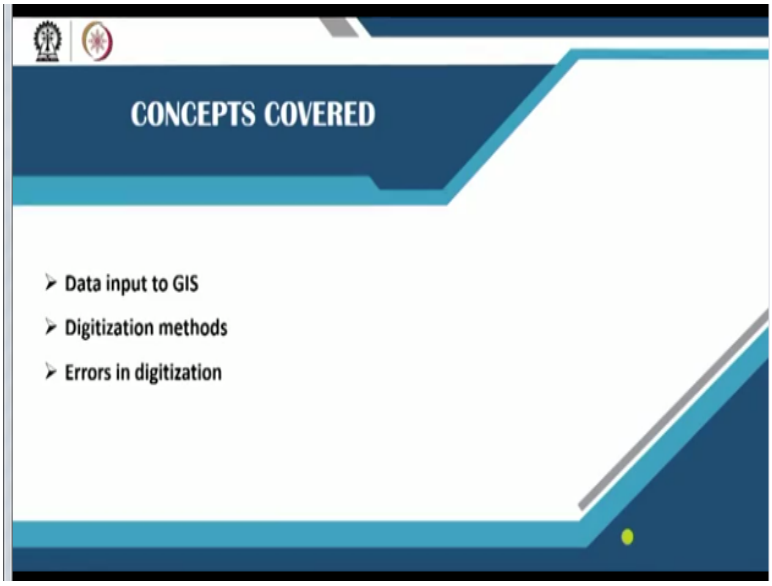
**Geographic Information Systems**
**Prof. Bharath H Aithal**
**Ranbir and Chitra Gupta School of Infrastructure Design and Management**
**Indian Institute of Technology-Kharagpur**
**Module No. 02**
**Lecture No. 10**
**Real World to Digital World through GIS (Continued)**

Hello, namaste welcome back we are in the last lecture of the module 2 where would the I would be showing you how different methods of generating a data for a GIS is there whether it is digitization whether it is manual digitization, automatic digitization what kind of different methods can be used. So probably this would be your first class of looking at actual data.

So, if you have certain data by your datasets by yourself maybe you have certain no module I mean certain equipments that I would mention it here, I would say that please start looking at how you can process the first set of data. So how do you put it in a geographic system, how do you convert it as a different tool to understand data to extract data that will go ahead in the next class but in order to create a input data this may be a first step, so in case you have certain data sets please start looking at it

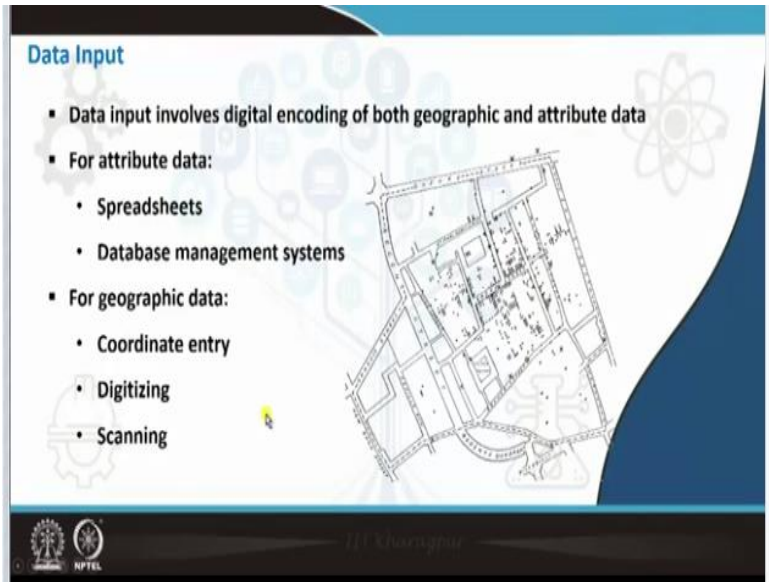
**(Refer Slide Time: 01:28)**




So, these are certain concepts that I would cover, one is the data input GIS what are the different digitization methods, what are the errors that is in this digitization, this is extremely important

for everyone to understand what are the different errors. So that it can be corrected in your entire data set.

**(Refer Slide Time: 01:48)**




So, when I say data input, data input involves digital encoding of both geographic and attribute data. So for example attribute data can be in a form of a spreadsheet it can be a data management systems. So any of those data management system elsewhere can also be an input to your data management system or it can be just as a spreadsheet, basic spreadsheet ok. When you look at the geographic data you will have to do a coordinate entry or you will digitize it and or you will scan it.

When I say geographic data, it is a data about the earth surface or a phenomena or an object ok. Whereas when I am speaking about the attribute data these are those data which adds value to this geographic data. So when you are looking at geographic data you are basically looking it as an entity something like this ok.

But if you add a value saying that this, this is maybe a regions quadrant or this maybe a particular road this is also a road it is a name of the road and length of the road etc. These becomes your attribute data, this can be in the form of a spreadsheet ok whereas geographic data has to be done with a coordinate entry, scanning or digitizing. So let us look at how we actually generate each of this geographic data.

