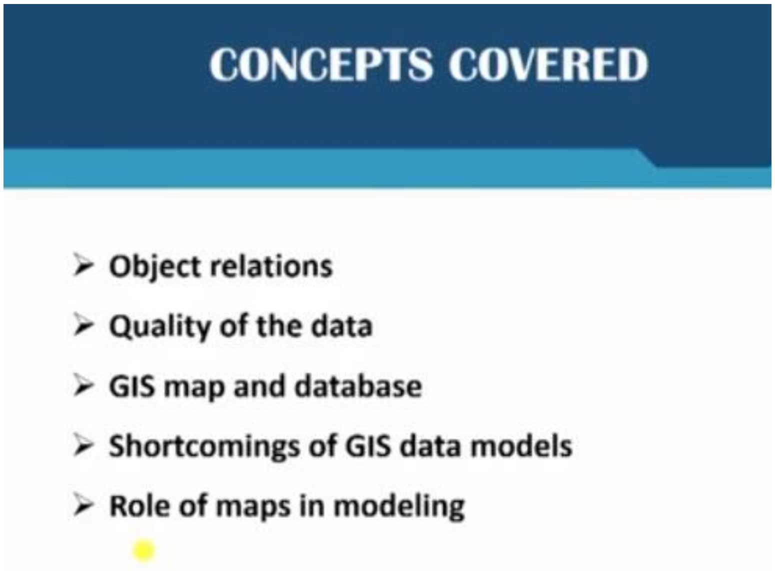
Geographic Information Systems
Prof. Bharath Aithal
Ranbir and Chitra Gupta School of Infrastructure Design and Management
Indian Institute of Technology – Kharagpur

Module No # 02
Lecture No # 08
Real World to Digital World Through GIS (Continued)

Hello and Namaste to everyone welcome back to understanding geographic information system I would be speaking on the same real world to the digital world through GIS in today's lecture but we will be able to understand in this what is how the quality matters what kind of quality we have to have or in a data and what are the different ways of understanding the data.

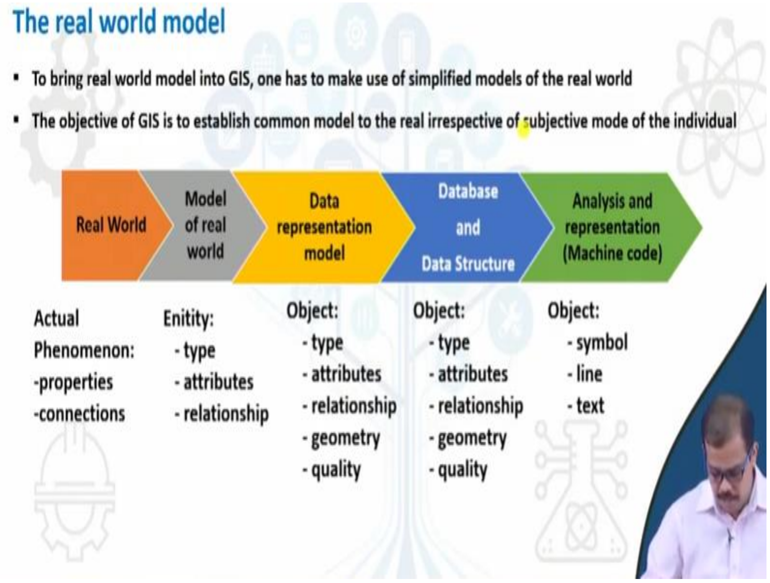
(Refer Slide Time: 00:51)



So when I say when I speak about the real world today I would be speaking about object relations which is same as entity relation the quality of data what kind of data we have? What is the quality of data and what do you mean by a GIS map and the database and map is just about representation. So it whenever I speak to someone so they say we know GIS but the final thing they understand is a map.

GIS is not just about a map it is about a tool it is a process of obtaining a last output so we will understand how GIS a map and how these two are different and GIS is finally all about a data base how do you store data in a database and retrieve it finally we will also see as any subject if you understand there are certain advantages certain disadvantages. So we will also see what are the short coming in this GIS data model whatever the data model that are existing and role of maps in modeling so it is the last part is the map what is the role of that map in modeling and map is not just modeling. So we will understand that in this particular lecture.

(Refer Slide Time: 02:15)

