Geographic Information Systems
Prof. Bharath H Aithal
Ranbir and Chitra Gupta School of Infrastructure Design and Management
Indian Institute of Technology-Kharagpur

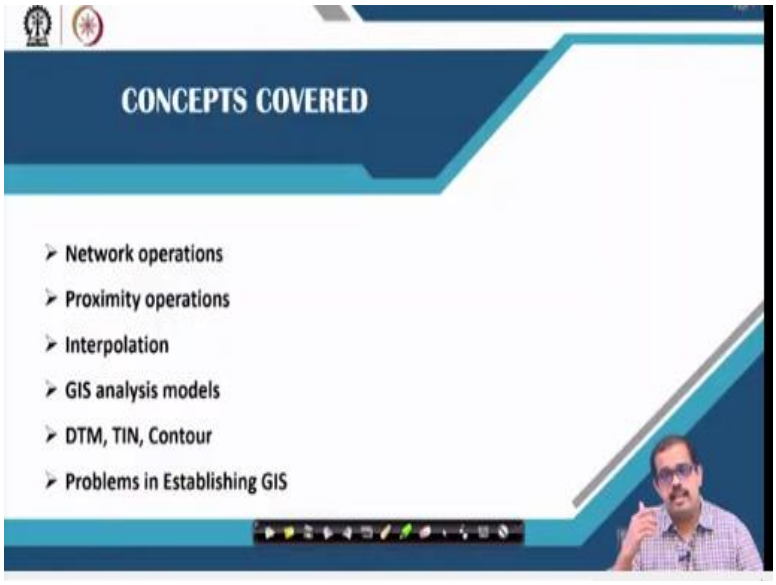
Module-09
Lecture-45
Advanced Spatial Analysis

Hello Namaste, welcome back to the course on geographic information system, this is a last session on how we will look at spatial analysis. Wherein I would be speaking on advanced spatial analysis in this particular class. Also I would give you some disadvantages or limitations that we are facing in today's GIS world. So when we are looking at this as a scenario we would also look at what are the basic limitations and how do you forego.

So probably at the end of this session I would also give you certain things that what should be look forward in terms of being a GIS user. So that maybe at the end of this course but in today's class let us understand what do you mean by advance spatial analysis, what are different tools in advance spatial analysis. We have done some of these analysis before but we have not already spoken it about as spatial analysis but just spoke it about a specific tool.

So but we will here we will speak each of these things as a tool ok, and then we would look at the limitations part of it.

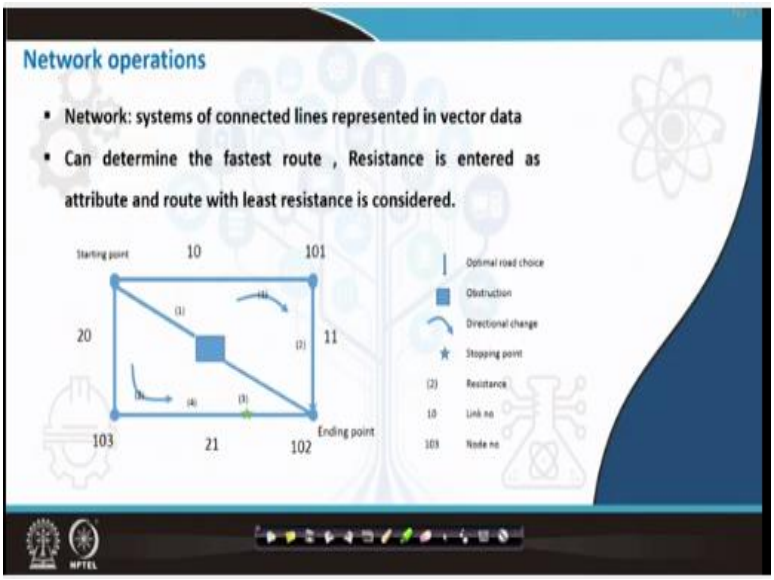
(Refer Slide Time: 01:24)



So as a rightly said we would look at network operations, proximity operations, how do we do an interpolation different methods in I have give you a very basic methods in interpolations not very advanced techniques ok. GIS analysis models, what are different models of analysis then looking at DTM. We have already spoken about what do you mean by DTM, how it is very different from your DSM.

Then looking at TIN model, the contour model etc. then problems in establishing GIS, that is the last that I would speak about in this particular class.

(Refer Slide Time: 02:00)



And when we have looking at network operations, networks it is a system of connected lines represented by a vector data ok, when you see this particular point. This is a particular point where there is a starting point and there is a ending point. There k there as I spoke at the previous class, if you want to reach at a particular from a particular of particular destination, there can be several roads.

This can be a shortest route that you reach here ok, there may be another route which has reaching here. There may be another route which has actually reaching that this point. So there are different nodes, there are different stopping points, there are different ways obstructions etc. so based on what is the best travel region that is how you travel for your particular operation, that is what is basically the network operators ok.

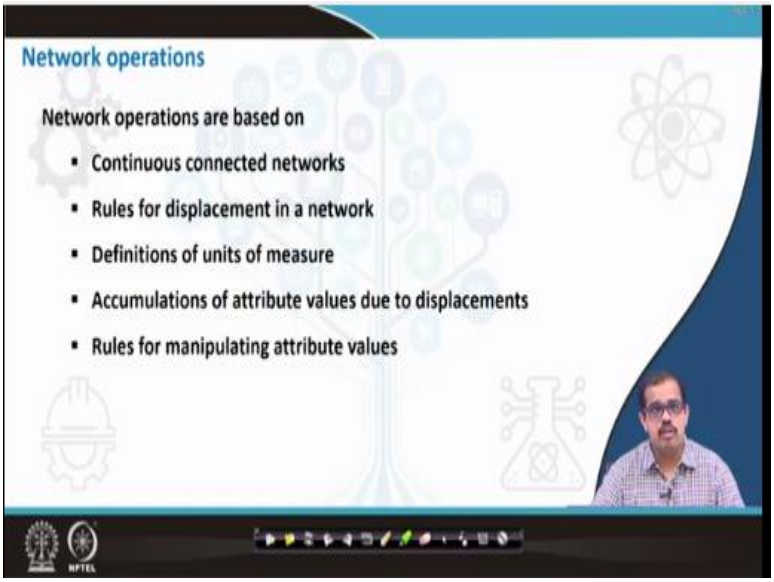
So it can actually help in determining the fastest route or shortest route or even resistance is entered as in the attribute and route with least resistance is actually considered. Let us say that I am actually solving the problem of electrical issue in a particular building ok, there are lot of wires that actually are moving out. So now I have to look at which wire may have been damaged because of it is life or it is different issues that may occur.

So what I try to do is that I will find out how these lines are actually connected where it is connected ok. If there are obstructions in that region or what is the kind of loading conditions in that particular lines and what is the amount of load. So once we have understood all of these things probably we will be able to understand. So if it was physically then you would have to said calculate each and everything separately and then combine it to a using certain mathematical issues.

Then come up with a particular equation and then find out what maybe a particular issue. But now I have a different layers, so once I have created these different layers which may not be an issue. Because most of the data is available just creating a layer is not an issue. Once you have created the layer then you overlay on it and you will said this particular line is actually as an issue ok, it is as simple as this.

That is how the network operations are done or one another operation is shortest route that I explain in the previous class ok.