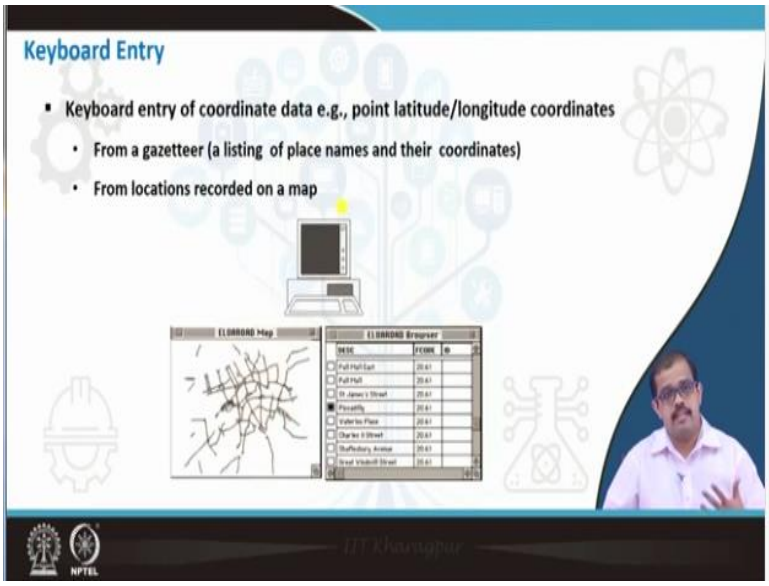
**(Refer Slide Time: 03:16)**


So when we look at data input to be harness it has the conversion of a hard copy to digital maps is the most time-consuming task in GIS. So it is almost 80% of the total project cost as per the estimate it is about US 10 billion annual market. So when you look at US dollar 10 million annual market it is extremely big conversion of any hard copy digital maps for example the maps that is in 1970s, 80s etc.

So these are extremely important in terms of understanding the evolution temporal evolution of that surface. Then these are quite labor-intensive, you need some extremely talented youngsters looking at the data and digitizing it is extremely tedious and error-prone. So you may have a lot of errors, so when you are creating data using this conversion of a hard copy then it is extremely-prone

Because most of the hard copy data or hard copy maps will have huge amounts of data maybe even you will not be able to look at it more I mean the way it has printed. So that is extremely important that creates an error basically. Then database development sometimes becomes an end in itself because it does not know how the database has to be created into a how the data model has to be created.

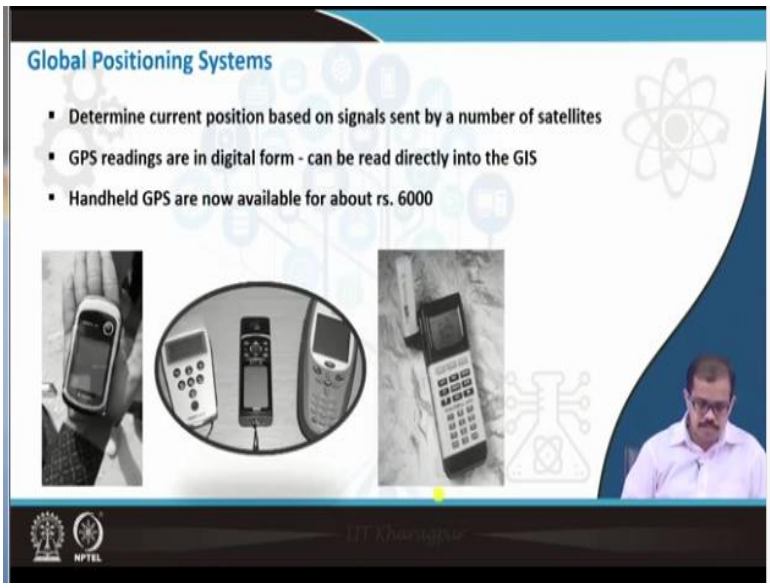
**(Refer Slide Time: 04:56)**




So understanding this, the first way of your entry, data entry is in a form of a keyboard. So when we are looking at a keyboard entry it is basically over the coordinate data maybe like a lat/long the coordinates that you enter that is shown something here ok. So from it maybe from a gazetteer listing of place names and their coordinates it may be from a gazetteer, it may be from your census data listing of name and coordinates from locations recorded on the map

So you take a you take a map, you start entering in your keyboard, different locations and different coordinates of each of these location the x, y value. So this becomes your first step of data entry it maybe from a keyboard just creating a excel sheet.

**(Refer Slide Time: 05:46)**



Or use sophisticated systems of something like a global positioning system. We will speak about global positioning system how it works, how different systems are there, India's own indigenous systems etc. But as of now GPS are handheld devices which can receive certain frequency of wavelength of spectrum in order to measure may a distance the geographical locations etc.

So these geographic say positioning system has also evolved over a period of time, so in order to give us better data better inputs of data. Now you have extremely efficient systems in order to measure this data, it is dependent on what frequency the actually if the signal that your device is receiving whether is a frequency that has extremely precise positioning or it is just a standard positioning frequency.

We will look at all of these in when we look at GPS but when you are looking at global positioning system this determines the current position based on signal sent by a number of satellites. So as in theory we need at least 3, 2, 4 satellites, normally it is a 4 satellite signals in order to give us a exact location on the earth's surface. So every place if you look at the forest car it is extremely different.

If you go into a forest and try to get a GPS signal it may be extremely difficult to get a GPS signal inside a forest because of it is cover. So looking at such instances there are certain issues that GPS can also provide, that is a disadvantage of GPS you should have a clear sky. So it may not be possible in a crowded region to get the GPS signals. Even for example in our urban form where it has completely crowded a your GPS may not work as efficiently as it has to be.