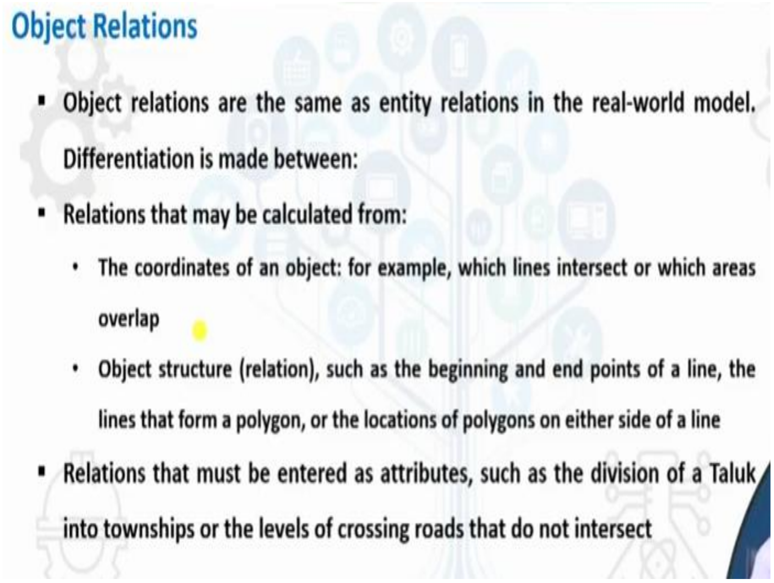


The first thing I would like to go back to my previous lecture wherein I had spoken about this particular representation. So when we look at the real world model into the digital world it would start from the real world then translate as the model of the real world into a data model then once we have a data representation model we have database and it is data structures and finally we look at the how it has to be analyzed and how it has to be represented.

So when we look GIS the main objective of GIS is established a common model to represent a real world irrespective of subjective mode of an individual. So now when we are representing a model we never think about individual this have even spoken about in my previous lectures so it is not just about individual it is a collective thought of individual it has to represent 100 thinking brains, 100 ways of thinking 100 ways of analysis otherwise the model is going to fail.

So it is starts in the real world ends with a data representation so this data representation can be mapped can be statistical analysis or it is just statistics or it can be anything else that is actually representing the output of a process. That is about a how real world model I have just considered this slide to give you a flavor or may refresh you with the same aspects that have thought in the previous class.

(Refer Slide Time: 04:02)



So when we look at objects and relations so we spoke about what do you mean by an object and what do you mean by relations. So when I say object relation this is exactly same as entity relations I am sure every student as understood what do you mean by entity and what do you mean by relation. So now some people use this convention of object relations so object relation is nothing but an entity relation in a real world model okay.

So when we look at entity it is in a data model but when we look at an object it is in a real world model. So object relations is a real world model whereas entity relation is in a real world data model so both are same but the representation is different so differentiation of this is made between relations that maybe calculated from so that when you look at the relations there can be calculated from coordinates of an object.

For example which lines intersect or which areas overlap so this is a very good example of when we look at the road so when we are actually drawing that entire thing as an object as an entity so we will look at is which slides intersect. That intersections become a very important point this can be also a very important point in terms of a reference point of a frame of data I will speak about frame of data in my probably in my next lecture now it can be a frame work data.

So it is extremely important to understand this as an relations then object structure again when we say objects structures is nothing but a relation such as a beginning of a i mean it may be beginning and the end points of a line. For example if I take the entire road structure beginning of a road to end point of a road that road only that section okay. So that lines that form a polygon or location of polygon on either side of the line that is very important.

I give you an example of a church street in my previous class wherein I explained how particular road is there are number of sections there is a intersection and there are huge number of polygons on both sides of the roads that have to be captured. So these are objects and the relationship the way it interact is nothing but a relationship this objects are in the real world. So when you convert into a data model they become entity and these entity will also have a relationship as that of an object.

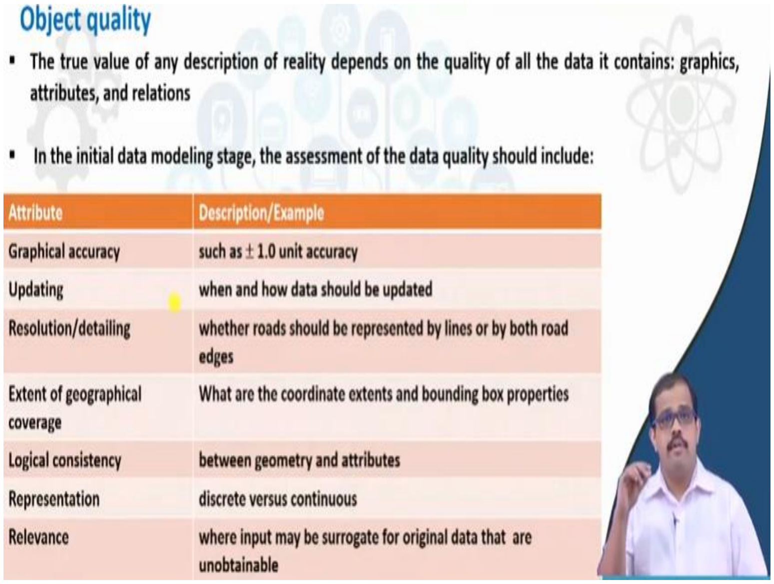
So now object relationship is equivalent to an entity relationship that is how an real world model is now converted into a database model or a data model. Now relations that must be entered as attributes so you have certain relations that has to be entered for example when you are putting out a division of a Taluk into a township or levels of crossing of the read that do not intersect. So when you are giving a division of a Taluk so you have into a township you have to give the certain details as attributes how it has been divided and what is the area of that particular division what is I mean there are lot of attributes that you can give.

Similarly when you are looking at the road what are the level of the road that do not intersect. So these is a very important database for example if you are looking at a Google there are so many roads which may not intersect but when you draw a line segment for example if you have fly over underneath a flyover there is a intersection so you have to show case a two line segments at least which wherein one line segment is representing a flyover which as no intersection whereas other line segment which is actually representing the road which has intersection.

