SPATIAL MODELS IN GIS

1. **Spatial analysis is defined as the process of studying entities by examining, assessing, evaluating, and modelling spatial data features such as locations, attributes, and their relationships that reveal the geometric or geographic properties of data.**

2. Geospatial analysis is a process of GIS data interpretation, exploration and modelling, from acquisition to understanding results. The retrieved information is computer-processed with spatial analysis software and varies depending on the number of tasks and their complexity. The simplest one is visualization, while a more detailed approach suggests comprehensive analytics with specific tools to elaborate actionable insights.

3. Examples of spatial analysis include measuring distances and shapes, setting routes and tracking transportations, establishing correlations between objects, events, and places via referring their locations to geographical positions (both live and historical).

Typically, spatial analysis consists of five key stages: understanding your goal, preparing data, choosing suitable tools and techniques, performing the research, and estimating results.

4. So, first of all, it is important to figure out what you are interested to know. The next will be to choose corresponding spatial analysis methods for data manipulation and interpretation. Knowing exactly what you want to get, you will be able to tailor the research techniques to your needs. Once you determine, you start the next stage of spatial analysis – data processing and interpretation. Last, you estimate the results to define if you achieved your goal or not.

Type of spatial analysis

Types of spatial analysis varies from simple to sophisticated.

a. **Queries and reasoning** are the most basic, in which GIS is used to answer simple question posed by user. No change occur in the Data base and no new data is formed.

b. **Measurement** are simple numerical values that describe aspect of geographical data. They include measurement of simple properties of objects such as Distance or direction.

c. **<u>Transformation</u>** are simple methods of spatial analysis that change data sets by combining them to obtain new data set and eventually new insight. Transformation use simple geometric, arithmetic or logical rules and they include operation that convert raster data to vector data or vice versa.

d. Descriptive Summaries

e. Optimization

f. Hypothesis testing

<u>SPATIAL DATA GIS REPRESENTATION OPTIONS (DATA MODEL FOR GIS)</u>

Spatial data can be expressed in several formats, including vector and raster. **Apart from vector and raster, non-spatial** data is represented by attributes. It refers to add-on information to spatial data, for example, images, maps, or additional information that adds value to the spatial data of the object under consideration. In GIS, the **spatial data models** handle where the features are and **Nonspatial data models** or Data base management system handle the feature description and how each feature is related to other.

a. <u>Raster based spatial models</u>. RASTER is a spatial data model that uses Grid and cell to represent spatial variation of a features. The raster model represents the two- dimensional location of phenomena as a matrix of grid cells. Each cell stores a data item defining the entity, the class or the value of the represented phenomena. Each cell is also known as a 'pixel'. raster gives data a pixel grid representation, wherein each pixel reflects specific information such as color, measurement unit, etc. Conventionally, raster denotes imagery; however, in the context of spatial analysis, it refers to aerial photographs or satellite images, known as 'orthophotographs'.

Raster data performs a discretisation of the geometric area of interest and the entire space is broken into grid cells of a fixed or uniform size. In this type of representation of geographic data, a set of cells located by coordinates is used, each cell is independently addressed with the value of an attribute by specifying values to each grid cell. The simplest raster data structures consist of an array of grid cells. Each grid cell is referenced by a row and column number and it contains a number representing the type or value of the attribute being mapped.

GENERAL CHARACTERISTIC OF RASTER DATA

a. Simple 'grid' structure of rows and columns.

b. Based on cells or picture elements (pixels).

c. Linear feature (e.g. a road) is a contiguous set of cells.

d. Resolution based on size of grid (cell) the smaller the cell, the higher the resolution.

e. Features are considered homogenous within a pixel.

f. Storage increases with the square of resolution.

RASTER DATA: ADVANTAGES

1. Due to the nature of the data(simplest form of data) storage technique data analysis is usually easy to use by GIS force program and quick to perform.

2. The inherent nature of raster maps, e.g. one attribute maps, is ideally suited for mathematical modelling and quantitative analysis.

3. Discrete data, e.g. forestry stands, is accommodated equally well as continuous data, e.g. elevation data, and facilitates the integrating of the two data types.

4. Grid-cell systems are very compatible with raster-based output devices, e.g. electrostatic plotters, graphic terminals. Also compatible with digital satellite imagery.

