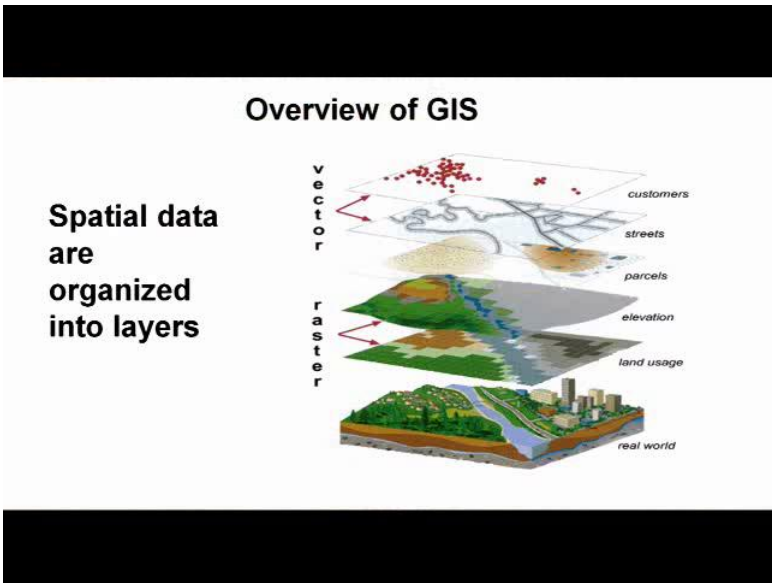
Introduction to Geographic Information Systems
Dr. Arun K Saraf
Department of Earth Sciences
Indian Institute of Technology, Roorkee

Lecture - 03
Different types of vector data and concept of topology

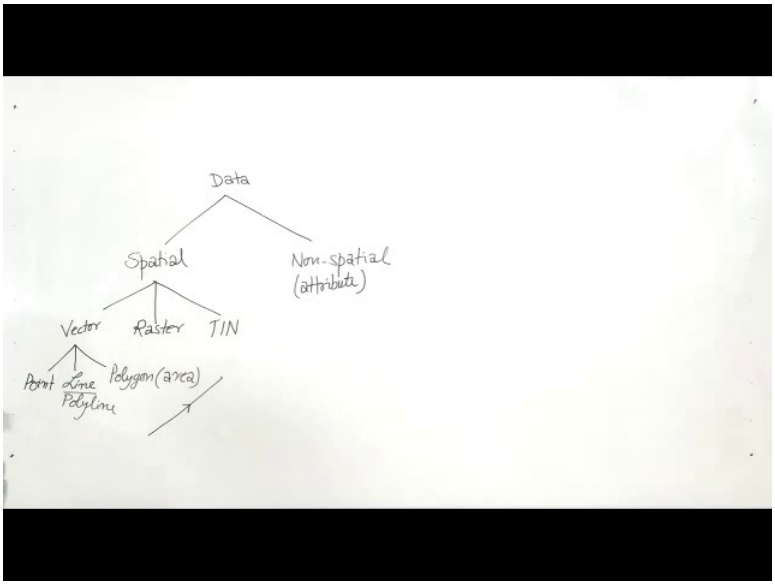
Hello, welcome to this 3rd Lecture in this series. In previous lectures I gave overview of GIS. Also touched little bit about the different types of data which we use in GIS. Now today we will be looking in much more details about the different types of data. And one more topic which is again branch of mathematics which is topology; why it is important, how it works in GIS, all those things we will discuss.

(Refer Slide Time: 01:08)



First about the different types of vector data, if you recall a this diagram which I have also used in lecture 1 and 2 that when you segment the real world into different layers or themes you create different types of vector data as well as raster data. So, first what I am going to do is just to bring these two major divisions in the data.

(Refer Slide Time: 01:28)

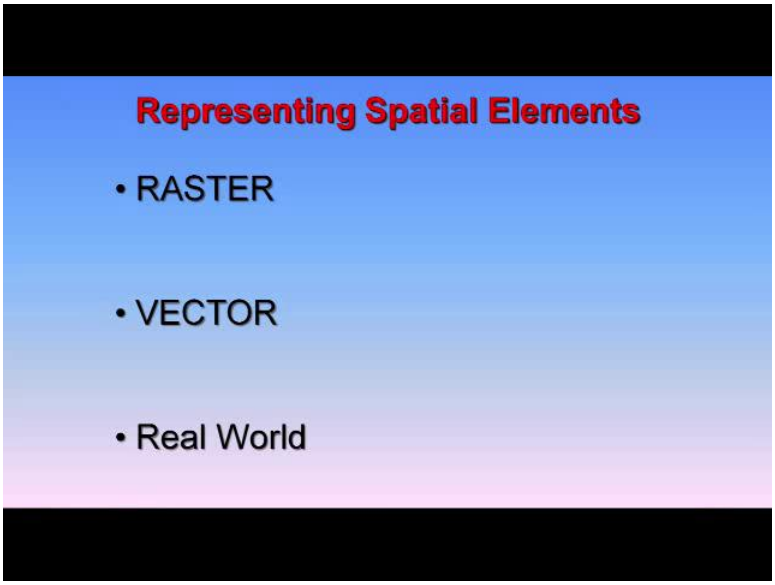


Data in GIS we say this as a spatial data and then we come to the non-spatial, which is our attribute data as well. We can also write here attribute data or in simple terms we have also say tabular data. Now a spatial data can be divided in 3 main categories; one is the vector data again this vector term and the concept of vector has also come from mathematics. We will see different types of vectors as the main topic of today. Then raster maybe in later lectures we will discuss different types of raster's and difference is between vector and raster advantages and raster and vector. And third is all together new different type of data structure which is called TIN; Triangulated Irregular Network.