So this becomes an very important aspect when you are looking at certain databases that has to be built. So if you let say it is just a polygon which is actually showing a road that is not intersecting let us say someone want to drive to that particular intersection he or she will not be able to understand until they reach the top of the flyover that intersection road should be reached to the underneath of the flyover.

So that is extremely important that is why you should have certain attributes which are representing that for example the attribute here as the flyover take a left or keep on your left keep on your right so all of these becomes and attribute and the flyover location fly over starting ending all these becomes an attributes. So always certain relations has to be entered as attributes so if it is entered as attributes then only your database becomes very significant.

(Refer Slide Time: 09:08)



Now when we look at an object quality so always your data output depends on the quality whether it is on the object quality or in entity quality. So you have to look at the quality in terms of what is it true value of description of that really I mean reality that depends on the quality of the data. So when you actually representing it in a form of a graphics or an attributes or any relation you have to very clearly describe that what kind of true value it actually carries.

In for example if you are trying to generate a data in your modeling stage the assessment of data quality always has to have certain aspects for example what is a graphical accuracy of that particular data what how it is been updated is it updated today is it updated temporarily which means to say every 10 days every 1 year every 2 year every 5 year every 10 year what kind of temporal accuracy has?

When you are modeling certain phenomena let us say urbanization so when you are looking at urbanization you let us say you have a data which is in 1970’s and 2010. So you just look at two data and you say over this is the huge growth of urban area that cannot happen. So you have to look at what are the different I would call it as agents which would have called for example industrial area would have set up in certain regions they would have been employment opportunity they have come up suddenly they are all sudden government issue that would have come up there are N number of issues that can be counted for an urban growth.

So if when you look at each of this issue they are associated with certain range of year so if you can collectively understand what is the range that you have to consider so you would be able to understand what is the growth of the urban region over a period of time that is exactly why need a temporal accuracy otherwise just capturing to the years you would be actually giving a mis information of about how the growth as happened.

Let us say of there if let me take an example small example of any city for example if there is a industrial setup in 1990’s then in 2000 it is all boom of a IT industry then in 2010 you may have same city you would have seen much big industries coming up and education institutes coming up and by 2017 or 18 it has become a hub of both educational institutions, economic hub and also extremely culturally and a mega polygon basically.

So when we are looking at this entire phenomena 1990 so I gave you in 2001, 2010 and 2017 so now if I have a data which is almost 6 years or 8 years or 10 years that is capturing the entire phenomena of how the change would have happen. I would suggest if it is 6 to 8 years when you are looking at the entire dates so it would give you how the urban growth happened over a period of time that is why it is very important.

Now let us say if I am looking at just a pattern of growth how the entire city is growing then I may not need a data which has exactly at a building level or even when I am seeing a foot path or even a car on street so I need a data which is giving a overall representation of the city. So this is where resolution comes into effect if someone is trying to understand what I mean what is the it may be the amount of space or the greenery that is there on the ground it may be amount of space that is pedestrian space that is there on the ground or it is maybe what is the distance that has to be calculated or what is the height of the building size of the building such cases you need to know the data that is to the building level or to even to the level of pedestrian.

So then you need a data which is which has very high details so now resolution also which comes a important part how you capture a data? What kind of data you are having you cannot just give a urban data if you are trying to process very detail resolution data in order to capture the urban area it may not be possible because you need such high computing part also and also it does need a very qualitative and sophisticated way of understanding otherwise it is not possible for understanding in most holistic way.

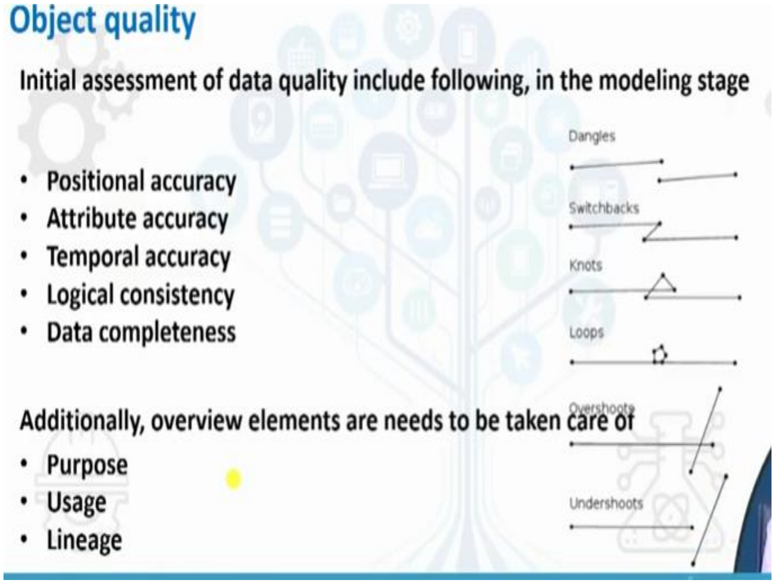
So looking at what kind of data is needed what resolution is needed (()) (14:13) important and extend of geographical coverage. So what is the size of the data that you need what is the geographical coverage that you need whether it is located only to whether you are trying to understand only the Kolkata part of the region or whether you are looking at much bigger region wherein 4, 5 districts of the entire region as clubbed on you are looking at a KMA part of it or whether you are looking at KMA under a buffer region of the maybe 10 kilometer, 20 kilometer.