5. The data form can be use to do various spatial analysis.

6. The model maintains uniformity when it comes to size and shape due to matrix and multi-array like structure.

7. Comparatively to its vector counterpart, the technology is far cheaper and affordable.

8. This makes the data livelier and presentable due to the involvement of colour codes, and hence when pairing with vector models, it gives proper relatable information.

RASTER DATA: DISADVANTAGES

1. The cell size determines the resolution at which the data is represented.

2. Processing of associated attribute data may be cumbersome if large amounts of data exists.

3. Raster maps normally reflect only one attribute or characteristic for an area.

4. Since most input data is in vector form, data must undergo vector-to-raster conversion.

5. Most output maps from grid-cell systems do not conform to high-quality cartographic needs.

6. projection transformation are time consuming

7. Data structure are not compact

8. More difficult to represent topology

DESCRIBE METHOD OF ENCODING RASTER DATA

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

a. Cell-by-cell raster encoding. This minimally intensive 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**."

b. 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 (e.g., discrete datasets such as land use/land cover or habitat suitability) and is less useful where neighboring pixel values vary widely (e.g., continuous datasets such as elevation or sea-surface temperatures).

3. **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."

b. **VECTOR DATA MODELS**

A representation of the world using points, lines, and polygons. Vector models are useful for storing data that has discrete boundaries, such as country borders, land parcels, and streets. Commonly referred to line drawings or illustrations where an object is represented as a point, line, or polygon. Vectors graphically represent the real world through points, lines, and polygons. With the help of points, one can create lines, which further give rise to enclosed spaces in the form of polygons. These vector data types allow the graphical mapping of physical objects found on the Earth's surface. The data is kept in shapefiles.

Vector data use X and Y coordinates to define the locations of points, lines, and areas (polygons) that correspond to map features such as fire hydrants, trails, and parcels. As such, vector data tend to define centers and edges of features.

In the second and more intensely developed approach to information integration, attribute information is associated with point, line and polygons – as spatial entities that describe features occurring in the real world

CHARACTERISTIC OF VECTOR DATA

1. Based on object (point, lines and area)
2. Constructed using arcs nodes and vertices
3. Resolution can be independent of details
4. Every point has a unit location

VECTOR DATA: ADVANTAGES

1. Data can be represented at its original resolution and form without generalization.

2. Graphic output is usually more aesthetically pleasing (traditional cartographic representation)

3. Since most data, e.g. hard copy maps, is in vector form no data conversion is required.

4. Accurate geographic location of data is maintained.

5. Because it recognizes entities, model allows for efficient encoding of topology, and as a result more efficient operations that require topological information, e.g. proximity, network analysis.

6. Require less disk storage

7. Topology can be completely described and easy to maintain

8. Better suited for Map output

9. Compact data structures for homogeneous areas

VECTOR DATA: DISADVANTAGES

1. The location of each vertex needs to be stored explicitly

2. For effective analysis, vector data must be converted into a topological structure. This is often processing intensive and usually requires extensive data cleaning.

3. Topology is static, and any updating or editing of the vector data requires re-building of the topology.

4. More complex data structures.

5. Simulation is difficult because each unit has a different topological form

6. Display and plotting can be expensive, particularly for high quality, colour and hatching

7. Spatial analysis and filtering within polygons are impossible

Algorithms for manipulative and analysis functions are complex and may be processing intensive

Often, this inherently limits the functionality for large data sets, e.g. a large number of features.

Continuous data, such as elevation data, is not effectively represented in vector form. Usually substantial data generalization or interpolation is required for these data layers

GIS DATA SOURCES

1. Not so long ago, all information used in a GIS had its origin in a paper map whose content was later transformed to adapt it to the particular nature of that GIS. Geographical data were obtained from the **digitalization** of printed cartography; that is, from the conversion of analogical maps into digital data that GIS can handle.

2. Apart from the fact that we can use them in a GIS, digital data have many advantages and represent an important qualitative improvement. Digital data are easier to update, easier to distribute (specially since the Internet was created), use less physical space and are easier to maintain (digital data do not degrade: their physical support does, but they are easy to replicate without losing quality).

