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 # 06
Real World to Digital World Through GIS

Hello and Namaste I am back here with the module 2 I will be speaking about how do we actually convert a real world data into to a digital world. So when we look at real world phenomena quite complex and the way it has to be represented has to be taken care of various issues. So this module is mainly looking at how do we handle the real world data how do we convert it to a digital world data then finally how does system or a GIS or geographic information system actually looks at this data.

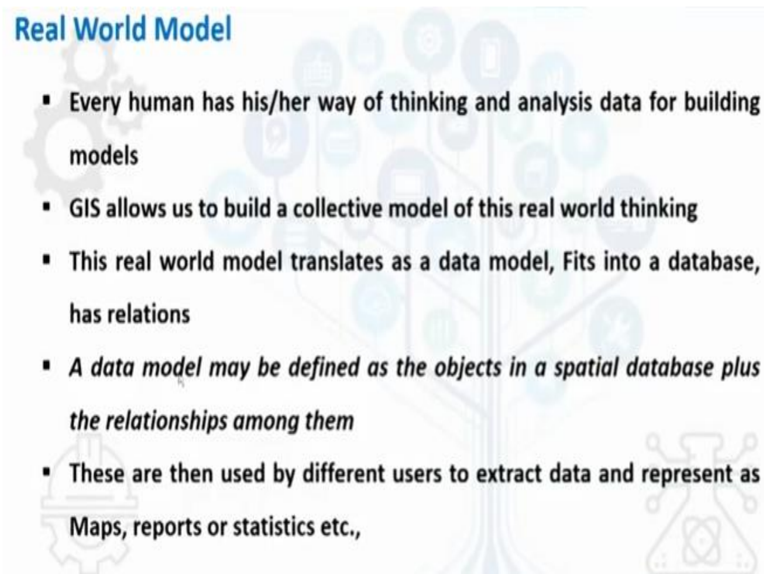
So that is what we have learned so when I am looking at this module the previous module was completely a introduction as a subject but this module is basically giving you of flavor of how GIS works. So this is the second part of the module which actually gives you a flavor is how to start working with geographic information system.

(Refer Slide Time: 01:27)

So let us look at the concept that will be covered in today's class it would be I would first look at what do you mean by a real world model then look at principles of real world model wherein I would give you an examples of how real world is how and what kind of modeling techniques we have to use. Then we will look at how do we feed the real world information in terms of digital world.

When I say entities here everything that is in the real world that can be represented in a digital world is nothing but an entity. So entity as I said can have huge number of attributes in the previous week we have learned that attributes can be both spatial and non-spatial. So entities can have huge number of spatial and non-spatial attributes we look at this then we will also look at how entities attributes and relationships are built what do you mean by relationships and how each of these a specific qualities are then linked.

(Refer Slide Time: 02:28)



So when I get into the real world model for example let us say that there this room as several human beings several students are sitting here. So when we look at each and every student they have a own way of thinking own way of analysis of data for building a models if I have to build a model if I have 100 students sitting in this class each of the 100 students will have the own way of putting it forward.

If I have looking at the urban model the economist will have the economic inputs into the model when as a geographies I have their own way of looking into the model whereas the (()) (03:09) could refer it to in a different way. So thinking the capability and the way that it is interpreted is very different so when I have to look at any real world phenomena this has to be taken into consideration and this is foremost important when we are looking at real world which means to say that we have to capture every information of every one of these otherwise your model may fail in the context when you are applying it to a real world it may work in your lab but when you go out and apply it on the land it may fail to a very large extent.

So then can GIS handle it yes of course GIS with a multitude of tools the science that is built behind this tools can handle the collective model real world thinking when I say collective model it is everyone's thinking can be captured only thing is that users should be able to interpret every detail of that particular model okay. And moreover when we are building a model we should start thinking independently than looking at only domain specific we should not remain domain specific but look at more aspects of how everyone perceives that particular model to be represented.

If let us say we are trying to capture may be how a particular series is growing so we have to look the physical aspects we have look at the economic aspect we have to look at how if there is an industrial area that is coming up how it as actually benefiting in all the aspects and whether physically can come into that aspect and if it has this structure what kind of structure it actually can have if it has this kind of structure will it affect the environment or through a emission of certain gases.