So look at what geographical locations you are so that tells you the accuracy because that is the area that you will be looking at. So if it is too bigger area a medium resolution or when I say medium course resolution let us say not very detailed resolution to very detailed resolution can be used when you looking at a data which has need to know the very specific then you need a detail is a very detailed resolution and extent of geographical location would be a much smaller.

The logical consistency so between geometry and attributes that is extremely important then representation. So how do you represent a data whether it is discrete data or whether it is a continuous data I will come to this what do you mean by discrete data what do you mean by continues data I will tell you with the examples. So it that also has to be there then relevance this is where most of the people use GIS tools for do not address relevance at all. Relevance is like either input maybe a surrogate of the original data that are unobtainable.

So you should look at whether they have a relevance to whatever thinking that you have whatever the output that you have if it is irrelevance never include that particular data okay. so these are the things that you have to look at have an object called they are many other many more things when you start working you will be able to understand what are the other things that has to be considered and should not be considered but as of now these are the major things that you would have to look at when you are looking at an object quality.

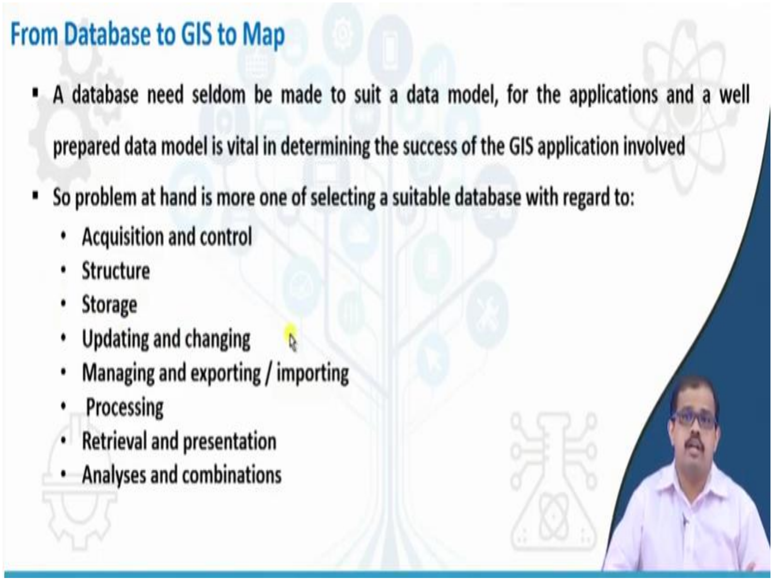
(Refer Slide Time: 16:25)



So and when you look at object quality in trying to put up the data so you need to look at the positional accuracy also the attribute accuracy, that temporal accuracy, the logical consistency and data completeness. So when I say data completeness is what I am referring to this okay I spoke about the dangles, switchbacks, the knots, the loops, overshoots, undershoots okay. I will also explain it with an example rather than explaining it here in the next slide but just understand that these are some of those errors that you may find in your data models that you would have created.

And additionally overview elements are needed to be taken care of for example purpose, usage and lineage how the data has given as a lineage it is how the data has evolved over a period of time what was the source and how it is evolved over a period of time okay.

(Refer Slide Time: 17:27)



So how do we look at from a database to a GIS map a database need seldom to be made to suit to a data model. So you cannot have a database which can does not suit a data model it has to suit a data model for the application as well as prepared data model is vital determining the structures of GIS applications that are involved. So when you are looking at this you have to look at suitable databases that have certain qualities.

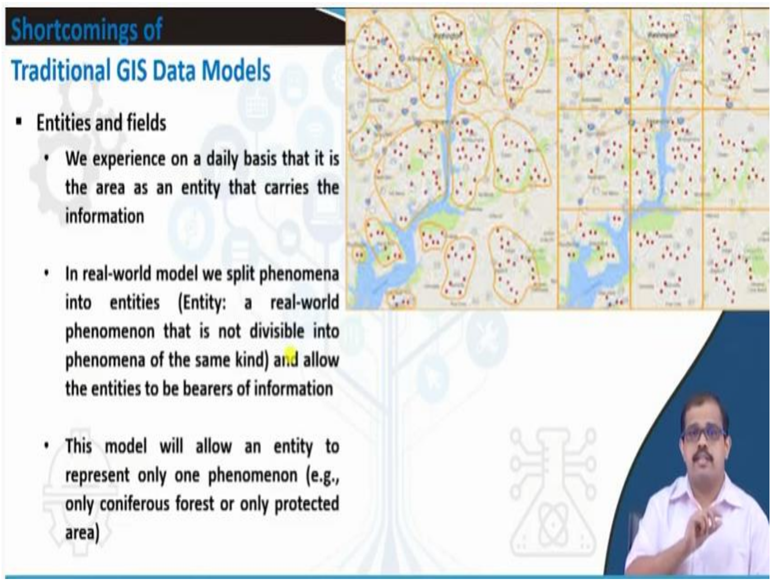
So when I say suitable databases for example how the data is acquired and controlled so that is where the first thing comes how you are acquiring the data at first so what is the source of a data. So that is the first way of looking at the suitable database then structure how do you want to store it? Whether it is highly structured database whether it is complex database it is a simple database so look at it structure then how do you store it? What kind of storage is needed?