(Refer Slide Time: 04:26)

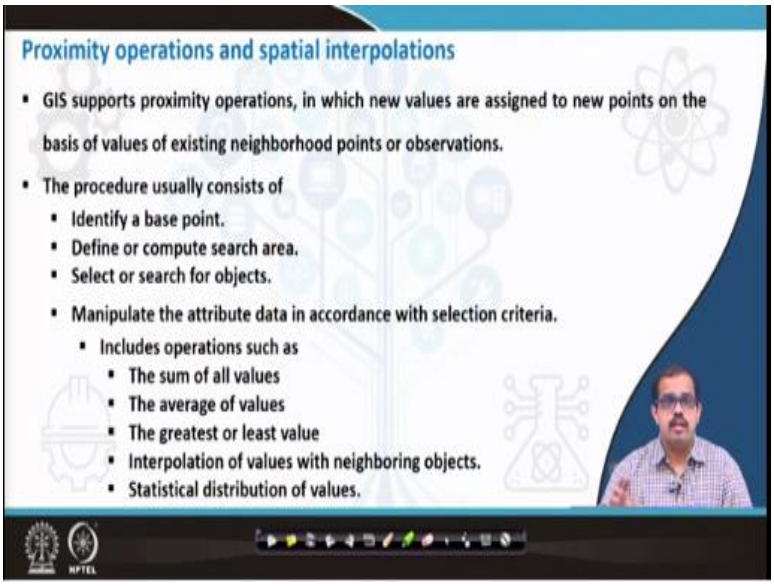


So network operations are actually based on the continuous connected networks, so always network operations are continuously connected. So please keep this is in mind, I have seen some people who try to put it on network operations where there are no connected entities ok or connected relations. So until unless it is connected network either through a entity or a relation you cannot do any kind of network operations.

So rules were displacement in network, so you should understand what are the different rules that you can have it on network for understanding the displacement units, definitions of units of measure, what kind of measurement you are trying to do. For example I gave you an issue of electrical loading and a road network. So both units does not match each other, so you have to maintain different units for different systems.

So please keep this in mind, accumulation of attribute values due to displacements there may be a large amount of displacement or new replacement that would have been happened over a period of time. So there may be a lot number of attribute values, so look at those attribute values and then do a network operation. Rules for manipulating attribute value, so you should create your own rules to actually manipulate the attribute values ok.

(Refer Slide Time: 05:50)



So this is about the network operations, so use that very carefully where probably when we are looking practical class we will look at how we do an network operation ok. Then there is a proximity operations. So proximity operations and special interpolations are very well known tools or systems then in any special GIS. So GIS supports proximity operations in which new values are assigned to new points on the basis of value of the existing neighborhood points or observations.

For example you have 10 meteorological station and you want to represent the entire what is the rainfall in the entire region. So you collect all the 10 values meteorological station. So you apply a proximity analysis in terms of interpolation, any kind of interpolation that you may use. So that depends on what kind accuracy you need, so based on that you will find out what is the rainfall in different regions which means to say that you are trying to look at the proximity locations from that particular point metrological stations to find out what is the value around that particular station.

The procedures that you it can be here is identify a base station. First thing is look at what is the base station ok, then define a compute search area. Then select and search of objects that can be there in that particular region. Once you have understood the objects there are there in that region then look at the values that you will have to look at.

It includes operations such as sum of all the values, average of values, the greatest of least value. Any kind of operations that is dependant on the user to use such operations, I have you an example of rainfall. So if I have want to look at the gradient of rainfall across the particular region where there is 10 metrological station probably I would look at interpolation of values with the neighboring objects.

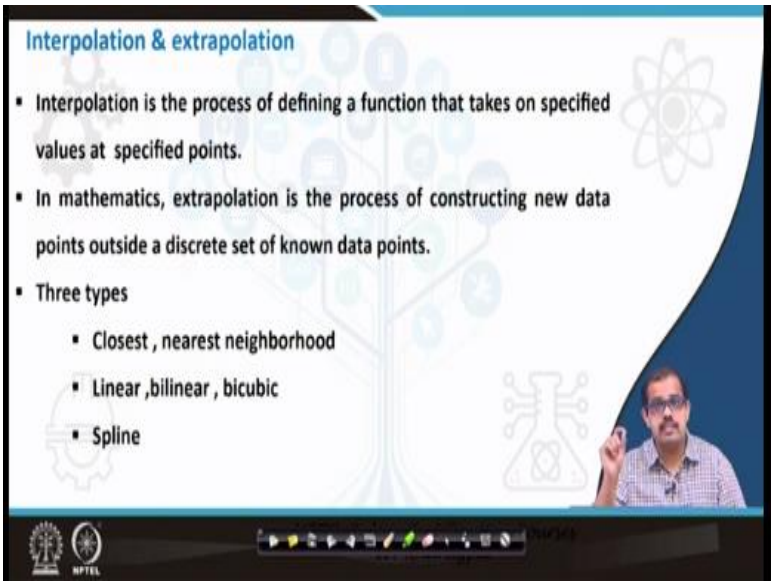
So I know in between those stations what is the change of values are happened, the gradient of change. So that once you have the gradient of change and represented as a thematic map probably it is easier for anyone to understand what kind of analysis is has been I am in, what kind of the rainfall is there in that particular region. So it may be useful in various kind of analysis, for example someone is trying to relate the vegetation with rainfall or different kinds of vegetation with rainfall ok.

It maybe evergreen forest, it maybe coniferous forest, it may be an just an open land. So if you want to establish a relationship, the first thing that you will do is that you will collect the meteorological station. Then you have the first you interpolate it and find out what else the region. So you have 2 let us say you have a 2 or 3 cosmids of that particular water body there.

So you have rainfall data then you will find out what is the amount water availability in that particular region. Then you will find out what maybe the amount of I mean if there is vegetation, what maybe the water availability in the summer season. And what maybe the water availability without vegetation or specific vegetation. So that is how different kinds of analysis can be done, so that is where your proximity analysis is essentially useful.

Then you have statistical distribution of values, how these values are distributed statistically, so that also can be done. So everything all kinds of analysis is dependent on user, there is no fixed rules of using any kind of tool or so it depends extremely on the way the user thinks ok.

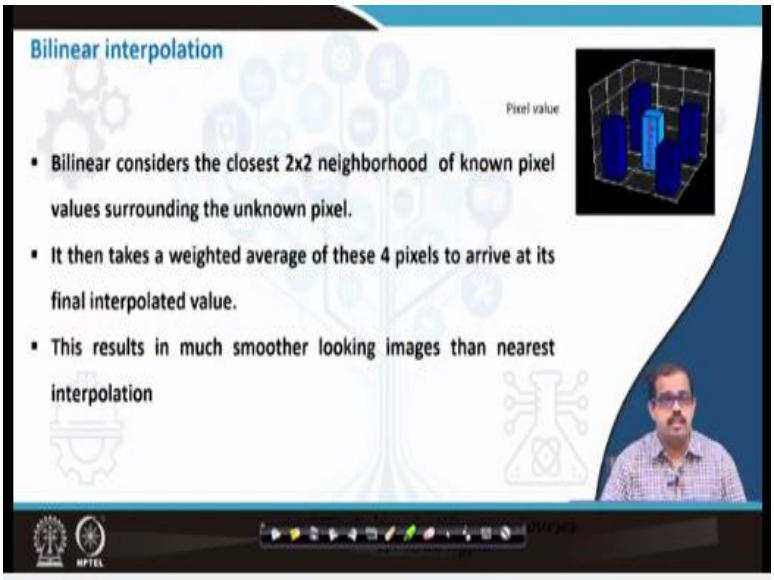
(Refer Slide Time: 09:25)



So interpolation when you are looking at a interpolation, as I said interpolation is a process of defining a function that takes a specified value at specified points. In when you are looking at mathematical part of it, extrapolation is a process of constructing new data points outside a discrete set of known points. An interpolation is creating a new set of data inside set of known 2 described points.

If there are 2 described points, so if you are finding out a point here then it is interpolation ok in between, extrapolation is outside these 2 points ok. So when you are looking at it the most easiest way least computational way is using a nearest neighborhood. Then you have linear, bilinear, bicubic, spline, so any kind of interpolations can be used in such analysis.