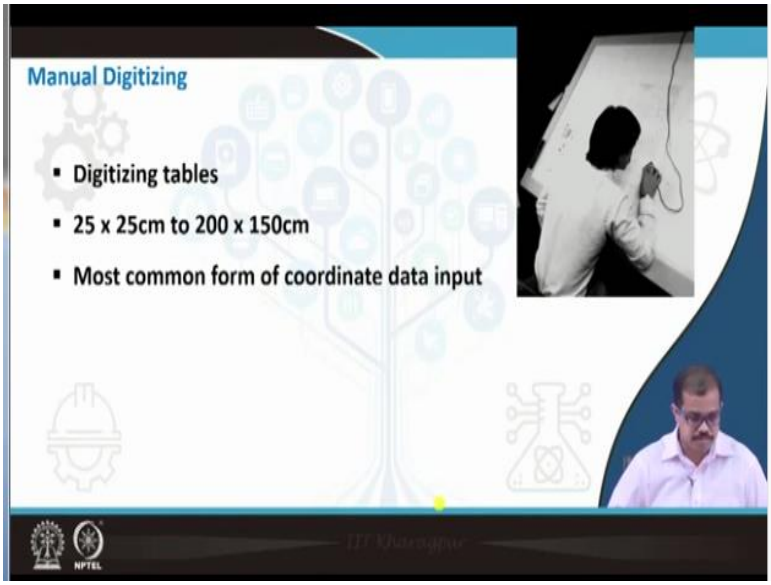
For example when you have certain mobile sets that you have assisted GPS built-in, you can see that your positional accuracy is completely different from what is shown on your map when you are it gives an accuracy of up to 1 kilometer, 500 meter, 600 meters etc. because the signal accuracy it may be certainly bad.

So GPS signals can be converted into a digital form you go to the field, you collect the data, you get back from the field the same can be translated into the coordinates or into an excel sheet by just bit of conversions. Then now you have lot of softwares which are bundled with this GPS

systems which can easily give you the geographical data as it is. So now as rupees 6000 to 8000 you get handheld GPS and moreover you have extremely good assisted GPS in your mobile phones.

But only thing is that this assisted GPS should be on a precise positioning which is normally not there and your whatever the data systems are there that should be extremely good. So most of these signals are actually calculated through your wireless stations sat that are there around you.

**(Refer Slide Time: 09:19)**

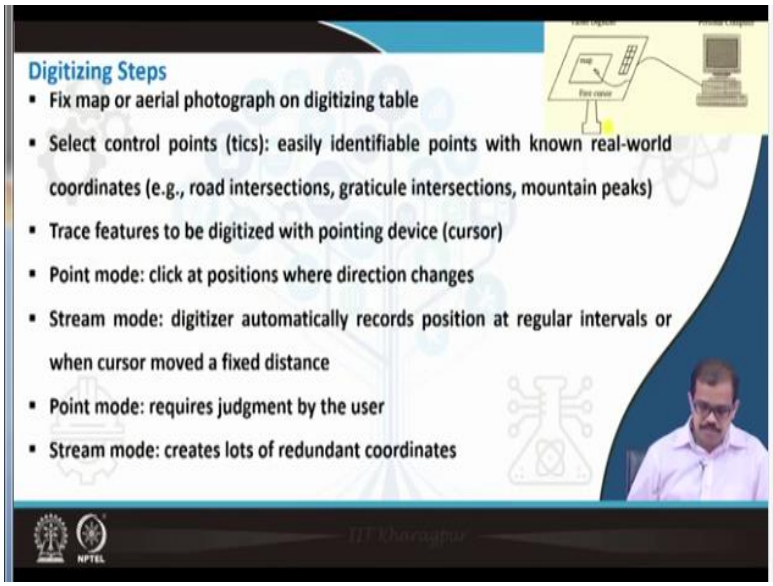



So the next part is manually digitizing, this is using a digitizing tables, so something like this probably when we started looking at GIS. So this is how we started with the digitizing tables in today's context this does not exist, no one uses such tables nor to digitize it. So this can be 25 by 25 centimetre of tables to 200 to 150 centimetre of tables.

So you have a digitizer if you seen the this person's hand this is a digitizer ok which has electronic signals. As you move and click on that particular position based on single click it gives you a point if you click and move on and stop at a particular point it gives you a line segment. If you move for the entire polygon it gives you the entire polygon, so that is how you start digitizing using the manual digitizing tables.

So this extremely tedious, if for example if you have let us say a city map it is extremely overgrown city like of any Indian cities, if you have a city map it is extremely difficult for you to identify an entity and put it using a manual digitizer and takes a huge amount of time and effort.

**(Refer Slide Time: 10:42)**




Then how do you if someone asks me how do you digitize yeah each of these first you fix a map or a aerial photograph on a digitizing table. Now in case someone wants to use it and has this facility please look at it looking at traditional ways of how it has been it started digitizing is also important and may be useful. Select control points that is using a tick ok easily identifiable points with real known world.

For example as I pointed out previously the framework data ok maybe an roads that are intersecting or you have very well known road segments or you know certain points which are extremely identifiable in a map. So in the real word so and that things that have not changed over a period of your study. So these can be your control points, so please look at it more carefully because control points define how accurate your data is otherwise you will not be able to get an accurate data.

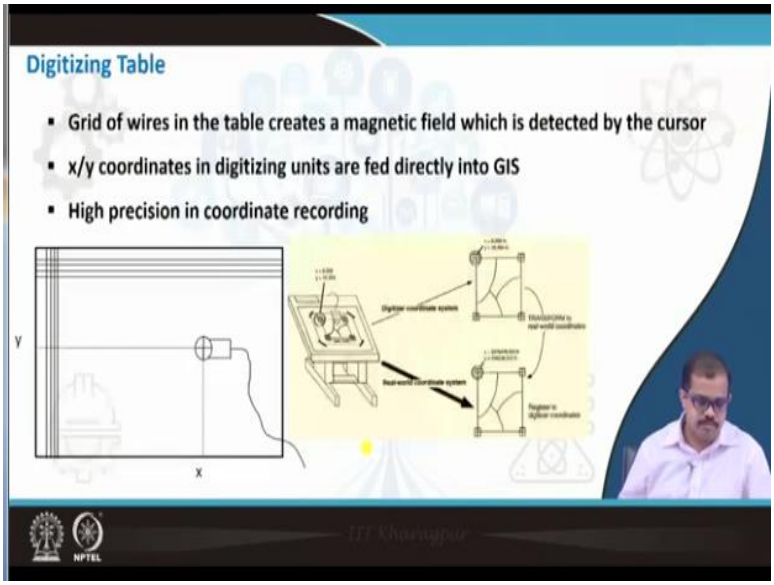
Then trace features to the digitize with the pointing device, so you go into for example let us say I am trying to digitize this I will start tracing this particular feature ok when I say point click, click a position where the direction changes where if it is a stream mode or a line mode your