3. Techniques for geographical data acquisition have advanced and it is possible now to create data that can be directly integrated into GIS. Data sources that produce data ready for use in GIS are called **primary** data sources. Those that generate data that has to be adapted or converted are called **secondary** data sources.

THE MAIN DATA SOURCES THAT PROVIDE DATA FOR GIS.

a. **REMOTE SENSING**. Remote sensing is the **acquisition of information about an object or phenomenon without making physical contact with it**. Instead of measuring the object itself, it measures the perturbations —**mainly the electromagnetic ones**—

that it causes on its surroundings. In our case, it is applied to objects on the Earth's surface.

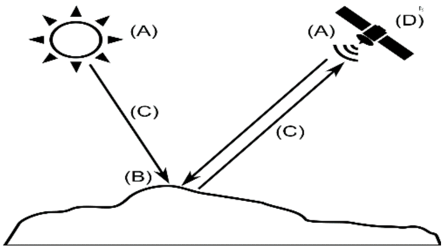

Figure 7.1: Elements in a remote sensing system.

(1) **<u>A source of radiation (A)</u>**. It can be natural or artificial. Radiation emitted by the source reaches the Earth's surface and it is altered by the presence of the objects on that surface. Remote sensing studies that alteration. Objects themselves can emit radiation as well.

(2) **<u>Objects (B) that interact with radiation</u>** or can emit it, as mentioned above.

(3) **<u>An atmosphere (C).</u>** through which radiation moves from the source to the objects. The atmosphere also interacts with the radiation and alters it.

(4) **<u>A receiver (D) which receives the radiation.</u>** once it has been emitted or altered by the objects. The receptor measures the intensity of the radiation coming from different points in the area being studied and, with them, generates its final product (in most cases, and image).

b. **<u>PHOTOGRAMMETRY</u>**. Photogrammetry is the technique used to study and precisely define the shape, size and position in space of any object, using measurements from photographs. Of special interest to GIS is the branch of photogrammetry known as **aerial photogrammetry**, which uses aerial photographs and it is mainly used for generating elevation data through a process known as **restitution**. Phogrammetry can be analogical or digital, the later being the one more related to the field of GIS.

Instead of single images, the branch of photogrammetry known as **stereophotogrammetry** uses pairs of images, each of them taken from a different point. These images form a **stereo pair** and with them a three-dimensional reconstruction of the original scene can be produced. This can be used by an operator to see the scene with **depth and volume** so that terrain forms can be identified and elevation information obtained.

If using satellite images, stereo pairs can be obtained from those platforms and sensors that allow **changing the angle of vision**, so in the same satellite pass, pictures of a given area can be taken from different points.

Photogrammetry can be **analogical** or **digital**, the latter being the one more related to the field of GIS.

Stereoplotters are used to combine and align the images that form the stereo pair. Current stereoplotters are called *analytical stereoplotters*, and contain elements from GIS, along with more specific elements. Among these, we find specific visualization software and peripherals such as 3D mouses or other mechanical elements found in analogical photogrammetric devices, making it easy for operators to adapt to this new type of tools.

c. **<u>PRINTED CARTOGRAPHY/DIGITIZATION</u>**. A large amount of cartography exists in printed form, such as maps or old analogical aerial photographs. To be used in a GIS, this cartography has to be **digitized**, which means creating raster or vector layers from them. In this latter case it also implies **separating the different types of information** that the map might contain, since the information in a single printed map would be stored in independent layers, in GIS

Digitizing a printed cartographic document involves three steps:

(1) **<u>Georeferencing</u>.** the original document. That is, setting a geographical context (coordinate system, control points, etc.), so the digitized elements produced are correctly referenced.

(2) **<u>Digitizing the spatial component</u>**. That is, creating the corresponding geometries.

(3) **<u>Digitizing the thematic component</u>**. Creating cell values for raster layers or attributes in the case of vector layers.

Digitization can be **manual** or **automatic**. If manual, an operator introduces the value, while an automatic process is done through an algorithm.

To create raster layers, the most common method is **scanning** the original document using a **scanner** which creates a digital image from an analogical one.

d. **<u>GPS</u>**. One of the most relevant advancements in geographical data sources have been **global navigation satellite systems** (GNSS). **For any given point and at any time**, these systems allow us **to know the exact location of that point** with an accuracy of a few meters or less. To do that, they use a constellation of satellites to which information is transmitted from the study point, and use that transmission to compute the coordinates of the point.