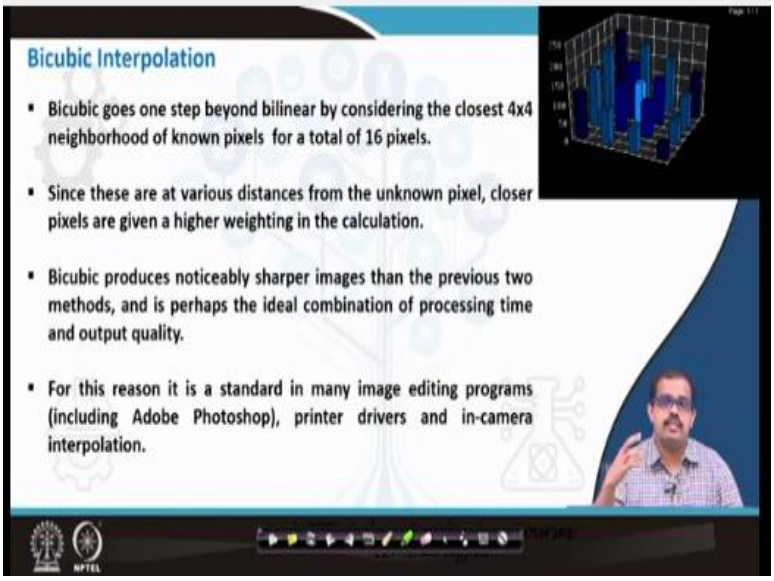
(Refer Slide Time: 10:18)



And if you look at this when I say nearest neighborhood, so you have the nearest value that is actually assigned to that particular. So let us say you have an neighborhood, so you will look at which is a nearest value that actually matches to this particular point. So nearest value is assign to that particular point. Whereas when you are looking at bilinear interpolation it considers a closest 2 cross 2 neighborhood which means a 4 neighborhood if it is an 8 pixel neighborhood.

I am speaking about a raster data ok surrounding any of the unknown pixel ok. So now if you uses 2 cross 2 neighborhood which is a 4 neighborhood ok. So it takes a weighted average of these 4 pixels at it is final interpolated value, the results is smoother looking images than any of the nearest interpolations, so this is more smoother ok.

(Refer Slide Time: 11:11)

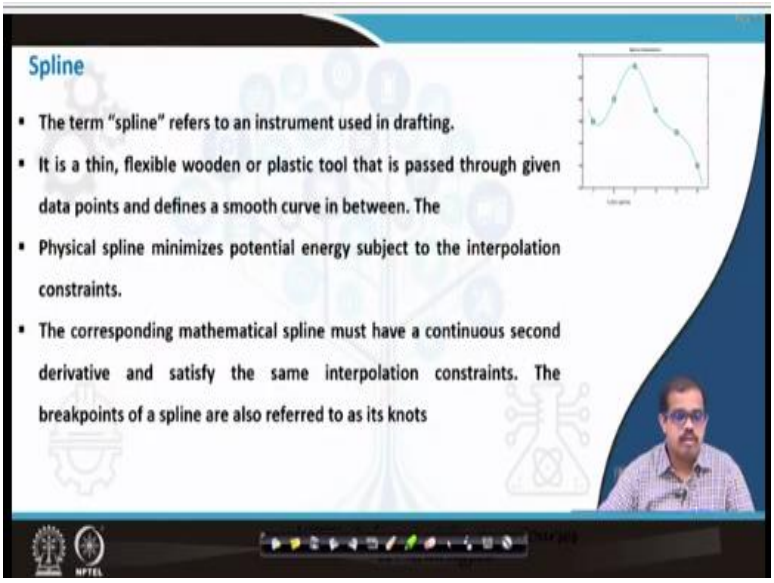


And if someone has good computational capability probably they look at cubic or bicubic, when I say cubic it looks at 8 pixel surround that particular point ok. If there is a central raster value and you have a number of pixels that you are considering, so 8 directions is 8 pixels. So **you** you take in 8 pixels and you find out what is the particular point based on maybe you will submit and then consider the mean of that value.

So you calculate the mean of that particular value you will get the cubic interpolated value of particular point unknown point. Similarly when you look at a bicubic we look at a 4 cross 4 neighborhood, it is a 16 pixels that you consider around the particular point. So since this is the distance we do not know based on unknown pixel, the closer pixels are given higher weights in terms of calculation and as you in a way it is a lower weights ok.

So bicubic produces noticeably sharper images than the previous 2 methods and is perhaps the ideal combination of processing time and output quality. For this reason it is a standard in many image at editing programs or printer drivers and in-camera are interpolated ok. So you have printer drivers is and in-camera drivers are more have been interpolated ok.

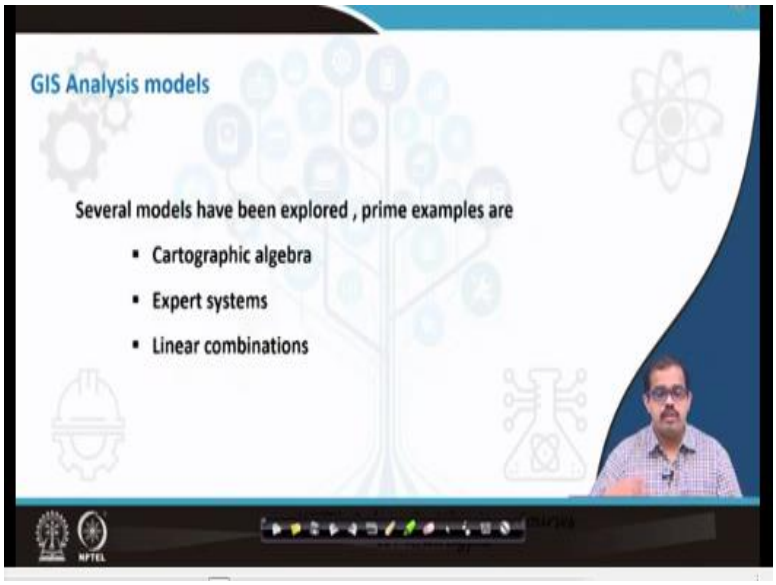
(Refer Slide Time: 12:38)



Then there are something called as spline, spline interpolation. So it refers to instrument used normally in drafting, many of the planners use plain interpolation which is most probably the standard for planning activities. So it is a thin, flexible wooden or a plastics tool that is pass through given data points and defines a smooth curve in between them. The physical spline minimizes the potential energy subject to interpolation constraints.

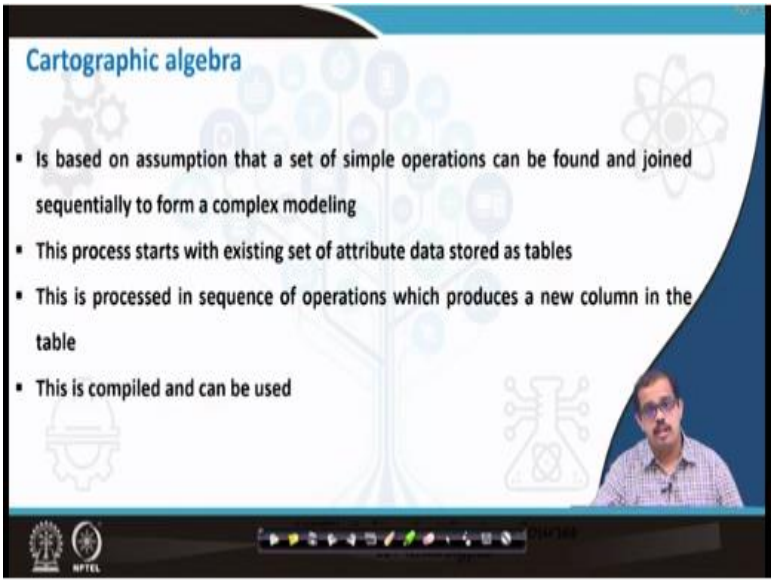
The corresponding mathematical spline must have a continuous second derivative and satisfy the same interpolation constraints. So it uses a second derivative value, the break points of the spline or also referred to as knots ok. So these are knots that you use, so whatever I am showing here is actually a cubic spline ok.

(Refer Slide Time: 13:35)



Now when you are looking at different models in a GIS there may be several models. For example there is based on cartographic algebra, there is expert systems, there is linear combinations. So you can even look at certain research articles recently where expert systems are used for different kinds of GIS analytical model.