digitizer automatically records at regular intervals. Then if it is a stream mode, only thing is that when you look at a point mode it requires extreme well judgment by the user.

For example let us say I am trying to click on yes here ok, if even a slight change will change its geographic coordinates ok. So that has to be looked at because that is a point it is a single point if you click it is done fine. So when you are looking at a stream mode it creates a lot of redundant coordinates because you are trying to push it something like this so in case you are not pushing it in a proper line segment it may create a lot of tangles.

Because of which you will have errors in your entire data set, so that is why the manual digitization is nowadays avoided but previously it was the thing that people use in order to build up the data set.

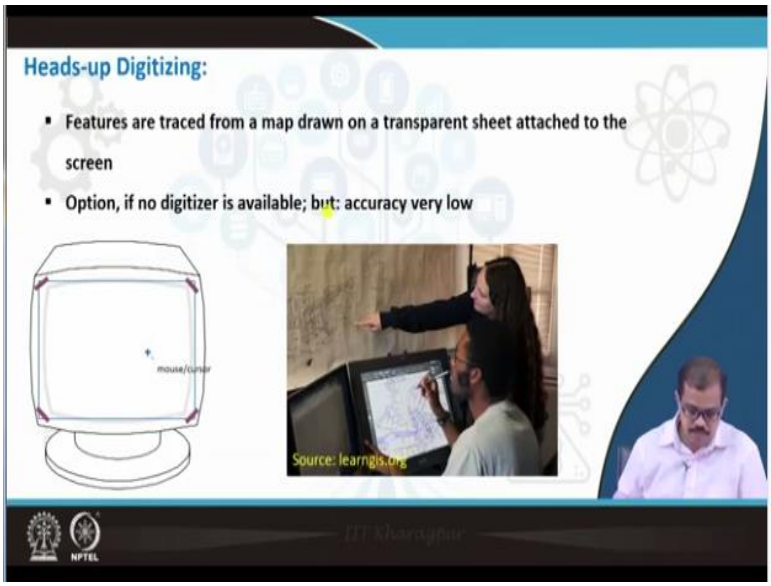
**(Refer Slide Time: 13:08)**




So when you look at the entire digitizing table you have a grid of wires in the table that creates a magnetic field something like this ok. So at the ends of this table you have the grid of wires then you have x, y coordinates in the digitizing units are fed directly into a GIS software. High precision you should have a very high precision in coordinate recording otherwise you are going or to be horribly wrong.

If you are any of your data set, you have do not have a precision you are surely getting a wrong coordinates into your data fed. So when you start analyzing it you will not sit at a extremely right point, you may have a shift of points which creates a geographically are wrong notion of representation.

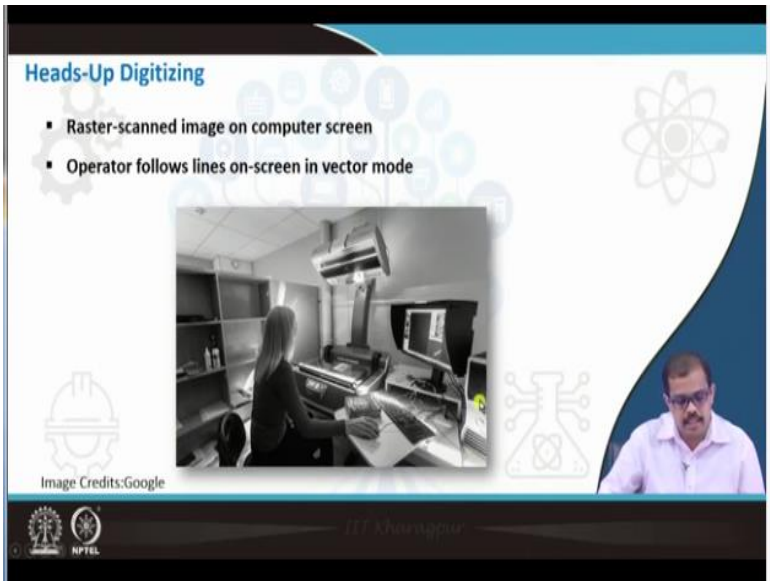
**(Refer Slide Time: 13:59)**




So the other method is heads-up digitizing, so for example if you do not have a digitizing table. So how do you do it, so use your own monitor which has capacity touch to use as a digitizer. So these features are traced from a map through a transparent sheet attached to the screen ok. So for example there is a map attached here you have a transparent screen. So you use this method but what happens with this is **is** extremely low accurate.

So if you are representing a very coarse resolution then you may use it but this accuracy is extremely low, so people avoided at measure of times. If you see there is a person here who is actually digitizing it using a heads-up digitizer, or for example to represent if your monitor is something like this there is normally these monitors are now I mean have a touch display. So if you put your map something like this starts clicking on this particular mouse cursor and start digitizing it. So that is how you do an heads-up digitizing but nowadays you donot have any heads-up digitizer used in many of the places ok.