Then updating and changing how often how frequently this is updated and stored and changed and managing exporting and importing what kind of management would happen what is the exporting criteria? What is the importing criteria? All of this stuff as to be understood in a better way in order to suit or look at what kind of database we will look at databases so when we are looking at databases you will get you will understand why a particular database has been chosen for particular purpose.

Then you have processing so what kind of processing is needed? How it is needed and if that processing is necessary for your particular purpose or is processing ever necessary for any user of that particular database so we have to look at that. Then retrieval and presentations if you many of the GIS model have huge number of ways of retrieval and presentations which are absolutely not necessary.

But there are certain things that are absolutely necessary so maintaining that could be more suitable instead of having every way of representation and presentation. Then analysis and combinations so this is very subjective and used in how it has to be analyzed each user as his or her own way of look analyzing it or combining the data so that is how that depends on user and probably it is not up to the looking at for entire group of users.

(Refer Slide Time: 20:11)



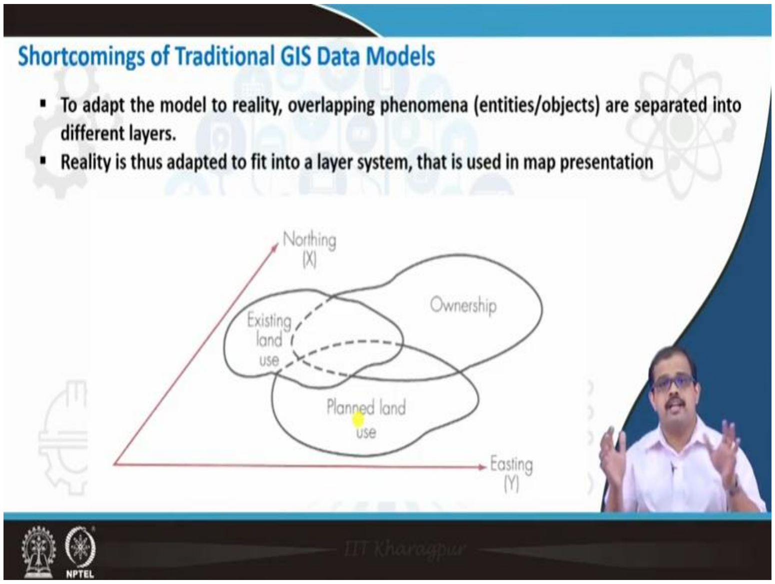
Now as I said every subject or every science has its own short comings of how it has been used or the way it has been used. For example let us consider entities when we are looking at entities in a real world if you look at it the entire area okay entire area that is a carrier of information okay. For example if a real world is considered maybe let us say a building with surrounding vegetation or car parking area etc., you have so entire area is nothing but information okay it gives you can information.

So those are carriers of information but when we look at the real world we do not split phenomena in different parts for example we do not say that this building is nothing but line say this particular trees nothing but a point we never do that we say it is a tree this is a building. So when we are actually creating a drawing of it has the complete area is represented in a (())

(21:22). But when we are looking at the model we represent each of this entities in its own ways if it is a vector model so we represent trees if the form of a points we represent roads in a form of lines we represent probably the building in the form of a polygon.

So this is what is one of the shortcomings of an traditional GIS we cannot really mimic the entire phenomena of a real world into a digital world but it can be a representative factor in terms of a real world to a digital world okay. So it is not a exactly similar okay.

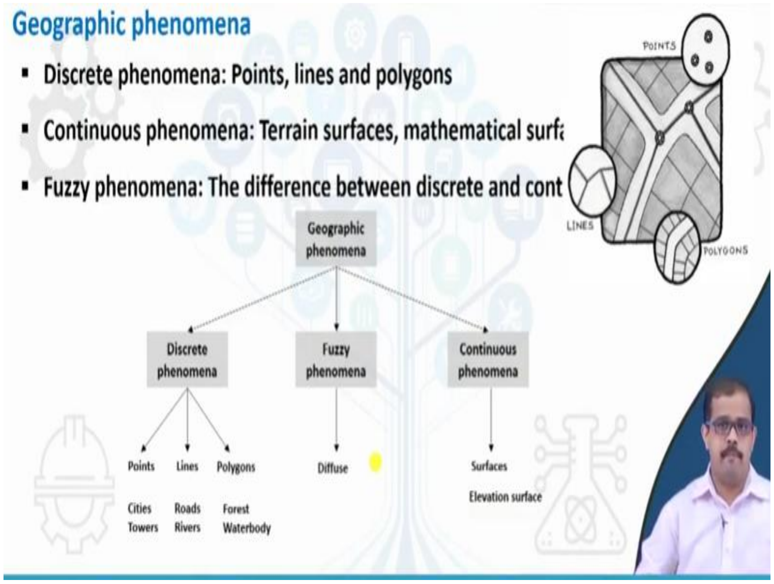
(Refer Slide Time: 22:09)



Now if we look at other aspect for example what I have shown here so there are certain overlapping phenomena's okay. So when we are looking at this overlapping phenomena's in the real world it is entirely adopted to fit to that system okay so you have an overlapping area you have for example there is this is one of the land use. The ownership belongs to a certain place and there is an existing land user it is already there which means the planned land use this is the only place where the ownership of this there is particular region of or may be a person is available for developing this planned land use.