The first and most popular of these systems is the **Global Positioning System** (GPS). It has 24 active satellites (the satellite segment) along with terrestrial stations to control them (the control segment) and it is based on **trilateration**. Distances are measured from a GPS unit (the user segment) to a certain number of satellites. Knowing those distances and the exact position of the satellites, the position of the unit can be computed. Position is computed with its *x*, *y* and *z* coordinates. The GPS system uses WGS-84 as its reference ellipsoid.

The satellite network is designed to guarantee that, from any point of the Earth's surface, and at any time, a GPS unit can locate the required number of satellites to compute its position.

Precision of the GPS system depends on the GPS unit. There are many classes of GPS receivers, but the main ones, **from the point of view of GIS, are two:**

(1) <u>**GPS for general use**</u>. This units are low-cost **small and portable** for outdoors activities, where a high accuracy is not required. The GPS receivers found in smartphones fall into this category.

(2) <u>**GPS for surveying**</u>. Larger units, usually with independent antennas that are connected to the receiver to increase accuracy. It also ensures better location of satellites in difficult conditions, such as under forest canopy. They are designed for professional use.

Regarding its connection with GIS, GPS units can collect coordinates and then create GIS layers with them. They can store points (called **waypoints** in the usual GPS terminology), or capture the path followed by the user (called a **track**). In this last case, the receiver stores points at regular intervals so the user can just move and does not have to manually store the coordinates along the route. This information can be later introduced in a GIS for further analysis or visualization.

e. <u>**VOLUNTARY GEOGRAPHICAL INFORMATION**</u>. The participative ideas of the so-called **Web 2.0**, when combined with tools such as recreational GPS units, or with simplified software for editing and digitizing, result in interesting initiatives in which people, with no specific training in cartography or surveying, can acquire and share geographical information.

Although this cannot be considered a different data source (the techniques and devices used have already been described in previous sections), there is an important change in the philosophy behind data collection and usage, which makes it worthwhile to treat this type of data separately in the context of this book.

The term **Volunteered Geographical Information** (VGI) refers to the use of the Internet to create, manage, and share geographical information which has been voluntarily contributed by a community of users, also using the Internet. The set of techniques and tools used by those users is termed **neogeography**.

Neogeography has changed some fundamental ideas in cartography, since it has modified the traditional concept of geographical information (which was created by a very skilled few), its characteristics, or the role it played in certain fields. The following are some ideas about these changes and neogeography itself.

- **Popularization and democratization**. Cartographic production has always been in the hands of governments or agencies and, in many cases, has been strongly censored due to its high strategical value. With the advent of VGI, geographical information becomes more democratic, and its creation is a free, participative and unrestricted process. The top-down approach that had dominated the production and use of cartography is thus inverted.
- Citizens become **sensors** and are more conscious of their geospatial reality.
- Cartographic production loses its mysticism.

The most relevant VGI project at this time is **OpenStreetMap** (OSM), a `` collaborative project to create a free editable map of the world.".

f. **Metadata.** Regardless of their origin, data might need additional data to be interpreted. For instance, if we have the coordinates of a point, to correctly interpret it we need, among other things, the coordinate system in which those coordinates are expressed. The data we work with (the coordinates) should be accompanied by some ancillary data (such as the EPSG code of the coordinate system).

This ancillary data are known as **metadata**. Metadata are **data about the data**, and their purpose is to **explain the meaning of the data**. That is, they help the user to better understand the meaning of the data and the information that they contain. Metadata are an additional document that accompanies the data and that allows for better management and use.

Within GIS, metadata are usually **associated to a layer and its content** and can be **referred to both components** (spatial and thematic).

The concept of metadata is not exclusive of digital geographical data. A printed map has metadata in a certain way. A legend or a text in its margin with information about the date in which it was created, are also metadata. In the case of digital metadata, the metadata are **independent from the data itself**. That allows us to perform operations **separately on the metadata**, which opens many possibilities and gives them a greater value.

Two of the main functions of metadata are **ensuring the correct use of data** and **facilitating its management, discovery and exploitation**.