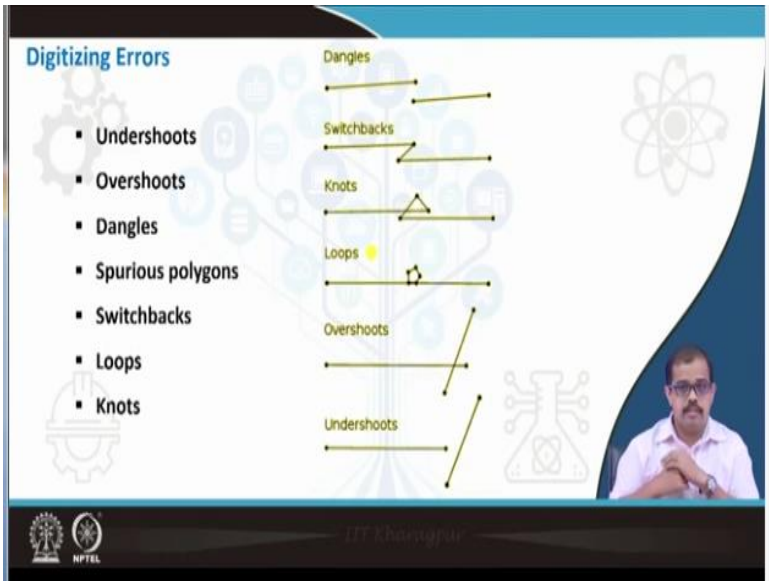
**(Refer Slide Time: 15:27)**




So heads-up digitizing can also be done something like this, there is a scanner that keeps running and create a raster image just is one of the images that have got in google. These systems are never there in last one and half decade people have stopped using such devices what operator does is he follow the lines on the screen in a vector mode. As it scans for example the a paper scan, if you have seen a scanner in any of those I mean any of xerox machine it scans smoothly from one end to the other end and moves back.

So the user also starts digitizing it from one end to the other end and moves back, that is how digitization is done ok. If it is a point he does it with the point if it is a line it is sat with the line and if it is a polygon he does it with a polygon. So it is does every part of a page one row by one row depending on what is the size of the scanning system is, so that is how it is a heads up digitization works but nowadays this is not used much.

**(Refer Slide Time: 16:42)**




But if someone has this facility also available today you can have a look at **it it is** it is actually a good experience of understanding how digitization happens, you can learn a lot of issues that you handle in digitization. Today it is more of automatic methods or manual digitization in your computers, I will also give you some examples of how you digitize, you have at least 1 practical class where you actually do the entire digitization of your map.

So that would be given as one of those I mean practice sessions that you may have to look at, if you have to understand it better. So when you look at digitization errors this I have already shown you before. The first error that you see is undershoot which you can see here, there is one line segment here there is the other line segment here. So but this line segment for to be exactly correct has to fit into this particular segment.

or if it is a road that is crossing something like this ok, this has to fit in something here. But it is undershooting by this amount, that is called an undershoot. So whenever you are having a digitization error as undershoot this has to be corrected before you processing it asset into a database ok. Then you have overshoots for example this particular line has overshooted this particular point.

Then you have dangles something like this you have 2 line segments that had to meet but it is dangling, it is separated by a space vertically or horizontally, so that is called dangling, so that

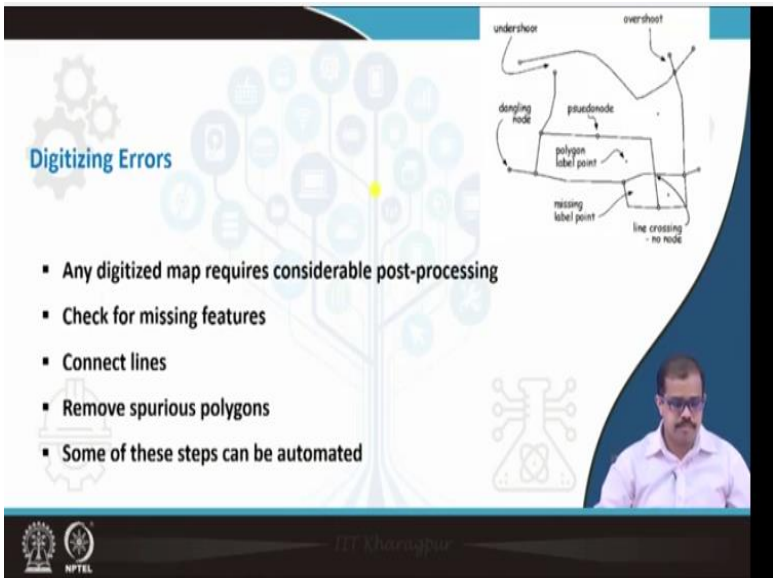
has to be corrected. You look at which the data reference is correct and correct and correct the other vector data another the line segment. Then you have spurious polygons I am not shown here certain polygons which may not be particularly there.

Then you have switchbacks something like this where it had to be something completely like a line segment. But when you are digitizing you went an overshoot and came back and you digitize the same line. So certain errors that occur, so switchbacks is very commonly seen in most of the digitized datasets. So look at this certain errors like this, so this has to be corrected then you have loops which may have fallen probably when for example if you are digitizing their at after certain point of time you push your mouse, so it creates a loop ok.

So with a click it creates a look, so such things are very commonly seen in most of the databases which has to be corrected. And you have knots, sometimes like this, so all of these errors have to be created and has to be rectified. But one specific thing that has been developed over a period of time there is most of the GIS software today has developed certain tools which can actually look at all of these errors in automatic way just by a click of the mouse.