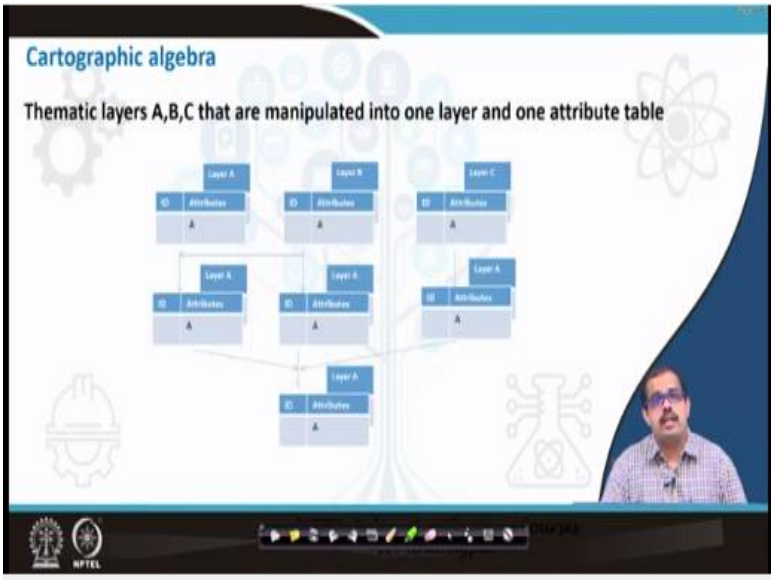
(Refer Slide Time: 13:54)



When you are looking at cartographic area, it is based on the assumption that a set of simple operations can be found and joint sequentially to form a complex model ok, which means you have a set different models which has simple algebra. Then combining it will form a complex model with a complex algebra. Then the process starts with existing set of attribute data stored in

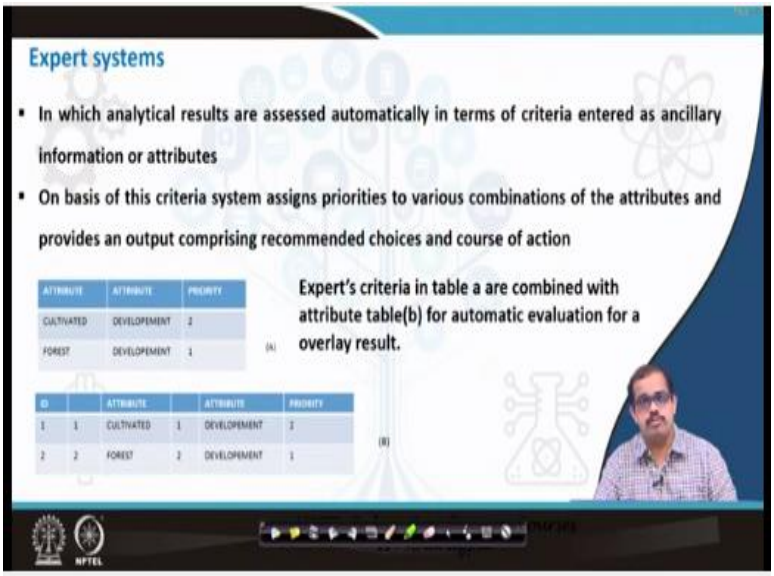
the tables, these are processed in a sequence of operations which produces a new column in the table, then finally these are compiled and can be used for various analysis.

(Refer Slide Time: 14:31)



Then this is example of an thematic layer of A, B, C that are manipulated into a single layer and a 1 attribute table at end to perform it as a model. So you have an layer with attribute value layer B with certain attribute value layer C with attribute value. So you perform certain analysis here than once you have found out that analysis then you will have the final layer which actually interacting with which is actually containing all the attributes that you would have produced.

(Refer Slide Time: 15:01)

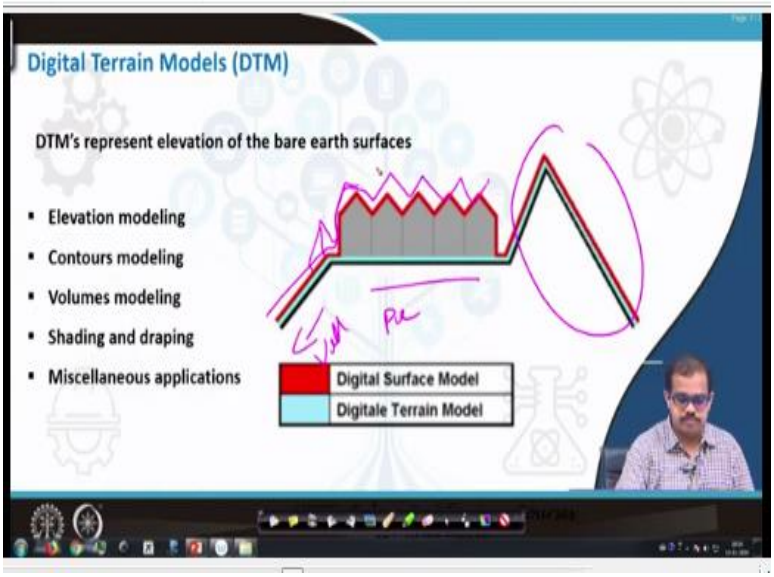


Then you have an expert system these in which the analytical results are assessed automatically in terms of criteria entered as in ancillary information of attributes. So you have attributes, you have a criteria based on that you try to look at the expert system. On based on the criteria system the well known expert is analytical hierarchical process.

So if you want to look at this probably the expert systems which if you look at any of the GIS papers you can find out AHP is well known way of looking at how an expert system can be used as an expert. When I say expert system these are the values that are maybe from the expert surveys or expert opinions that you taken and create your own table listing where you actually compare across variables.

Once you compare across variables you start defining it with weights, and once you have define it weights, it automatically evaluates to give you a particular value. So based on that value if the value is less than 0.01 or 0.1, so normally it is concerned to be the analytically possible for any kind of analysis. So this kind of using an expert system may be useful in when you are trying to use a huge number of variables in particular analysis or different entities in your analysis.

(Refer Slide Time: 16:30)



And the next kind of modeling maybe using a digital terrain model, I have given you a good example of what do you mean by a surface model and a terrain model. For example here if I let us say I have an data that planimetric data that I have considered. So now I can do any kind of

analysis, for example if I can do elevation modeling, I can look at contour modeling, I have volume modeling, shading and draping modeling.

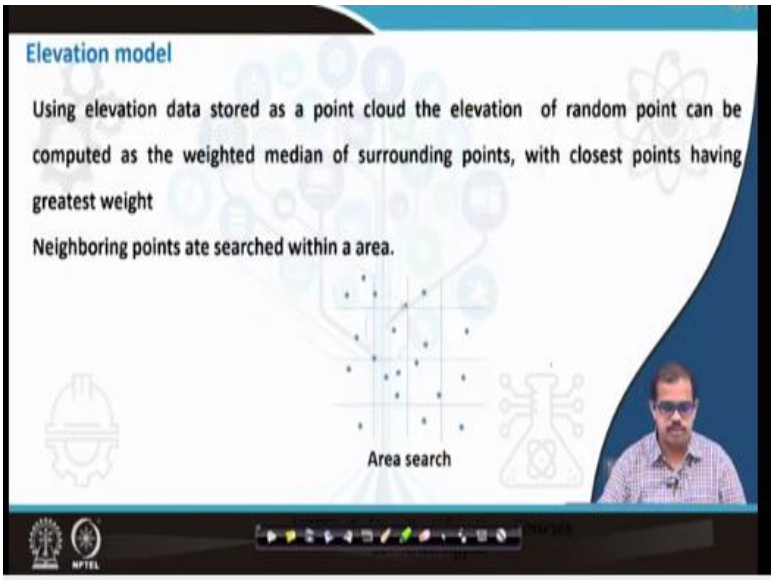
And all of these can be performed using these surfaces, when I look at DTM digital terrain model, it is only the surface of the earth that is measured in this particular model. And when you look at the digital surface model here let us say that see this is the surface of the earth that is being mentioned here ok. So now this is what this is the valley and this is the hill and this is the plain area.

So when you are looking at this if you are if I want to mark it here this is a hilly area, this is a planar area, plain area and this is the valley ok whatever is thing. Now if you look at this follow this blue line that is representing the terrain, how the terrain of the earth surface is there. So without human interference what kind of terrain of the earth surface is there, that is exactly the digital terrain model.

If with the interference of the human interference is there or the natural way of development is there, what it maybe a tree here, it maybe I am in houses that have been developed. So if that is also use then you try to find out what is on the surface, so you get this kind of model like this. Let us say if there is a tree it would have comes something like this, so that is nothing but a digital surface model.

