitoring an area's physical features by measuring its reflected and emitted radiation from a distance is known as remote sensing. Remote sensing is the science and art of getting knowledge about an object, region, or phenomena by analyzing data collected by a technology that is not in direct touch with the object, area, or event under exploration. You are using remote sensing while you read these words. The light reflected from this page acts as a sensor, and your eyes respond to it. Your eyes collect "data" in the form of impulses that correlate to the quantity of light reflected from dark and bright parts on the page. These data are examined (or interpreted) in your brain computer, allowing you to explain the dark spots on the page as a collection of letters that form words. You also realize that the words form sentences and that the information conveyed by the phrases is interpreted. In a variety of ways, remote sensing may be thought of as a reading process. We gather data remotely using a variety of sensors that may be analyzed to understand more about the objects, places, or phenomena under investigation. Remotely collected data includes variations in force distributions, acoustic wave distributions, and electromagnetic energy distributions, to name a few.

Geo-spatial technologies have mirrored web-based computing approaches toward the interoperable geo-web since the mid-2000s: software as a service, asynchronous particle update, mashup, open sources, rich user experiences, collaboration tagging, open interoperability, and structured information. Open source technologies, as well as their releases and uses in the geospatial field, are nearing maturity in both academia and industry.

5.3 INTRODUCTORY IDEA ABOUT RS SOFTWARE

The role of software in remote sensing and GIS technology is very significant. All the activities and functions performed by it with these efficient and highly capable software tools. However, there is a lot of software available for the enhancement of performance of these tools but most of them are Open Source remote sensing software. Though we can find also licensed software but the open source software plays greater role in remote sensing & GIS, some of the major open source software are-

- a) Arc GIS
- b) E Cognition
- c) E foto-
- d) ERDAS Imagin
- e) ENVI
- f) g vSIG
- g) Google Earth
- h) GRASS GIS
- i) IDRISI
- j) InterImage
- k) ILWIS
- l) OpenEV
- m) Opticks
- n) Orfeo tool box
- o) OSSIM PolSARPro
- p) QGIS
- q) SAGA GIS
- r) The Sentinel Toolbox
- s) SNAP
- t) SOCET SET
- u) TNTmips
- v) Whitebox GAT

Arc GIS

ESRI developed ArcGIS, a general-purpose GIS software system. It's a large and well-integrated software platform for creating operational GIS. A geographic information model for modelling aspects of the real world; components for storing and managing geographic information in files

and databases; a set of out-of-the-box applications for creating, editing, manipulating, mapping, analyzing, and disseminating geographic information; and a collection of web services that provide content and capabilities (data and functions) to networked software clients are all included in ArcGIS. Mobile devices, laptop and desktop PCs, and servers may all run parts of the ArcGIS software system. ArcGIS offers a broad range of features bundled in a basic set of menu-driven GIS apps that execute core geographic operations.

E Cognition

The goal of eCognition is to make geographic data interpretation easier, faster, and more automated. It enables remote sensing companies to extract precise and trustworthy geoinformation from any type of geospatial data fast and easily. E Cognition software has capabilities for a wide range of application domains, including Urban, Forestry, and Agriculture applications, as well as a number of use cases including Feature Extraction and Change Detection. The software can approach a picture in the same method that a human translator would approach an image analysis task, thereby simulating the cognitive ability of the human mind. Its one-of-a-kind method of converting your mental model - why you see something in the data and what you perceive - into computer-readable code known as a "rule set." The software can analyze not only single pixels but also picture objects. It is possible to integrate all powerful processing phases such as Image Object Creation (Segmentation), Knowledge Classification, Machine Learning, Template Matching, and Object Modification to create an analytical solution. It shows a high methodology and adaptability in mixing intelligent features and smart classifiers reveals the entire information potential of your data, resulting in high performance output.

E foto

In the category of open source remote sensing software, E-foto is the digital photogrammetry expert. The goal of the project is to create an instructional digital photogrammetric workstation. The GNU/GPL principles for free software are followed in this implementation. The E FOTO software was not designed to compete with commercial digital photogrammetry systems. Users may study the principles, tools, and algorithms contained in a commercial solution for a digital photogrammetric workstation in the E-FOTO environment. The major operations are: -

- Photo Triangulation.
- Stereoscopic Modelling.
- Digital elevation model extraction.

ERDAS Imagin

ERDAS Imagine is a raster-based software programmed dedicated to extracting data from images. ERDAS IMAGINE comes with a complete collection of tools for creating correct base imagery for use in GIS and ESRI Geodatabases. ERDAS IMAGINE includes capabilities including image orthorectification, mosaicking, reprojection, categorization, and interpretation that allow users to

study image data and show it in a number of formats, from printed maps to 3D models. It supports numerous processes and is the most powerful package for derived information likewise;

1. Map and report generation and printing through the map composer, PowerPoint, or Word
2. Land-cover mapping and terrain categorization
3. Spatial modeling and analysis
4. Data conversion
5. LiDAR editing and classification.
6. Color balancing, mosaicking, and compression
7. Orthorectification
8. Feature capture and update
9. Terrain creation, editing, and analysis

ENVI

ENVI is the ideal tool for visualizing, analyzing, and presenting digital images of any kind. ENVI's comprehensive image-processing solution includes complex yet simple spectrum tools, geometric correction, terrain analysis, radar analysis, raster and vector GIS capabilities, as well as significant support for pictures from a variety of sources. It may be used on a desktop, in the cloud, or on mobile devices, and it can be configured via an API to fit unique project requirements. It combines cutting-edge spectral image processing and image analysis technology with a simple, user-friendly interface to let users extract useful data from photos.

gvSIG

gvSIG is well-known for its extensive set of features. This is also relevant of its distant sensing capabilities. Users can utilize supervised classification like band algebra, and decision trees. It is easy to work in a variety of formats with gvSIG Desktop, vector and raster files, databases and remote services. Some of its diverse tool like-

- TASSLED CAP: Ideal tool for monitoring health of vegetation and urban growth.
- VEGETATION INDICES: It helps in analyzing chlorophyll and plant health for multispectral data

Google Earth

It is a computer application that creates a 3D picture of Earth based mostly on satellite images, previously known as Keyhole Earth Viewer. The application maps the Earth by superimposing satellite photos, aerial photography, and GIS data onto a three-dimensional globe, allowing users to view towns and landscapes from multiple perspectives. It simplifies the exploration of world by Entering addresses and coordinates, with the help of modern peripherals. it is also a Web Map Service client that can display many types of imagery layered on the earth's surface. In addition to Earth navigation, Google Earth's desktop programme offers a number of other features,

including a distance measurement tool. It includes photorealistic 3D images created using photogrammetry, as well as 3D architectural models in select cities. The imagery from Google Earth is shown on a digital globe that shows the planet's surface as a single composited image taken from a large distance.

GRASS GIS

GRASS GIS is one of the most widely used freeware GIS programme. It's been around for almost a quarter-century. It was designed by the United States Army Corps of Engineers as a land management and environmental planning tool. This open source software is now used by academics, government entities (NASA, NOAA, USDA, and USGS), and GIS practitioners since the code can be scrutinized and modified to their needs. GRASS GIS stored the processed data from several projections to be stored in different folders. The tools of grass gis are-

- Image classification and PCA.
- Edge detection.
- Radiometric correction.

Geomatica

It is a remote sensing desktop software suite for processing earth observation data. Geomatica 2014 is the most recent version of the programme. Geomatica is primarily designed to speed up data processing by allowing users to load satellite and aerial photos for sophisticated analysis. Many educational institutions and research projects throughout the world have utilised Geomatica to study satellite images and trends, including the GlobeSAR Program, which was sponsored by the Canada Centre for Remote Sensing in the 1990s.

IDRISI

Clark Labs at Clark University created IDRISI, an integrated geographic information system (GIS) and remote sensing software for the study and presentation of digital geospatial data. IDRISI is a PC grid-based system that provides tools for academics and scientists studying earth system dynamics in order to make more effective and responsible decisions on environmental management, resource development, and resource allocation.

InterImage

InterIMAGE employs a unique picture interpretation method. A technique based on and led by a semantic network-structured hierarchical explicit description of the interpretation problem. The InterIMAGE software's purpose is to analyse and classify remotely detected pictures automatically. Object-based image analysis is used to accomplish this (OBIA). To categorise objects, OBIA divides pixels into groups with comparable spectral, geometric, and spatial resolutions. The InterIMAGE programme allows users to establish personal data networks, allowing the software to identify images automatically based on the user's preferences.

ILWIS: Integrated Land and Water Information System

The Integrated Land and Water Information System (ILWIS) is a software programme for remote sensing and GIS that combines image, vector, and thematic data in a single, comprehensive application. Import/export, digitization, editing, analysis, and presentation of data, as well as the creation of high-quality maps, are all included in ILWIS. The ILWIS programme is well-known for its functionality, ease of use, and inexpensive cost, and it has amassed a large user base over the years. It is one of the most user-friendly vector and raster software programme available in the contemporary market.

OpenEV

OpenEV is a raster and vector geographic data viewing and analysis software library and application. Private firms, colleges, governments, and non-profit organizations all across the world use OpenEV. It's a frontend for a geospatial toolkit that's open-source, that provide both-

- An application for displaying and analyzing geospatial data.
- A developer library from creating new applications.

Opticks

Imagery, video (motion imagery), synthetic aperture radar (SAR), multi-spectral, hyper-spectral, raster math, radar processing and other forms of remote sensing data are all supported by Opticks. Opticks distinguishes itself from other remote sensing programme by supporting the processing of remote sensing video in the same way that it supports images. It may also be used as a development framework for remote sensing software. Opticks' plug-in design and accessible application programming interface allow developers to increase its features (API).

Orfeo tool box: Optical and Radar Federated Earth Observation

It's an open-source remote sensing project that was much demanded by time frame. It can handle high resolution optical, multispectral, and radar pictures at the terabyte scale and was built on the shoulders of the open-source geospatial community. The photos from Earth Observation Satellite can be processed easily by using OTB software library. There are several applications accessible, ranging from ortho-rectification and pan sharpening to categorization, SAR processing, and much more. It is a mixed bag of remote sensing tools like-

- Image segmentation, classification and filtering.
- Change detection
- Radiometry, PCA and pan sharpening

OSSIM: Open Source Software Image Map

OSSIM is a set of geospatial libraries and apps that may be used to handle images, maps, topography, and vector data. It has a lot of versatility. It can handle over 100 raster and vector files, as well as over 4000 different projections. The following are some of OSSIM's features:

1. MPI libraries provide parallel processing capabilities.
2. Sensor modelling that is meticulous.
3. Models of Universal Sensors (RPCs)

4. Support for a wide range of map projections and datums
5. Image chains based on parameters that are non-destructive
6. Precision Correction of the terrain and orthorectification
7. Mosaicking, compositing, and fusions at a higher level
8. Tonal balance and histogram matching

PolSARPro

PolSARpro (Polarimetric SAR Data Processing and Teaching Toolbox) is a tool for high-level education in radar polarimetry and promotes scientific exploitation of polarimetric SAR data. The programme includes a collection of well-known algorithms and tools, lays the groundwork for the use of polarimetric approaches in scientific study, and encourages research and application development. This programme can handle SAR from satellites with dual and full polarization, such as:

- ENVISAT-ASAR
- RADARSAT-2
- Terra SAR-X
- ALOS-PALSAR

QGIS: Semi-Automatic Classification Plugin (SCP)

QGIS is a geographic information system (GIS) programme that allows users to analyze and update geographical data as well as create and export graphical maps. Vector data is saved as point, line, or polygon features in QGIS; raster data is stored as point, line, or polygon features. The programme can georeference pictures and supports many raster image formats. Shapefiles, coverages, personal geodatabases, dxf, MapInfo, PostGIS, and more formats are all supported by QGIS. Web services, such as the Web Map Service and the Web Feature Service, were also supported, allowing data from external sources to be used. Multiple layers can be displayed in QGIS, each comprising distinct sources or visualizations of sources. Print Layout is used to create a printed map using QGIS. It may be used to add numerous map views, labels, legends, and other elements to a map. Other open-source GIS software that QGIS connects with include PostGIS, GRASS GIS, and Map Server.

SAGA GIS: (System for Automated Geoscientific Analyses)

The System for Automated Geoscientific Analyses (SAGA GIS) is a computer software that edits geographical data in a geographic information system (GIS). SAGA GIS is designed to provide scientists with a powerful yet simple framework for executing geoscientific methodologies. The application programming interface is responsible for this (API). SAGA contains a rapidly expanding library of geoscientific methodologies, which are packaged in interchangeable module libraries. It may be used in conjunction with other GIS applications such as Kosmo and QGIS to achieve more information in vector files and higher-resolution map-making capabilities.

The Sentinel Toolbox

The SENTINEL Toolbox is new open source software for scientific learning, research, and exploitation of SENTINEL's and heritage missions' vast archives. The Toolbox is built on the tried-and-true BEAM/NEST architecture, and it inherits all existing NEST features, including multi-mission compatibility for the majority of civilian satellite SAR operations. The research is funded by the European Space Agency's Scientific Exploitation of Operational Missions program (SEOM). The ERS, ENVISAT, and Envelope programs have built up a substantial worldwide research community over the previous 20 years, and SEOM aims to federate, support, and extend that community. It has three separate operations-

- Sentinel-1 Toolbox (SAR applications).
- Sentinel-2 Toolbox (High-resolution optical applications).
- Sentinel-3 Toolbox (High resolution optical applications)

SNAP: (Sentinel Application Platform)

All Sentinel Toolboxes, as well as toolboxes that handle data from additional satellite sensors, are built on the SNAP architecture. Due to the following technical advancements, the SNAP architecture is excellent for Earth Observation processing and analysis: Extensibility, Portability, Modular Rich Client Platform, Generic EO Data Abstraction, Tiled Memory Management, and a Graph Processing Framework.

SOCET SET

It is a software program that performs photogrammetry activities. Photogrammetry programs were mostly analogue or custom designed for government agencies prior to the emergence of digital alternatives. It takes stereo (binocular) digital aerial pictures and automatically creates a digital elevation model, digital feature (vector data), and orthorectified images from them which is known as orthophotos. Customers utilize the generated data to construct digital maps, as well as for mission planning and targeting. Film or digital cameras might be used to create the source photographs. The cameras might be deployed on a satellite or in an aircraft. The imagery must have two or more overlapping photos obtained from various viewpoint points, which is a fundamental criterion. This "binocular" feature is what allows the 3-D topography and feature data to be extracted from the picture analytically.

TNTmips

TNTmips is a geospatial analysis system that includes a full-featured GIS, RDBMS, and automated image processing system, as well as CAD, TIN, surface modelling, map layout, and data dissemination capabilities. TNTmips is a single integrated system that works on both Mac and Windows operating systems and has the same interface, functionality, and geodata structure. TNTmips' interface, database text content, communications, map creation, and all other internal components have all been translated into a variety of languages, including Arabic, Thai, and other languages.

Whitebox GAT

Whitebox Geospatial Analysis Tools (GAT) is an open-source, cross-platform GIS and remote sensing software suite provided under the GNU General Public License. Covering all of the Romance languages each of the more than 410 analytical tools has a user-friendly graphical user interface (GUI) with help and documentation embedded into the dialogue boxes. Users can also get support from a variety of offline and online sites. Raster and vector (shape file) data formats are supported in the current release. There is also a lot of functionality for processing data from laser scanners (LiDAR) that is stored in LAS files. It is surrounded by a hydrological theme. In actual this phenomenon occurs due to the fact that it superseded the Terrain Analysis System (TAS), a tool of Esri for hydro-geomorphic applications.

5.4 SUMMARY

The enormous rise of the world's population, as well as their diversified requirements, has gradually raised the need for optimal resource usage. The globe now requires appropriate amounts of supplements to feed and satisfy its massive population, as well as diverse raw materials for a strong industrial basis and the provision of enough work possibilities for the vast majority of the required work force. The methodical planning and correct exploitation of our resources can play a significant role in resolving these issues. Sustainable controllable technologies need consistent and reliable data on resource status and usage potential. In order to achieve these needs, satellite remote sensing data combined with collateral data proven to be quite successful. The Geographic Information System (GIS) proved to be an extremely useful tool for storing, manipulating, analyzing, integrating, and retrieving data. The combined use of these front-line technologies aids in the development of a strategy that may be beneficial in preparing our golden future with resource management that is sustainable.

5.5 GLOSSARY

EARTH OBSERVATION: Looking down at the Earth from aircraft and satellites using various sensors which make images that are afterwards used to study what is happening on or near the Earth's surface

ENHANCEMENT: Anything that you do to an image to make it simpler, faster or more accurate to analyze and interpret by eye is a form of "enhancement"

IMAGE ANALYSIS: This is the process of studying an image in order to explain, measure, map, count or monitor what is on the Earth's surface.

LANDSAT: Owned and launched by the United States, this is a series of remote sensing satellites that use the visible and infrared parts of the spectrum to record images of the Earth's surface.

LIDAR: The acronym for Light Detection and Ranging, formally known as LaDAR (Laser Detection and Ranging)

MOSAIC: A big image made by combining smaller images

Pixel: The smallest spatial unit in a digital image. A satellite image is made up of a matrix of pixels, each having its own vector of digital values for each band.

RESOLUTION: Spatial resolution describes how clearly you can see detail in a picture.

SENSOR: A sensor is the device that records a remote sensing image, much like a camera.

5.6 ANSWER TO CHECK YOUR PROGRESS

Practice Questions whether they are true or false.

- 1) Open source software is software that is made available to the specialized communities. (True/False)
- 2) ESRI developed ArcGIS. (True/False)
- 3) ERDAS Imagine is a vector-based software programme. (True/False)
- 4) 3D picture of Earth is created by E foto software. (True/False)
- 5) ILWIS is Integrated air and Water Information System. (True/False)
- 6) Vector data is saved as point, line, or polygon features in Orfeo toolbox. (True/False)
- 7) The Toolbox is built on the tried-and-true BEAM/NEST architecture. (True/False).
- 8) Object-based image analysis is used to accomplish this (OBIA) used byInterImage. (True/False)
- 9) ARC gis is a PC grid-based system. (True/False)
- 10) Geomaticasponsored by the Canada Centre for Remote Sensing. (True/False)

ANSWERS: -

1. False.
2. True.
3. False.
4. False.
5. False.
6. False.
7. True.
8. True.
9. False
10. True.

Fill in the blanks.