Now, if we see different types of vectors there are 3 entities are there. But, generally the concept of vector which was given to us in mathematics is you are having 2 points; one is beginning point or origin another one is destination and then you are having direction. So, you are having basically magnitude and direction. That is typical vector. But in GIS we use 3 different types of vectors. The first one is the point and the second one is line or also we say poly line and last one is polygon I we will be defining very soon all this or we can also write area.

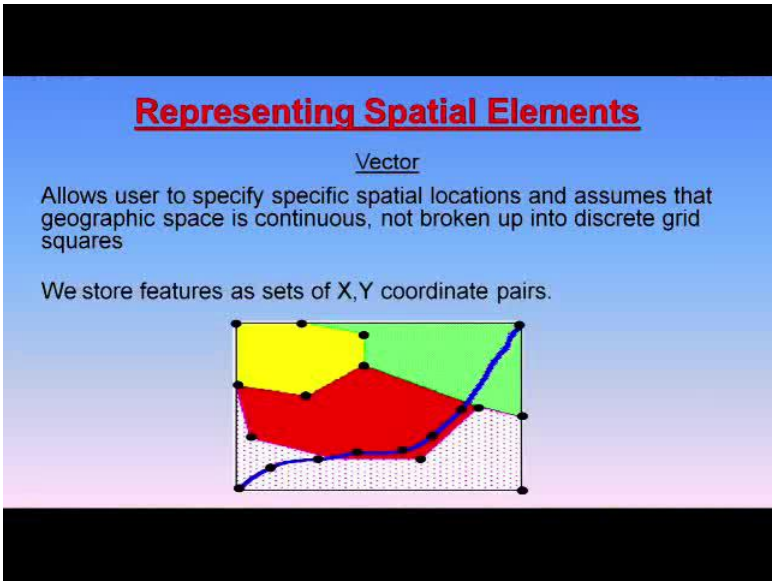
Now, point a is a geo dimensional entity which we will see very quickly.

(Refer Slide Time: 03:53)



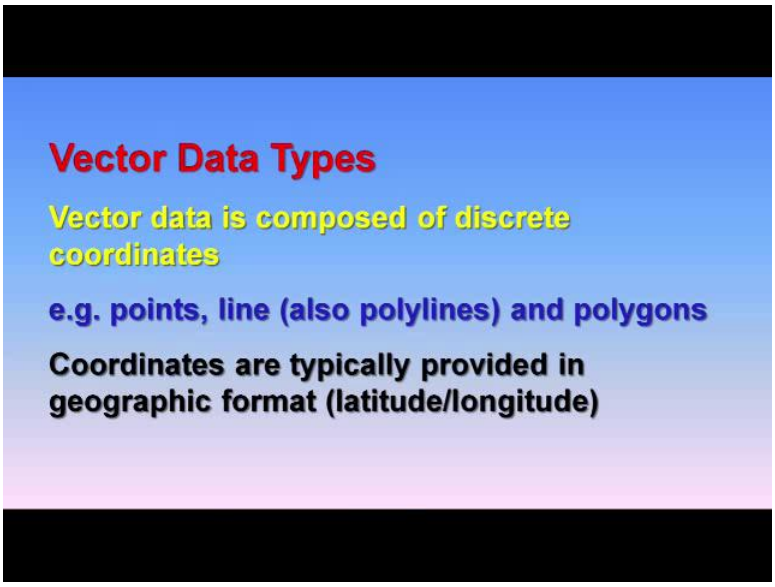
So, the real world which is a repeat of the slide, so now vector data as I have said that it is a very simple way of storing our real world in different layers of vector. So, vector can be point data which we use the geographic coordinates that is x and y. And here in this example 2 different types of vector data are there.

(Refer Slide Time: 04:18)