But and you can actually correct most of the errors that I have shown. But there are certain digitize places where the software cannot interpret ok or correct it. So that is where you are the manual error correction has to be done, so there needs at least 5% of manual effort in order to correct that entire dataset.

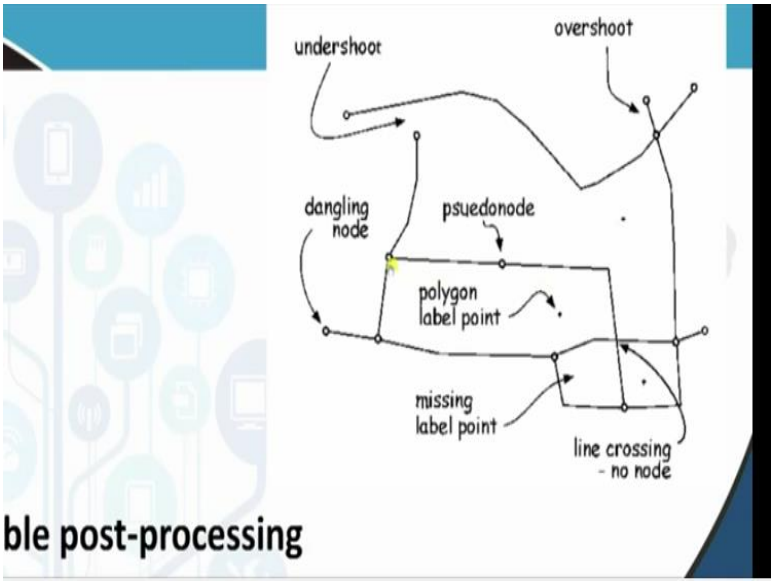
**(Refer Slide Time: 20:29)**




So when we look at digitizing errors I have given another example something here. So if you look at it, this is the entire vector dataset that I have develop, when you see this this is a dangling node, see this had to be connected somewhere but it has dangling out ok, how do you find out this is a dangling node, if you see the original data set you can find out why it is a dangling node

There is if you see most of this are actually labelled but there is a label point that is actually missing here the point if you see here all these have labels ok. Whereas this particular polygon does not have a label ok or connecting line segment does not have a label, so this has to be corrected, that is missing.

**(Refer Slide Time: 21:21)**



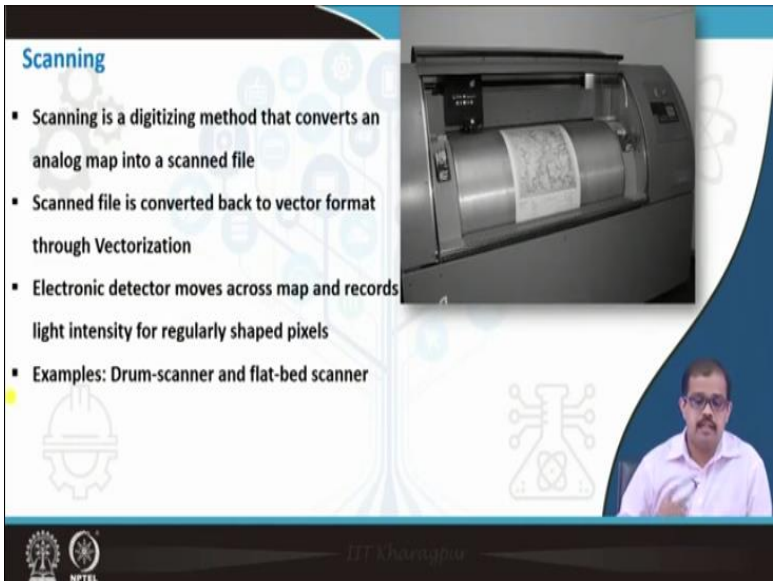
So when you see very carefully here ok if you see this in a more careful way, this particular line is crossing and not setting up the node, that is also a mistake, that has to be corrected, each of these line segments have a node. So this node has to be connected, so there is that that is another mistake and when you see this entire thing when you are creating a particular line segment you have clicked somewhere in between and create a node that becomes a pseudonode which has to be corrected.

So these are certain errors which are surely would be there in your entire in your data set. So please look at some of these nodes there is overshoot here it was suppose to stop here but it went on across here, this is an overshoot. And when you see this there is an undershoot, so all of these things has to be looked at here if your data set has to be proper. Then when you are look at the digitizing errors any of the digitize map requires post processing.

It cannot be without manual intervention at least 5 to 10% of post processing has to be done manually. Check for any missing features that is extremely important, I am sure in a large map you would have missed a huge amount of feature. So you please do a secondary checking with the missed features, connect lines that is very important aspect but it can be done automatically nowadays.

Then you have removal of spurious polygons this is also done by certain tools automatically. But have a secondary check, so that you are sure that your dataset is clean. Then some of these steps also can be automated but always have a look at with the secondary way of looking.