So all of these as to be taken into consideration if the real world model as to be built perfectly. So that is very important when you are looking at GIS is a tool GIS is a system GIS is science but the way the data is fed into the system that is where it produces output. If there is the data is selfish junk then it produces a output that is completely garbage. So if the data is real good quality data I would also speak about how the data quality has to be assessed.

So if it is a good quality data with certain way of collecting the data then GIS can give you extremely good results. So the real world model can translate as a data model so when you are looking at digital world it is represented as a data model so each and every aspect is converted as a data and this data model sits in a data base okay. So this fits into a database and it has relations so now the first thing is you have a real world if you putting it into the digital world then it converts it as a data model.

So when I say data model for example if we have 10 rooms in a particular building of you want to represent all the 10 rooms in GIS. So every room is an entity so please understand this every room is an entity okay so this entity fits in a database with a specific ID okay so each ID will be linked for example the size of this room the capacity of this room the lighting in this room the video visual capacity of this room all these are qualitative quantitative data the spatial non spatial data so these are attribute information for this particular room which has a particular ID.

So this attribute information can be linked to your entity through a relationships that is what I means by a real world data converting into a digital world through a data model which fits into a database and has relations. So please understand every dataset should have which is a being represented from a real world should always have a data model should look at database relations and each of these things which are trying to represent will be an entity right.

So when we look at a data model it may be defined as objects in a spatial database plus the relationships among them that is what is called a data model fine. So when we are looking at data models we have different kinds of data models which will look at it vector way data and vaster data model vector way data model has different kinds of representative models whereas aster data models as its own different types of representations which we will look at probably at the end of this week or maybe in the next week and we will also look how this data models are better than each data model has advantages and disadvantages.

So we will look at each of them and how we can try to use these data models for our for this analysis. So let us look at it bit later but please keep in mind we have two types of data models one is vector other one is vaster any photograph for example you would have taken a photograph zoomed in to the photograph. So you will find it as a number of square boxes with certain color density yes so these square boxes are nothing but pixels in those images okay pixelated data.

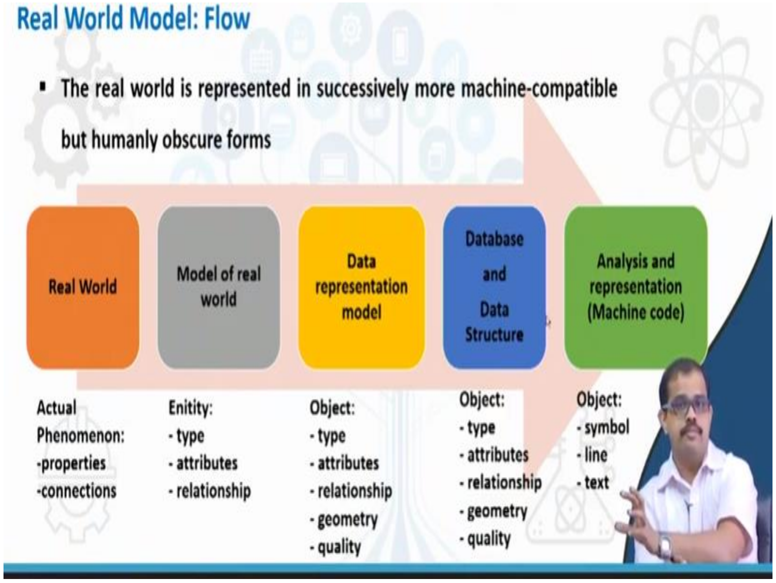
So when you see that is nothing but a raster data okay if you are representing a data in a form of a point line and polygon. A point line and a polygon for example if I am trying to map this particular region around a building there is certain number of trees there is a building there is a road in front of a building now if I have to represent the trees the road and a building so trees can be captured in a point because these are at one particular location road is entire line segments.

So it can be captured in form of a line whereas as a building is at polygon it is an area right so we capture it in the form of a polygon so each point number of points can form a line number of line segment form a polygon. So this is how a vector data is built okay so you have vector data have a raster data so it think all of you understand what are different types of data models that is vector and raster.