- 1) The software can analyze not only but also picture objects.
- 2) Photo Triangulation is the goal of.....
- 3)is the ideal tool for visualizing, analyzing, and presenting digital images of any kind.
- 4) Google Earth includes photorealistic created using photogrammetry.
- 5) created IDRISoftware.

- 6) OSSIM can handle over raster and vector files, as well as over different projections.
- 7) libraries provide parallel processing capabilities.
- 8) Print Layout is used to create a printed map using.....
- 9) The Sentinel Toolbox is built on the tried-and-true architecture.
- 10)software suite provided under the GNU General Public License.

Answers: -

1. Single pixel
2. E Foto
3. ENVI
4. 3D image
5. Clark lab
6. 100,4000
7. MPI
8. Q gis
9. BEAM/NEST
10. White Box GAT

Attempt these questions: -

- 1) What is open source remote sensing softwares?
- 2) What do you mean by “Rule Set”?
- 3) What are the major operations of E Foto?
- 4) What are the tools of Grass GIS?
- 5) What is full form of OBIA?
- 6) What is ILWIS software?
- 7) Which software is single integrated system that work on multiple operating system.?
- 8) What do you mean by binocular features?
- 9) Which software is a tool for high-level education in radar polarimetry and promotes scientific exploitation?
- 10) Application of Orfeo tool box is?

5.7 REFERENCES

Rose, L. C., 2004. Geospatial Portal Reference Architecture: A Community Guide to Implementing Standards-based Geospatial Portals, OpenGISDiscussion Paper, OGC 04-039. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, Volume XXXVIII, Part 8, Kyoto Japan 2010

Wiki.GIS.com

<https://gisgeography.com>

<https://www.whiteboxgeo.com>

Orfeo Toolbox url<http://www.orfeo-toolbox.org/otb>

OSSIM urlhttp://www.ossim.org/OSSIM/OSSIM_Home.html

Opticksurl<http://opticks.org>
<https://docs.ecognition.com>

5.8 TERMINAL QUESTIONS

1. Define various open source remote sensing software's.
2. Define sentinel toolbox software's along with its various operations.
3. What do you understand by ERDAS Imagin software's? Elaborate its various functions.
4. What is the objective of remote sensing sand GIS software's?
5. Do you think the remote sensing software's is fulling the needs of 21st century, elaborate.

UNIT 6 - OVERVIEW OF OPEN SOURCE SOFTWARE

6.1 OBJECTIVES

6.2 INTRODUCTION

6.3 OVERVIEW OF OPEN SOURCE SOFTWARE

6.4 SUMMARY

6.5 GLOSSARY

6.6 ANSWER TO CHECK YOUR PROGRESS

6.7 REFERENCES

6.8 TERMINAL QUESTIONS

6.1 OBJECTIVES

The objectives of this unit are to introduce learners/students about the Open Source software and its importance. Students will study some common and basic definitions of software, its types in details, importance and types of licenses for operating any software and further the advantages of Open Source software. In the glossary section some common definitions are given, which may give the basics of applied terms in this unit.

At the end of this unit, it is supposed that, the students themselves be encouraged to operate their respective applications in the open source environment. This all would also explore their research and innovative thinking.

6.2 INTRODUCTION

Computer manufacturers during 1950 to 1960 used to distribute software along with computers. They used to produce their own software or hire software companies to do so. It was sold as a finished product and there was no trend of outsourcing software. Typically software was embedded and the distribution of the software along with computer was a normal practice. At that time source code (human readable programming instructions) was circulated along with the software and users used to alter the software to fix bugs and add new features by themselves.

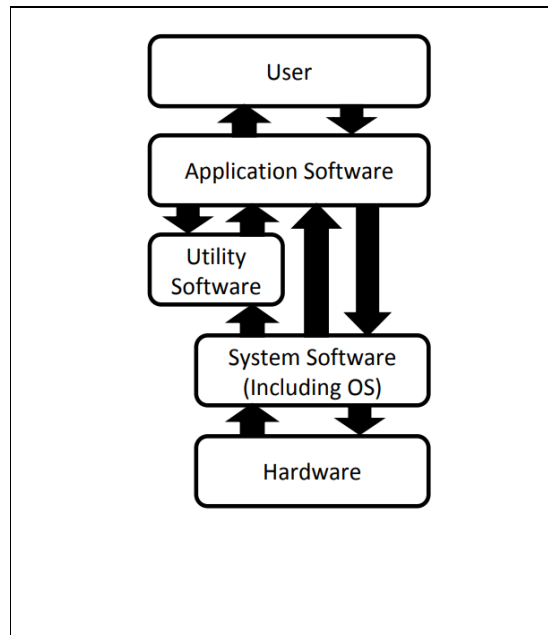
Since Open source has progressed to its present state as a result of many technological breakthroughs in the field of digital era. After the advancement of the Internet, open source grew exponentially that allows thousands of programmers/researchers/innovators around the globe to collaborate collectively to design the applicable software with freely available source code. So open source software has become a favorite alternative to commercial/proprietary software.

Computer software, or simply software, is a collection of data or computer instructions that tell the computer how to work. This is in contrast to physical hardware, from which the system is built and actually performs the work. In computer science and software engineering, computer software is all information processed by computer systems, programs and data. Computer software includes computer programs, libraries and related non-executable data, such as online documentation or digital media. Computer hardware and software require each other and neither can be realistically used on its own.

In its most general sense, computer software is a set of instructions or programs instructing a computer to do specific tasks. Software is a generic term used to describe computer programs. Scripts, applications, programs and a set of instructions are all terms often used to describe software.

6.2.1 Types of Software

The software characteristics include performance, portability and functionality. Developing any software requires the understanding of the operational, transitional and revision characteristics of the software under development. Based on the objectives, computer software can be categorized into:



6.2.2 Application software

The domain of information technology belongs to one of the most rapidly developing areas on the world. Nowadays, software organizations, which are intend to be successful in the software development market, have to keep their knowledge up to date. The task of picking up right methodologies, techniques and tools is critical when we are talking about delivering high - quality and maintainable software products and still keep time and money costs in reasonable limits. Current modern trends in the development of applications software could help organizations in their work and their respective business. As of application software are those that uses the computer system to perform special functions beyond the basic operation of the computer itself. They include packaged and copyrighted software, software developed for custom purposes and application domain, for trial or beta version of specific software, like more.

Application software may be categorized into broadly four categories. Out of these broad categories, number of sub-categories is presented here, but these are just indicative categories/sub-categories, the range of these categories/sub O categories is very wide and not limited to any list. These are:

- **Business**
 - Word Processing
 - Spreadsheet
 - Database
 - Presentation
 - Note writing
 - Business software for mobiles, e-mails
 - Project Management
 - Accounting

Document Management
Enterprise Computing

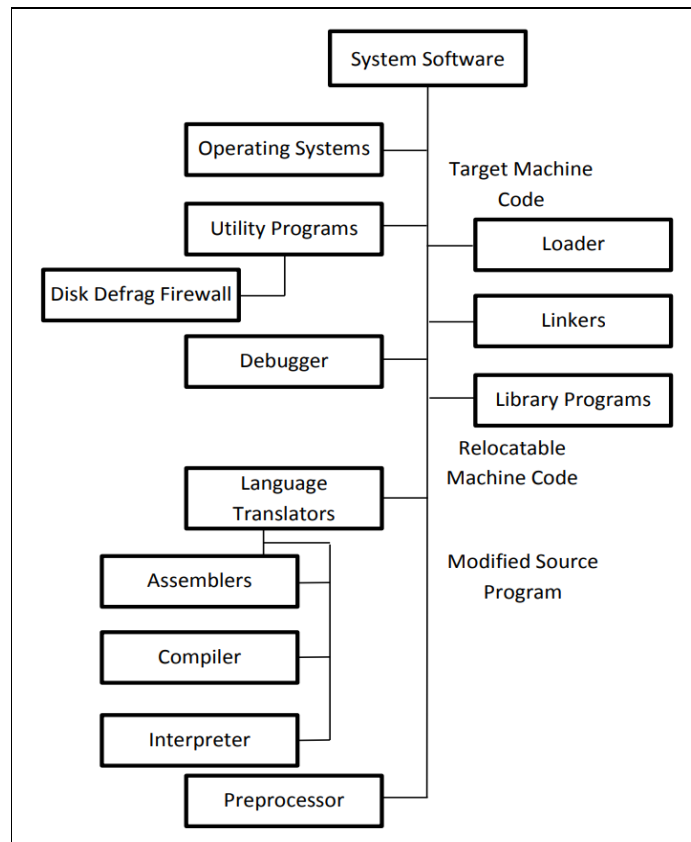
- **Graphic and Multimedia**
 - Computer Aided Design (CAD)
 - Desktop Publishing (DTP)
 - Paint/Image editing/enhancing
 - Audio/Videography
 - Multimedia authoring
 - Web page authoring

- **Home/ Personal/Education**
 - Software PC suite
 - Financing/Tax/GST/EMI calculation/preparation
 - Audio/video/image editing
 - Navigation and Mapping
 - Entertainment
 - References and Educational project preparation

- **Communications**
 - Web Browser
 - E-mail and Messaging
 - Chat Rooms
 - Blogging
 - File Transfer Protocol
 - IP addressing and subnetting
 - Sharing the network resources

6.2.3 System software

System software is software designed to provide a platform for other software. Means software that directly operates the **computer hardware**, and to provide basic functionality needed by users and other software.



6.2.4 Features of the system software

- It is difficult to design.
- It is written in the low-level language, or written in machine language which is only understood by the machine.
- It is difficult to manipulate.
- System software is very close to the system.
- The speed of the system software is fast.

6.2.5 Components of system software

System software comprises of the following components:

- **Device Driver**

It is a computer program. Computer hardware can interact with the higher level computer programs with the help of device drivers. These device drivers act as a translator between the application of the user and the hardware device. Due to this, it simplifies programming.

- **Operating System**

The operating system is the system which manages the hardware of the computer. The operating system consists of program and data. An operating system also provides us with services so that we can execute the application software.

6.2.6 Types of Operating Systems

- **Real-Time OS:** Is installed in special purpose embedded systems like robots, cars, and modems.
- **Single-user and single-task OS:** Are installed on single-user devices like phones.
- **Single-user and multitask OS:** Are installed on contemporary personal computers.
- **Multi-user OS:** Is installed in network environments where many users have to share resources. Server OSs is examples of multi-user operating systems.
- **Network OS:** Is used to share resources such as files, printers in a network setup.
- **Internet/Web OS:** Is designed to run on the browser that is online.
- **Mobile OS:** Is designed to run on mobile phones, tablets and other mobile devices.

Examples: Ubuntu, Mac OS X, Ubuntu Server, Red Hat Enterprises, Chrome OS, iPhone OS, Android OS, Linux, Unix, Microsoft Windows 10 etc.

- **Server**

It is a program which works like a socket listener in a computer networking system. A server computer is a series of computers which link other computers with it. The server can provide some essential services to both private users and public users via the internet across a network.

- **Utility Software**

It performs small tasks which are used to manage computer hardware and application software.

Example: system utilities, virus scanners etc.

- **Windowing System**

It supports the implementation of window managers. Windowing system provides essential support to graphics hardware and pointing devices like keyboard and mice. The windowing system is a component of the Graphical User Interface (GUI).

- **Malicious Software or Malware**

These software are developed to harm and disrupt computers. As such, malware is undesirable. Malware is closely associated with computer-related crimes, though some malicious programs may have been designed as practical jokes.

- **Computer Virus**

A computer virus is a type of computer program that, when executed, replicates itself by modifying other computer programs and inserting its own code. When this replication succeeds, the affected areas are then said to be "infected" with a computer virus.

- **Antivirus software**

Antivirus software, or anti-virus, also known as anti-malware, is a computer program used to prevent, detect, and remove malware. It was originally developed to detect and remove computer viruses, hence the name. However, with the proliferation of other kinds of malware, antivirus software started to provide protection from other computer threats. In particular, modern antivirus software can protect users from malicious browser helper

objects (BHOs), browser hijackers, ransomware, keyloggers, backdoors, rootkits, trojan horses, worms, malicious LSPs, dialers, fraudtools, adware and spyware. Some products also include protection from other computer threats, such as infected and malicious URLs, spam, scam and phishing attacks, online identity (privacy), online banking attacks, social engineering techniques, advanced persistent threat (APT) .

- **Firmware Software**

It is the operational software which is already embedded in flash, ROM, EPROM, EEPROM memory chips so that the OS can identify them quickly. The task of the firmware is to directly manage and control all the activities of any single hardware. Traditionally, firmware was installed on the non-volatile chips. We can upgrade them by changing with the new programmed chips. However, nowadays, firmware was installed on the flash chips. So, now we can upgrade them without swapping the chips.

The main difference between the firmware and the driver is that firmware will reside within the devices whereas the drivers will install in the operating system.

6.3 OVERVIEW OF OPEN SOURCE SOFTWARE

Free and open source development models have made tremendous contributions to computing, sustaining both research and commercial projects and making it easier for large groups of people, who may not even be acquainted, to help each other. Open source licensing and development approaches have been challenging and transforming software development for decades.

Open Source Software means, software which is freely available to anybody and of which source code (use to create a program) is free to view, use, modify and redistribute without any discrimination. It's a collaborative effort where programmers can make changes, improve the source code and share between peers, allow further modifications and incorporate changes within the community. Further, the term open source refers to something people can modify and share because its design is publicly accessible. The term originated in the context of software development to designate a specific approach to creating computer programs. Today, however, "open source" designates a broader set of values—what we call "the open source way." Open source projects, products, or initiatives embrace and celebrate principles of open exchange, collaborative participation, rapid prototyping, transparency, meritocracy, and community-oriented development.

Richard M. Stallman established the Free Software Foundation in 1985. It's a non-profit organization solely dedicated to support and promote the free software movement, freedom to study and modify the software. He defines the term, 'Free Software' means a "software that respects users' freedom and community. It means that the users have the freedom to run, copy, distribute, study, change and improve the software. Thus, 'free software' is a matter of liberty, not price"

A software license agreement is the legal contract between the licensor and/or author and the purchaser of a piece of software which establishes the purchaser's rights. A software license agreement details how and when the software can be used, and provides any restrictions that are imposed on the software.

Some common categories for software under copyright law, licenses, proprietaries are also need to understand by the developer, these are proprietary license.

- **GNU General Public License**

The GNU General Public License (GNU GPL or GPL) is a series of widely used free software licenses that guarantee end users the freedom to run, study, share, and modify the software. Historically, the GPL license family has been one of the most popular software licenses in the free and open-source software domain.

- **End User License Agreement (EULA)**

An end-user license agreement (EULA) is a legal contract entered into between a software developer or vendor and the user of the software, often where the software has been purchased by the user from an intermediary such as a retailer. A EULA specifies in detail the rights and restrictions which apply to the use of the software. Many form contracts are only contained in digital form, and only presented to a user as a click-through which the user must "accept". As the user may not see the agreement until after he or she has already purchased the software, these documents may be contracts of adhesion.

- **Workstation licenses**

A workstation license is a license that can be used on a workstation computer (master computer), where the host application (3ds Max, Cinema 4D, ...) is running in "full" mode - with its graphical user interface, material editor, and all other features enabled, as opposed to a node license, where the computer is used only for rendering. Workstation licenses are used on computers where you are working with the host application to create scenes, apply materials, adjust lighting, and perform other tasks which require you to actively use the host application.

- **Concurrent use license**

A software license that is based on the number of simultaneous users accessing the program. It typically deals with software running in the server where users connect via the network.

- **Site licenses**

A site license is a type of software license that allows the user to install a software package in several computers simultaneously, such as at a particular site or across a corporation. Depending on the amount of fees paid, the license may be unlimited or may limit simultaneous access to a certain number of users.

- **Perpetual licenses**

A perpetual software license is a type of software license that authorizes an individual to use a program indefinitely. Generally, outside of termination, a perpetual software license allows the holder to use a specific version of a given software program continually with payment of a single fee.

- **Non-perpetual license**

For starters, non-perpetual licenses generate cash flow for the vendor, and this helps the vendor to innovate and further improve the product. Every vendor has its own licensing terms, but in many cases, non-perpetual licenses entitle customers to free upgrades for as long as the license remains in effect.

- **Proprietary software licenses**

Proprietary software consists of software that is licensed by the copyright holder under very specific conditions. In general, you can use the software, but you are not allowed to modify the software or distribute it to others.

The original source code for the software is not available, which means you can't see the actual code written by the programmers. Proprietary software is, therefore, also referred to as closed-source software. This is done on purpose to protect the intellectual property invested in software development. If the source code were released, even with copyright restrictions, competitors could benefit from using this code.

Many proprietary software applications are also commercial, meaning that you have to pay for a license. However, many other proprietary software applications are free. The fact that software is free does not mean it is not proprietary.

There are numerous examples of proprietary software. Both the Windows and Mac operating systems are proprietary, and so are many of the typical software applications used in organizations, such as Microsoft Office. Many specialized software applications, such as those used for database management and various types of enterprise information systems, are also proprietary. In many cases, software companies have invested many years of software development into a product. By making the software proprietary, they are protecting their investment and making it possible to commercialize their software. The revenue from software sales can then be used to continue developing the software.

- **Free/Open Source Software Licenses**

The important tenacity of open source licensing is to reject anybody the right to exclusively exploit a work. The term license is used to define the legal way copyright and patent holder award permission to others to use his intellectual property. An open source license is a way to use its intellectual property in such a way that software freedom is protected for all. It allows the source code to modify, share under defined terms and conditions. This allows end users to customize according to their needs and necessities. Open source licenses are mostly free of charge though there are some exceptions.

- **Open Source Initiative (OSI)**

The Open Source Initiative (OSI) is a non-profit organization dedicated to promoting Open Source Software which came into existence in the year 1998, founded by Eric S. Raymond and Bruce Perens. The free software movement of Stallman did not catch the mainstream. Eric Raymond and Bruce Perens thought that an important portion of the problem existed in Stallman's term "Free" software. Their intention was to encourage OSS in commercial arena and they thought that both Free/Open Source Community and the

commercial industry will benefit from broader OSS distribution; therefore, along with other hackers, they started the “Open Source Movement”. The licensing terms were on the same lines as governed by the free software foundation. The difference was OSI preferred practical benefits rather than philosophical benefits that endorsed granting users moral freedom offered by free software foundation. The ‘Open Source’ label was coined at a strategy session held on 3rd February 1998 in Palo Alto, California after releasing the announcement of the Netscape source code. “In 1998, a group of individuals advocated that the term free software should be replaced by ‘Open Source Software’ (OSS) as an expression which is less ambiguous and more comfortable for the corporate world.” The OSI developed the Open Source Definition (OSD). The definition is not a license itself, but a guideline and trademark for OSS software licenses other than GPL. Licenses according to the OSD guaranteed several freedom to software users, including commercial users. In order to raise acceptance of OSS in the business world, the term Open Source Software instead of Free Software was established and widely accepted. The 1990s experienced a significant rise in attention paid to Open Source projects. Many companies from the IT industry began supporting the project.