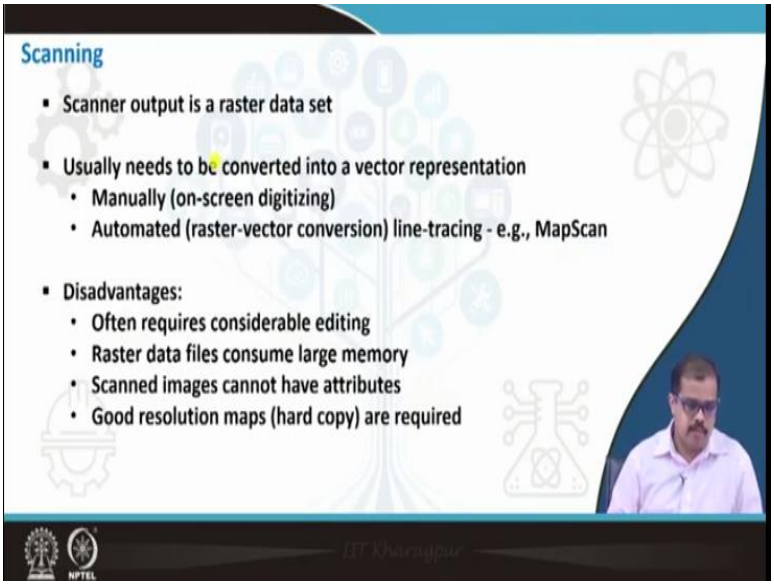
**(Refer Slide Time: 23:11)**




Then other method is scanning, scanning it converts analog map into a digital map. So now most of them are electronic they are drum scanners and flatbed scanners, most of scanners now use are the flatbed scanners. So the scanned file can be easily converted back to a vector format and digitize the raster to vector conversion can be done easily and vectorization there is now of popular instead of digitizing.

This is the most popular method now used by various researchers across the globe or practitioners across the globe who are converting the hard maps into a digital format. So this is more easier and less cumbersome in order to convert a hard map into a digital map.

**(Refer Slide Time: 24:03)**



So normally a scanner output is a raster data, it usually needs to be converted into a vector representation. Manually is done by on-screen digitization where you have we will also look at how manually we can create will use in a practical class will create one map manually. So **so** that you understand how the vectorization is done but nowadays once you have a raster data most of the processes automated.

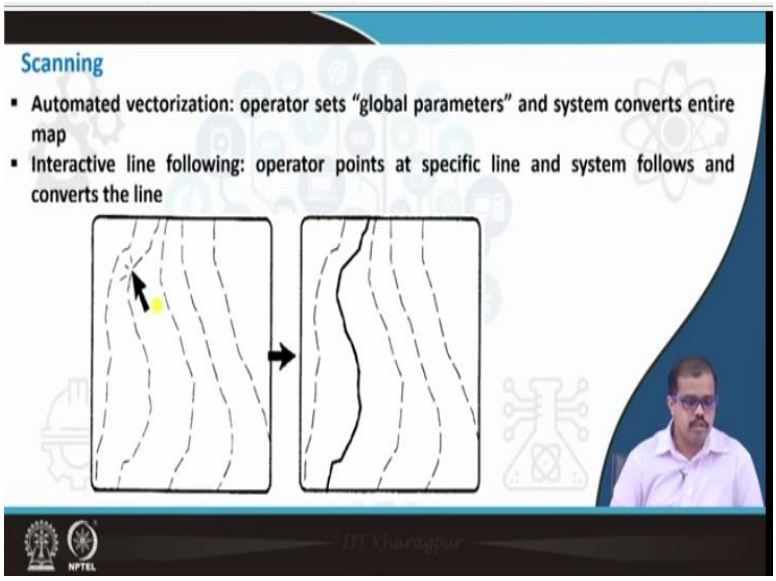
So raster to vector conversion is done then you have light line tracing there is an software which is called map scan can be extremely useful in such issues. So and you have certain disadvantages often requires considerable editing I am sure most of the automated things needs considerable editing you need to put in a lot of efforts in editing the secondary or have a secondary check.

Then raster data files normally consumes huge space normally the data that you have created maybe from your own data sources maybe even through digitizing can sit maybe in about 50 60MB depending on what size of data or the area that you have considered. But when you look at the raster data it may go up to a GB or 2GB or 3GB size of data or even hundreds of GBs of data depending on what area that you have considered.

So but scan images cannot have attributes keep that in mind, so scanned images until unless these are geographically connected then the database is created only then you can have attributes. You can extract attributes on the scanned images but they cannot have attributes by themselves once you have scanned it ok. Good resolution maps are extremely important in order to create a scanned data.

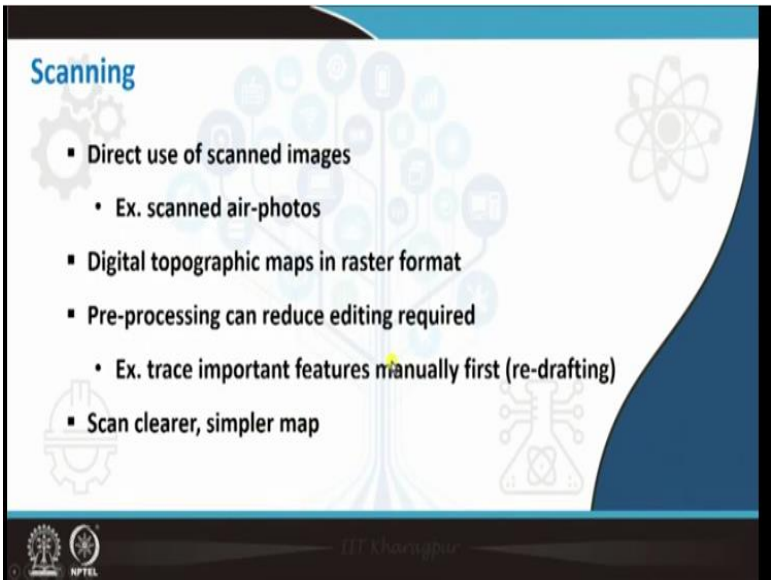
If you have a very bad resolution that is not repair that is not giving you a data model in a proper way you will not be able to or entities are not represented in a proper way, you will not be able to identify what entity is there in the map. So you need extremely good resolution, extremely clear map in order to scan it and convert it into a digital map. So this is one of the big disadvantages that you can see as a scanning.