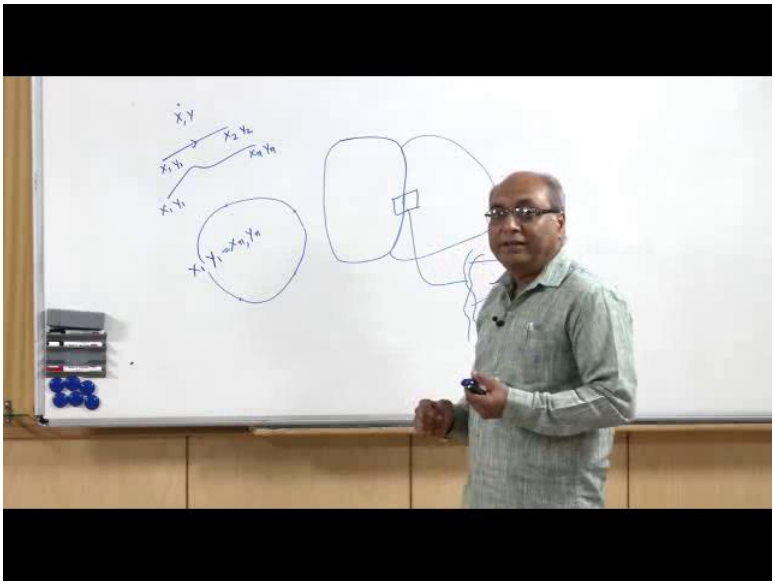
But, remember that a in the beginning I have said that each layer each type of data is kept in different themes or different layer. There are 2 layers which are being displayed here; one is in the background is polygon layer, and another one is line. Point data is not there, but the coordinates or nodes or internodes for each line and polygons are also displays at point.

(Refer Slide Time: 04:52)



So, a vector types what this is composed of discrete coordinates, because we call them vector data is also called as discreet data it is not continuous as raster data which we will see later on. And then I have already mentioned that 3 different vector entities existing GIS and we have extensively used them is point line and polygons.

(Refer Slide Time: 05:21).



Now in case of a point; basically when we say point, point is having just one set of x and y and it is does not have any dimension. So, we say 0-dimensional entity vector entity in GIS. The rest of the data like may be the z value or some other values may go as a non-spatial or attribute value. So, point is a 0-dimensional entity, whereas line will have one origin; x 1 and y 1. If it is the straight line it will have x 2 y 2, but if it is a polyline then it will have a set of x 1 y 1 and here x n and y n.

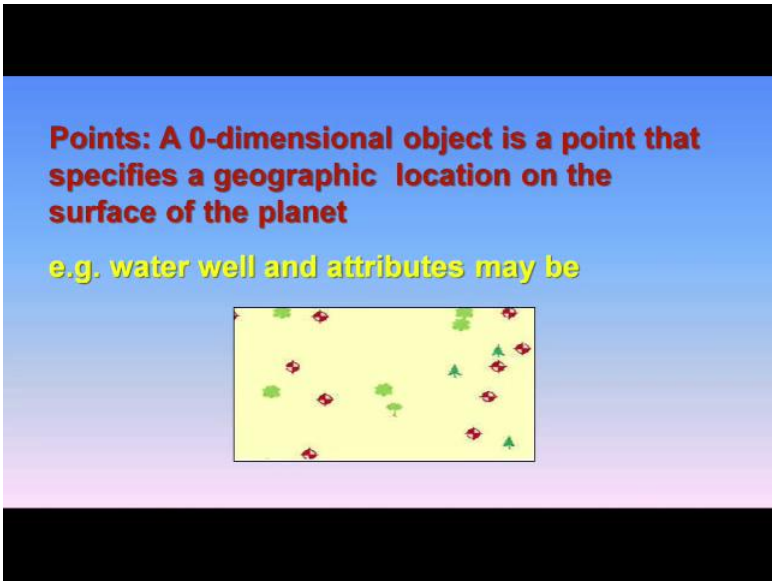
So, this way it is having a series of x and y. In case of a straight line you will have just simple two pairs; begin point and end point, but in case of a polyline you may have various point data which are interconnecting. So, point becomes a 1-dimensional entity and sorry point becomes a 0-dimensional entity and line becomes or polyline becomes 1-dimensional entity. And if we go for polygon then the x 1 y 1 has to be the same or equal as x n and y n, because it is interconnected here because the first origin and destination has to be the same and rest of the points would be in between. Once the origin and destinations are same then you get an area and you can also measure the perimeter and that is why we say 0-dimensional because it does not have the dimensions. So, you do not have an area for point and you do not have any perimeter for the point, but for the line it is the 1-dimensional entity therefore you can measure the length of a line.

I will give you the examples of features which we represent using different vector entities. The third one is the polygon, because the begin point and end point are same and therefore you can have a perimeter and you can have an area. Depending on the scale of a map sometimes in a small scale maps when entire India is being represented then even a large city will be represented as a point data. But only in a larger scale map say 50,000 scale of survey of India topo cities and big city may appear as a polygon rather than a point data.

So, depending on the scale different features may appear differently. And if they are recorded as point then we have to accept though in attribute we may have the area for that city. So, point a is 0-dimensional entity. Only for display we use some dimensions, but just some features or circle or triangle but it does not have any dimension. And that is why in GIS softwares when you zoom a point it does not enlarge, it remains of the same size. And where line is a 1-dimensional entity and therefore you can measure the length like road network, rail network, telephone lines, and power grids everything we represent in a vector. This is typical vector, because of begin and end points you are having the magnitude and then a direction is also available to you.

