Primary Data :-

- Primary Data are those that are developed by the individual or organization that intend to use them or, in some cases, are contracted out with specific guideline how they will be prepared.
- Primary Data are generally going to experience a higher level of quality control than secondary data. They may required both intensive collection (Field Collection) and conversion (Digitizing) than secondary data.
- As a result they are often better quality for the specific application and more costly to produce.
- There are many forms of primary data ranging:
  - Field Observation
  - Biological Specimen Sampling
  - Human Interviews and Surveys
  - Arial Photograph and Satellite Remote Sensing
  - Field Mapping
  - GPS Surveys
  - Police Calls and Accident Reports etc...
- It would require volume to summarize all the guidelines and criteria for all possible primary data collection methods.

Input Device :-

1. Digitizer :-

- For inputting the spatial data manually, the use of the digitizer is standard.
- The digitizer is somewhat more advance and much more accurate version of an input device used in all modern PCs.
- Inside the Mouse of Digitizer, the Sensors that are responds to the motion of rubber ball encase in the frame of device.
- All the locational information is contained within the mouse itself.
- The electronically active grid inside the digitizing table will record the degree of movement for the digitizer.
- The device like mouse called puck is connected to the table and move along the table and it moved to the different location on the map that is attached to the table.
- He digitizer puck contain the crosshair device encased in glass or clear plastic, that allows the operator to place the puck exactly over individual map element.
The puck has a buttons that indicate the starting and ending of lines or polygons or explicitly define left and right polygon and so on.

Modern digitizer can provide resolution of 0.001 inch, for an area of 42 inches by 60 inches.

Some of the useful factors are there for selecting a digitizer.

- **Stability :-**
  Deal with the tendency of the exact reading of the digitizer to change as the machine warms up. The solution is to allow the digitizer to come up to operating temperature before using it.

- **Repeatability :-**
  How close will be the first and second readout be? 
  Good digitizer should be repeatable to about 0.001 inch

- **linearity :-**
  Digitizer is able to measure a specific distance of a correct value as the puck moved over a large distance.

- **Resolution :-**
  Digitizer should be able to record a increment of the space.

- **Skew :-**
  Measure of the squareness of the results on a tablet
  Do coordinates located at the four corners of your digitizer produce a true rectangle, as intended?

Input of Raster Data, Vector Data or Both :-

- No matter which approach is chosen for the GIS Input, but it is necessary to determine at the outset whether you will be using a Raster or Vector GIS.
- Program, especially those that evolved primarily to handle remotely sensed data, operate on the grid data structure where as others operate primarily on vector data structure.
- Line following scanners tends to produce Vector Output.
- Drum scanners will produce raster Output.
- Although conversation between Vector to Raster and Raster to Vector are fairly common, there are something should be remember.
**Conversation :-**

- **Vector To Raster :**
  - Results are visually Satisfactory
  - But the attribute result is not satisfactory for each grid cell. This is particularly true along the edge of area.

- **Raster to Vector :**
  - Result preserves the vast majority of attribute data.
  - Visual result will often reflect the blocky, step like format on the grid cell from which the conversion proceeded.
  - Algorithms are available for smoothing this blocky appearance with the use of mathematically based graphic technique called Splines.
  - That smooths out the jagged lines and sharp edges.

**Reference Framework & Transformation :-**

- Digitizer input the map from aerial photography or other analog product of remote sensing equipment.
- Maps are representation of a three-dimensional reference globe projected on a flat surface.
- There are three primary processes:
  1. Translation.
  2. Scale Change.
  3. Rotation.

**Translation :-**

- Translation is simply movement of part or all of a graphic object to a different location on the Cartesian surface.
- This is done by adding and subtracting the coordinate values necessary for the X and Y coordinates for the object.
- In other words, the new X coordinate $X'$ for each graphic object will be equal to the original X coordinate plus some value $T_x$, and the new Y coordinate $Y'$ for each graphic object will be equal to the original Y coordinate plus some value $T_y$.

$$X' = X + T_x \quad \text{and} \quad Y' = Y + T_y$$

Where $T_x$ and $T_y$ are the Translation amount.
- **Scale change** :-

  - Scale Change is also relatively useful because of the need to compare differently scaled maps and to output in different scale as well.
  - This is done by multiplying the overall X coordinate extent by a scale factor $S_x$, and each set of Y coordinates by Y scale factor $S_y$.

    \[ X' = X \times S_x \quad \text{and} \quad Y' = Y \times S_y \]