**(Refer Slide Time: 26:28)**




So for example there is an automated vectorization here this is your scanned map ok operator sets the global parameters and the system converts the entire map ok. So you can see here these are certain maps that is create, this is line that has been created using this ok. So now how do you do this operator points to a specific line for example this line if I keep bonding and systems start converting that as a line, so that is how the automatic vectorization works ok.

**(Refer Slide Time: 27:05)**

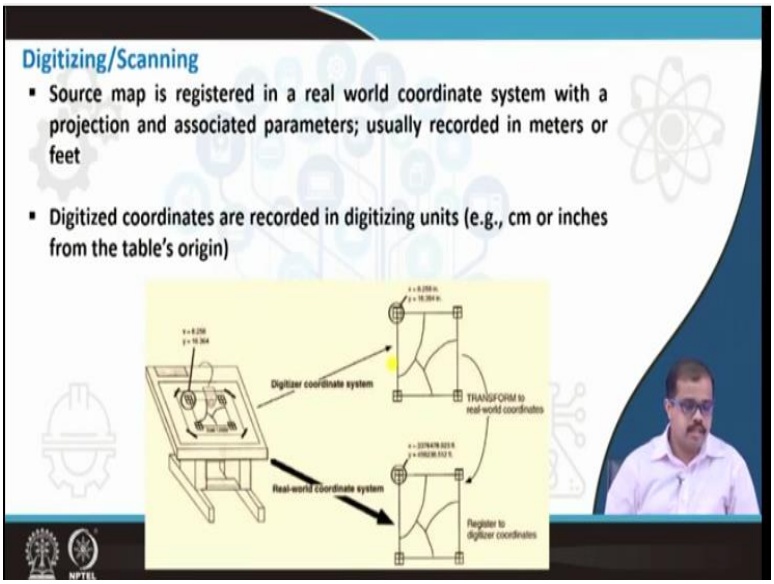



So as scanning it can be direct use of any scanned images it can be scanned air photos, digital topographic maps, pre processing can reduce editing required. So lot of pre processing of the photo can be done in order to reduce errors, any trace the important features first that is my

advice to anyone who is using such thing. First trace important features manually, then give the rest for automated extraction, do not do automatic extraction always.

You would miss a lot of features trace at first then use an automated extraction, scan clearer simpler maps that is the best way ok otherwise it is going to be tedious.

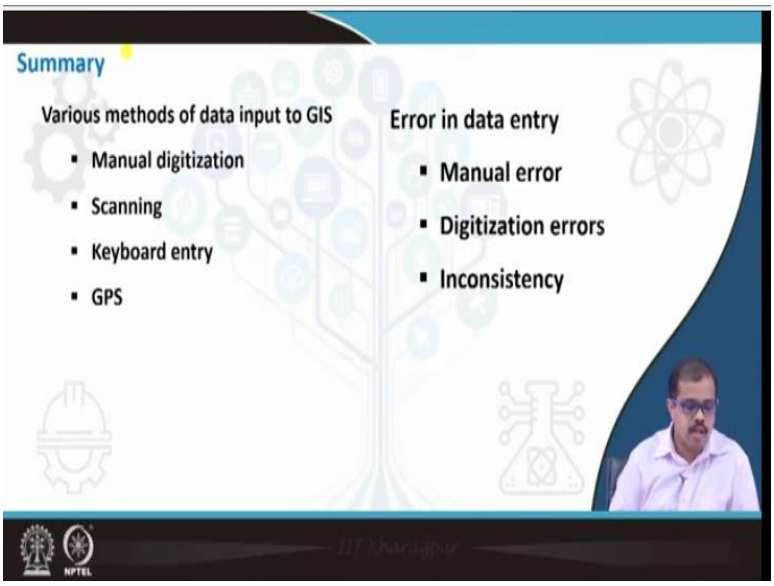
**(Refer Slide Time: 27:52)**




So when you look at digitization and scanning the source for a map is registered in the real word coordinate system with projection and associate parameters usually these are recorded in meters or feet ok. So digitize coordinates are recorded in digitizing units, so that depends on whether it is a centimetre, inches from the tables origin. So that has to be very carefully looked at, how do you convert the real word phenomena into digitize word phenomena

So when you are using a digitizing table, please keep this in mind. So digital the way of digital measurement is very different from your real world measurement. So the conversion has to be significantly different, so please look at this and record it in your map or in your digitize format. So that the unit conversion is done in a much proper way.

**(Refer Slide Time: 28:48)**




So in order to summarize we have various methods of data input to GIS or in a manual digitization, scanning, the keyboard entry and GPS. We have looked at all of these, we look at GPS in much more details on how the signal is there, what are different type of systems do we have any parallels to GPS, indigenous ways of which has a parallel to GPS also has been developed, we look at all of these.

Then error in data entry, this is extremely very important and very crucial in developing a dataset. So the manual error that may creep in then you have digitizing errors and inconsistency. So all of this has to be addressed before you have a data set but my suggestion is that please go into certain data depository which have already explained even it can be your data.gov.in, download certain datasets, small datasets.

I do not expect it to a extremely big dataset if possible look at how the data is there I look at the data get a feel of a data. So that is the first way of learning your data, if you understand the data get a few have got a feel of data then you know what is there on the ground. If possible visit that place, that is how you start looking, you start being a GIS person. So maybe take up this work look at how you can get different data sources different, different data and look at how you can manage the data.

Thank you very much, let us meet in the next class wherein I would look at the next module of lectures, thank you.