- **Advantages of Open Source Software**

The Open Source initiative has given the definition and characteristic of OSS distribution terms that must fulfill the criteria such as ‘free redistribution; source code; derived works; integrity of the author’s source code; modification; redistribution, no discrimination against any person or group; no discrimination against fields of endeavor; distribution of license; license must not restrict other software and license must be technology neutral’.

Students, professionals, academia, and different user community prefer open source software to proprietary software for a number of reasons that includes:-

- **Free Redistribution**

The license shall not restrict any party from selling or giving away the software. As a component of an aggregate software distribution containing programs from several different sources. The license shall not require a royalty or other fees for such sale.

- **Source Code**

The program must include source code, and must allow distribution in source code as well as compiled form. Where some form of a product is not distributed with source code. There must be a well-publicized means of obtaining the source code for no more than a reasonable reproduction cost, downloading via the Internet without any charges. The source code must be a preferred form in which a programmer would modify the program.

- **Derived Work**

The license must allow modifications and derived works, and must allow them to be distributed under the same terms according to the license of the original software.

- **No Discrimination against any Person/Group/Fields**

The license must not discriminate against any person or group. The license must not restrict anyone from making the use of the program in a specific field of endeavor.

- **License Must Not Be Specific to a Product**

The rights attached to the program must not depend on the program's being part of a particular software distribution. If the program is extracted from that distribution and used or distributed within the terms of the program's license, all parties to whom the program is redistributed should have the same rights as theirs, who are granted in conjunction with the original software distribution.”

- **Training**

Learners like open source software because it helps them become better programmers. Because open source code is publicly accessible, students can easily study it as they learn to make better software. Students can also share their work with others, inviting comment and critique, as they develop their skills. When people discover mistakes in programs' source code, they can share those mistakes with others to help them avoid making those same mistakes themselves.

- **Security**

Some people prefer open source software because they consider it more secure and stable than proprietary software. Because anyone can view and modify open source software, someone might spot and correct errors or omissions that a program's original authors might have missed. And because so many programmers can work on a piece of open source software without asking for permission from original authors, they can fix, update, and upgrade open source software more quickly than they can proprietary software.

- **Stability**

Many users prefer open source software to proprietary software for important, long-term projects. Because programmers publicly distribute the source code for open source software, users relying on that software for critical tasks can be sure their tools won't disappear or fall into disrepair if their original creators stop working on them. Additionally, open source software tends to both incorporate and operate according to open standards.

- **Community**

Open source software often inspires a community of users and developers to form around it. That's not unique to open source; many popular applications are the subject of meetups and user groups. But with open source, the community isn't just a fan-base that buys in to an elite user group, it's the people who produce, test, use, promote, and ultimately affect the software they love.

6.3.1 Free Software versus Open Source Software

Free Software is also known as Open Source Software (OSS) or Free/ Open Source Software (FOSS) or Free/Libre Open Source Software (FLOSS). Developers have used these alternative terms, consequently to describe Open Source Software (OSS) which is also Free Software. The Free Software Foundation defines the distribution terms and says Free

Software means software that respects users' freedom and community. It means that the users have freedom to run, copy, distribute, study, change and improve the software. Thus, Free Software is a matter of liberty, not price. The Open Source Initiative on (www.opensource.org) has given the distribution terms of OSS that must comply that it should allow modification, redistribution, open source code, no discrimination against any person or group or fields or endeavor, distribution of license who receives the program; license must not be product specific; restrict other software and must be technologically neutral. The Free Software and Open Source Software are not only two different terms and ideologies but also the founders who propounded these two terms advocate two different explanations and stand firm on their philosophies, although the members of both communities collaborate intensively on practical projects. These two terms carry a different legacy, variance and hold a separate identity altogether. The term Free Software came into existence in the year 1984 is the older one of the two terms and the term "Open Source Software" was invented in 1998.

- **Disadvantages of Open Source Software**

The benefits of open source are tremendous and have gained huge popularity in the field of IT in recent years. This is mainly because the advantages of **open-source software** are that it's free to use – its greatest advantage. As it is developed by a non-profit community, it has some disadvantages as well. These are:-

The main disadvantage of open-source software is not being straightforward to use. Open-source available operating systems like cannot be learned in a short time. They require efforts and trainings before you are able to convenient or expert on them. You may need to little bit or sometime much effort to make things easier.

There is a shortage of applications that run both on open source and proprietary software; therefore, switching to an open-source platform involves a compatibility analysis of all the other software used that run on proprietary platforms. In addition, there are many ongoing parallel developments on open source software. This creates confusion on what functionalities are present in which versions.

Many of the latest hardware are incompatible to the open-source platform; so you have to rely on third-party drivers.

Migration of data from one system to OSS creates a problem because may be the users are not technically equipped.

No vendor policy, no accountability and nobody is responsible for it.

Sustainability and scalability are questionable till the developers are developing the product.

Most libraries are using proprietary software. If they are switching to OSS they have a fear that they will lose the data and they have to bear the migration cost. Some of the proprietary software ensures that a customer should not move to OSS and they have snapped certain privileges to allow easy migration to OSS.

6.4 SUMMARY

The evolution of free software began from the moment the first mechanical computer came into existence. It was cult and cultures that time to provide software free of charge. At later stage that got discontinued due to decentralization of software industry. Introduction of the Internet gave the biggest boost to free software developers to freely share their work among peers and make collaborative work stronger than before. Libraries have become major beneficiaries of the OSS projects. Open source community made hundreds of software that can be utilized to render the library services in this age of information technology effectively. Today there are many OSS tools and standards which are freely available for developing different applications. These OSS applications are available in a varied range that includes library automation software, digital library, content management, journal archiving, journal publishing, e-learning management, etc. When there are benefits of some phenomenon there are also some drawbacks. It's a part and parcel that goes hand and hand. At the end of this unit the researcher has given a comparison of OSS to CSS just to assess quickly the benefits and drawbacks of OSS and CSS. .

6.5 GLOSSARY

- **ROM-** Read-only memory is a type of non-volatile memory used in computers and other electronic devices. Data stored in ROM cannot be electronically modified after the manufacture of the memory device.
- **EPROM-** An EPROM, or erasable programmable read-only memory, is a type of programmable read-only memory chip that retains its data when its power supply is switched off. Computer memory that can retrieve stored data after a power supply has been turned off and back on is called non-volatile.
- **EEPROM-**EEPROM stands for electrically erasable programmable read-only memory and is a type of non-volatile memory used in computers, integrated in microcontrollers for smart cards and remote keyless systems.
- **FOSS4G-** FOSS4G is the acronym for Free and Open Source Software for Geospatial. It is the annual recurring global event hosted by OSGeo since its inception in 2006. Its predecessors were rooted in the GRASS and MapServer communities and can be traced back to the beginning of this millennium.
- **Low Level Programming Language-** A low-level programming language is a programming language that provides little or no abstraction from a computer's instruction set architecture—commands or functions in the language map closely to processor instructions. Generally, this refers to either machine code or assembly language.

- **Machine Level Programming Language-** Machine Level Programming language, or machine code, consists of binary code and is the only language that is directly understood by the computer. Both machine code and assembly languages are hardware specific.
- **Scripts-** A computer script is a list of commands that are executed by a certain program or scripting engine. Scripts may be used to automate processes on a local computer or to generate Web pages on the Web.
- **Applications-** An application, also referred to as an application program or application software, is a computer software package that performs a specific function directly for an end user or, in some cases, for another application. An application can be self-contained or a group of programs. The program is a set of operations that runs the application for the user. Examples are: Microsoft suite of products (Office, Excel, Word, PowerPoint, Outlook, etc.), Internet browsers like Firefox, Mozilla, and Chrome.
- **Algorithms-** An algorithm is a step by step method of solving a problem. It is commonly used for data processing, calculation and other related computer operations.
- **Computer Programs-** A computer program is a collection of instructions that performs a specific task when executed by a computer. A computer requires programs to function. A computer program is usually written by a computer programmer in a programming language.
- **FOSS4G-** FOSS4G is the acronym for Free and Open Source Software for Geospatial. It is the annual recurring global event hosted by OSGeo since its inception in 2006. Its predecessors were rooted in the GRASS and MapServer communities and can be traced back to the beginning of this millennium.
- **Source code** -Source code is the part of software that most computer users don't ever see; it's the code computer programmers can manipulate to change how a piece of software—a "program" or "application"—works. Programmers who have access to a computer program's source code can improve that program by adding features to it or fixing parts that don't always work correctly.

Examples of Open Source Software

Linux, Mozilla Firefox, Liber Office, QGIS and Apache

- <https://www.fsf.org/>- The Free Software Foundation (FSF) is a nonprofit with a worldwide mission to promote computer user freedom. It is working to secure freedom for computer users by promoting the development and use of free (as in freedom) software and documentation—particularly the GNU operating system—and

by campaigning against threats to computer user freedom like Digital Restrictions Management (DRM) and software patents.

- **GNU-** GNU is an operating system that is free software—that is, it respects users' freedom. The GNU operating system consists of GNU packages (programs specifically released by the GNU Project) as well as free software released by third parties. The development of GNU made it possible to use a computer without software that would trample your freedom.

6.6 ANSWER TO CHECK YOUR PROGRESS

1. Differentiate between open source and proprietary software.
2. Can Open Source software be used for commercial purposes?
3. Differentiate between free and open source software with suitable examples.
4. What are the roles of an operating system?
5. Mention some Applications of Open Source software systems.

6.7 REFERENCES

1. www.google.com
2. www.wikipedia.org
3. <https://opensource.com/resources/what-open-source>
4. Understanding Open Source and Free Software Licensing By Andrew M. St. Laurent
5. <https://www.fsf.org/free-software-supporter/2016/>
6. <https://coronarenderer.freshdesk.com/>
7. <https://study.com/>
8. <https://wiki.osgeo.org/>

6.8 TERMINAL QUESTIONS

1. Write in details about the Software Development Life Cycle, keeping in mind about open source software.
2. What are the reasons for preferring open source software over the proprietary software?
3. What is software freedom? Explain its components.
4. Differentiate between open source and closed source.
5. List out various open source software's in the following domains: Communications, Database, Education, Games/Multimedia, Desktop publishing.

UNIT 7 - OVERVIEW OF QUANTUM GIS SOFTWARE

7.1 OBJECTIVES

7.2 INTRODUCTION

7.3 OVERVIEW OF QUANTUM GIS SOFTWARE

7.4 SUMMARY

7.5 GLOSSARY

7.6 ANSWER TO CHECK YOUR PROGRESS

7.7 REFERENCES

7.8 TERMINAL QUESTIONS

7.1 OBJECTIVES

The uses and the need to explore of Open Source Software is arising in current time. As it nourishes the temperament of research amongst the user community, whether they may be professional software developers, school – college going students, innovators and others. So, using Open Source Software, and exploring and extracting the desired results will provide the benefits to various stakeholders. Moreover, by connecting/involving with these open source activities, provides platforms to connect with similar nature communities of nature that also enhance the team spirit and team management.

Further along with other subjects associated with Geo-informatics curriculum, students/researchers/academia required to build their hand-on or practical training of these geo information. For the QGIS, which are easily downloadable and also supports various online help within website or within GIS community group is always recommended by everyone and for everyone, associated with GIS community.

After going with given information about QGIS, data basic, one can enhance its knowledge about the GIS software, and moreover he/she surely enhance their skill in

- Basic, practical understanding of GIS concepts
- Techniques and real world applications
- How GIS is utilized in the larger context of rural and urban applications
- Understand the basic concepts of geography necessary to efficiently and accurately use GIS technology.
- Building the ability to perform basic GIS analysis of concepts.
- Using/demonstrated a practical application of GIS
- GIS mapping development.

7.2 INTRODUCTION

Computer software, or simply software, is a collection of data or computer instructions that tell the computer how to work. This is in contrast to physical hardware, from which the system is built and actually performs the work. There are few broad categories of the software that includes

Application Software: is a type of software that uses the computer system to perform special functions or provide entertainment functions beyond the basic operation of the computer itself. There are numbers of application software available as per the need of users.

Systems Software: is the software which manages the computer hardware, as to provide basic functionalities that are required by users, or for other software to run properly .System software is also designed for providing a platform for running application software.

Software License: A software license is a legal instrument (usually by way of contract law, with or without printed material) governing the use or redistribution of software. Under United States copyright law, all software is copyright protected, in both source code and object code forms, unless that software was developed by the United States Government, in which case it cannot be copyrighted. Authors of copyrighted software can

donate their software to the public domain, in which case it is also not covered by copyright and, as a result, cannot be licensed.

A typical software licence grants the licensee, typically an end-user, permission to use one or more copies of software in ways where such a use would otherwise potentially constitute copyright infringement of the software owner's exclusive rights under copyright.

Proprietary Software License

The hallmark of proprietary software licenses is that the software publisher grants the use of one or more copies of software under the end-user license agreement (EULA), but ownership of those copies remains with the software publisher (hence use of the term "proprietary"). This feature of proprietary software licenses means that certain rights regarding the software are reserved by the software publisher. Therefore, it is typical of EULAs to include terms which define the uses of the software, such as the number of installations allowed or the terms of distribution.

The most significant effect of this form of licensing is that, if ownership of the software remains with the software publisher, then the end-user must accept the software license. In other words, without acceptance of the license, the end-user may not use the software at all.

The most common licensing models are per single user (named user, client, node) or per user in the appropriate volume discount level, while some manufacturers accumulate existing licenses. These open volume license programs are typically called open license program (OLP), transactional license program (TLP), volume license program (VLP) etc. and are contrary to the contractual license program (CLP), where the customer commits to purchase a certain number of licenses over a fixed period (mostly two years). Licensing per concurrent/floating user also occurs, where all users in a network have access to the program, but only a specific number at the same time.

Software licensing often also includes maintenance sometimes includes technical support. This, usually with a term of one year, is either included or optional, but must often be bought with the software. The maintenance agreement (contract) typically contains a clause that allows the licensee to receive minor updates

Free and Open-Source Software

There are several organizations in the FOSS (**Free and Open-Source Software**) domain who give out guidelines and definitions regarding software licenses. The Open Source Initiative defines a list of certified open-source licenses following their The Open Source Definition.

Free and open-source licenses are commonly classified into two categories: Those with the aim to have minimal requirements about how the software can be redistributed - permissive licenses, and the protective share-alike - copy left Licenses.

An example of a copy left free software license is the often used GNU General Public License (GPL), also the first copy left license. This license is aimed at giving and protecting all users unlimited freedom to use, study, and privately modify the software, and if the user adheres to the terms and conditions of the GPL, freedom to redistribute the software or any modifications to it. For instance, any modifications made and redistributed by the end-user

must include the source code for these, and the license of any derivative work must not put any additional restrictions beyond what the GPL allows.

7.3 OVERVIEW OF QUANTUM GIS SOFTWARE



Geographical Information System (GIS) is a collection of software that allows you to create, visualize, query and analyze geospatial data. Geospatial data refers to information about the geographic location of an entity. This often involves the use of a geographic coordinate, like a latitude or longitude value. Spatial data is another commonly used term, as are: geographic data, GIS data, map data, location data, coordinate data and spatial geometry data.

Further Quantum GIS (QGIS) is one such software which is a freely downloadable and open source software of GIS applications. QGIS provides excellent capabilities, stable operation and support numerous vector, raster, database, formats and functionalities. Like other GIS applications, QGIS provides a geographical user interface wherein different layers of spatial data can be displayed and analyzed. Many commercial GIS software available in the market are very expensive and cost a huge financial burden on users not only in purchasing licenses but also for their annual maintenance.

QGIS is a user friendly Open Source Geographic Information System (GIS) licensed under the GNU General Public License. QGIS is an official project of the Open Source Geospatial Foundation (OSGeo). It runs on Linux, Unix, Mac OSX, Windows and Android and supports numerous vector, raster, and database formats and functionalities. Its latest release is QGIS 3.4 to the date. Apart from them there are several other reasons of using QGIS. These are:

- It's free, as you installing this QGIS
- If you need extra functionalities in QGIS, a group of GIS community happy to assist you
- By the way you can add/embed the certain tool/application, for that programming knowledge is required
- Very It's constantly developing. Because anyone can add new features and improve on existing ones
- QGIS never stagnates. The development of a new tool can happen as quickly as you need it to.
- Extensive help and documentation is available. If you're stuck with anything, you can turn to the extensive documentation, your fellow QGIS users, or even the developers.

- Cross-platform. QGIS can be installed on MacOS, Windows and Linux.