    Where $S_x$ and $S_y$ represent the amount or percentage of scale change.
• **Rotation :-**
  
  o **Rotation** is used frequently during the process of projection and inverse projection.
  o It is also accomplish by using basic trigonometry.
  o For **new X** coordinate locations, the new X coordinate X’ is found by multiplying it by the cosine of the new angle (θ) and then adding that value to the original Y coordinates multiply by the sin of theta (sinθ).
  o The **new Y** coordinate location Y’ are found by multiplying the negative of the original X value by the sin of the angle and again adding that to the product of Y coordinate and sin θ.

  \[ X' = X \times (\cos \theta) + Y \times (\sin \theta) \text{ and } Y' = -X \times (\sin \theta) + Y \times (\sin \theta) \]

  *Where* \( \theta \) is the angular displacement.

---

**Map Preparation and Digitizing Process**

- We cannot begin the digitization process until we have provided the GIS software with information regarding the projection, the grid and so on.
- So it is important to developing the useful database.
- In this process you should prepare the information beforehand and keep it handy so that you will always know what it is and where to find it.
- It is also a good idea, before you place your map on the digitizing table to being the input process, to prepare a map by making appropriate marks directly on the map document or on firmly attached clear plastic covering to identify exact location you want to digitize.
- If you know all the points, it is better to digitize free hand.
➤ Because digitization is a tedious work, you probably will digitize the map in portions rather than all at once.

➤ **Registration Points / Tic Marks :-**

☑ Because you will probably digitizing the map in multiple sessions, and sometimes you will have to remove the map to allow other to use the equipment, you will need to tell the computer software *where your map area is and what its coordinates*.

☑ These first points, called Registration Points or Tic Marks, they will be entered in digitizer inches as well as in map coordinates.

☑ They also should be marked on the map as part of your map presentation, so that you as well as your computer know your starting points.

☑ *The registration points provide an outside frame for the document.*

☑ Usually three points need to locate at the corners of a rectangle to define the map area for the software. Some software can get by only two points if they are located on diagonal.

☑ In this case the software assume that the boundary is a rectangle and infer the other two points.

☑ The accurate location of the registration point is absolutely essential to ensure good quality.

☑ *Special care should be taken to locate your Tic Marks precisely.*

☑ It is good idea to *double-check* these because if registration marks are misplaced, virtually all the remaining digitization marks will be erroneous.

☑ And if you select that tic marks the software will provide you with an error measurement using a method called **Root Mean Square (RMS) error**.

☑ *The lower the RMS error the more accurate the digitization.*

☑ Other map preparations include a clear definition of the order in which you intend to digitize; a systematic approach to identify the portion of the map to digitize at each session has been mentioned.

☑ It is also good idea to develop a method to identify which area have already been digitize.

☑ Most digitizing software provides editing capabilities to help you to identify your mistake.
✓ Most digitizing software includes a feature that allows for the occasional shaky (अस्वस्थ) hand, namely, a fudge factor sometimes called **fuzzy tolerance**.

✓ The inclusion of this feature reflects the assumption that you will not be able to place the crosshair of digitize puck over exactly the same location twice.

✓ Small fuzzy tolerances are less forgiving of digitizing errors and many results in gaps between points that were meant to be connected.

➢ Final map preparation deal primarily with the tendency of the source material to shrink and swell with change in temperature and humidity (हवामान विज्ञान नाम). The material you are about to digitize should be allow to stay in the room for several hours, unrolled.

➢ It is good idea to avoid using unfold maps because creases reduce the accuracy of such a document.

❖ **What to Input :-**

➢ We have some basic guidelines on how to digitize, especially how to avoid errors during digitize, we can begin to select the appropriate data to input.

➢ A **major factor** guiding what cartographer put on the map and how it is produce is the target audience, called **Users**.

➢ The same can be said for producing the cartographic and geographic database for map analysis within the GIS.

➢ Different rules are there for deciding that what the inputs of any particular GIS application are.

➢ **Rule : 1 (Purpose)**

  o *To decide why you are building the GIS database in the first place.*
  o This will at least limit the input to coverage that is likely to be used.

➢ **Rule : 2 (Goal)**

  o *Results to the first, is that you must define your goals as specially as possible before selecting the layer.*
  o Because GIS begun uncertain guidelines to produce reasonable result without substantial reworking and correctness.
Rule : 3 (Avoidance)
- Avoid the use of exotic sources of data when conventional sources available.
- With known spatial information products, there will be multiple ways to obtaining the available data in some cases.