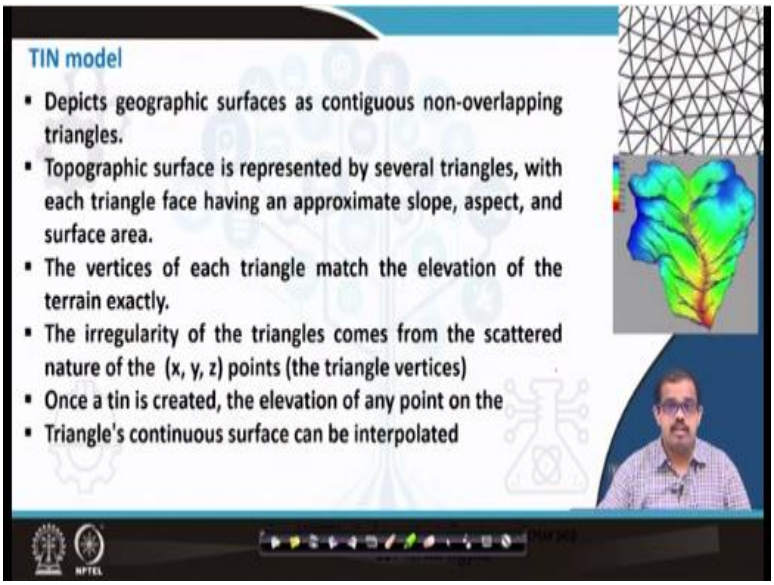
There are various ways extracting a terrain model and the surface model probably if someone is interested please explore how it is done ok. So that is one of the well known research methods in various kinds of analysis research inputs in various kinds of analysis that can be used.

(Refer Slide Time: 18:43)



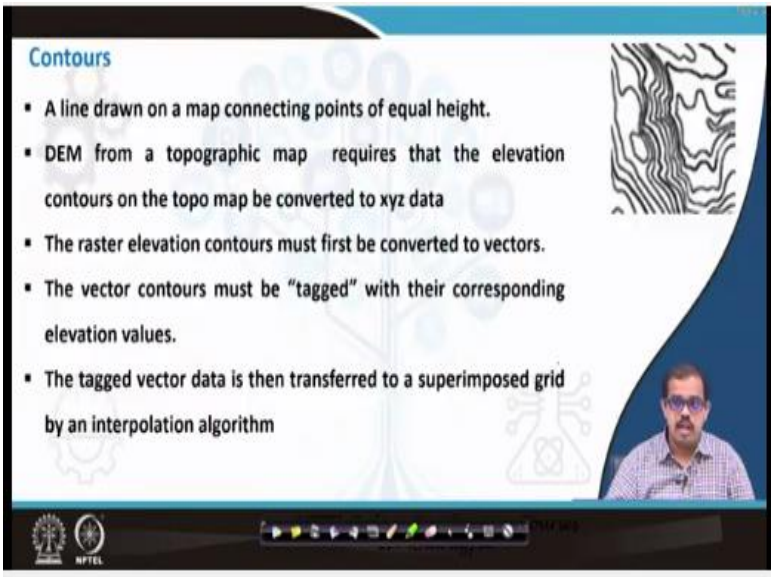
Then you have an elevation model this is you have an elevation data that is stored maybe in random points. So based on the computed weighted means of the surrounding points using a closest point you would try to find out what is the greatest weight. So nearing points are searched within that particular area and you will find out what is the particular elevation at that particular point, so that is a elevation model.

(Refer Slide Time: 19:11)



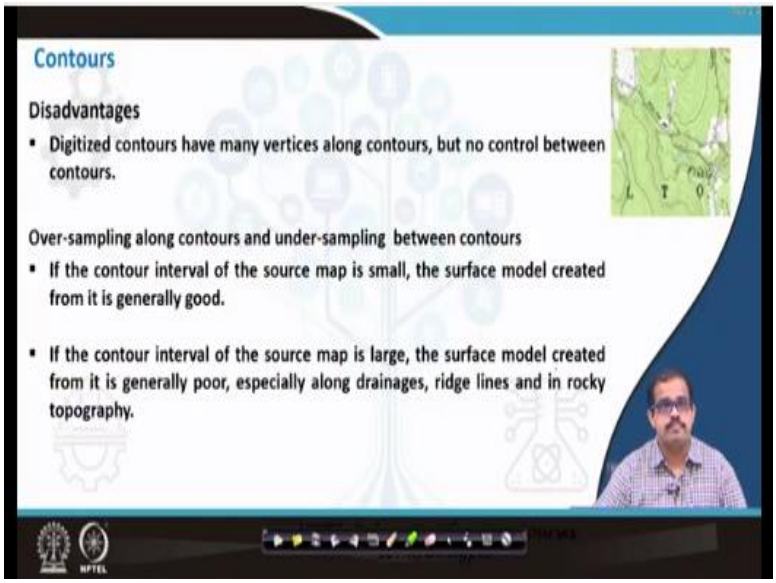
Then you have a TIN model, I have already explain the TIN model previously, so I would not get into details of what do you mean by a TIN model here.

(Refer Slide Time: 19:20)



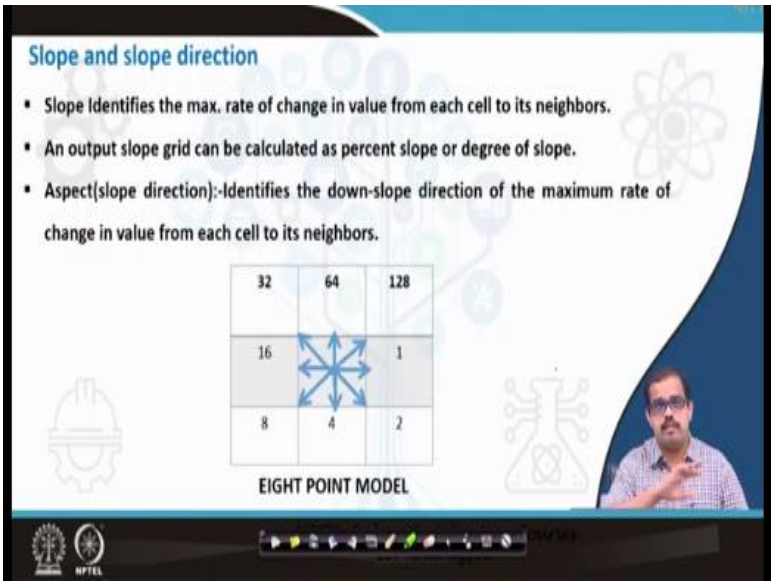
Then is the contour model, so again the same thing I have spoke about in the previous sessions maybe in the fifth week or sixth week. So please refer to those sessions when you are looking at the contour model.

(Refer Slide Time: 19:34)



So I have spoke about both the disadvantages and advantages, so you can probably refer there.

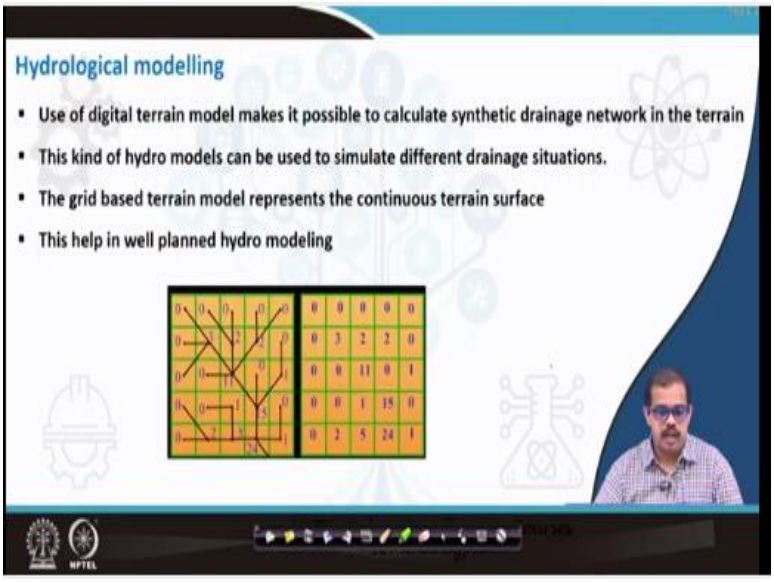
(Refer Slide Time: 19:39)



And there is another model which is called as slope and slope direction model. So this actually if you are looking at a hydrological application, this is extremely useful kind of model. It identifies the maximum rate of change and which is the direction of rate of change that is what is important in a slope and a slope direction model. Any output slope grade can be calculated using the percentage slope or degree of slope ok.