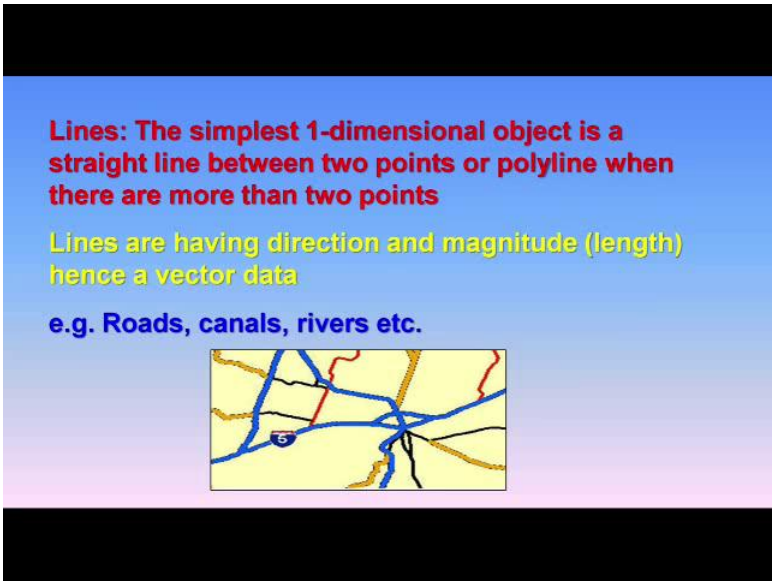
Therefore, in the network analysis of GIS we use the line features. So, that we know what is the direction of flow and how much things have to flow or travel, Whereas, in case of polygon like forest lithological unit, a rock type or a city in a large scale map can be represented as an area. Like agriculture field can also be represented as an area, even in a house in a larger scale map can be represented as an area. So, these are the 3 main vector entities which are handled in GIS.

(Refer Slide Time: 09:56)



As I have already mentioned that it is a 0-dimensional entity and examples sometimes depending on the scale suppose, a water well location is there and pump is there, on a map you display like this using point data. Then line a next simplest object is the 1-dimensional straight line if you are having then only two points begin point end point.

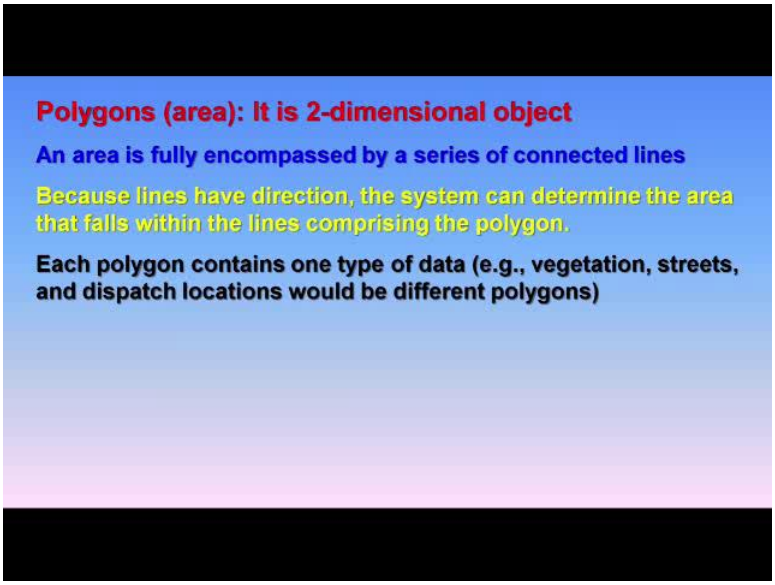
(Refer Slide Time: 10:15)



If you are having polyline then you may have many points. And it is a typical vector so you are having direction and magnitude. Examples are road, canal, river, streams, and other features. Now within a map if you are seeing here there are 3-4 types of line features are presented here, but in computer or in GIS software all these line features are kept in separate layers. So, the advantage is whenever we want to display together we can display and whenever we want to use just say streams or river network we will only use river networks and we will not use road and other thing.

So, that is the advantage of having data in different layers or themes. In case of polygon 2-dimensional object it is encompassed because of the begin point and end point is same.

(Refer Slide Time: 11:09)



Therefore, you can have the perimeter area and you can have other thing. Now there might be some question that supposes in reality we know road though in map it is represented in a line that means 1-dimensional entity. That these we cannot have the width of the road, but in reality yes - we know that a road might be single lined, double lane, third or four lane then it will have an area on the earth.

If you want to store the width of the road then it can go as a attribute as non-spatial data and it can be stored very easily. And they I have also mentioned in previous lecture that

there is no limit for attributes data you can have n number of attributes associated with single point or many lines, one line or polygon. So, there is no limitation of storing non-spatial or attribute information related with any vector entities.

(Refer Slide Time: 12:22)