Now these are used by different users to correct data and represent extract data and represent as maps reports and statistics. So once you have put anything in the database you can as I said in my previous lectures you can do any kind of queries any kind of analysis that is possible with that data so it is not that you can do every kind of analysis but you can do any analysis that is really reasonable and can be represented in the form a map, a report or statistics I did give you an example of a statistics on a map.

So you can use it for any kind of analysis that is that the data can support well so that is where your database and the data model helps.

(Refer Slide Time: 10:55)



So when we look at the flow of a real ball so though I will speak about this in my maybe in next lecture but I would like to give you some information before because it would give you some understanding of how the real world is translated into an normally and the database or in data model so now for example when you look at the real world if you look at here so you have real world which is based on actual phenomena.

So if you are looking at a particular building so you have an actually phenomena as a building and it may have certain connection number of students visiting a building number of process in the having their offices in the building or it may be number of recording centers there or it may be number or recording centers there or it maybe number of parking areas that are there. So these are certain connections that as real world will have or the number of trees around the building.

So these are certain connections so now you have to convert it to data model the first thing that we will do is if you want to represent the entire building so you have the building as a polygon right. So then each and every room in this building where maybe professors office or the students class or students maybe workshop area so all of these will be each of them will be an entity. So each of them entity will have a type will have attributes so how people are so how many students are there in that particular lab particular workshop how many people are actually I mean in that entire building.

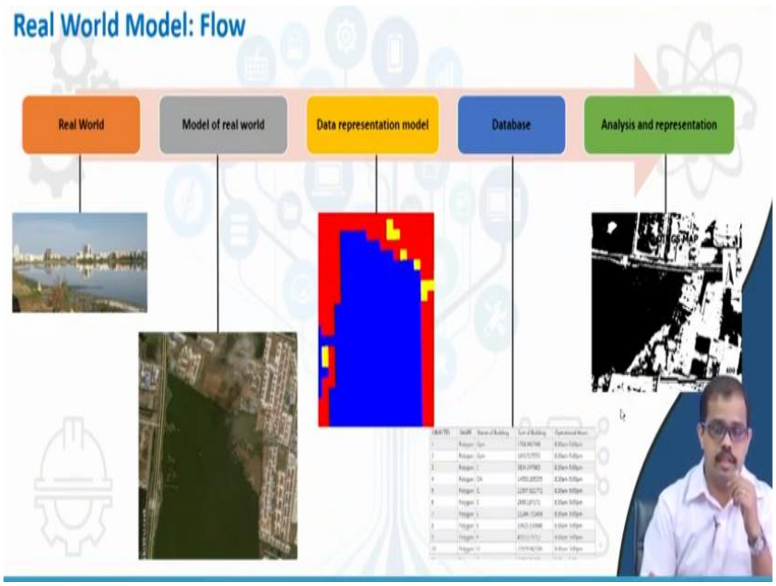
So these are the attributes that are there how many come at this specific time how many are coming at the other now maybe a bit later so all of these are the data that you can enter into as an attribute. So each of these attributes are then related as relationships so every attributes talk to each other with relations. So when you look at this once there is particular model is built then you have a data representation model.

So how this data is represented it is represented as type it can be represented with attributes okay this can be represented with geometry, quality relationships so we will all of this in my further weeks so how you actually represent each of representation models. And you have once you have converted into a model that has as I said two types of model. So this model will be stored in a database with a certain data structure.

So when data structure also has a same kind of a representation of a data representation model and when you look once you have a database you can as I said you can do any kind of analysis and representation this will be in the form of a machine code which can be in form of a symbols which can be in form of a line which can be in form of a text. Once you have any kind of analysis and representation this can be represented in a form of a map it can be represented in form of a reports what not?

So everything that can be kind of representation can be submitted for any further kind of policy decision making can be done once the analysis and representation can be easily performed. So this especially the way you built your model where and the data representation model and your database forms a crux of converting a real world data into a digital data.

(Refer Slide Time: 14:34)



So to give an example of how it may look for example the first one here there is a image which actually representing a real world. A real world phenomena where you have a lake you have a building around a lake you can see there are huge number of building there are come over on the pheriphery of the lake this some part of it some of them are encroached regions here this as a road that is just through the lake.