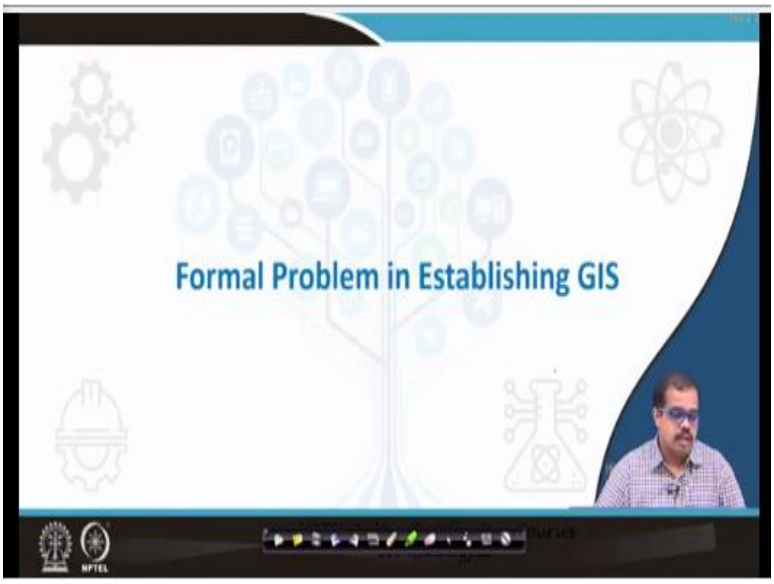
So aspect or a slope direction identifies the down slope direction at the maximum rate of change in a value from each cell to it is neighbors.

(Refer Slide Time: 20:19)



For example I am showing you this, this is one of those hydrological models that you can see. So when you are looking at this it is actually defining how the hydrological model is connected, how that particular drainage system is connected. So this kind of analysis can be performed using this slope modeling.

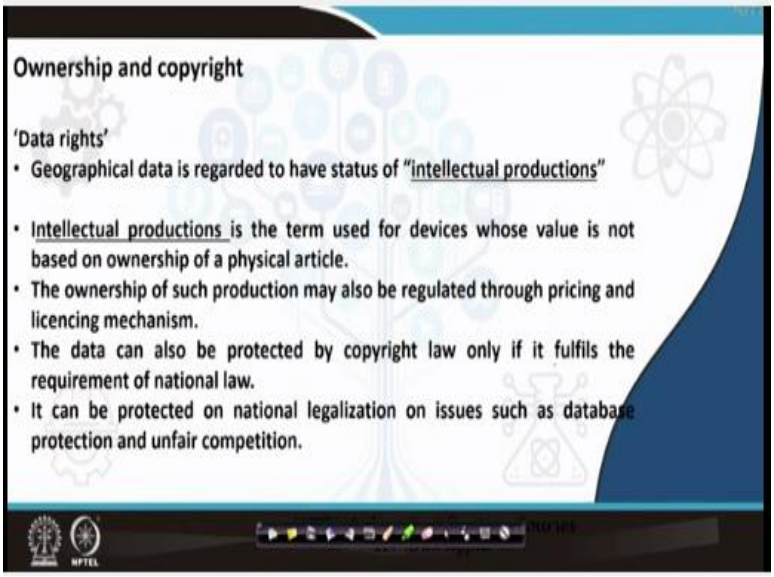
(Refer Slide Time: 20:40)



Then once these are the different advanced kind of analysis that you can do other than this you have fuzzy analysis, you have different kinds of expert systems that you can utilize, you have artificial intelligence techniques that you can apply it on GIS. So extreme whatever you think can be used in with GIS ok but I would stop at this aspect.

So that you will be able to understand it if someone wants to look at advanced part of it probably you should look at some advanced GIS in terms of developing it. Now let us look at what is the formal problem in establishing here, if you want to look at GIS as the establishing factor what maybe the issues.

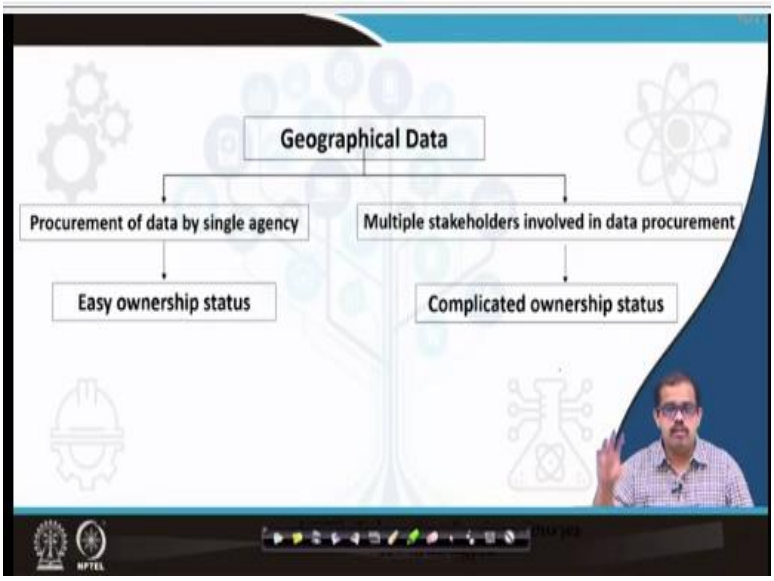
(Refer Slide Time: 21:23)



The first issue that you would come across is a ownership and a copyright ok. So you would have heard about data rights ok. So geographical data is always concerned to be an intellectual production ok, so it is regulated by IPs. So in when I say intellectual production it is termed used for devices whose value is actually not based on the ownership of a physical article ok but it can be a digital article.

So the ownership of such productions maybe regulated through it maybe through pricing or it may be through the licensing mechanism. The data can be also protected by a any copyright law only if it fulfills the requirement of a national law. It can be protected on an national legislation or issues such as database production and unfair computation. So all of these are issues when you termed of ownership etc.

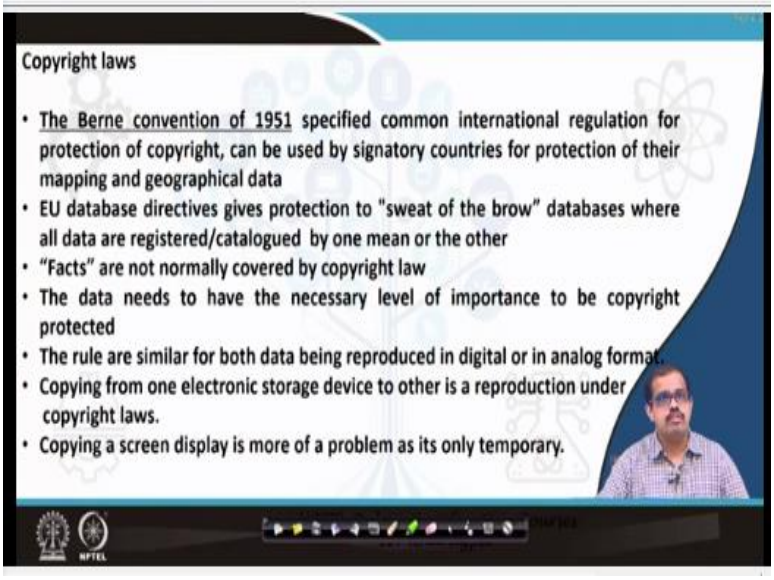
(Refer Slide Time: 22:20)



When you look at a geographical data you have procurement from a single end agency or you have multiple stakeholders when you are actually procuring such data. So if you have a single agency which you are trying to look at the geographic data who has a ownership, then you can easily get that ownership status. But if you have multiple stakeholders then it is extremely complicated.

That is where today's GIS data is actually failing the consumer or the user who really want who has to consume this data for his or her daily related activities ok.