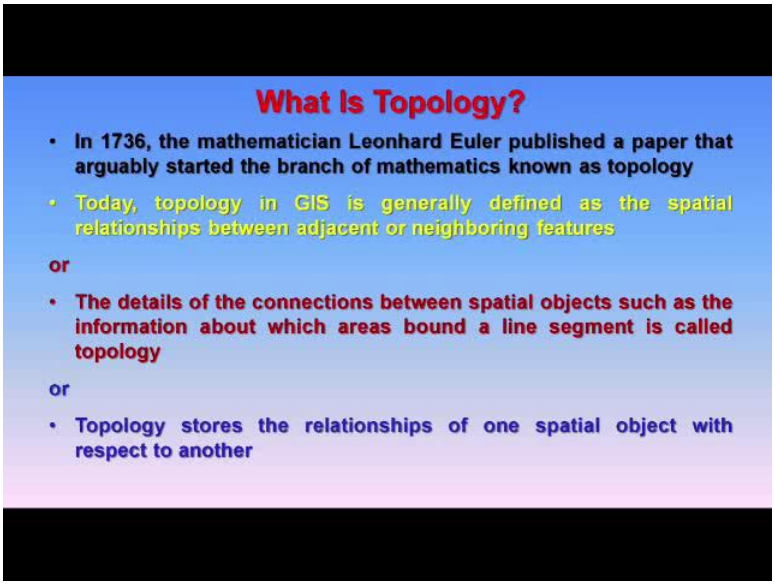
Now another topic which I want to touch upon here about the topology which is associated with vector data sets and that is why I thought that I will discuss first this is one, rather than going for different types of raster, advantages and disadvantages raster, and vector and later on we will discuss then as well. So, topology basically in GIS the neighborhood operation or neighborhood information is very much required. And in order to establish that spatial relationships between neighboring or adjacent features a concept of topology which has come from mathematics is very much implemented into GIS.

For example, if it is a vector line then one should know that what is one the left side and what is on the right side, when we say like in case of a river or canal we say left bank or right bank. So, how you be determined these direction, just because we put ourselves in the direction of flow and whatever on the right side we say this is right bank and whatever on the left we say left bank. Similarly, in order to know the spatial relationships between adjacent on neighboring features topology is required. In earlier version older

versions of GIS software it was later on it used to be constructed topologies. So, first the data raw data will come into the system and topology used to be constructed and it gave lot of problem.

So, now in modern GIS software's the topology is created once you bring the data into the GIS database. And there are different ways of defining topology which I will come very soon. So, using such a data which is having a proper structure of topology it becomes much easier to handle in GIS, it becomes much easier to analyze such data, their relationship, planar relationships and allows us to discover relationship between a different data layers as well. So, topology is must with vector data.

(Refer Slide Time: 14:50)

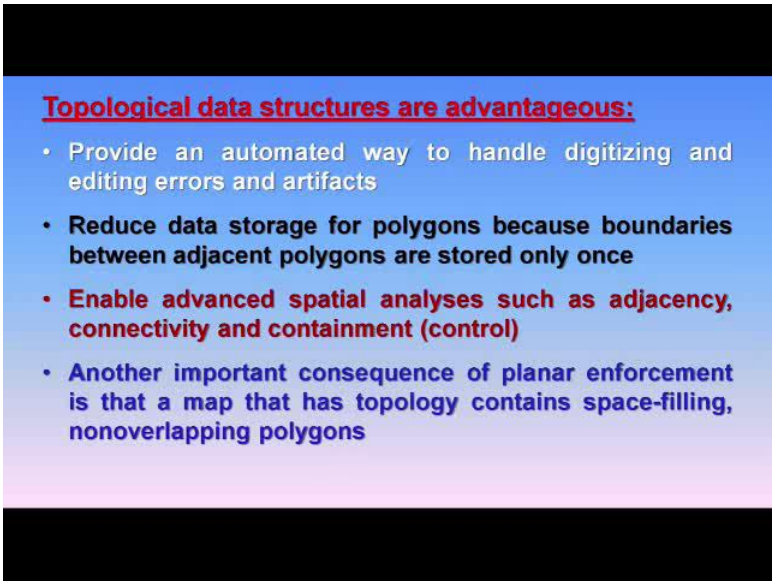


As I have already mentioned that is say the concept of topology was very old in 1936, it was given by mathematician Leonhard Euler who published a paper and then later on it became a complete branch of mathematics which is topology. Topology in GIS is as defined as I have already mentioned special relationship between adjacent or neighboring features. In different books literatures or on net you would find there it is like varying definition of topology, but the concept wise all are same like details of connections between these spatial objects such as information about which areas boundary line

segment is called topology. Or may be a very brief definition of topology that topology stores the relationship of one spatial object with respect to another.

In very simple terms you can think that topology is keeping information about neighbors. In our day today life also this is human tendency we always keep information, try to give as much as information about neighbors. Those are same thing in our GIS science or GIS technology then once the data is having information about neighbors then this relationship later on during analysis can be exploited and some good results can also come.

(Refer Slide Time: 16:18)



So, topology is a structure or data structures are having advantages; provided an automatic way to handle digitization and editing errors and artifacts. Artifacts and digitizing errors let me allow edit here, because what happens that when you digitize or when you convert the analog data into a digital data using a GIS software during digitization there might be some errors which may come in your data. Suppose, I have to digitize two polygons at two adjacent polygons, remember this I am using word two neighboring or adjacent polygon and therefore the topology is very much required.

So if I am digitizing two adjacent polygons, the one way or without having concept of topology I would be digitizing the common boundary between these two polygons once more and like this I will be digitizing another polygon. While digitizing the common boundary between two neighboring polygons twice I am bound to bring some errors. If I say zoom this part what I will find that the one line is going like this another line is going like this, and these are nothing but the artifacts. These we never intended to have these errors or these islands of information or extra polygon, but because of line have been digitized twice these errors will come.