Using Quantum GIS software

Here are step by step guide to using QGIS

- Installing QGIS
- Rectificaion/Geo-referencing of toposheet
- Data downloading in QGIS
- Creation of *.shp file in QGIS
- Opening a Satellite image in QGIS
- Map Composition in QGIS

- First of all install the QGIS software from the <https://qgis.org/en/site/>

The first interface of the website is like this



1. Click to For Users tab, it gives you further three steps, under the namely Get Started Using QGIS

Step 1: Download QGIS

Step 2: Check our documentation

Step 3: Connect with the QGIS community

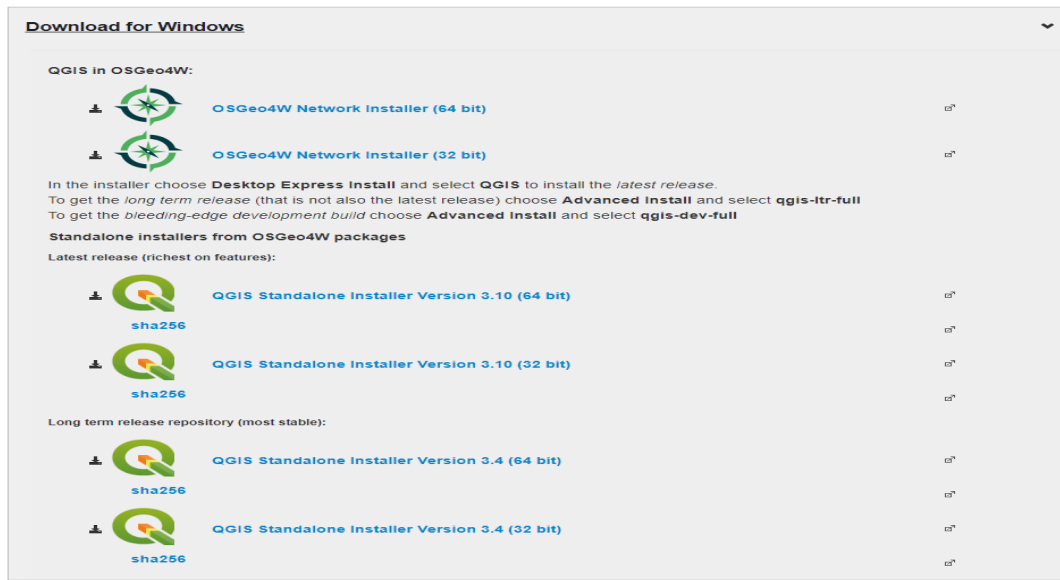
Then choose the Step 1, i.e. download QGIS, net Window will be like this



Note: As the website <https://qgis.org/en/site/forusers/download.html> gives number of options for installing QGIS for different platforms, here in this chapter the description and steps are showing only for Windows based version of QGIS. Students are free and advised to explore it on the different platform for enhancement of their skills.

2. When you click on download for Windows, the next window show/gives you number of options. These are:

(i) Installer for 64 bit and 32 bit

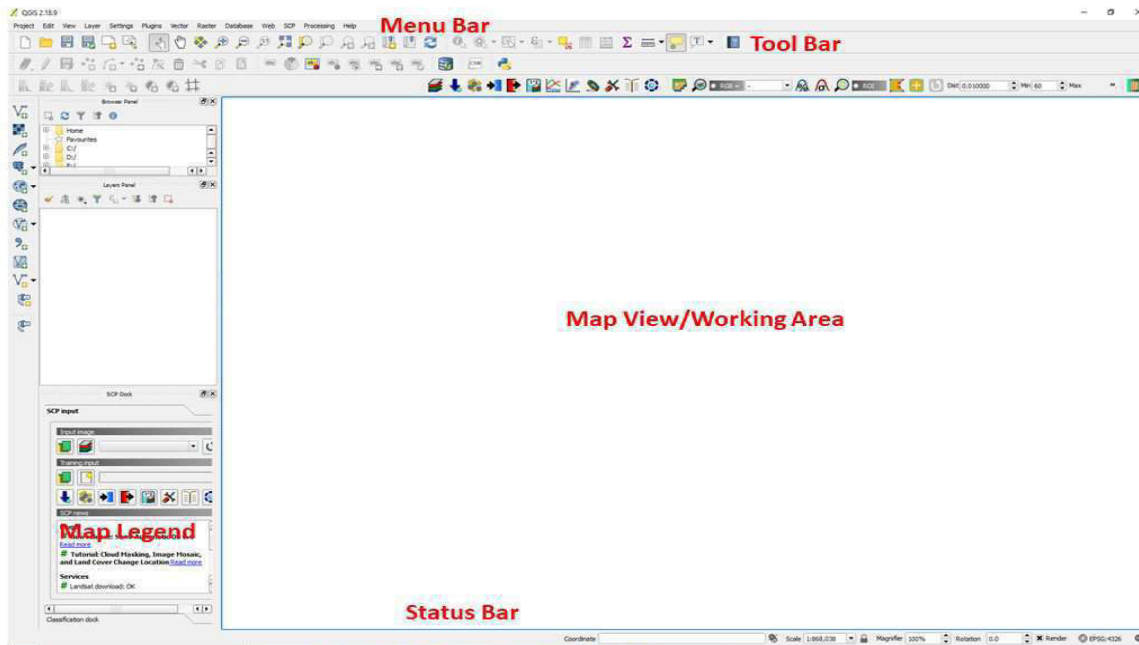


(ii) In the further options, it gives you the number of QGIS setups, which includes latest developed versions and previous versions.

If the user is new, for that recommendation is that user should choose most stable version of the software. It is because, in the most stable version lots of user community is associated and help will be easily available.

Once you download any version, you are ready to work on it.

The first interface of the software is look like



Projects

The state of your QGIS session is considered a Project. QGIS works on one project at a time.

Settings are either considered as being per-project, or as a default for new projects.

QGIS can save the state of your workspace into a project file using the menu options

File > Save Project or File > Save Project As.

The kinds of information saved in a project file include:

- Layers added
- Layer properties, including symbolization
- Projection for the map view
- Last viewed extent

The first interface of QGIS comprises of basically five parts, the details are:

Menu Bar

Provides access to various QGIS features using a standard hierarchical menu.

Tool Bar

Provides access to most of the functions as the menus, plus additional tools for interacting with the map. Each toolbar item has popup help available. Hold your mouse over the item and a short description of the tool's purpose will be displayed.

Every menu bar can be moved around according to your needs. Additionally every menu bar can be switched off using your right mouse button context menu holding the mouse over the toolbars.

Map Legend

Lists all the data layers in the project. This area also allows to change style or label the layer.

Map View

The map displayed in this window will depend on the vector and raster layers you have chosen to load.

Status Bar

The status bar shows you your current position in map coordinates (e.g. meters or decimal degrees) as the mouse pointer is moved across the map view. To the left of the coordinate display in the status bar is a small button that will toggle between showing coordinate position or the view extents of the map view as you pan and zoom in and out.

Progress Bar

A progress bar in the status bar shows progress of rendering as each layer is drawn to the map view. In some cases, such as the gathering of statistics in raster layers, the progress bar will be used to show the status of lengthy operations.

By using above tools/work area user can add number of data layers, zoom in and out to a specific area in the map, check off and check on the relevant layers and see the coordinates, scale and projection of displayed map.

Opening the existing Raster or Vector files.

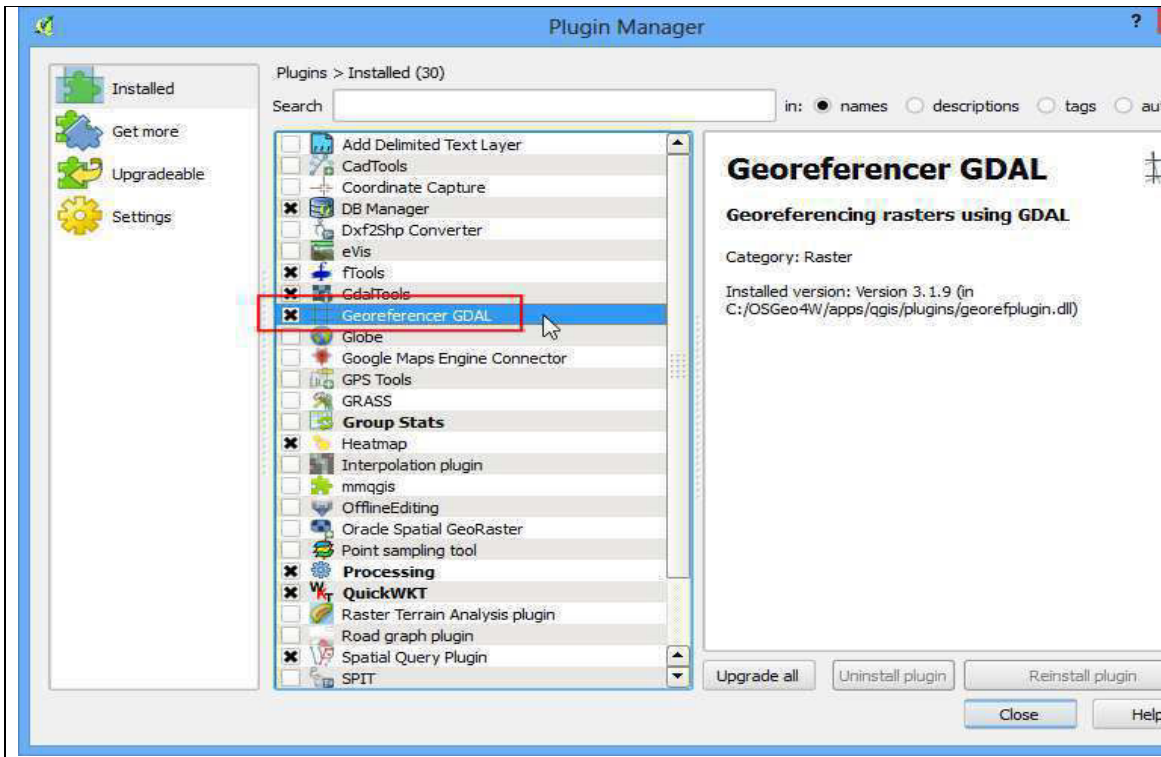
- (i) Click on Add Raster Layer option
- (ii) Click on Add Vector Layer option

In each tool presented on the toolbar, tool tip appears, when mouse pointer is on the particular tool.

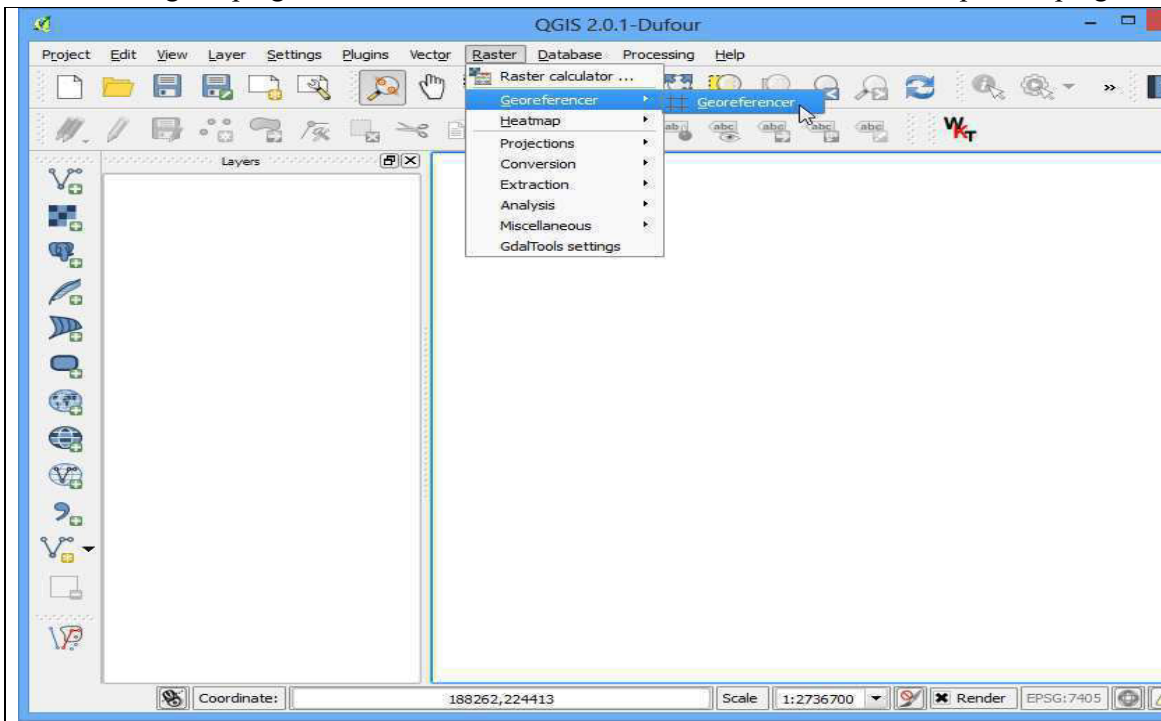
b. Now we have to first **georeference** the scanned topo - map or whatever the data we are handling. As we should be aware about the importance of geo-referencing or rectification - GIS projects require georeferencing some raster data. It is the process of assigning real-world coordinates to each pixel of the raster. These coordinates (Latitude, Longitude) can be obtained by doing field surveys - collecting coordinates with a GPS device for few easily identifiable features in the image or map. In some cases, where you are looking to digitize scanned maps, you can obtain the coordinates (Latitude, Longitude) from the markings on the map/image itself. Using these coordinates or GCPs (Ground Control Points), the image is warped and made to fit within the chosen coordinate system.

Georeferencing in QGIS is done via the 'Georeferencer GDAL' plugin. This is a core plugin - meaning it is already part of your QGIS installation. You just need to enable it.

Go to Plugins > Manage and Install Plugins and enable the Georeferencer GDAL plugin in the Installed tab.



For installing the plugin Click on Raster > Georeferencer > Georeferencer to open the plugin.



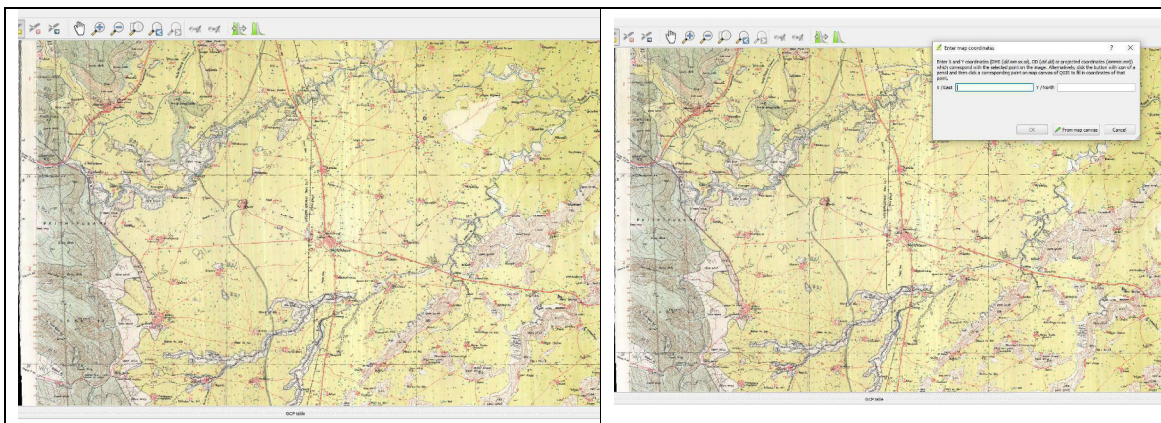
The plugin window is divided into 2 sections. The top section where the raster will be displayed and the bottom section where a table showing your GCPs will appear.



Now for opening any Raster, click on File > Open Raster

In the next screen, you will be asked to choose the raster's coordinate reference system (CRS). This is to specify the projection and datum of your control points. If you have collected the ground control points using a GPS device, you would have the WGS84 CRS. If you are geo-referencing a scanned map like this, you can obtain the CRS information from the map itself. Looking at our map image, the coordinates are in Lat/Long. There is no datum information given, so we have to assume an appropriate one. Since it is India and the map is quite old, we can bet the Everest 1830 datum would give us good results.

You will see the image is opening just as left side image of given table.



Now for rectifying/georeferencing it, you have to Click Add point button of the toolbar, and then Click on the point for which you know the co-ordinate. A pop – up window will appear as shown in the right side of table, asking for entering latitude and longitude of the point.

Once your entered lat. /long. Then proceed similarly for the next points, it must be noted that for a good rectification you need to take at least four points and there is no limit of number of points to be taken.

Once all points taken, then Go to > File > Click on Start Georeferencing

Once the process finishes, you will see the georeferenced layer loaded in QGIS.

c. Data Downloading from GPS in QGIS

To transfer user data, such as waypoints, routes, and tracks from the GPS devices, some additional equipment will be required.

The first item needed is a PC cable that will allow this device to connect to the 9-pin serial port on a computer. If the computer you are using does not have a serial port, you will also need a Serial to USB converter cable. The converter cable is only compatible with 32-bit versions of windows.



Enabling the Device to Receive User Data

- With the device powered ON, press the **Menu** button twice to access the **Main Menu**
- Highlight **Setup** and press the **Enter** button
- Highlight the **Interface** tab at the top of the screen
 - If the **Interface** tab does not appear, keeps scrolling to the left until it is visible
- Press the **Down** button one time
- Press the **Enter** button
- Highlight **GPS** and press the **Enter** button

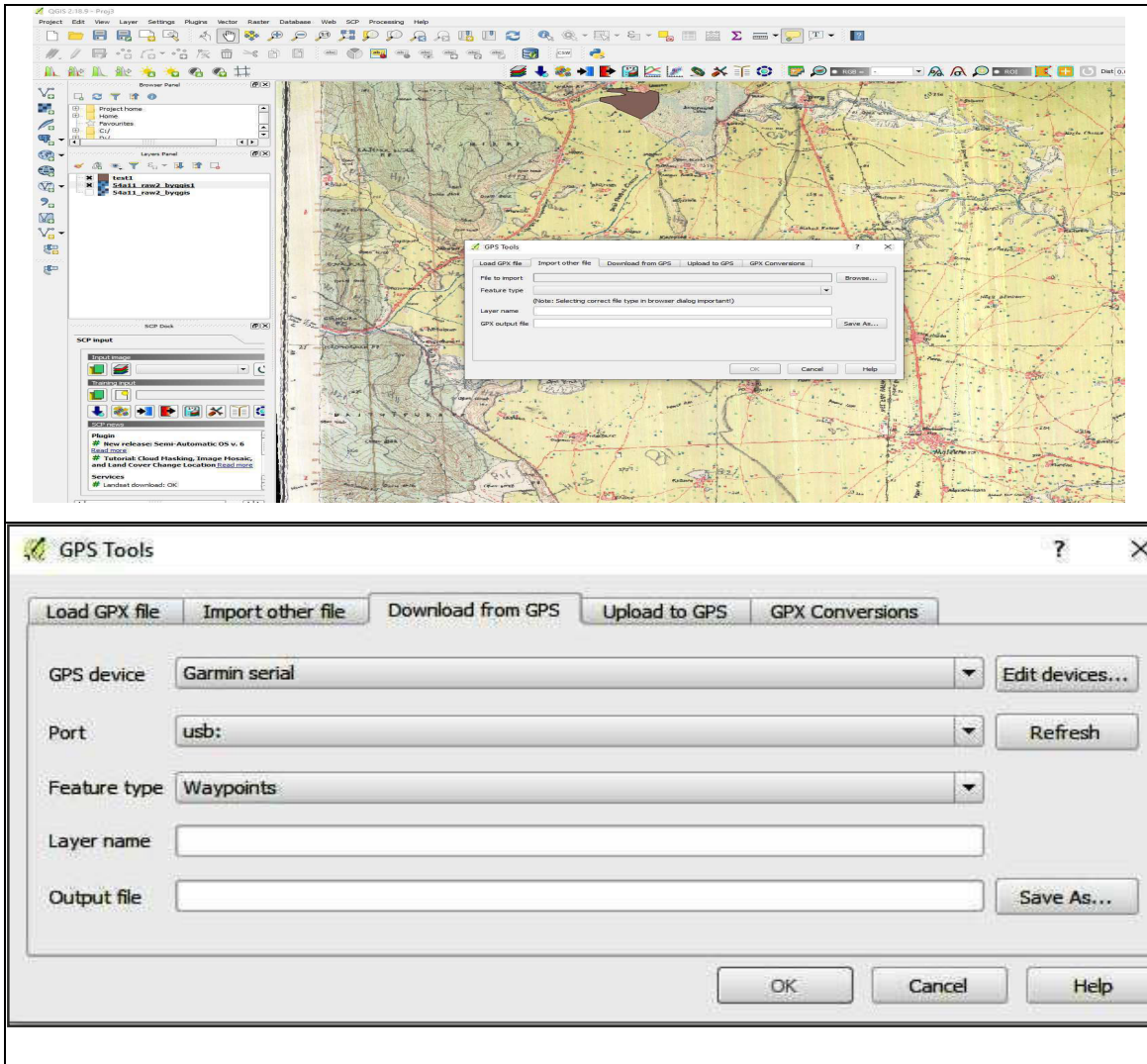
Global Positioning System (GPS)

The Global Positioning System (GPS) is a space-based satellite navigation system that provides location and time information in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites. The system provides critical capabilities to military, civil and commercial users around the world. It is maintained by the United States government and is freely accessible to anyone with a GPS receiver.