So this so now when you are actually creating a model you represent this particular road which is here as a line right each of this buildings will become if you see the curser each of this cursers actually become a polygon and when you this lake also is a polygon okay and you have certain points for example here you have vegetation here certain trees here etc., So these becomes a point so now this how your model is represented in form of a model of a real world or a data model.

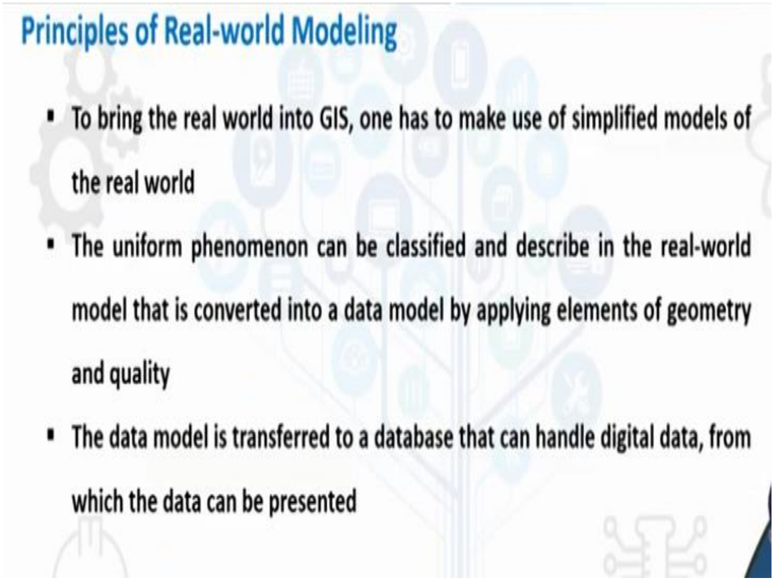
Now data representation model is then converted something like this you have for example if you representing it has a if this is if you representing as a raster model. So when you look at this is nothing but a water body you have open spaces here you have complete urban area and you have small maybe vegetative and certain open areas that are here so this is nothing but a data representation model in a raster model in for the real world.

Now then once you have done it you have a database so now conversion of this for example there is polygon a one of those polygons is nothing but a gem other polygon is again a gem then you have a certain letter that have been given for maybe lettering that in terms of each of this building is given for particular letter okay. Then we represent it is in a form of a size of a building then operating hours.

So the if this is an may be a office building so we also represented as a operating hours so these are nothing but your database. So when you look at this is these are the entities the shape the polygons here and had given a name so it is associated with an entity. Now then these are the object ID's which we refer always in order to connect with or represent a particular data then these are the attributes.

So this is the these are the qualitative information's that are there so these are the attributes so now each of these is represented by a relation in terms of size relation in terms of operation hours so that is about how you have an entity you have an relation and you have a object so or a database. So this is how the entire flow of the real world model will be then once you have done it you will analyze only you may crop a certain part of a region analyze it and you can show this how different maps different representation are there.

(Refer Slide Time: 17:54)