Whereas, if we use the topological concept in GIS then what would happen that we will not be digitizing that common line twice, so what will happen that I have digitized this and the then generally in GIS software what we will do we over step a little bit, we will come and digitize from here and here. And what the system will do, it will remove this extra part and it will develop a node inter node in this polygon at here and will connected. Now advantage here that the two neighboring polygons will share a common boundary with a agreement that if anyone you know destroys own polygon or one polygon is removed then there will not be any harm to the adjacent polygon it is a good neighborhood I understanding. So, if that understanding is not there then what would happen in real life we also see that two people are having a two adjacent land plots and one has already constructed his house.

Now, the wall which is going to be common between another neighbor if they agrees they can share the wall without constructing another wall by the new neighbor, but if they do not agree then what they will do the new neighbor will construct another wall and then in between they will leave the gap and that is called no man’s land, and it is a wastage. In topological concept, because in order to avoid editing errors and artifacts the common boundary between two polygons is shared equally without any problem and that is a good neighborhood understanding.

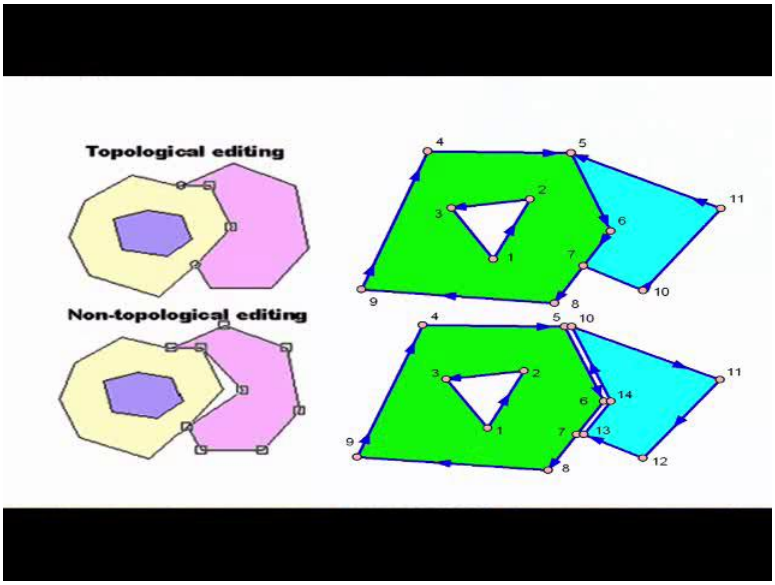
And this is what one of the biggest advantage of topology in GIS. Once a that kind of understanding is there it is very obvious to think that this common boundary has not been stored twice and therefore it will not bring errors in artifact. At the same time it will

reduce the data storage requirements, so the reduction in the data storage for polygons because boundaries between adjacent polygons are stored only once.

This is another very important point, that no human if there is an arbitrary line and a human is asked to digitize the same line twice then it is not a possible for him to digitize exactly the same way. So, like if this is the line and if I have to do digitize again what I will do no matter how to much care I will take I will be doing something like that. And therefore bringing artifacts island of information and lot of problems are will be there. So, in GIS nowadays while digitizing we construct the polygon we construct the topology and keep avoid all these complications later on in our analysis. And once a topology has been constructed which also enable us to advance spatial analyses such as adjacency, connectivity, and contentment. When we come to the discussion on GIS analysis part all these things we will be discussing in details, but if topology has not been constructed while digitizing the data or bringing data from analog format to digital format then these operations cannot be performed in GIS.

Another important consequence of planar enforcement that is the construction of topology that maps that has topology contains a space filling non overlapping polygons. Because if two adjacent polygons have been constructed using concept of topology then these problems will not arise at all, same example here given here.

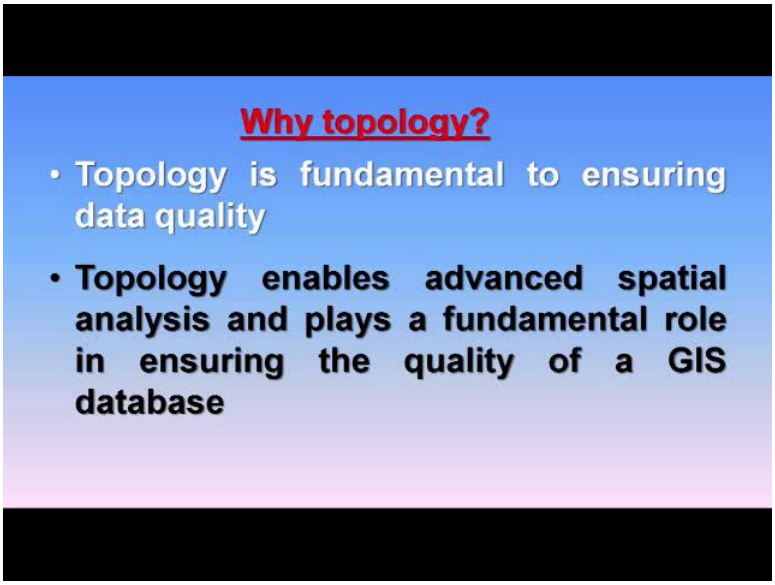
(Refer Slide Time: 22:01)



That here there are software's I can give example like coral draw and other graphical software's which do not have the concept of topology implemented, and therefore if you have to digitize you might be digitizing like this and you may leave some gap here. But, if the GIS software's will never have such problems because while digitizing we will go and keep constructing topology simultaneously.