<u>LEGACY GIS</u>

1. Legacy data refers to information that is stored in outdated or obsolete systems, formats, or technologies. This includes **structured** and **unstructured data**. It includes data **generated by humans** as well as **machine generated**. While legacy data may not be áctively used, at least, if it is used, it's not used regularly, it is often crucial- for legal, regulatory, or historical reasons. Managing legacy data can be challenging, but it is critical for ensuring compliance, improving data management, and making better business decisions

Examples of legacy data include:

customer records, financial data, emails, documents, databases, spreadsheets and presentations.

The challenge with legacy data is that it can be difficult to access and use, especially if the systems and technologies used to store the data are no longer supported or compatible with current technology. Sometimes more up to date technology simply can't access, or even open,the legacy data file formats. Effective management of legacy data is critical for ensuring compliance, improving data management, and making better business decisions. Legacy data can be stored in a variety of formats, such as paper records, magnetic tape,floppy disks, and early versions of

file formats (e.g. early versions of Microsoft Excel). It can also be stored in outdated software or hardware systems that are no longer supported or used (e.g. software like Lotus 1-2-3).

In some cases, the data may be stored in proprietary formats that require specific software or hardware to access. Managing legacy data requires a careful approach to ensure that the data remains secure, accessible, and compliant with legal and regulatory requirements. It often involves migrating, archiving, and preserving the data in a way that ensures it remains accessible and usable over time. Effective management of legacy data can help organizations unlock valuable insights and information that can support better decision-making and improve overall business operations.

Importance of Managing Legacy Data

Managing legacy data can be important for several reasons:

a. **Compliance**: Legacy data may be subject to legal or regulatory requirements that mandate how it should be stored, managed, protected and expired. Failure to comply with these requirements can result in legal and financial penalties, reputational damage, and loss of business.

b. **Improved data management**: Legacy data can be a valuable source of information that can be used to inform business decisions, identify trends, and support strategic planning. Effective management of legacy data can ensure that it remains accessible and usable, even as new technologies emerge.

Better decision-making: By analyzing legacy data, organizations can gain insights into past business operations and use this information to improve future decision-making. Legacy data can also help organizations identify patterns and details that can help product development, define marketing strategies, and other key business decisions.

Enhanced customer service: Access to historical data is critical for providing high-quality customer service. For example, having access to a customer's purchase history can help a business tailor its products and services to better meet their needs.

Cost savings: By managing legacy data effectively, organizations can avoid the costs associated with maintaining outdated systems and technologies. This can include costs associated with hardware and software maintenance, data migration, and compliance-related expenses.

Types of Legacy Data

1. There are several types of legacy data that organizations may need to manage. Including:

a. **Structured data**: This is data that is stored in a structured format, such as a database or spreadsheet. Structured data may include customer records, financial data, inventory data, employee records, sales data, medical records, website analytics and more.

b. **Unstructured data**: This is data that is not organized in a structured format, such as emails, documents, images, and videos. Unstructured data may be stored in a variety of formats and will probably require specialized tools to manage and analyze.

c. **Machine-generated data**: This is data that is generated by machines, such as log files, sensor data, and machine-generated reports. Machine-generated data may be produced in large volumes and may require specialized tools to process and analyze.

d. **Human-generated data**: This is data that is created by humans. It includes emails, chat logs, documents, spreadsheets, presentations, and social media posts. Human-generated data may be stored in a variety of formats and may require specific software to access, review and update.

Challenges of Managing Legacy Data

1. Managing legacy data can be challenging for many reasons, how they affect each individual organization may vary and the importance of these challenges often also varies between

organizations and geographies. Here are some of the key challenges:

a. **<u>Compatibility</u>**: Legacy data may be stored in outdated formats (such as spreadsheets stored in Lotus 1-2-3 file formats) or systems that are no longer compatible with current technologies (such as data stored on CDROM). This makes it difficult to access, read, or transfer the data.

b. **<u>Accessibility</u>**: Even if the data is accessible, it may be difficult to find or retrieve specific pieces of information within the data. This can make it time-consuming and resource-intensive to use the data effectively.

c. **<u>Data quality</u>**: Legacy data may contain errors, inconsistencies, or inaccuracies that can make it less useful for analysis or decision-making. Ensuring data quality may require significant effort and resources.