So now if this has to be represented in your database so what happens is that this becomes one polygon this becomes the second polygon and this becomes the third polygon. So now all these three are different layers that fit into the entire layer system okay then represented as map and once it is represented as map you overlay it an show it as one of those systems but in a real world it does not fit as different overlay structures but it is just a connected system.

(Refer Slide Time: 23:34)

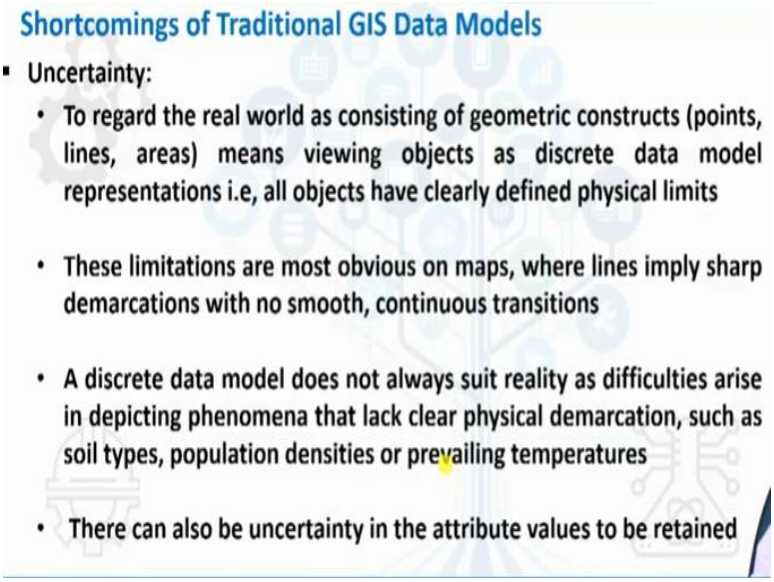


So that is one of the draw back and similarly when we look at geographic phenomena. So when as I said geographic phenomena can be discrete can be continuous okay but when I said discrete phenomena you will represent it in the form of a point and a line and a polygon but when you are looking at continuous phenomena it may be a terrain surface it may be a mathematical surf. So when I say terrain surface it is a raster basically a raster so and when I said discrete has been basically a vector model.

So there may exist in a real world which certain spaces which are difference between discrete and continuous which are in between discrete and continuous or which may have characteristic of discrete and a continuous these are nothing but a fuzzy phenomena. So this is where the data models representations falls short the fuzzy phenomena is extremely difficult to the represent in a data model.

So that is one of the drawbacks of the traditional GIS way of looking at it but with recent literature people have overcome they have looked at different sciences in understanding how they can use this fuzzy measures to represent both the discrete and continuous together has been developed maybe last 2, 3 years.

(Refer Slide Time: 25:10)



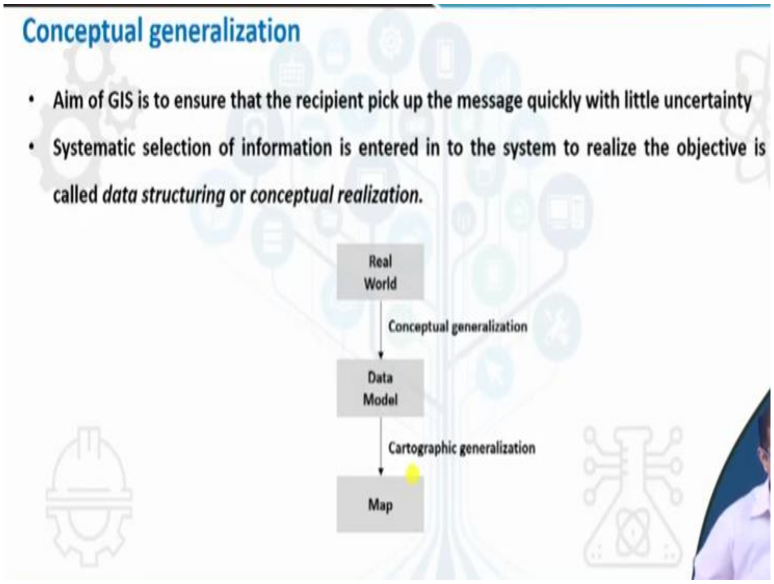
So another shortcoming of a traditional GIS model is uncertain now to regard the real world of consisting of geometric constructs for example if it is a points, lines and a polygon which means you are actually viewing the real world as a discrete data model okay of representation all the time that all objects are clearly defined by the physical limits. So when you have a physical limit as is said you have a boundary then you start limiting your space that only to that region.

These limitations are obviously on maps whereas the lines imply there are sharp demarcations which no smooth, continuous transitions if you draw a map let us say you have drawn a map of Kolkata city so but exactly next to it there is a another district which is actually overlaying on the city part so it is continuous phenomena but what are trying to do if you are actually demarcating Kolkata city you have a sharp demarcation between those two places which actually makes it a difficult to understand the continuous transition it will be a discrete data all the time.