Rule : 4 (Selecting the Best)
- If all your multiple data sources for the particular them are in traditional form then Use the best, most accurate data necessary for your task.
- If you and another team member are familiar with the given set of data and can comfortably use it correctly and if it increases the utility of accuracy of database, it should be applied.

Rule : 5 (Diminishing)
- Remember the law of diminishing return when deciding on data accuracy levels.
- Difficult to separate categories over an area that is essentially all grain fields.
- Computing the human resources needed to provide clarification would increase the overall cost of system.

Rule : 6 (On Same Map)
- Whenever possible, and when the quality of data dictate it, you should input these data as separate coverage of the same map sheet.
- Advantages:
  - Data are on the single map, you did not need to go to multiple map sheets and then repeat all the preliminary steps in map preparation.
  - Because the data are on the same map sheet they are already geo referenced, reducing to perform this difficult task later.

Rule : 7 (Specification)
- Each them should be as specific as possible.
- The more specific a them, the easier it is to search if you need to know something.
- When you perform an operation like overlay, it is easier to keep a track of the processes if you are completely familiar with the data.
How Much to Input :-

- “How much?” is the question that is related to the data types you will input.
- For Example: As you prepare for your journey you will need to know how much food to pack, not just what kind.
- As with preparing packs for a trip, the input of data into a computer is a **sampling process**.
- In vector GIS, each line you input will likely have some curves. To produce reasonable facsimile of the line using straight-line segments, you will decide thousands of themes where to place the digitizer puck and where to record the data.
- The location of straight line can be recorded accurately with only two points: one at beginning and one at ending.
- Line and polygon complexity can be compared to information.
- The more the line changes direction the more information it convey.
- The higher the information content, the higher the sampling rate needs to be known as **Information Theory**.
- The idea of information content can also be applied to raster data. Once again the general rule is this: the smellier the object to be identified in your database, the smaller the grid cells need to be.
- Resolution is the principle often determines the selection of grid cell size for the entire database.
- The Information Theory can also be applied to the input of raster data, but keep in mind that grid cell are two-dimensional rather then one-dimensional.
- Whether in raster or vector, sampling is depending on he amount of area covered by the map and the use for which the data are input.
- **Small scale maps, those covering large amount of space, contain the much more abstract view of the land surface.**
- The amount of errors contain in the symbol is depending on the scale of the map on which it placed.
- **Lines on the small-scale maps take up more land surface than small-size line on the large-scale maps. This physical condition called scale dependent error** is an indication that the amount of error is directly related to the scale of the map and need to be considered during the map preparation phase prior to digitizing.
Methods of Vector Input :-

- After preparing the map and putting it on the digitizing table, you will need to use the digitizer puck to locate and record the registration marks.
- Software documentation and the software itself will indicate the requirements.
- The software packages will indicate which numbered keys you need to enter for specific object types.
- Some numbered keys will be used to indicate the location of the point entity, others for beginning and ending of line segment, and till other for the closure of polygon.
- Many digitizing errors, especially those made by invoices, are due to pushing the wrong numbered button.

Methods of Raster Input :-

- In the first place, we must decide how much area should be occupied by each grid cell. This decision must be place prior to digitizing.
- We must decide whether it is appropriate to use a method of encoding that shortens the process, such as the run-length encoding.
- Although compacting methods are good at reducing data, their use in encoding may be even more important because of the reduction of the input time. Once you have selected a method, you will need to decide how each grid cell will represent the different themes that will occur.
- Beyond the grid cell resolution, this may be the most important decision you will have to make.
- The entities and the attributes are entered simultaneously. This approach required vector-to-raster conversation.
- Most often difficulties extended from digitizing adjacent areas using vector lines which are then converted to two separate polygons. In such a case software must decide which polygon will contain the grid cell through which the line runs.
- The decision is sometimes based on the “Last Come, Last Coded” rule. This is when the same line digitize first for one polygon, and then again for the second polygon.
- Different four methods are there:
  1. Presence / Absence Method
  2. Centroid-of-cell Method
  3. Dominant Type Method
  4. Percent Occurrence Method
1. **Presence / Absence Method :-**

✓ A decision is made on the basis of whether the selected entity exists within the given grid cell or not—hence the name is Presence/Absence.

✓ **Advantages :**
  
  o This method is that decisions are easy.
  
  o No measurements are necessary.
  
  o Simple Boolean operator either “is” or “isn’t”.
  
  o This method is useful for coding points and lines for grid systems because these entities do not take up a large portion of a cell’s area.
  
  o Thus if a road crosses through a grid cell, its presence is recorded with an attribute code (a number); if does not, it is ignored.