(Refer Slide Time: 23:01)



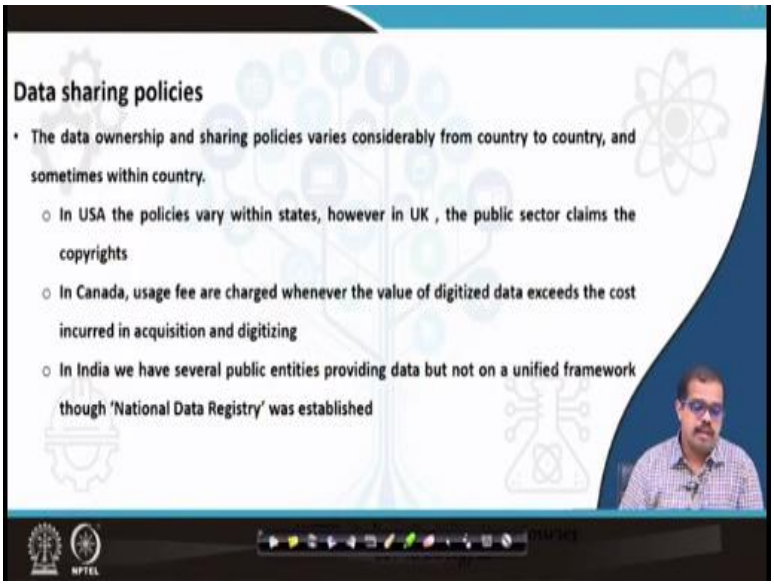
So when you look at the copyright laws that normally go on such digital data you have the Berne convention of 1951 which said which specify the common international regulations of protection of copyright ok. So then you have an EU database directives that gives protection to sweat of the brow databases where all data registered or catalogued by one or the other mean.

Then facts are normally not copyrighted under the copyright law, only the digital data is under the copyright law ok. The data that needs to have a necessarily level of importance to be a copyright protected. So you should understand that there are only certain data which are actually can be copyright protected. So what is that data probably you can look at the Berne convention entire report or you have Indian rules also wherein you can look at similar activities.

The rule are similar for both data being produced in digital or analogue form ok. Copying from one electrical storage to other is a reproduction under the copyright law ok. So when you are actually reproducing it you are suppose to either credit the author and also produce certain information from that author as necessary. So copying a screen display is more of a problem but for example if you see the very recent issue is that.

If you actually put an presentations on online ok or maybe your videos online, people take a screenshot of that video and added it to their presentation. So that is a more a problem, that is really a problem ok. But now you have reverse image search where you can find out from where that screenshot has done. But it is really a problem in terms of copyright laws, so be specifically careful. So when you are producing data you are governed under the copyright laws.

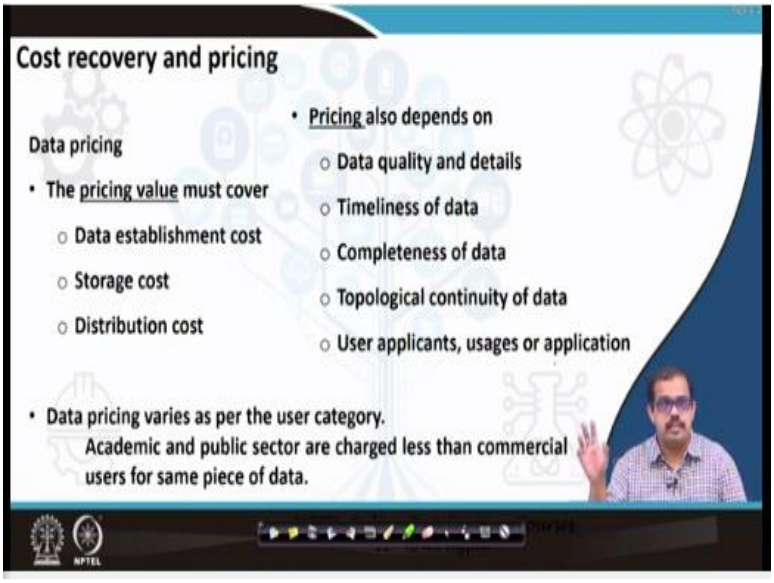
(Refer Slide Time: 25:09)



So sharing also certain policies ok, so for example in USA the policies vary with states, however in UK the public sector claims the copyright. In Canada you may have certain usage fees whenever charge whenever the value of a digitized data exceeds the incurred. I mean if you have a digital data, the value of it is much higher than when it is incurred in acquisition and digitizing, then you have to pay up some value as a fee in Canada.

Whereas in India we have several public entities providing data but these are not on a unified framework. We do not have any value network but we have a huge dataset in India which is publicly available only thing is that we do not know the exactly the source ok. But government is really doing a excellent work in terms of developing a national data registry, so that the more of this spatial data is available to the user upfront ok.

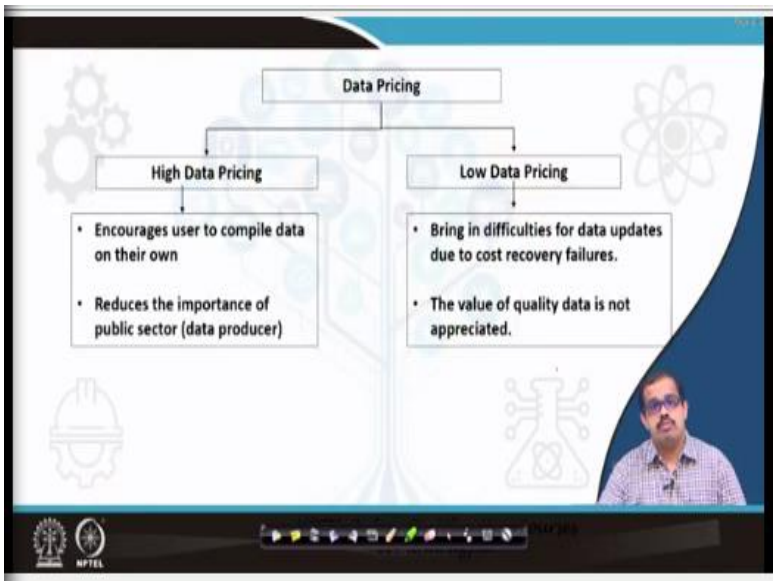
(Refer Slide Time: 26:19)



So you have cost recovery and pricing, I said the best way of maintaining the data is using the cost recovery and pricing. When I say data pricing the pricing value must cover data establishment cost, storage cost and distribution cost. When I look at pricing, pricing depends on data quality and details, timeliness of data, completeness of data, topological continuity of data, user applicants who will use that particular data or using certain applications.

So all of these are things are considered when you are looking at pricing ok. So it can be different for user categories. For example in India the academic and the public sector are charge less whereas the commercial users use the same piece of data is they have buy it at a more higher prices. So it depends on what kind of users are there ok, so there are certain relaxations and when you are looking at as an academic and a public sector ok.

(Refer Slide Time: 27:15)



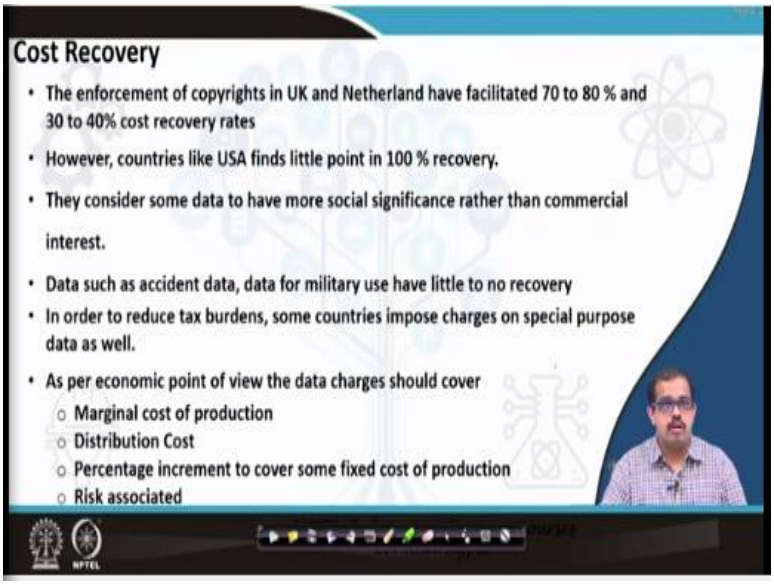
So when you look at data pricing, you have high data pricing versus the low data pricing. High data pricing it will actually help encourage the user to compile data on their own. Basically that is exactly what users of today are trying to do, there are lot of data that is available which is actually proprietary data. If you have to buy any of those data you have to pay a large amount of upfront.