So, there might be some overlaps, there might be some gaps if we are not using topology but if like this we are using topology then we will not have any gaps what so ever. So, we have understood, but anyway let us go more in details about why topology is fundamental to ensuring data quality.

(Refer Slide Time: 22:57)



That is the most important because remember one thing in GIS error propagates in GIS. GIS cannot remove by itself. This is one of the very fundamental things which one has to remember in GIS; that GIS cannot make your maps accurate if they are already having errors inbuilt or inherent errors. And therefore it is always said that after each and every step in operation in GIS one should check for errors. And if any error is detected corrected immediately, because if you leave that error for correction later on then you will propagate that error and that error which might be a very small error in the beginning can become a very large error in your analysis. If you create an output which is having say accuracy something like say correct 50 percent wrong it is a useless product.

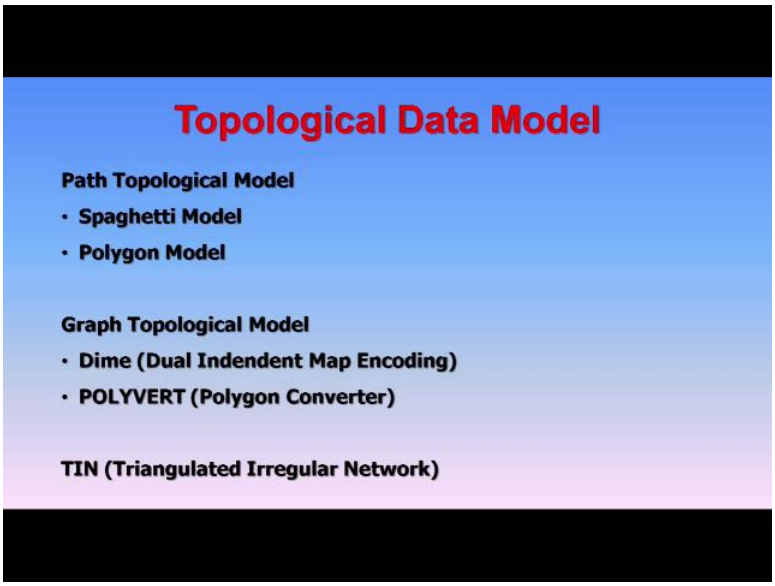
So, it is better always to check errors after each and every GIS operation though it may be a simple digitization analog to digital. So, at that time if errors are checked, topology is checked, everything is then moved further next operation. So that is why to a topology is fundamental to ensure data quality. If data is good then your output through GIS is going to be very reliable, but if data itself is having error then you cannot help. However, there might be some scenarios there we do not know because the data all data I am not going to collect you are not going to collect, somebody else some other organization might have collected and it is always good to know what kind of errors that data is

suffering. So, at least the aim in GIS of that data analysis that we should contain those errors to it is their minimum level.

If we do not know the errors which are coming in the data and we keep analyzing the data, as I have mentioned error propagates in GIS and can bring all together highly erroneous unreliable results to you and that is not the aim of GIS. Aim of the GIS to integrate data in a manner that the results become very reliable, but if it is already having errors, errors have not been controlled within the operations of GIS then your results would have very low level of confidence.

So, topology also enables advance spatial analysis connectivity and other functions, plays the fundamental role in ensuring the quality of GIS database. Now mathematicians have further developed this branch of topology they have brought different topological data models. All topological data models have not been implemented in GIS probably they are not required. But we will very quickly we will see the two major categories of a topological data models.

(Refer Slide Time: 26:17)



One is called Path topological model and which has got two different types; one is Spaghetti Model and Polygon Model. And then another type of topological model is

Graph topological model, some graphical software's have been implemented. These topological models like Dime and Polyvert.

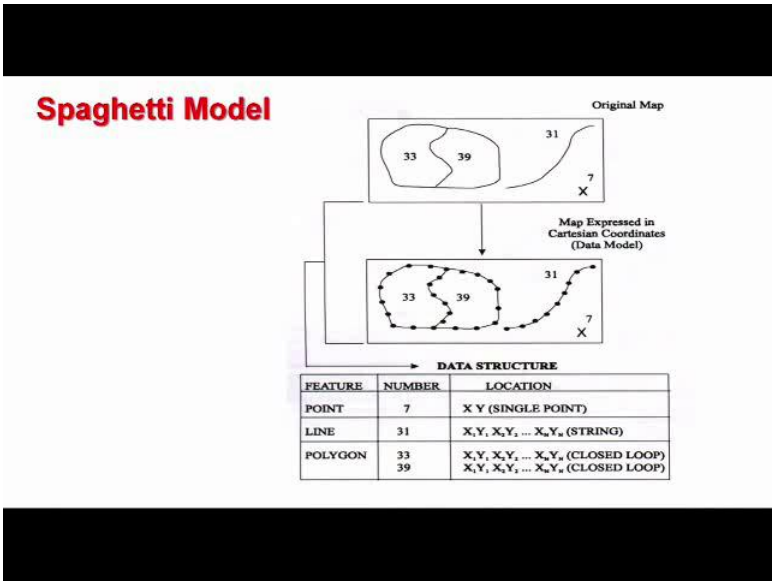
Now, here I would like to spend little bit time on path topological model and there are two different; types spaghetti model and polygon model. The path polygon topological model that is the second in path topological model has been implemented in GIS extensively. The reason is because as I gave the examples of adjacent polygons and therefore this type of topological models are very suitable for GIS. Spaghetti model in which if there is hundreds of line features they are lined in isolation. Let me give you one example in spaghetti in Italian means, these are the noodles which in Italy it is called spaghetti. That means, if you are having plate which is having noodles you can pick a single noodle and no other noodles will have any problem about that.