d. **<u>Security</u>**: Legacy data may contain sensitive or confidential information that needs to be protected from unauthorized access or breaches. As security threats evolve over time, ensuring the security of legacy data can be a significant challenge especially if some of the legacy systems or software are no longer supported by the original manufacturer.

e. **<u>Cost</u>**: Managing legacy data can be expensive, particularly if it involves migrating data to new systems or technologies. The cost of maintaining outdated systems or finding specialized tools, or employees to manage the data can also be significant.

f. **<u>Complexity</u>**: Managing legacy data can be complex, particularly if the data is stored in a variety of formats or locations.

g. **<u>Legal and regulatory compliance</u>**: Legacy data may be subject to legal and regulatory requirements that mandate how it should be stored, managed, protected and expired. Ensuring compliance with these requirements can be challenging.

8. In order to overcome these challenges, organizations may need to invest in specialized tools and expertise to effectively manage legacy data. Both of these are likely to incur significant cost.

However, by doing so, you can ensure that the data remains accessible, accurate, and secure over time, and can unlock valuable insights and information that can improve business operations and drive success. It can also lead to a reasoned approach to migrate the appropriate legacy data to a more modern, more accessible system.

COMPLEXITY OF LEGACY SYSTEMS

a. **Technical debt**: Over time, legacy systems can accumulate technical debt, which refers to the cost of maintaining outdated software, hardware, and other technologies. This can make it difficult to upgrade or migrate legacy systems to newer technologies.

b. **Integration challenges**: Legacy systems may not be designed to work with newer technologies or systems. This can make it challenging to integrate legacy systems with newer tools and technologies, which can result in compatibility issues and errors. Customization: Legacy systems may have been heavily customized to meet the specific needs of an organization, but those needs will likely have changed over time. This can make it difficult to transfer the system to a new environment, as the organization may need to replicate customizations in the new environment or go through a lengthy (and costly) mapping process between legacy and more-modern system.

c. **Documentation**: Legacy systems may have outdated or incomplete documentation, which can make it challenging to understand how the system works and how it should be managed. If the legacy system is no longer supported by the original manufacturer then clarifying the issues with the documentation will be difficult.

d. **Knowledge transfer**: Employees who were familiar with legacy systems may retire or move on to other roles in your organization or move to a different organization. This makes it challenging to transfer knowledge about how the legacy system works and how it should be managed by new employees.

e. **Security risks**: Legacy systems may be more vulnerable to security risks than newer systems, as they may not have been designed with modern security threats in mind and the legacy system may no longer be receiving security updates. This can make it challenging to ensure the security of the system and the data it contains. In some organizations this may lead to increasingly restrictive access to potentially valuable legacy data.

Using Legacy Data in GIS Modeling there exist large datasets from previous excavations, published and unpublished, whose digitisation, spatial mapping and re-analysis can greatly facilitate investigations of social behaviour and changing environmental conditions. Added to this there are also archives of historical maps and excavation plans, whose digitisation and combination with these other archaeological data for digital analyses give us greater understandings of the past. This volume presents a number of projects that demonstrate the usefulness of digital environments for analysing such non-digital data. These projects use these 'legacy data' within true GIS,pseudoGIS, or other digital environments to answer specific questions concerning social behaviour and particularly the social use of space.

This process generally involved five steps:

a. **Legacy reports**: In many cases; onlythe legacy survey & SI reports are available. Although basic, georeferencing the report enclosures and harvesting any important tabular data can at least give an indication of what / where to lookfor data, and who to ask to obtain the actual datasets.

b. **Public datasets**: Publicly available datasets & reports can be obtained from providers such as the Marine Data Exchange and the National Data Repository (in the UK). These can be used to fill gaps in existing datasets or initial screening purposes.

c. **Visualization**: Where geospatial data is available, it's often in multiple different formats easily bring these together to visualize centrally with figure production and 3D modelling tools.

e. **Open-source**: Quality, free and open-source GIS software such as QGIS makes data access very easy for other disciplines enabling expert insight and visualization. Where legacy data formats are encountered, numerous tools are available for converting into those that are readily useable.

f. **Data-sharing**: Data sharing - GIS can provide the foundation for presenting data via web providing maximum stakeholder engagement on multiple platforms.