The discrete data model not always suit the reality as difficulties arise in depicting phenomena that has physical demarcation such as it may be soil types it is extremely difficult to demarcate soil types then population densities prevailing temperature. So if you have a ward if you trying to represent population density in a ward you can do it. But if you are not trying to represent a ward wise and if you are trying to make it more smoother so that the representation is much better it may not be possible.

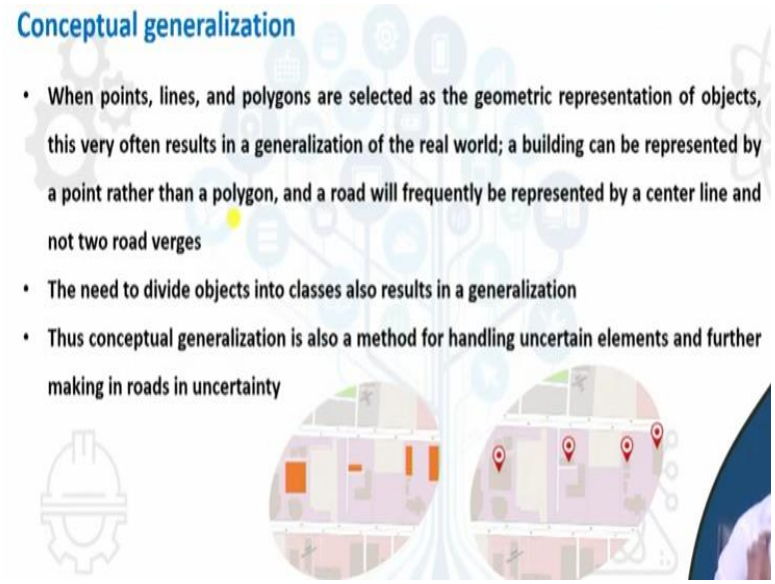
Prevailing temperature is another good reason of how the representation may go wrong there may be also uncertainties in absolute value to be retained sorry attribute value to be retained so these different attribute values also can have certain issues.

(Refer Slide Time: 27:22)



So when you are looking at generalization it is a real world conceptual generalization terms of concepts to the data model than we look at cartography generalization that is in (()) (27:37) the form of point line and polygon or different representative factors or different cartography instruments and finally put it on a map. So that is how the conceptually generalized the entire GIS part is generalized. So with the generalization has what I spoke about the limitation in the last 3, 4 slides.

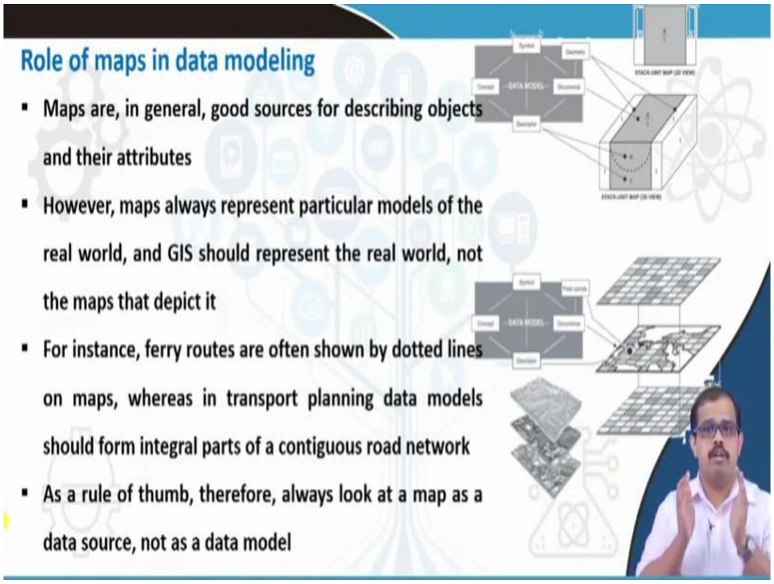
(Refer Slide Time: 27:57)



So when point lines and polygons are selected as geometric representation objects they often results in generalization of a real world. Building can be represented by a point also as a polygon and the road will be frequently represented as a central line and not the 2 or 3 lane or a 4 lane road that is existing it is just one line that will be represented if you seen in Google map until and unless you zoom in you will not be able to find out whether it is 2 line or 3 line or a 4 line or 6 line or 10 line.

So it is only represented by a single line so that is where the generalization happen the need to divide objects into classes also results in generalization everyone who create a database always create it in a (()) (28:51) class. So that objects dividing into classes will create a huge generalization when you are actually looking at conceptual analysis. Thus conceptual generalization is a method of handling uncertain elements very uncertain elements and further making the road I mean any of those data very uncertain.