That means, there is no relationship, no over heading operation, no over neighborhood relationships existing and that is why individual noodles can be picked. If we are having line features in our data, if there is no topology then they do not know the relationship and once there is no relationship then it is not going to work. So, the software's the example is gave like coral draw and many graphical software's which uses the vector data they have implemented only path spaghetti topological model, not path polygon topological model.

In GIS software's this is path polygon topological models have been implemented. I will elaborate further on this, but before that the third type of data which we use in GIS I have also mentioned earlier in the board that apart from vector and raster there is a third type which is TIN; which is Triangulated Irregular Network. In some literature you will find the people have kept that as a vector data some people have kept as a raster data, but in my opinion it is neither vector nor raster. It as to be different data sets, when you will see details about TIN you will also realize that this is neither vector nor raster anyway.

So, we come to the details about first step spaghetti model and how it is stored in the system, and then path polygon topological model.

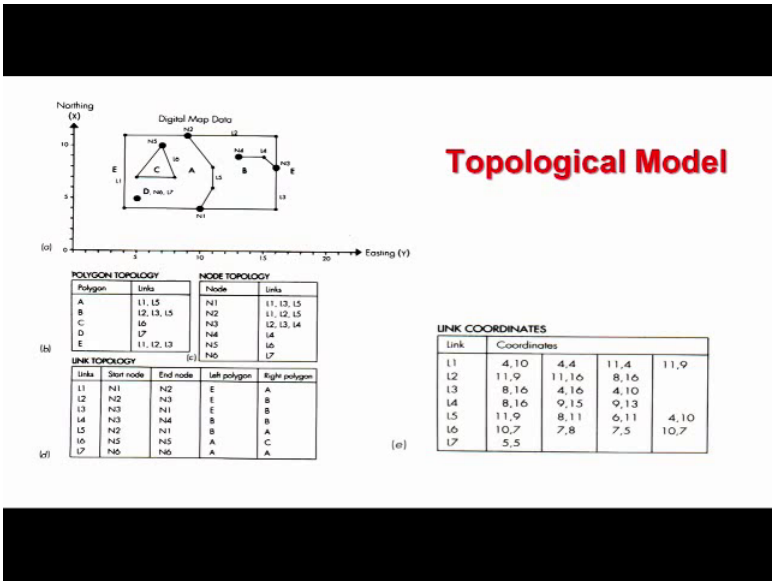
(Refer Slide Time: 29:14)



Here, all three types of vector entities are being stored here as point line polygon and one mappler is simplified map is given here, where you are having a the 7 crossed is marked for the point data 31 id is a line data or polyline data and then 2 polygons are also shown here. If we have to store in the maps then all these will have their nodes and internodes that means the series of coordinates for point data. There will be only single coordinate for line data there will be multiple and then the polygons again there will be multiple. So, this is how in a spaghetti data model this is how it is stored, in a very simple way.

Since there is a no construction of topology and therefore the storage part is very simple, but during analysis you cannot take advantage of that, the neighborhood relationship information.

(Refer Slide Time: 30:11)



Whereas, in topological model almost the same example that you are having one point data, you are having again line data and then you are having polygon data. No in order to store in topological model instead of having just a one table for all kinds of data here now we are having four tables. So, one table is called polygon topology table another one is a node topology, link topology and then link coordinates. If I take the example of a say a first about the node topology then these are the nodes, so N1 will have links L1, L3 and L5. N1 is this one it is having link this L1 which is the outside polyline then L3 which is another outside and the in the middle L5. N2 link two will have the similarly it will have L1, L2, L1 and L5 likewise. Further for each link then the set of coordinates are also stored.

For example for link 1 which is the outside link it is having coordinate as per this system easting and northing or coordinates are marked these are the set of coordinates for link 1. Similarly for link topology now that has to be stored that which nodes connect the link. So, the link 2 is connected with N1 and link 1 is also connected with N2. And see this information about neighborhood who is on the left side, who is on the right side that information is also stored.

And then finally, comes the polygon topology table where like a polygon a which is this one it is having information it is constructed using two links one is L1 and L5. And likewise, information including neighborhood operation has been stored in the system. Once it is stored topology is constructed. And then you can take advantages of this.

So, I come to end of this brief presentation about different types of vectors and associated concept of topology.

Thank you.