Once you have collected/recorded data on a GPS receiver, the optimum use of this collected data must be possible when you transfer this to computer system, so that required

mapping objectives may be achieved. QGIS provides very easy utility for transferring/importing/downloading the GPS data into the computer or QGIS. Here are the steps:

On the Menu bar Click > Vector then GPS and choose GPS Tools



- Browse to the folder which contains the *.GPX file and select it. Use the browse button to select the GPX file, and then use the checkboxes to select the feature types you want to load from that GPX file. Each feature type (waypoint, track and routes) will be loaded in a separate layer (i.e. waypoint list as a point layer) when you click OK.
- Here the layer created is a temporary, for converting it into *.shp or feature class, we will need to create a new shapefile from it. For that, right click on the layer and choose the option Save as from the layer properties menu. This will bring a new dialog box where we can set the name of the new output file and its format (ESRI shapefile). Enable the option Add saved file to map.

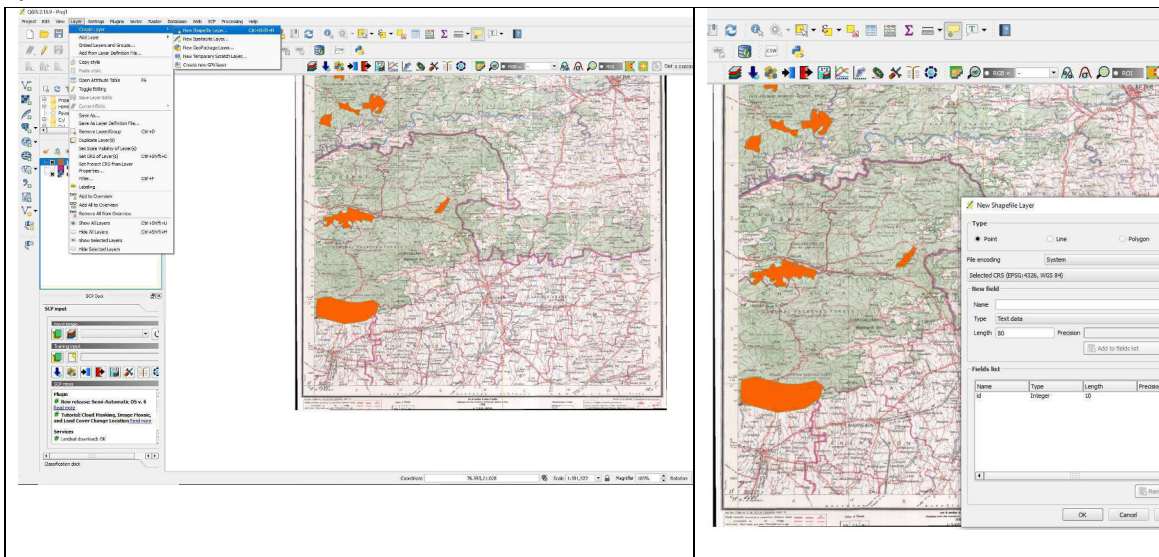
d. Now our next step will be creation of new *.shp file.

OR

Drawing Lines, Points or Polygons (Vectors) in QGIS

Open the georeferenced image (Layer > Add Raster Layer) or WMS layer (Layer > Add WMS Layer) in QGIS.


Select Layer > New > New Shapefile Layer to create the new empty layer for your vector feature. You will be prompted to confirm the Coordinate Reference System for the layer. By default, the Coordinate Reference System for the project is used, but you can override this if you wish.



You will then be prompted to confirm:



- The **Type of feature** you are tracing - point, line or polygon
- The **Name of the Attribute** you are tracing
- The **Type of the Attribute** - eg, Text, Whole Number
- The **Width** of the field for the Attribute. Select Add to Attribute list
- The **Shapefile Name** to save the layer as

The shapefile will be added to the Table of Contents panel on the left hand side. RIGHT click on the layer to bring up a menu of options connected to it.

To add or edit features, select the Toggle editing button  or RIGHT click on the shapefile layer in the Table of Contents and select Toggle editing.



For capturing **points**:

Click on the Capture Points button ,

For capturing **lines** Click on the Capture Line button  and For capturing **polygons** > Click on the Capture Polygon button 

Click a set of points along the perimeter of the polygon. RIGHT click when you have reached the final point to finish, and choose a name for the polygon and save it using toggle editing button.



For editing in the existing shape file containing features points, lines and polygons:

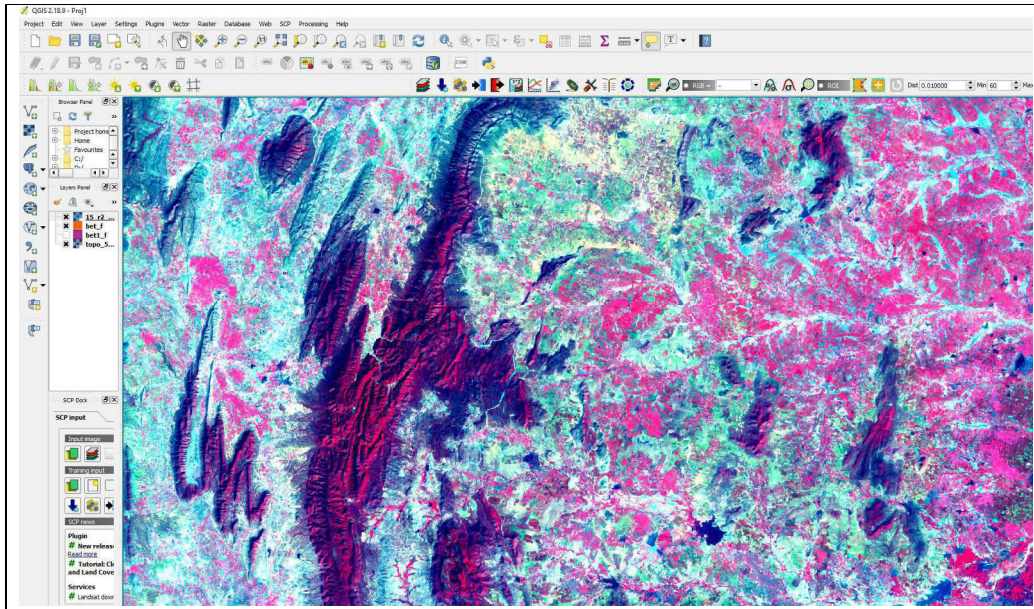
- The Move Feature button  allows a whole feature to be moved to a new position
- The Node tool  allows you to select the feature and move one of the points to a new position
- RIGHT *click* on the layer in the Table of Contents and select Open attribute table to view and edit the information connected to the features
- RIGHT click on the layer in the Table of Contents and select Query to select features in the layer
- RIGHT click on the layer in the Table of Contents and select Save As to save the layer in a number of formats, including KML.
- You can save your current open layers and features at any time by using the Save Project button

e. Opening a Satellite image in QGIS

Raster data in GIS are matrices of discrete cells that represent features on, above or below the earth's surface. Each cell in the raster grid is the same size, and cells are usually rectangular (in QGIS they will always be rectangular). Typical raster datasets include remote sensing data, such as aerial photography, or satellite imagery and modelled data, such as an elevation matrix.

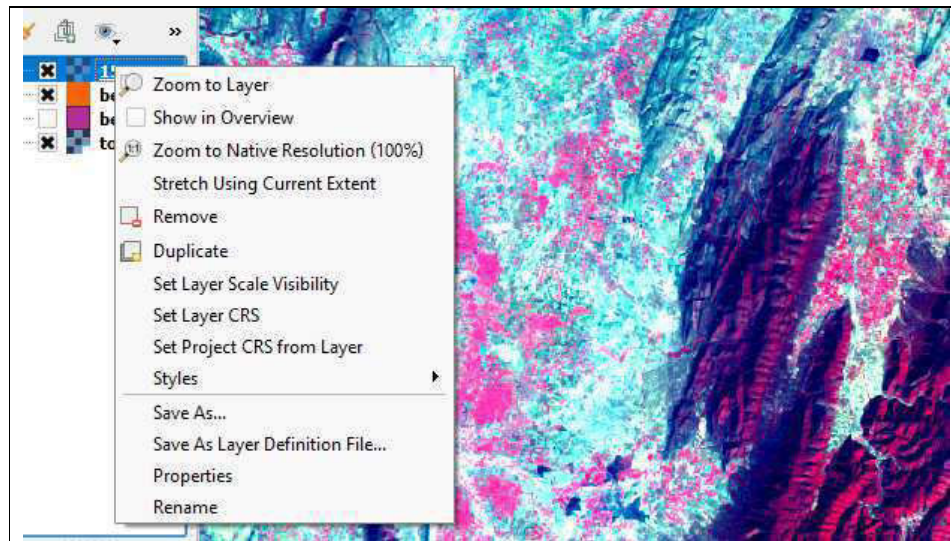
Unlike vector data, raster data typically do not have an associated database record for each cell. They are geocoded by pixel resolution and the x/y (Latitude, Longitude) coordinate of a corner pixel of the raster layer. This allows QGIS to position the data correctly in the map canvas.

Raster layers are loaded either by clicking on the  Add Raster Layer icon or by selecting the Layer ▸  Add Raster Layer menu option. More than one layer can be loaded at the same time by holding down the Ctrl or Shift key and clicking on multiple items in the Open a GDAL Supported Raster Data Source dialog.



Once a raster layer is loaded in the map legend, you can click on the layer name with the right mouse button to select and activate layer-specific features or to open a dialog to set raster properties for the layer.

Right mouse button menu for raster layers



f. Map Composition in QGIS

A map is a symbolic depiction emphasizing relationships between elements of some space, such as objects, regions, or themes. Many maps are static, fixed to paper or some other durable medium, while others are dynamic or interactive.

Map composition means the layout of the basic structural map elements. Their composition is a result of creative abilities of the author with the aim of follow cartographic rules and it depends on many circumstances such as name, scale, legend, map field and

themes etc. Before going to creating map composition, just go through in brief of some of these mentioned elements.

North arrows

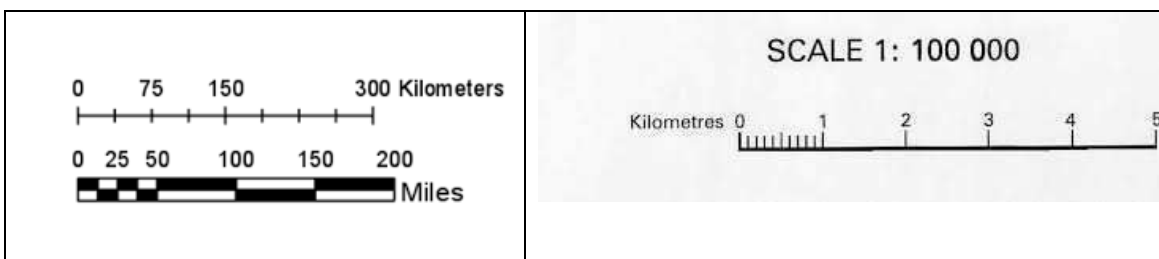
A north arrow is a graphical representation indicating the direction of north in an Area. And a true north arrow points to geographic north pole where the lines of longitude converge. A magnetic north arrow points to the magnetic north pole, which is the direction indicated by a magnetic compass.



Scale bars and Scale text





































Scale bars provide a visual indication of the size of features, and distance between features, on the map. A scale bar is a line or bar divided into parts. It is labeled with its ground length, usually in multiples of map units, such as tens of kilometers or hundreds of miles.

Scale text indicates the scale of the map and features on the map. Scale text tells a map reader how many ground units are represented by a map unit—for example, "one centimeter equals 100,000 meters". Scale text can also be an absolute ratio independent of units, such as 1:24,000.



Legends

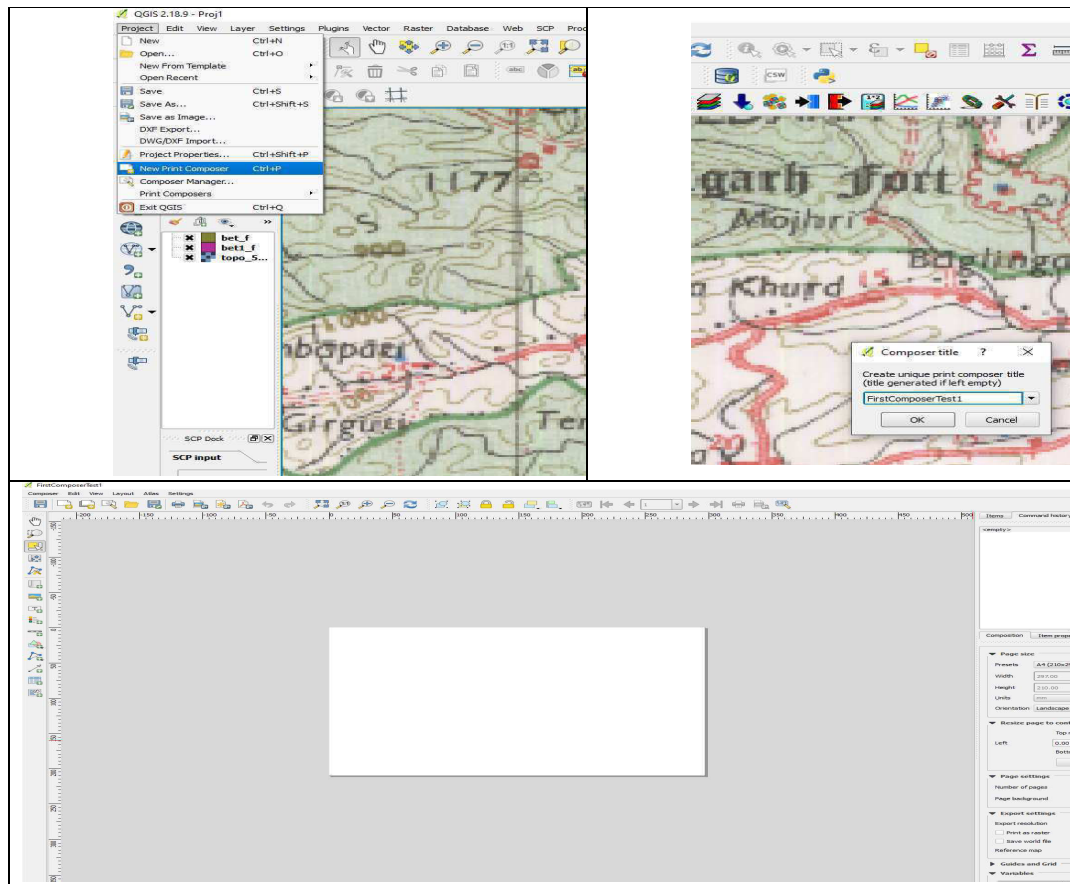
A map legend or key is a visual explanation of the symbols used on the map. It typically includes a sample of each symbol (point, line, or area), and a short description of what the symbol means. The legend is a graphical representation of information, with design principles similar to a map or any other graphic.