(Refer Slide Time: 29:16)

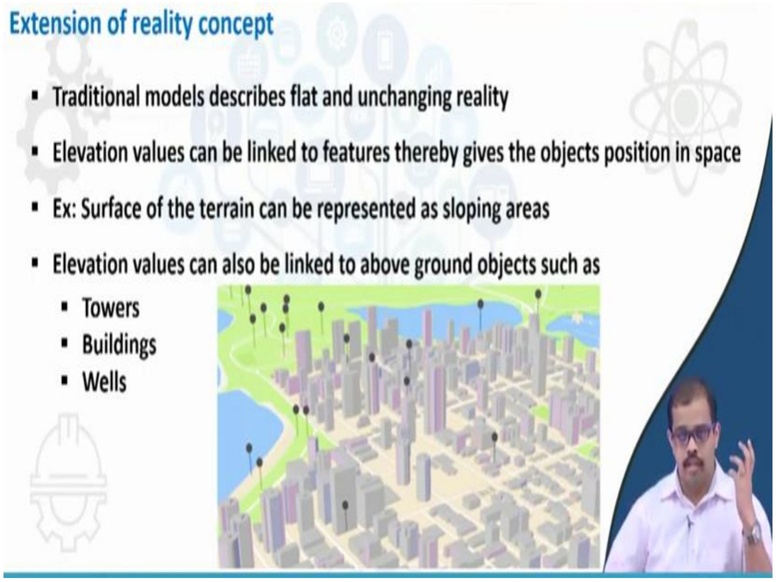


So you have to keep this in mind when you are looking there are certain limitation but it can be always looked away with it and when you are maps as I said they are good source in describing data but they represent a particular model of the real world and GIS should represent a real world and not maps depicted. So as I said GIS is a tool it is a science, science behind the tool should be representing the real world and not the maps.

So when we look at the ferry routes for example that is what I have explained here or these are often shown in dotted lines on maps whereas in a transport data models they represent the route in a proper way as a integral part of contagious network. So that is extremely important when you are representing a proper GIS model. So as a but in a map it may not shown that is what I am trying give you an example so GIS data and the map data two different aspects map data is just an end point whereas GIS is as science.

So when you look at the GIS data with geographic information then it would be much different from a map data. So always look at a map as a data course and not as a data model it is only a source or it is just a visualization aspect not a data model so people try to put in that map is always a data model no it is never a data model. It is just a source or an end point you have a data model in between creation of a data to that.

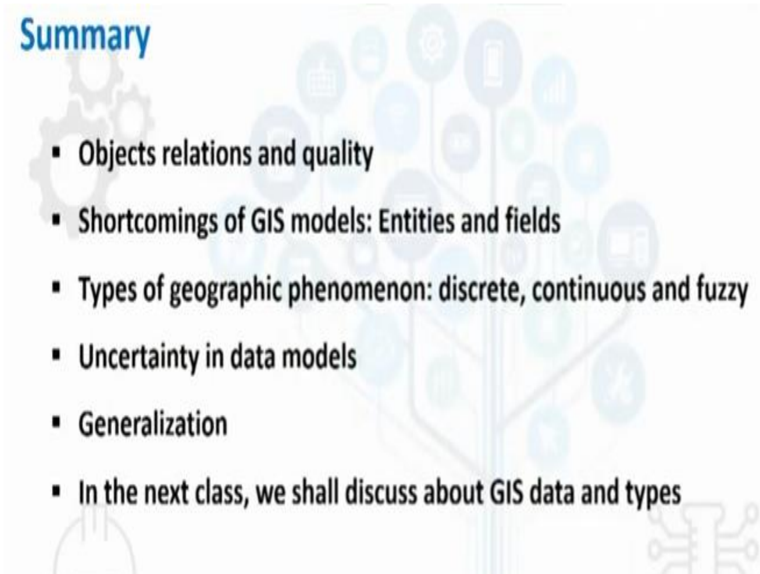
(Refer Slide Time: 30:59)



So when we look at the reality concept so you have elevation value for example to upper ground objects such as towers, building, wells all of these aspects but when you are looking at a reality concept that traditional models are describes the unchanging reality okay but your concept when you build up a GIS data should always consider a dynamic way of changing of any of the information on the earth surface.

So this is the extremely important so when you are actually building a models you have a I have a understanding how the dynamic process happens so if you can build a dynamic model which can actually input a real time data then it becomes a efficient data model in terms of handling data any kind of a data. So it may not be a specific data it can be any of data that can be handled.

(Refer Slide Time: 31:54)



So to summarize this class we looked at what you mean by object relations or the entity relations and most importantly the quality. The quality and the data quality that was there the data quality that has to be there in terms of handling of data we looked at shortcomings of GIS models as entities and fields. We looked at types of geographic phenomena 3 types discrete, continuous and fuzzy how discrete and continuous is represented whereas fuzzy is extremely difficult.

We looked at uncertainty in data models in various ways all this uncertainties as addressed when you are looking at the data model. So look at the data model in a much rather than looking at in a traditional way we will look it in a holistic way many of these issues can be addressed when you are actually converting the data into a data model. Then you have generalization so always when we are trying to create a model from the data we create generalization of the particular data set.

So first it is conceptual generalization then we look at cartography generalization so both of these have to be addressed in a proper way so that becomes a limitations of your model. Then in the next class we will look at what are different GIS data types now until now we have understood the concepts of GIS so we have not entered what GIS data types are there I am always explaining you about what is the vector what is the raster etc.,

So now we will get into that mode of understanding so we will smoothly transition into what do you mean by vector what do you mean by raster how it is represented what are the different models that are represented. So may be next few classes may be 4, 5 classes I will transition into that kind of modeling and types of data representation. So thank you very much let us meet in the next class.