So instead of that it is better to develop our own data, that is what people are trying to do. If you let us say and more importantly it reduces the importance of any public sector or a data producer ok. That is why many of those companies which are commercially putting out the data are may not be well known because the data are extremely costly which cannot be accessed by the maybe a public agency or from or academic or a general public.

If you let us say in that contracts if you have a low data policy, data pricing policy. So it brings difficulty in data updation due to cost recovery failures. So you have your data has to be updated every now and then, so it has to has certain value. So if you have a cost that is actually very less in terms many of the government agencies today charge extremely less for the data cost. So if that is a case then it may not be feasible in updating the data that is why most of the government data is outdated.

So if you really have to update, so your pricing should be competitive enough but not very high ok. The value of quality data is not appreciated, so that is why if it is a low data especially when you are looking at the Indian data, so the value of quality data is not appreciated if it is a low price data ok.

(Refer Slide Time: 29:05)



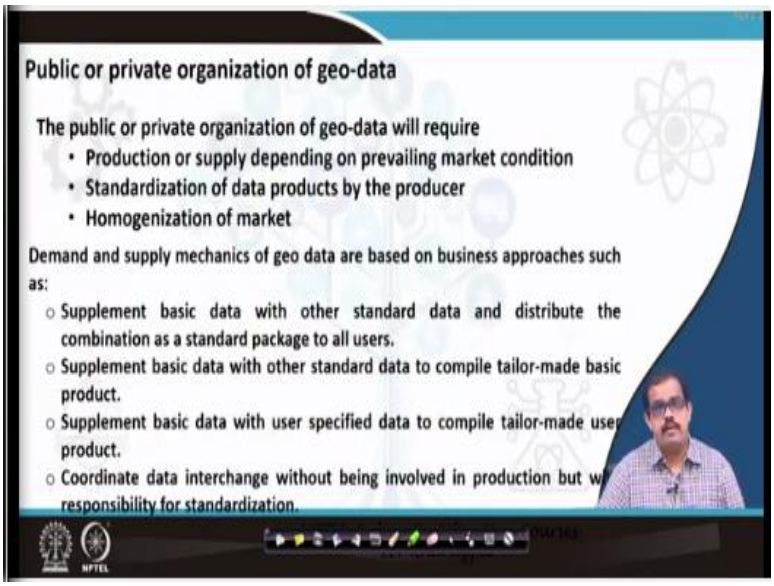
Then there are cost recovery, different cost recovery methods when you are looking at enforcement of copyrights and states of united kingdom and Netherlands they are facilitate 70 to 80% and 30 to 40% of cost recovery rates are that are available in those countries. However if you look at USA the cost recovery rates for any data is 100% always, whatever data is produced by them by different agencies is 100% consumed.

So they do not have any loss but when you look at Indian scenario as per certain reports we have some are marginal cost about 68% recovery. So that is where the Indian data our in Indian context the data is not very much publicly available or it is not utilized in a larger context. They consider some data have to be more social significance than the of commercial interest in when you are looking at US data.

Data such as accident data, data for military use may have little or no recovery at all. So those should not be in pricing bracket, so keep that in mind. So as an economic point of view you should have a marginal cost of production, distribution cost, percentage increment to cover some

fixed cost of production and the risk associated. So all of these put in together maybe your digital data charge with should be quite marginal ok.

(Refer Slide Time: 30:30)



When you are looking at public or a private organization of a geo-data the public or a private organization of geo-data will require production or supply depending on prevailing market conditions. So it everything is decided on the market condition wither on standardization of data products by the producer or homogenization of the market. So depending on all of these the public or the private organization of geo-data will be required to produce supply the demand of the data.

And when you look at demand and supply mechanics of geo-data it is basically a business approach ok. So you should look at supplement basic data with standard data if it is possible and distribute the combination as a standard package to all users. So that is what most of the public companies to till now or supplement basic data with other standard data to compile a tailor-made basic data.

So this is what the private companies do till now or any of those companies which are providing data. Supplement basic data with user specified data to compile tailor-made user product is what most of the enterprising companies do till now. So all of these are different ways of business

approaches that you can do ok. So it depends on a particular company and a particular user to choose what kind of data and how from which source that particular data is considered.

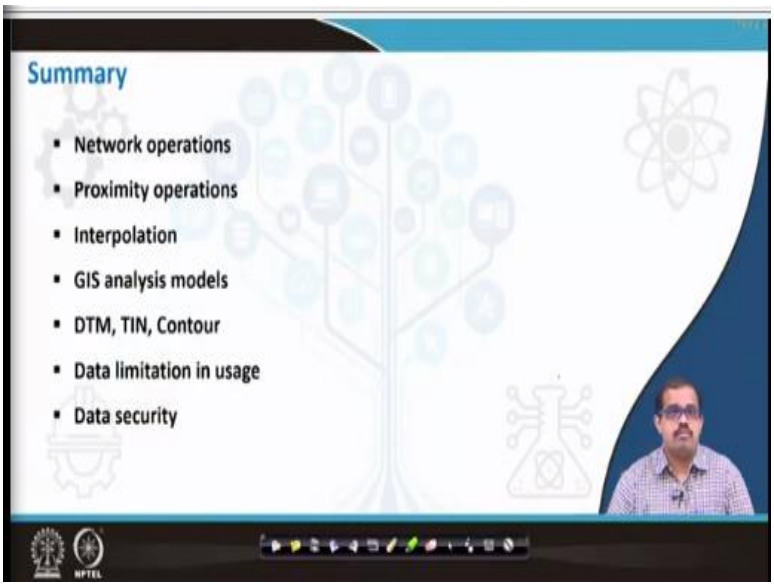
(Refer Slide Time: 31:56)

And the final part is data security, so different countries have it is own regulations. So once a data is crossing the boundary it is bound to define that particular countries data regulation. So in USA the financial services, modernization acts permit organization to acquire distribute substantial personal information unless an individual specifically ask to be removed it from registry ok.

Sensitive data such as maps or border areas, disputed territories, endangered species, rare plants are considered to be of national interest and restricted accordingly. You can see such restrictions even in India, there is a national data registry which can be flexibly used as one of those security tools in when you are looking at data security. You can find out a huge amount of security concerns or rules that have been put forth in terms of data productions.

And in some cases limited information is openly available and combining data from open registers maybe consider to be an registry and security risk, so be careful with this. In future it is recommended to share data in a coded format to avoid tampering. So this is what now the GIS companies are actually looking at. So they are coding out all the GIS data and so that they avoid any tampering of data or misrepresentation of the data ok.

(Refer Slide Time: 33:26)



So this is about today's class, we looked at network operations, we looked at proximity operations finally different types of interpolations. Then we looked at DTM DSM I have spoke about TIN and contour. So I did not look at much but it is also comes under the advance data models. So if someone is interested please look at how the expert system work, how the artificial intelligence can be clubbed with GIS for performing advanced analysis.

Then probably you can look at fuzzy analysis and other systems that are available. So that is what is advanced GIS, then we looked at data limitations in usage how it can be used and finally we ended it with data security. So data security is extremely important in context of data production and data usage. So that is the first and a foremost thing that any data user or any GIS user should look at it when you are looking at context of producing or distributing data ok.

So let us meet in the next class, so next probably 2 weeks or 2 and half weeks will be completely looking at hands on. So we would be using QGIS software, so people who have not downloaded QGIS software please download it now. So in the next 2 weeks we will be completely giving you how that particular software is organized, what are different operations that can be done on those softwares.

How to use a raster data, how to use a vector data, how do you develop a thematic map, what kind of maps that you can prepare, how do you actually present a map to the user, how do you develop a final map. And probably looking at different aspects of it maybe I will give you some examples of interpolations etc. My TAs would actually accompany me in looking at all of these aspects as a practical sessions.

So till then have a nice time, so please download QGIS and install it, so let us meet in the next class, thank you very much.