 Quaternary	 Devonian	Legend	 Interstate	 NPDES Facilities
 Neogene	 Silurian		 US and State Highway	 Dams
 Paleogene	 Ordovician		 Local Thoroughfare	 ESA Points
 Cretaceous	 Cambrian		 Toll Road	 State Boundary
 Jurassic	 Neoproterozoic		 Ramp	 Public School
 Triassic	 Mesoproterozoic		 Railroad	 Private School
 Permian	 Paleoproterozoic		 USGS 100K Index	 Airports
 Carboniferous	 Archean		 Municipal Boundary	 ESA
			 Water Bodies	 Tribal Land
			 Rivers	 USCG Jurisdiction

Frames

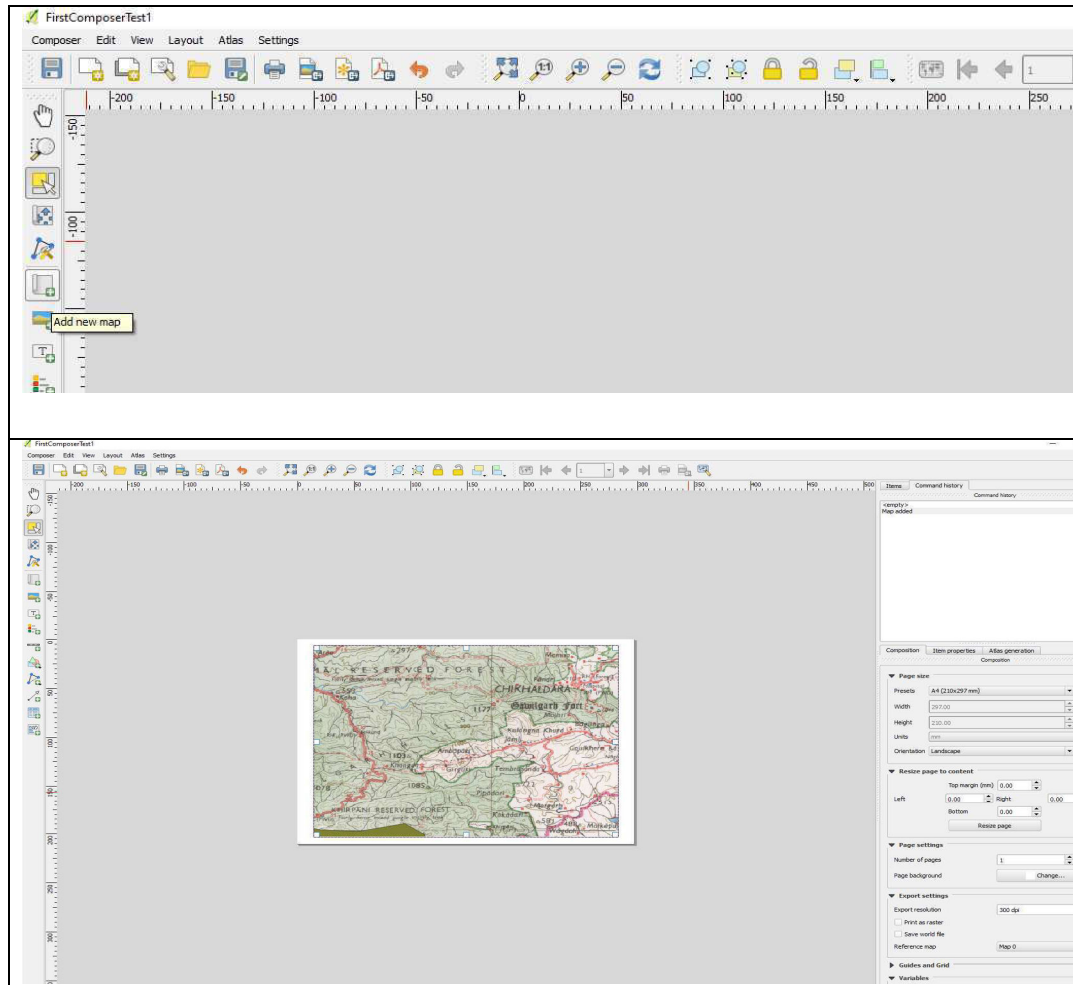
Map frames are containers for maps on your page. They can point to any map or scene in your project. They can even be empty, which most often occurs when creating templates. It is important to note that the extent of the map inside a map frame is unique and independent of any map view that may be open in the project.

To compose a map in QGIS >go to menu toolbar > click project> select new Print Composer.

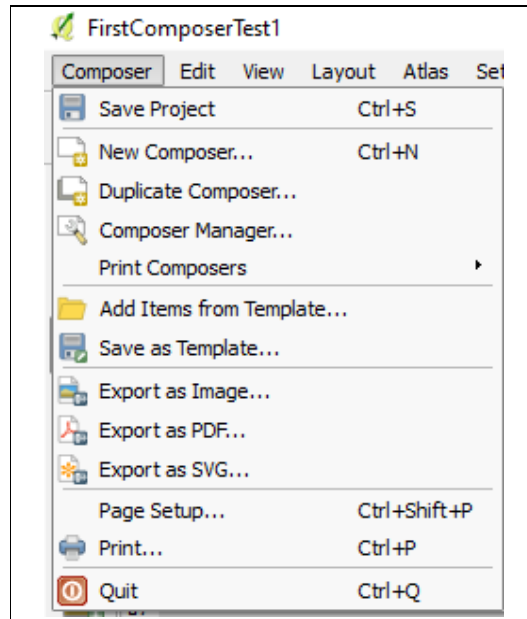
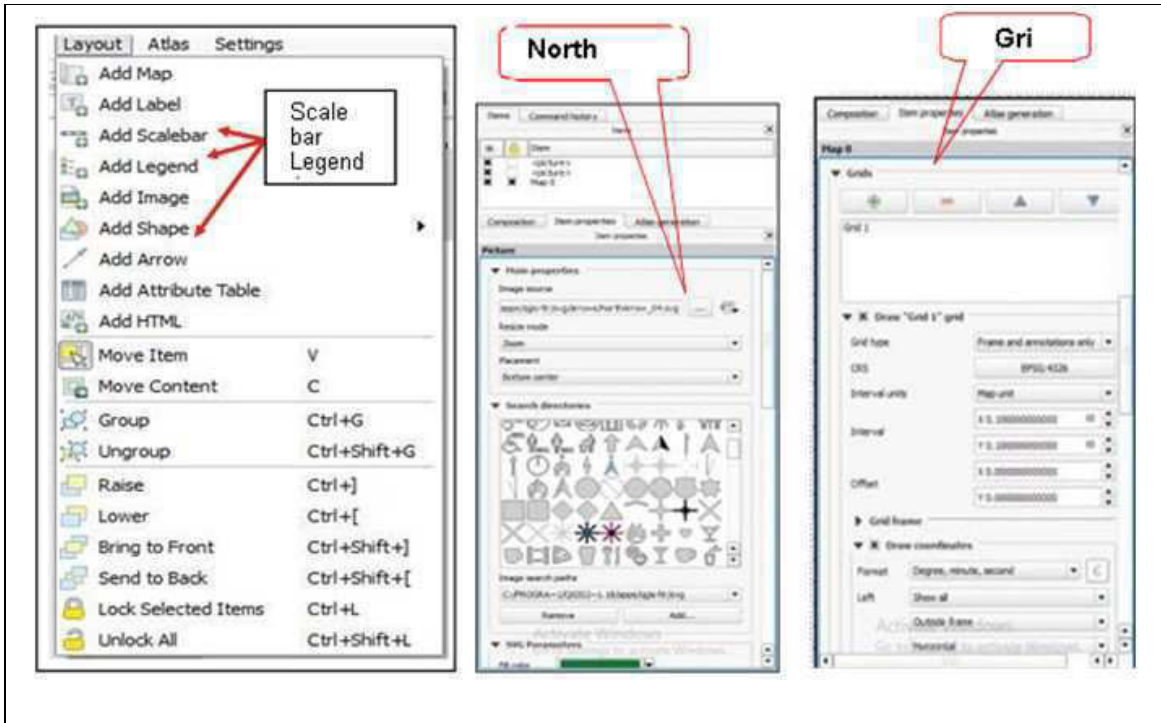


After clicking on the new print composer one dialog box will appear >compose title > name the Composer File Title.

Opening the Print Composer provides you with a blank canvas as mentioned in above figure, which represents the paper surface when using the print option> click on add new map> drag it in the map canvas.



Go to layout toolbar to add label, scale bar, legend and north arrow >to generate or add grid > click on item properties on the left side of map canvas > click on grid > add grid (+) > specify the CRS >adjust the interval > select the draw coordinates option > change the format into degree decimal or degree minute second.



7.4 SUMMARY

Quantum GIS (QGIS) is one such software which is freely downloadable and open source software of GIS applications. QGIS provides excellent capabilities, stable operation and support numerous vector, raster, database, formats and functionalities. Like other GIS applications, QGIS provides a geographical user interface wherein different layers of spatial data can be displayed and analyzed. Although much efforts are done for compiling of this unit by taking few examples, but students/researchers are advised and recommended to explore the QGIS up to optimum level. Lots of online help, communities are associated, which are always available and ready for extended assistance.

7.5 GLOSSARY

Datum: A geodetic datum or geodetic system is a coordinate system, and a set of reference points, used for locating places on the Earth.

WGS 84: WGS84 is an extensively used term in mapping. It stands for World Geodetic System 1984 and comprises of an ellipsoid, a horizontal datum, a vertical datum and a coordinate system.

Plug-in

In computing, a plug-in is a software component that adds a specific feature to an existing computer program. When a program supports plug-ins, it enables customization

Python

Python is a general purpose and high level programming language. You can use Python for developing desktop GUI applications, websites and web applications. Also, **Python**, as a high level programming language, allows you to focus on core functionality of the application by taking care of common programming tasks.

Shapefile

The shapefile format is a geospatial vector data format for geographic information system (GIS) software. It is developed and regulated by ESRI as a mostly open specification for data interoperability among ESRI and other GIS software products. The shapefile format can spatially describe vector features: points, lines, and polygons, representing, for example, water wells, rivers, and lakes. Each item usually has attributes that describe it, such as name or temperature.

Open Source Software (OSS)

OSS is computer software with its source code made available with a license in which the copyright holder provides the rights to study change and distribute the software for any purpose. Open source software is developed in a collaborative public manner.

Raster

One type of geospatial data is called raster data or simply "a raster". The most easily recognised form of raster data is digital satellite imagery or air photos. Elevation shading or digital elevation models are also typically represented as raster data. Any type of map feature can be represented as raster data, but there are limitations.

Raster data in GIS are matrices of discrete cells that represent features of the earth's surface. Usually satellite or a real imagery is represented in Raster data format like GeoTIFF, img, pix, hdf (Hierarchical Data Format) etc.

Multiple overlapping rasters are used to represent images using more than one colour value (i.e. one raster for each set of red, green and blue values is combined to create a colour image). Satellite imagery also represents data in multiple "bands". Each band is essentially a separate, spatially overlapping raster, where each band holds values of certain wavelengths of light. As you can imagine, a large raster takes up more file space. A raster with smaller cells can provide more detail, but takes up more file space.

Vector

There are various ways of representing these geographic coordinates depending on your purpose. Vector data is a coordinate-based data structure commonly used to represent point, linear and enclosed map features. Each linear feature is represented as x, y coordinates. Attributes stored in a database are associated with the features. Vector data includes:

Point features

A single coordinate (x y) represents a discrete geographic location (Temple, Monuments, Forest Chowkies, Forest Fire Point, etc.)

Line features

Multiple coordinates (x1 y1, x2 y2, x3 y4, ... xn yn) strung together in a certain order, like drawing a line from Point (x1 y1) to Point (x2 y2) and so on. These parts between each point are considered line segments. They have a length and the line can be said to have a direction based on the order of the points. Technically, a line is a single pair of coordinates connected together, whereas a line string is multiple lines connected together. (Roads, Railways, Fences, Cart-tracks, Pack-Tracks, etc.)

Polygon features

When lines are strung together by more than two points, with the last point being at the same location as the first, we call this a polygon. A triangle, circle, rectangle, etc. are all polygons. The key feature of polygons is that there is a fixed area within them. (Land Parcels, wetlands, forests, etc.)

OSGeo

The Open Source Geospatial Foundation (OSGeo) is a non-profit non-governmental organization whose mission is to support and promote the collaborative development of open geospatial technologies and data.

WMS

A Web Map Service (WMS) layer is based on data from an Open Geospatial Consortium (OGC) WMS server. The WMS layer allows you to access these services over the Internet and add their content to a map.

KML

Keyhole Markup Language (KML) is an XML-based markup language designed to annotate and overlay visualizations on various two-dimensional, Web-based online maps or three-dimensional Earth browsers (such as Google Earth).

7.6 ANSWER TO CHECK YOUR PROGRESS

1. State the available versions of QGIS.
2. Name the few proprietary and open source software.
3. What is raster and vector, give suitable examples.
4. Make a list of GIS software available, comprises of proprietary and open source separately.

7.7 REFERENCES

1. <https://www.qgis.org>
2. <https://en.wikipedia.org>
3. T. Mitchell. Web mapping illustrated, published by o'reilly, 2005.
4. https://www.qgistutorials.com/en/docs/georeferencing_basics.html
5. <https://www.google.com>
6. <https://www.garmin.com/en-US/AboutGPS/>
7. https://docs.qgis.org/2.8/en/docs/user_manual/working_with_raster/supported_data.html
8. Open Knowledge Foundation (2012)
9. <http://lists.osgeo.org/mailman/listinfo/qgis-user>
10. <http://lists.osgeo.org/mailman/listinfo/qgis-commit>
11. <http://lists.osgeo.org/mailman/listinfo/qgis-community-team>
12. <http://wiki.qgis.org>

7.8 TERMINAL QUESTIONS

1. What are the limitations of QGIS?
2. Write few file formats supported by QGIS as per Raster and Vector.
3. What are the advantages of importing GPS data over the collecting this manually?
4. Explain the buffer in any feature class (Point, Line, and Polygon) and also explain its uses.

UNIT 8 - OVERVIEW OF GIS PORTALS AND OPEN DATA ARCHIVES

8.1 OBJECTIVES

8.2 INTRODUCTION

8.3 OVERVIEW OF GIS PORTALS AND OPEN DATA ARCHIVES

8.4 SUMMARY

8.5 GLOSSARY

8.6 ANSWER TO CHECK YOUR PROGRESS

8.7 REFERENCES

8.8 TERMINAL QUESTIONS

8.1 OBJECTIVES

When any user or developer or any developing unit starts its respective production/development, there are some common basic essential elements/components required to develop/produce things. These things may be from raw material (physical), any service generating task, software development and many more, the list is endless. When we look for the development of some data (i.e. information) related services for society/individual/organization, out of many things, few things come first in mind, that are capital, space of operation or platform, resource person, backup of business operation, its marketing, selling, and maintenance with post-delivery operations. Like these many things are involved, and when we are considering the geospatial domain so out of many, satellite data/data use and procurement raises the key concern of the developer. Satellite data/data gives information as per specifications and they also involve cost according to its specifications. These costs may vary from sensor to sensor, as per timeline of the data, availability of data, types of data and depends on various factors. Sometimes, it is difficult to occur or arrange the cost for the individual/organizations, so that thing restricts the exploration of many innovative brains. By acknowledging the fact, efforts have been made by the Government, Non Government sides, which provide various Open Data Archives for students, researchers, academicians, developers and concerning user community. In this unit some common open data archives have been discussed, that ensure you/students/learners to overcome the limitations of financial issues that restrict you from exploration.

8.2 INTRODUCTION

Data is distinct pieces of information, usually formatted in a special way. Since the mid-1900s, people have used the word data to mean computer information that is transmitted or stored. Strictly speaking, data is the plural of datum, a single piece of information. These are characteristics or information that is collected through observation. In a more technical sense, data is a set of values of qualitative or quantitative variables about one or more persons or objects.

A Geographic Information System (GIS) is designed to capture, manipulate, store, analyze, and manage data. It is an extension of cartography — the science of making maps — and allows individuals to visualize, analyze, question, and interpret data.

An archive is an accumulation of historical records or the physical place they are located. Archives contain primary source documents that have accumulated over the course of an individual or organization's lifetime, and are kept to show the function of that person or organization.

When we consider geospatial domain, it enables us to acquire data that is referenced to the earth and use it for analysis, modeling, simulations and visualization. This technology allows you to make informed decisions based on the importance and priority of resources most of which are limited in nature. Through this discipline user can focus on using information technology to understand people, places, and processes of the earth. Spatial analysis of human and physical variables is fundamental to the discipline.

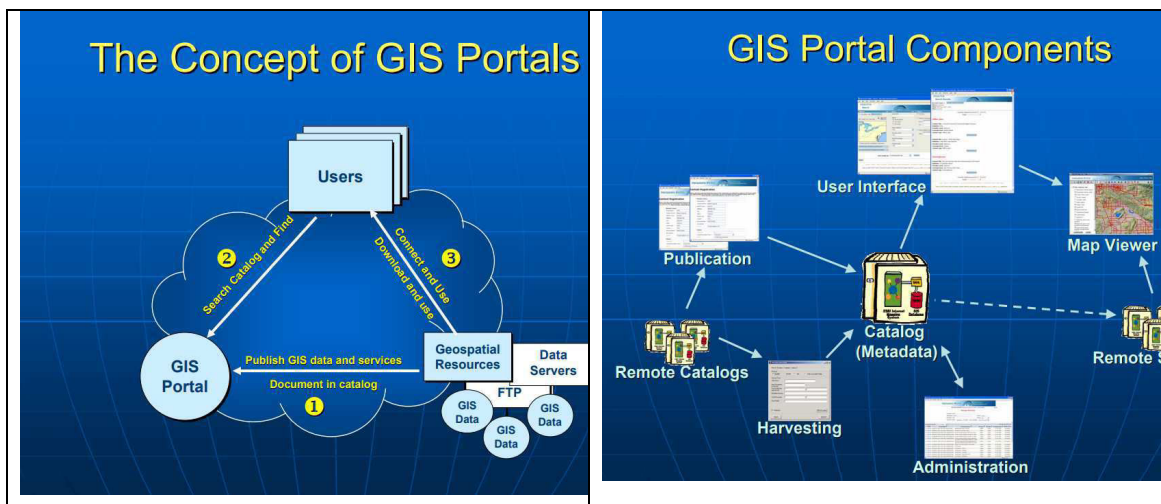
These data combines location information (Latitude and longitude), attribute information (the characteristics of the object, event, or phenomena concerned), and often also temporal information (the time or life span at which the location and attributes exist).

A web portal is a specially designed website that brings information from diverse sources, like emails, online forums and search engines, together in a uniform way. Usually, each information source gets its dedicated area on the page for displaying information; often, the user can configure which ones to display.

8.3 OVERVIEW OF GIS PORTALS AND OPEN DATA ARCHIVES

The GIS Portal gives authorized users access to a wealth of Geographical Information System (GIS) resources. These resources help users make good reports, smarter decisions, cultivate stronger relationships, and enhancing their knowledge.

Geographic information providers, including government agencies and commercial sources, use geoportals to publish descriptions (geospatial metadata) of their geographic information. Geographic information consumers, professional or casual, use geoportals to search and access the information they need. Thus geoportals serve an increasingly important role in the sharing of geographic information and can avoid duplicated efforts, inconsistencies, delays, confusion, and wasted resources.



GIS Portal Resources include:

- **Geospatial layer documentation**

Layers make up the primary content and are logical collections of geographic data that can be used to create maps and applications, and can be broadly categorized as either feature or tile layers of any applications you will build with any GIS software. These layers can be based on different services or other files available on disk or over a network. Layers can also be created in your applications from arbitrary data loaded over a network or from disk using any open source or proprietary software. Individual graphics can also be displayed on maps using graphics layers.

This table summarizes some of these core layer types and their common use cases.