2. **Centroid-of-cell Method :-**

✓ *Here the presence of an entity is recorded only if a portion of it directly at the center point of each grid cell.*

✓ Clearly, this requires substantial calculations, since each central point will needed to be calculated for each grid cell, and then the object will have to be compared with the location of that point.

✓ You must evaluate each grid cell Centroid against each entity.

✓ Therefore the use of this method will be restricted to polygon entity.

3. **Dominant Type Method :-**

✓ This is the more common and best method of coding polygon data.

✓ *The presence of an entity if it occupies more than 50% of the grid cell.*

✓ Under the most circumstances, the decision is straightforward and coding is reasonably representative of what is there.
It seems logical that if you are restricted to a single category for each grid cell, the one that occupies the most space should be coded.

Computationally, this requires the computer to determine the maximum amount of each polygon for each grid cell.

In two cases this approach is satisfactory.

- **First**, there may be highly irregular polygon shape. It is difficult to decide such cases by visual inspection.
- **Second**, three or more polygons types coverage in an irregular pattern within a single grid cell.

4. **Percent Occurrence Method:**

- This method is also used exclusively for polygon data.
- This idea is to give more detail, not by coding just the existence of each attribute but rather by separating each attribute out as a separate coverage and then recording the percentage of the area of each grid cell that occupies.
- For Example, a map of land use divided into urban and rural categories would be separated into two more specific themes one urban and one rural.
- The percentage urban and rural would be recorded for each grid cell.
- This method offers the advantage of more detail data about each attribute.

Remote Sensing Special Case of Raster Data Input: -

- Remote Sensed data being either primary or secondary data sources.
Sources for Acquiring Remotely Sensed Data:-

- The users may request specific **satellite** or **aircraft over flights**, for predefined areas, at predefined time, with limited amount of cloud cover.
- Or user might have their own **airborne or ground based sensing system** for acquiring their data. Under such circumstances, the remotely sensed data would definitely be considered primary.
- Some users currently employing some rather unique primary remote sensing data approaches for small study areas, including the use of **tethered balloons with digital cameras**.
- These data are not dominant compare to the any other source of input, such as traditional cartographic products, digital elevation data, digital line graphs from topologic map, digital land used data and digital soil data.
- The raster appearance of remotely sensed data may give the impression that GIS is software designed to manipulate these raster data.

Arial Photography :-

- **Aerial photography has long been a primary source of base map data for many common products**.
- A special types of areal photograph deserves mention because the images do not contain the scale, relief, and tilt distortion normally characteristics of aerial photographs. These products called **orthophotographs** are photographic image of the earth that resemble maps in that they have a single scale.
- The orthophotographs are subjected to process called **Differential Rectification**, which involves point-by-point correction of the scale and relief displacement normally caused by difference in elevation between aircraft and the topography over which it files.

Technical Difficulties :-

- Satellite Data **require processing to remove geometric and radiometric flaws** resulting from the interaction of two moving bodies. Techniques for correcting radiometric difficulties are readily available with most digital image processing software, and necessary equations are quite easily obtained.
For GIS input the major problem related to processing is a need to obtain geometrically correct ground position for the imagery. The geometric correction requires a number of **Ground Control Points (GCPs)** within the image, to place it in a correct spherical coordinate space on the surface of the earth.

Obtaining adequate GCPs may be quite difficult especially in area like forest GPS field unit has no direct line of sight to satellite.

- Care should be taken when using imagery lacking in GCPs because their absence degrades the coordinate accuracy.
- The issue of GCP accuracy should indicate the importance of coordinate accuracy to the overall functioning of GIS.

**Institutional Issues :-**

1. **General lack of availability of remotely sensed data.**
   - Requires the user to be familiar with the process of obtaining the data in the first place
   - Once the procedures are understood the problems could cover.

2. **Difference in the atmospheric condition.**
   - Change in the ground resulting from the flowering and dying off of vegetation.
   - In other cases two or more contiguous satellite images may have to be pieced together to cover a large study area completely.

3. **The Hardware and Software Cost.**
   - The ready availability of lower priced image processing software running on standard computing hardware has improved this situation considerably.