Layer Type	Service	Use Cases
Feature Layer	Feature Service	Querying, rendering, and editing vector geographic information.
Tile Layer	Map Service	Basemaps and other complex datasets that change infrequently. Tiles can be kept in sync with feature layers for operational data.
Vector Tile Layer	Vector Tile Service	Basemaps and other complex datasets that change infrequently. Tiles can be kept in sync with feature layers for operational data.
Dynamic Map Layer	Map Service	Complex data sets that change frequently or need complex rendering requirements.
Image Layer	Image Service	Rendering and analyzing satellite or other imagery data.
Scene Layer	Scene Service	Displaying and rendering 3D datasets.

- **Product documentation**

Product documentation is developed according to standards and best practices and is accurate, complete, and clear for front-end and back-end for continued development and sustainment efforts.

- **Product/software downloads**

Some web portals allow you to do some really amazing things in spatial and non spatial domain. These includes generation/downloading various spatial available layers, remote sensing and GIS software, d software/suits related to photo editing, sophisticated computer gaming, video streaming—all of these things are possible because of different types of software are available in the cloud also. Developers are always creating new software applications, which allow you to do even more with your computer.

- **Geospatial Enterprise Real-Time**

Real-Time GIS gives you the ability to simultaneously tap into, analyze, and display streaming data from many sensors, devices, and social media feeds. You can define filters and location-based analytics that automatically refine and focus real-time data on events that matter most to you.

- **Developer resources**

Developers/communities develop/produced numerous software/data resources for different user communities. These resources are further extensively used by researcher of user communities for developing new assignments.

- **Etc.**

Open Data includes

Since knowledge is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness. Open data is data that

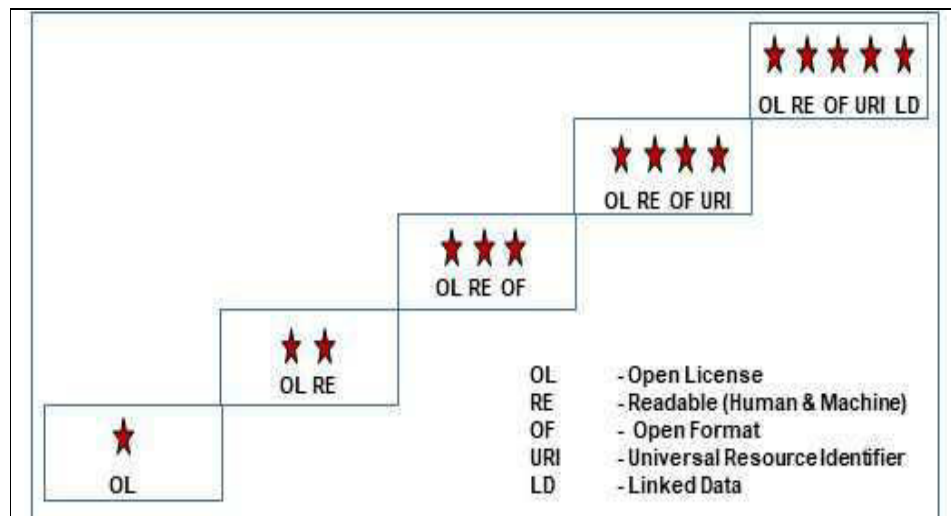
can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.

To summarize the most important:

- **Availability and Access:** the data must be available as a whole as and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
- **Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of Endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

Five Star Open Data

Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5 star deployment scheme for Open Data. Here, we give examples for each step of the stars and explain costs and benefits that come along with it.



The cost and benefit of these open data levels are detailed below:

★ Available on the web (whatever format) but with an open license, to be called as Open Data. The benefits are:

- You can look at it.
- You can print it.
- You can store it locally (on your hard drive or on a USB stick).
- You can enter the data into any other system.
- You can change the data as you wish.
- You can share the data with anyone you like.

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

It includes

- You can directly process it with proprietary software to aggregate it, perform calculations, visualize it, etc.
- You can export it into another (structured) format.

★★★ As (2) plus non-proprietary format (e.g. CSV instead of excel)

This includes

- You can manipulate the data in any way you like, without being confined by the capabilities of any particular software.

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff using an URI

- You can link to it from any other place (on the Web or locally).
- You can bookmark it.
- You can reuse parts of the data.
- You may be able to reuse existing tools and libraries, even if they only understand parts of the pattern the publisher used.
- Understanding the structure of an RDF "Graph" of data can be more effort than tabular (Excel/CSV) or tree (XML/JSON) data.
- You can combine the data safely with other data. URIs are a global scheme so if two things have the same URI then it's intentional, and if so that's well on it's way to being 5 star data!

★★★★★ All the above, plus: Link your data to other people's data to provide context

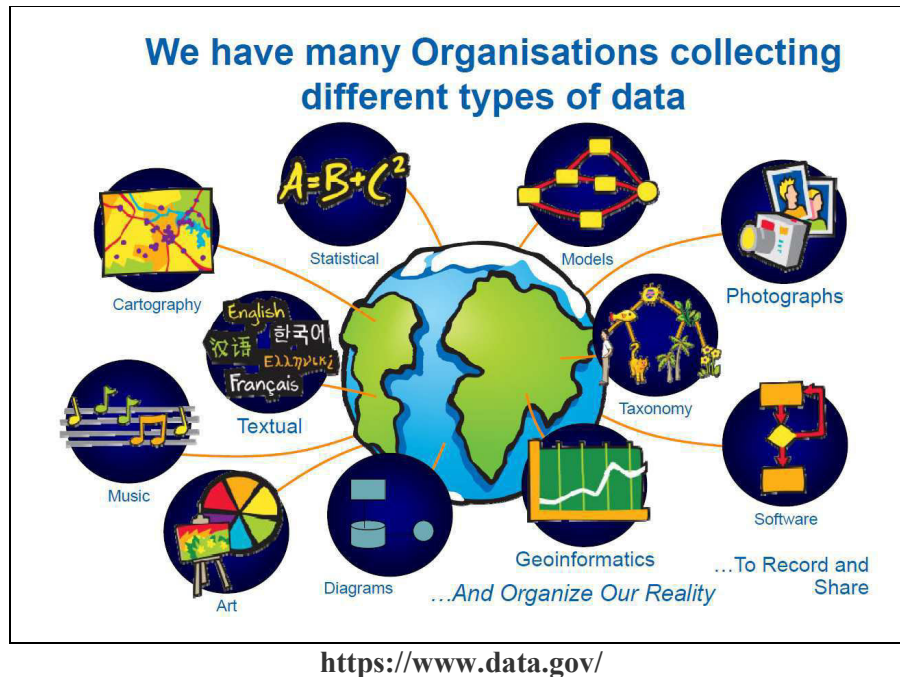
- You can discover more (related) data while consuming the data.
- You can directly learn about the data schema.
- You now have to deal with broken data links, just like 404 errors in web pages.

Data Classification

Different types of data sets generated both in geospatial and non-spatial form by different ministries /departments are to be classified as shareable data and non-shareable data. The types of data produced by a statistical system consist of derived statistics like national accounts statistics, indicators like price index, data bases from census and surveys. The geospatial data however, consists primarily of satellite data, maps, etc. In such a system, it becomes important to maintain standards in respect of metadata, data layout and data access policy.

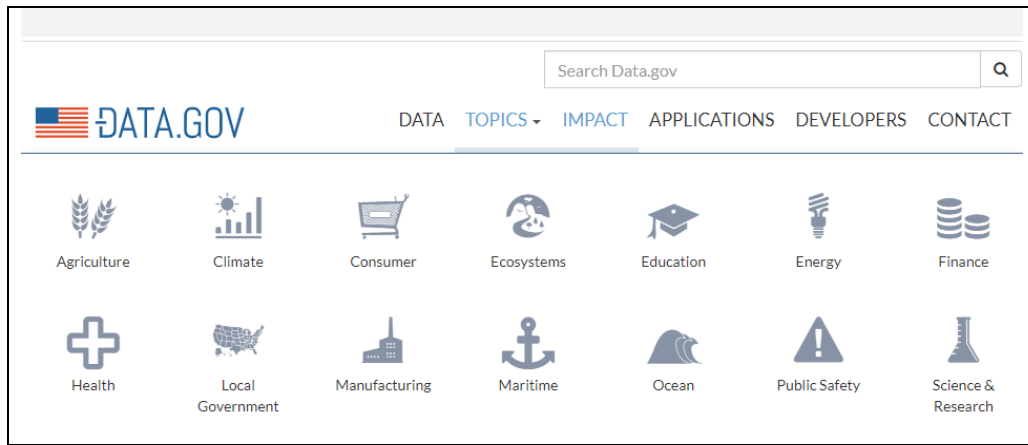
For sharing, there are numbers of websites available, you can browse them and download different types of data/software for developing their skills in the required domain. It must take utmost care from an arbitrary link as fact is as risky as letting people include content from any website in your pages. Caution, trust and common sense are all still necessary.

Here below some description of open data archives websites are given, you can explore these sites and try as much as possible extraction related to your objectives. While downloading/sharing and data or credentials, it is highly recommended that students/users/researchers ensure that the source/website should be authentic.



Data.gov is a rich resource for civic hackers, tech entrepreneurs, data scientists, and developers of all stripes. Here you'll find information about APIs, open source projects, and relevant developer resources across government. You'll also find updates on the data.gov infrastructure itself, an open source project managed on GitHub.

Data.gov also syndicates data from state and local governments. However, non-federal data sources are governed by different terms of service and often different licenses than Federal data. When using or harvesting data from Data.gov, please note this distinction. When harvesting large volumes of data or metadata through Data.gov, site recommend you filter for Federal sources and separate non-federal sources to avoid comingling metadata without making this distinction.



A little description of few topics in the archive is

Here you can find data related to climate change that can help inform and prepare America’s communities, businesses, and citizens. You can currently find data and resources related to coastal flooding, food resilience, water, ecosystem vulnerability, human health, energy infrastructure, transportation, and the Arctic region. Over time, you will be able to find additional data and tools relevant to other important climate-related impacts.

Explore education datasets, apps, resources for the classroom, and details about paying for college. These datasets cover education at all levels.

Explore datasets, tools, and applications related to health and health care. These resources come from across the Federal Government with the goal of improving the health and lives of all Americans.

Discover high-value public science and research data from across the Federal Government.

<https://data.gov.in/dataset-group-name/nrsc-open-data-archive>





About the portals (Open Government Data (OGD) Platform India)

Open Government Data (OGD) Platform India - data.gov.in - is a platform for supporting Open Data initiative of Government of India. The portal is intended to be used by Government of India Ministries/ Departments their organizations to publish datasets, documents, services, tools and applications collected by them for public use. It intends to increase transparency in the functioning of Government and also open avenues for many more innovative uses of Government Data to give different perspective.

The base Open Government Data Platform India is a joint initiative of Government of India and US Government. Open Government Data Platform India is also packaged as a product and made available in open source for implementation by countries globally.

The entire product is available for download at the Open Source Code Sharing Platform "GitHub".

Open Government Data Platform India has 4 (four) major modules, as detailed below, implemented on a single Drupal instance – An Open Source based Content Framework Solution

- **Data Management System (DMS)** – Module for contributing data catalogs by various government agencies for making those available on the front-end website after a due approval process through a defined workflow.
- **Content Management System (CMS)** – Module for managing and updating various functionalities and content types of the Open Government Data Platform India Platform.
- **Visitor Relationship Management (VRM)** – Module for collating and disseminating viewer feedback on various data catalogs.
- **Communities** – Module for community users to interact and share their zeal and views with others, who share common interests as that of theirs.

<https://deemed.ugc.ac.in/>

The UGC is determined to introduce accountability and transparency in the processing of applications for grant of the status of deemed to be university under Section 3 of the UGC Act, 1956. The UGC has, therefore, placed all the information regarding status of such applications on its website alongwith the performance and academic outputs of the existing Deemed to be Universities. Keeping in tune with the initiatives of Government of India to ensure that the Government services are made available to citizens electronically, the UGC has launched a Web Portal for Deemed to be Universities.

The following facilities are available in the portal:

- (i) Detailed information about Deemed to be Universities;
- (ii) Existing Deemed to be Universities will submit their academic outcomes and performance on line;
- (iii) Applications for Deemed to be University status will be received on-line by the UGC;
- (iv) Applications for off-campus centre(s), off-shore campus centre(s) /new courses/new departments/new schools will be received on-line by the UGC;
- (v) Students. Parents and general public can approach the portal for redressal of their grievances related to Deemed to by Universities.

<https://nsdiindia.gov.in/nsdi/nsdiportal/index.jsp>

The India Geoportal at NSDI, New Delhi <https://nsdiindia.gov.in> provides a single window gateway access to the Data Nodes of the NSDI partnering agencies from the Central

Ministries/ Departments and the State Governments. Considering the importance of geospatial data and its varied applications in the present day environment, Government of India constituted a Task Force to suggest the implementation modalities. To meet the needs of users, planners, policy makers, industry and academia and to reap the benefits of higher-end technologies like GIS, GPS, high resolution satellite sensors, a single source of information infrastructure is a long felt need. Moreover, fast computers and ICT technologies provide ways and means for supplying spatial data to the users on their desktops. With this objective in view, a web portal named National Spatial Data Infrastructure has been developed. National Spatial Data Infrastructure (NSDI) is a national endeavour that has come to be seen as the technology, policies, criteria, standards and people necessary to promote geospatial data sharing throughout all levels of government, the private and non-profit sectors, and academia.

NSDI Vision “That current, accurate and organized geospatial data will be readily and continuously available and will be accessible on a national, state, district and village level basis to contribute to the economic, environmental and social growth of the country”.



Governance: Establish the governance structure required to optimize the benefits from the government’s geospatial resources.

Data: Ensure the capture, preservation, and maintenance of fundamental (priority) geospatial data sets, and set guidelines for non-fundamental geospatial data.

Access: Ensure that government geospatial information and services can be readily discovered, appraised, and accessed.

Product Space: Coordinate with BIS, product developers, hardware manufacturers, smart phone manufacturers, telecom service providers, and users.

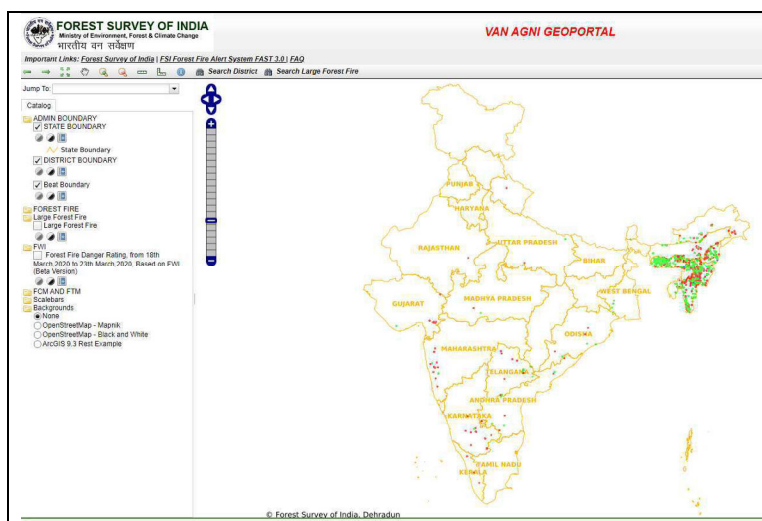
Interoperability: Ensure that geospatial data sets, services, and systems owned by different government agencies can be combined and reused for multiple purposes.

<http://fsi.nic.in/>

(Forest Fire Geo Portal of Forest Survey of India – Van Agni Geoportal)

Forest fires are one of the major drivers of damage caused to forests in the country. Uncontrolled forest fires can lead to significant losses of forests and ecosystem services. Studies suggest that climate change influences forest fire frequency and intensity which results in forests becoming increasingly inflammable. Technology such as satellite remote sensing based forest fire detection in near real time is of critical help in controlling forest fires.

Recognizing the need to revamp forest fire management in the country, the Ministry of Environment, Forest and Climate Change (MoEF&CC), Government of India has come up with the National Action Plan on Forest Fires, 2018. One of the main objectives of the action plan is to reduce the incidences of fires by informing, enabling and empowering forest fringe communities and incentivizing them to work in tandem with the State Forest Departments (SFDs).



Forest Fire alerts disseminated by Forest Survey of India (FSI) are based on the inputs from two collaborative arrangements between NASA-ISRO and ISRO-FSI. The Fire hotspots detected by MODIS (1km x 1km resolution) and SNPP-VIIRS (375m X 375m resolution) sensors are received at Shadnagar Earth station of NRSC and processed using NASA's algorithm. The fire hotspots are transmitted to FSI electronically, which are then processed automatically by FSI and alerts are generated and disseminated to the registered end users. The information and the analysis generated in the form of Table and Maps are also uploaded online on the website and a dedicated Geo portal for the purpose.