4. **Data and in most cases the satellite systems themselves were originally designed as experimental rather then operational system.**

5. **Education and Training.**
   - Today for both, GIS and Remote Sensing, the technical institutions and image processing and GIS vendor companies offer programs that provide extensive, hand-on experience with the use of particular remote sensing or GIS system.
6. **The final institutional problem of remotely sensed data input features organizational infrastructure.**

**GPS Data Input :-**

- An increasingly important sources of primary data input to the GIS database is the GPS data.
- It is often linked explicitly to the input of other primary data source such as field data.
- Software packages such as **PENMAP** allowed for the simultaneous input of GPS locational information.
- With the advent of the new handheld computer technology in the form of PDAs with full color displays and software packages like **ArcPad** we are now able to display digital map layers in full color in handheld unit, to link those same layers with GPS coordinates, to include a complete set of field notes.
- The computer and GPS technologies have already been integrated to allow out-of-the-box single-unit-solution.
- Although such mobile solutions are available, they are not common place because of the high price tag.
- The obvious **advantages** of the mobile GPS/GIS solutions include:
  1. Reduction in time and cost in development of geographic database.
  2. Rapid GIS and remote sensing data verification.
  3. Easy transfer of these data to full GIS database composes of both primary and secondary data.
  4. Incorporation of ancillary data and descriptions.

**Secondary Data :-**

- The efficient method of building a GIS database is to limit the amount of time and costs necessary to develop the database.
- Fortunately, the digital databases are becoming increasingly available. (Digital Elevation Model, Digital Line Graph)
- The U.S. department of agriculture makes soil maps available in digital form.
- But the availability of database also introduces the other **problems**.
  1. **The physical format of media.**
     - Countless hours can be spent trying to obtain digital data in the proper format. So many formats of CD-ROM technology are available.
You must be able to get the data in a format compatible with your retrieval equipment.
Even if the media are same, data can be provided for numerous formats.

2. Quality of the Data.

- Though data vendor may provide easier access for data, service may not provide data in original format.
- Some data will be filled easy viewable errors, some systematic and correctable, some not.
- You need to be aware of quality control procedures used by each vendor.

3. Use of an External Database.

- External databases require information about their own content such as: Data Dictionary. (Passive and Active)
- Passive Data Dictionary might include scale, resolution, and the name of data field s in the database, the codes used, and what they mean.
- Active Data Dictionary operates on the GIS database by performing checks for correctly coded inquiry (Check Constraint).
  For example, if your vector GIS database management system is set up to allow only a four digits code for a particular entity, the active data dictionary could check each inquiry to determine whether these four-digit limitation is uniformly met. Such checks are quit useful for preventing erroneous input.

Metadata and Metadata Standards :-

- Although data dictionary are useful, they are not enough for the growing set of GIS users, particularly where the data are shared among many users.
- In such a case organizations are now adopting the needs for rigorous set of standards that provide a detailed record of the dataset that they use and share with others called Metadata.
- Metadata describe in some detail the content, quality, condition, and other important characteristics of the data thus helping others to locate and understand.
The use of Metadata:

1. Organize and maintain an organization's investment in the data
   In this case, the integrity of the data is maintained because they contain and appropriate use of the data is cataloged.

2. Provide information to Data Catalog and Clearinghouse:
   Among the most difficult tasks of GIS professionals is to identify what data are available for their needs. The Federal Geographic Data Committee (FGDC) is developing a network of geographic data sharing called National Geospatial Data Clearinghouse (NGDC) through which the metadata can be shared.

3. Provide information to aid in data transfer:
   Any time the actual datasets are to be shared with others it is vital the metadata be shared as well. This allows the user to process and interpret data correctly and to match the data with its own data holdings.

Following are the necessary things used to complete the Metadata:

1. Identification
   - What is the database called?
   - Who developed it?
   - What geographic area are covered?
   - What category or them of data are included?
   - What if any restrictions are their for data access?

2. Data Quality
   - How good the data are?
   - How do the user know if the data are applicable for their use?
   - How complete are the data?
   - Was data consistency verified?

3. Spatial Data Organization
   - What data model was use to encode spatial data?
   - How many spatial objects are there?
   - Are methods other then coordinates used to encode location?
4. Spatial References

- Are coordinate locations encoded using Latitude/Longitude?
- What, if any, projection is used?
- What geographic datum was employed?
- What parameters should be used to convert the data from one coordinate system to other to preserve integrity?

5. Entity and Attribute Information

- What specific information is included?
- How is the information encoded?
- Were codes used?

6. Distribution

- Where and from whom do I obtain the data?
- What formats are available?
- What media types are available?
- Can I obtain the data online?
- How much does it cost?

7. Metadata Reference

- When were the data completed?
- Who complete the data?