This Forest Fire Geoportal is used to view forest fire related data along with other thematic layers. Along with this Web Map Service (WMS) has been made available for integration with web portals/geo portal of State Forest Departments (SFDs) and other disaster management agencies.

https://bhuvan.nrsc.gov.in/bhuvan_links.php

(ISRO's Geo-portal Bhuvan: Gateway to Indian Earth Observation)

Bhuvan, is the national Geo-portal developed and hosted by ISRO comprising of Geo Spatial Data, Services and Tools for Analysis. It has many versatile features, for Example

- Visualization of Satellite Imagery and Maps
- Analysis
- Free Data Download and
- Download Reports to name a few

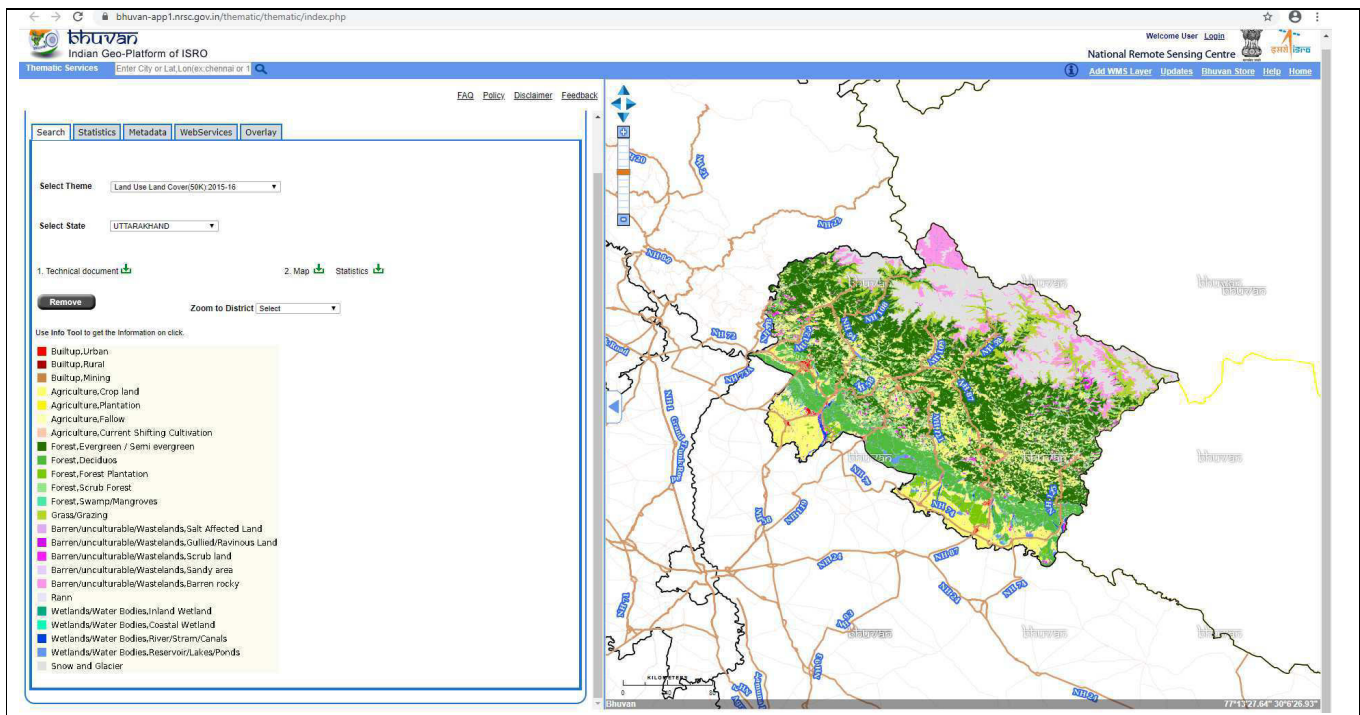
The Satellite Imageries are of Multi-sensor, Multi-platform and Multi-temporal in nature can be visualized in 2D and 3D. Varieties of thematic maps are also hosted for analysis of various natural resources themes. Bhuvan platform supports many applications that address Governance and other Geo Spatial applications that are being used by the Central and State Governments Departments, Academia and Industry.

It is today reckoned as the unique GIS platform of the country that is freely accessible on the Internet and used by wide varieties of user community including school children.

The screenshot displays the Bhuvan Indian Geo-Platform of ISRO website. The header includes the Bhuvan logo, the text "Indian Geo-Platform of ISRO", and the National Remote Sensing Centre logo with the text "NRSC | Search | Bhuvan Store | Bhuvan Collaborators".

The main content area is organized into several sections:

- Visualisation & Free Download:** Includes links for Bhuvan-2D, Bhuvan-3D, Open Data Archive (Free Download), and Climate & Environment (EO derived Products).
- Maps & OGC services:** Includes Thematic Services, Disaster Support Services, Management, Ocean Services, and Create a Map / GIS (My Maps | My GIS).
- Governance/Central Ministries:** Features a g-Governance Dashboard and various services like Flycatchers Distribution, Island Information, School Bhuvan, Toll Information, Ground Water, Pipeline Grid, Watersheds, and Urban.
- Application Sectors:**
 - Agriculture:** Forestry, E-Governance, Water, Tourism, Urban, Rural.
 - Rural:** Ground Water, Watersheds, MGNREGA.
 - Special Applications:** Data Discovery, Decision Support Dashboard, Hydrological Products, Image Gallery, International Disasters, Multilingual Translation, Online Shapefile Creation, OSCAT-3D, Smart Tracking, and Suvidha.
- State Portals/Applications (Click on any State):** A map of India with a dropdown menu for "State" (currently set to ANDHRA PRADESH) and a link to "Asset Mapping". Below this is a "Updates" section with a list of recent news items, such as "Bhuvan Training is postponed", "Release of Bhuvan JAINOORJA Portal", and "Update of Orthorectified AWiFS and LISS-3 datasets in NOEDA Portal".
- Map View:** A satellite imagery view of a city area with a central circular structure, overlaid with a grid and various map controls.



Bhuvan portal gives users an easy way to experience, explore and visualize satellite images from Indian Remote Sensing Satellites (IRS). This dataset has also integration capabilities with many open source GIS software viz., Q-GIS which offers researchers and planners to develop and orient spatial data to meet their specific requirements and objectives.

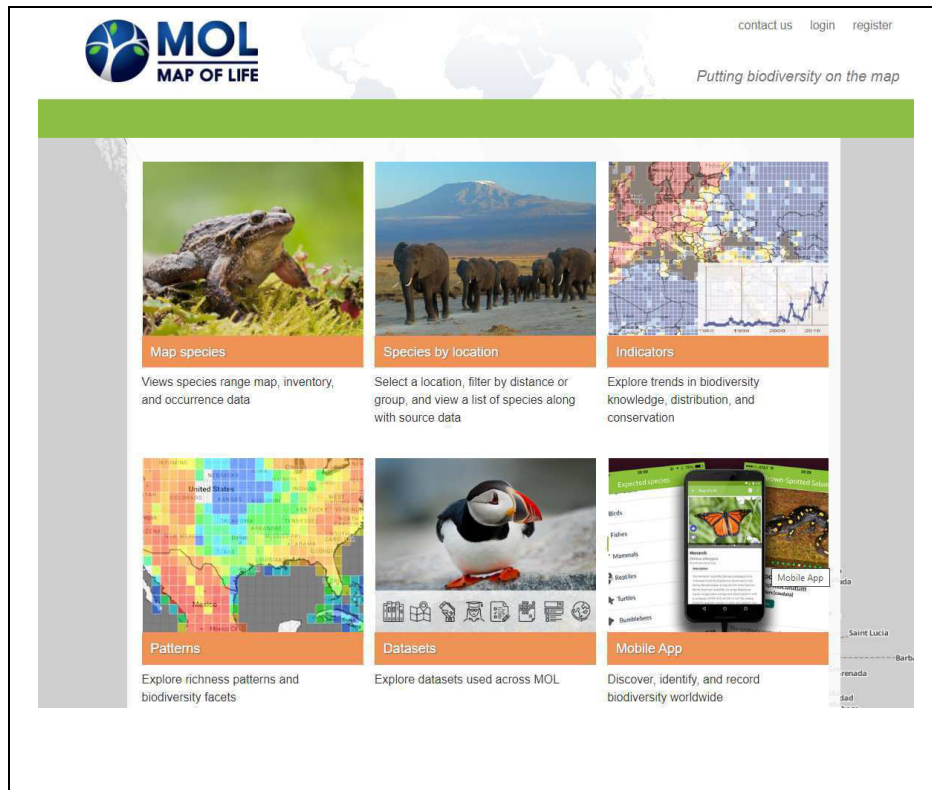
Among the various resources of the different regions this web based service is very useful at large for planning of ecosystem management, and to address specific management of important sectors like water, land and wasteland. It has an important role in mapping urbanization and growth of cities particularly on the face of climate change and disaster. Bhuvan has been upgraded with high focus on management services for disasters (cyclone, drought, earthquake, flood, forest fire, landslide etc.), tourism and with high resolution databases.

This portal has many spatial tools generally available in commercial GIS software. Any user having internet connection is capable to use online resources and to view earth features in 3 dimensional formats. A GIS layer can be created by adding their own point of interests and may be shared with others. Other facilities include chart routes, plot areas, calculate distances, and overlay separate images. Various layers can be shown or hide in any combination for required display. Using the scale and the robust measurement and terrain analysis tools, user can plot mileages, calculate elevation differences and measure slope angles between desired points. Further 3D view allows for terrain elevation profile along a path, and find places of interest along the way.

<https://www.mol.org/about>

Geographic information about biodiversity is vital for understanding the many services nature provides and their potential changes, yet remains unreliable and often insufficient. Built on a scalable web platform geared for large biodiversity and environmental data, Map of Life endeavors to provide 'best-possible' species range information and species

lists for any geographic area. Map of Life aims to support effective and global biodiversity education, monitoring, research and decision-making by assembling and integrating a wide range of knowledge about species distributions and their dynamics over time.



It assembles and integrates different sources of data describing species distributions worldwide. These data include expert species range maps, species occurrence points, ecoregions, and protected areas from providers like IUCN, WWF, GBIF, and more. All data assets are stored, managed, backed up, and accessed using a hosted cloud instance.

<https://www.gbif.org/>

GBIF—the Global Biodiversity Information Facility—is an international network and research infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth. It enables the users to navigate and put to use vast quantities of biodiversity information, advancing scientific research, serving the economic and quality-of-life interests of society, and providing a basis from which our knowledge of the natural world can grow rapidly and in a manner that avoids duplication of effort and expenditure.

Get data How-to Tools Community About

GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

Search

WHAT IS GBIF? ABOUT GBIF INDIA

Occurrence records 1,390,571,865	Datasets 51,271	Publishing institutions 1,574	Peer-reviewed papers 4,337
--	---------------------------	---	--------------------------------------

News: Call for proposals: Analysis of biodiversity data needs in the post-2020 framework

News: Call for data papers from European Russia

News: Call for nominations to the 2020 GBIF Young Researchers Award

2020 Ebbe Nielsen CH seeks open-data innov biodiversity

News: Virtual workshop: Advancing the Catalogue of the World's Natural History Collections

Data use: Biodiversidata publishes first occurrence dataset from Uruguay. Native amphibian, reptile, bird and mammal species recorded in Uruguay (photo by Pablo Balduvino - CC BY-NC 4.0)

Data use: Wikipedia pageviews reflect observable patterns in phenology

New species: *Liljeborg associata*. A new species of the genus Spence Bate, 1862 (Crust Amphipoda: Liljeborgiida with the burrows of the species *Urechis uncinatus* in the

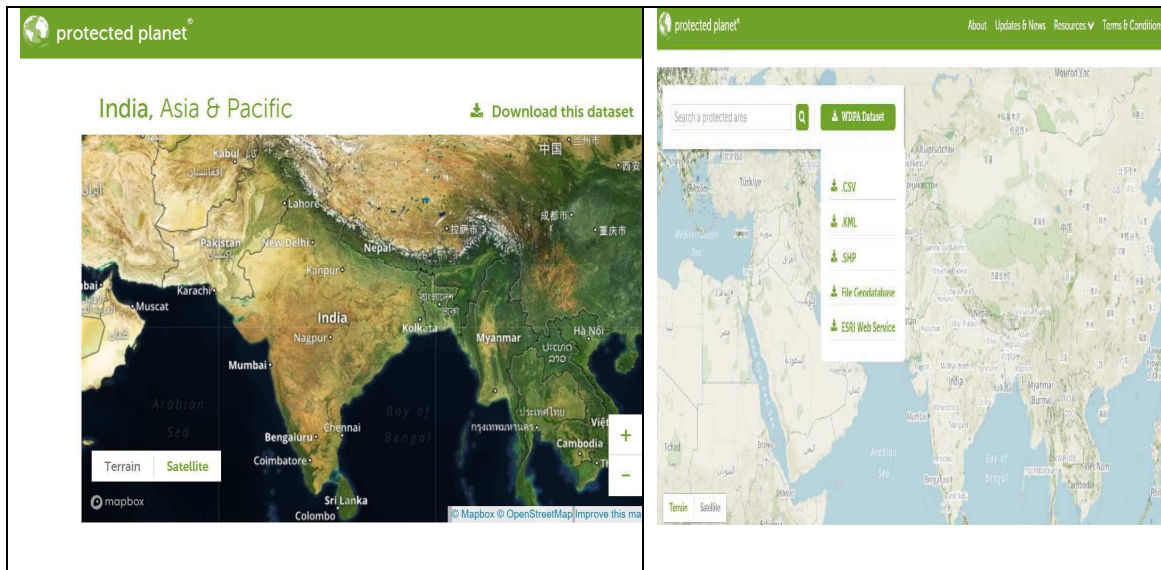
Working through participant nodes, provides data-holding institutions around the world with common standards and open-source tools that enable them to share information about where and when species have been recorded. This knowledge derives from many sources, including everything from museum specimens collected in the 18th and 19th century to geotagged smartphone photos shared by amateur naturalists in recent days and weeks.

<https://www.protectedplanet.net/>

Protected Planet is the most up to date and complete source of information on protected areas, updated monthly with submissions from governments, non-governmental organizations, landowners and communities. It is managed by the United Nations Environment World Conservation Monitoring Centre with support from IUCN and its World Commission on Protected Areas.

It is a publicly available online platform where users can discover terrestrial and marine protected areas, access related statistics and download data from the World Database on Protected Areas.

Protected Planet enables a spectrum of users to use existing protected area data for information-based decision making, policy development, and business and conservation planning. Businesses in a range of sectors including mining, oil and gas, and finance can use WDPA data to identify biodiversity risks and opportunities of a given project. Conservation planners can use the information to predict the outcomes of various proposals and focus on initiatives and areas that are most likely to result in positive impacts.



The goals of Protected Planet are that:

- The Protected Planet Initiative informs decision-making and enhances action;
- The Protected Planet Initiative is the global platform for the communication, exchange, acquisition and analysis of knowledge and data on the status and trends of protected areas;
- The Protected Planet Initiative provides the world's policy-makers with the best possible information on protected areas and their value for conserving biodiversity and ecosystem services, and supporting human communities.

8.4 SUMMARY

Achieving independence through archiving can simplify IT decisions and systems. As we are well aware that data archiving/data backup protects older data that is not needed for everyday operations of an organization and that is no longer needed for everyday access. It reduces primary storage requirement, and allows an organization to maintain data that may be required for regulatory or other requirements. Data archiving benefits includes the reduction in the cost of primary storage as well as data security.

On the other-hand we found that the web portal may have a positive effect on attitudes towards sharing and searching for different required information. It is important to continue experimenting with web-based resources in order to increase user participation in different sector of decision-making. In this unit, the importance of different data archival as well web portals are given. You as a explorer just think that the given web portals are just to widening your thoughts that provide just the diversity of web which are available free of cost. There are numbers of web portals are available and list is endless. You are advised to search out, explore the web and come out with feasible ideas/innovation for the society.

8.5 GLOSSARY

IUCN

The International Union for Conservation of Nature is an international organization working in the field of nature conservation and sustainable use of natural resources. It is involved in data gathering and analysis, research, field projects, advocacy, and education.

WWF

The World Wide Fund for Nature is an international non-governmental organization founded in 1961, working in the field of wilderness preservation, and the reduction of human impact on the environment. It was formerly named the World Wildlife Fund, which remains its official name in Canada and the United States.

GBIF

The Global Biodiversity Information Facility is an international organization that focuses on making scientific data on biodiversity available via the Internet using web services.

Web Map Service

A Web Map Service is a standard protocol developed by the Open Geospatial Consortium in 1999 for serving georeferenced map images over the Internet. These images are typically produced by a map server from data provided by a GIS database.

Data Node

A node in hadoop simply means a computer that can be used for processing and storing. There are two types of nodes in hadoop Name node and Data node. It is called as a node as all these computers are interconnected.

Website

A website is the group of web pages which are placed in a location on the internet and accessed through a web address. Content on a website is globally visible, publically used, remains same for the different individuals. Users need not to login for accessing the website. The user can perform any specific task, and the website supports it.

A website could be industry-specific, product specific or services specific etc.; these websites are intended to educate their site visitors about their industry, products or services information. There is no use of a personalized database, and the website does not usually reference it.

Web Portal

A web portal is a typical knowledge management system that delivers the facility for organization or companies to build, share, interchange and reuse knowledge. It is the private location on the internet retrieved through a unique URL (web address), and probably login id and password. Web portal content is login protected and user specific and its interface could be public and private.

It allows access to multiple user roles. Contents in a web portal are dynamic and changed frequently. The visibility of one content changes from person to person which means content could be unique to a user based on group member settings. Contents are collected from the different and diverse sources.

Downloading and Uploading

Downloading is how you copy a file from a website to your computer, tablet, or device. An upload is when you send a file to a website. Downloads are mostly used for program installations, apps, documents, and anything else that you need a local copy of.

Metadata

The information that describes the data source and the time, place, and conditions under which the data were created. Metadata informs the user of who, when, what, where, why, and how data were generated. Metadata allows the data to be traced to a known origin and know quality.

NDSAP

The Union Cabinet approved the National Data Sharing and Accessibility Policy (NDSAP) on 9 February 2012. The objective of the policy is to facilitate access to Government of India owned shareable data and information in both human readable and machine readable forms.

It aims to provide an enabling provision and platform for proactive and open access to the data generated by various Government of India entities.

The objective of this policy is to facilitate access to Government of India owned shareable data (along with its usage information) in machine readable form through a wide area network all over the country in a periodically updatable manner, within the framework of various related policies, acts and rules of Government of India, thereby permitting a wider accessibility and usage by public.

8.6 ANSWER TO CHECK YOUR PROGRESS

1. Define web portals in details.
2. What are the key differences between a website and a portal?
3. Explain the different theme based web portals with their uses.

8.7 REFERENCES

1. <https://developers.arcgis.com>
2. <https://www.google.com/s>
3. <https://en.wikipedia.org/>
4. <https://www.data.gov/developers/archiving>
5. <https://opendatahandbook.org/guide/en/what-is-open-data/>
6. <https://opendefinition.org/od/2.1/en/>
7. <https://data.gov.in/sites/default/files/NDSAP.pdf>
8. https://data.gov.in/sites/default/files/NDSAP_Implementation_Guidelines_2.2.pdf

9. <http://5stardata.info/>
10. <https://www.isro.gov.in/earth-observation/bhuvan>
11. Jetz, W., McPherson, J. M., and Guralnick, R. P., 2012.
12. Flannigan M., Stocks B.J., Turetsky M. and Wotton M. (2008) Impacts of climate change on fire activity and fire management in the circumbo real forest, *Global Change Biology*,15 (3): 549 – 560.

8.8 TERMINAL QUESTIONS

1. What are the best practices for Catalog/ Resource (Datasets/Apps) contribution?
2. What is metadata?
3. Explain the role of data contributor.
4. What is crowd sourcing?
5. Write about three web portals with brief introduction and their social assistance to society as a whole.