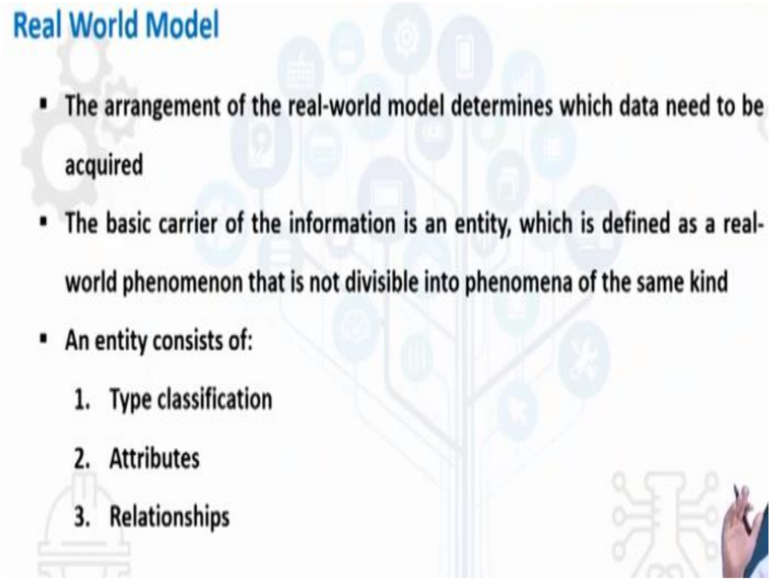


Now when we look at this as a real world model so we have to look at how the principles actually work in real world model. So when we look at a real world model and has it has to be represented in GIS it as to be extremely simplified model otherwise it cannot be easily interpreted in terms of the common users who want to just get a information about what data has been collected and how it has been represented.

The uniform phenomena can be classified and described in the real world model that is converted as a data model by applying elements of geometry and quality. So we look at what is elements of geometry but just for your information it is the real world model is converted to a data model by having elements of geometry. So geometry I would speak about what are different elements of geometry and what is the quality important the very important aspect as I said before it is the quality of the data if you have the good data so you get the good output.

So if you have the bad data when I say bad data it is not the data quality is bad the way it is collected or the way it is represented is a bad data. So having a bad data so will give you a junk output. So the data model is then transferred to a database that handle digital data from which the data can be normally presented.

(Refer Slide Time: 19:24)



So when we are looking at the real world model you should you normally have the real world that determines the data need to acquired so when you physically look at the real world you know what are the data needs that are required in order to convert it into a data base or a data model. So when I say a real world model let us say you are looking at a Google image and you are trying to mark that you are trying to look at this particular region in the Google map as I had shown before.

So you will have to get the boundaries of each of the buildings there so that they are represented as entity. Then attributes of that building how many number of floors are there size of the building etc., but you can even calculate it on GIS that is not a big problem but number of floors certain attributes of data which cannot be calculated as to be actually populated into the database. So once you have populated in as an attribute data into the database now that is what actually the real needs are.

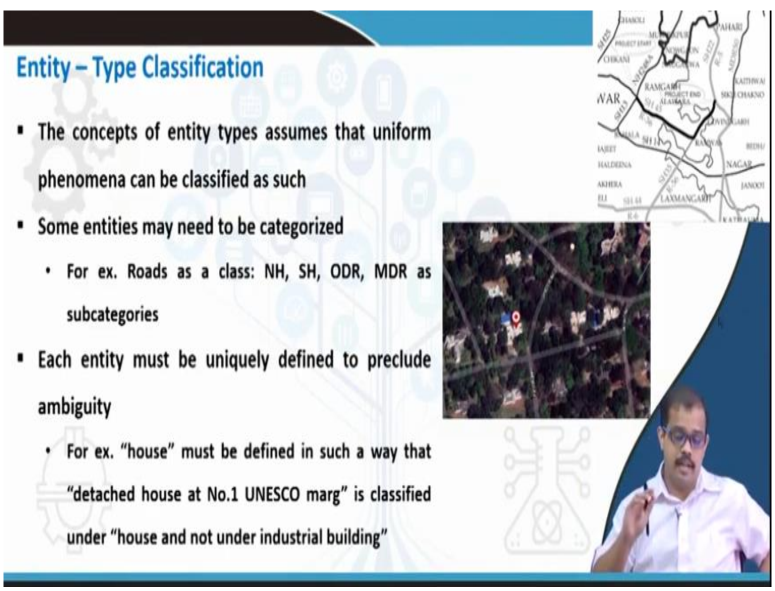
So it depends on the data user whether you need a height of a building whether you need the size of the building whether you need number floors in a building the occupants of the building population of the building. So it is dependent on the user what kind of information he needs s once he looks at the physical information from the earth surface then he can understand okay these are the specific aspects that I have to collect.

So once he goes he can to the field he can collect all those information and he or she can collect all the information and come back. Once the information is collected this the information is now represented as an entity that is what I have spoke about when I gave and example as an building and number of rooms in an building. So always a real world model is represented as an entity so now when we look at real world model it is absolutely not an possible to device it as a phenomena okay.

So how it is not devisable as a phenomena of the same kind I would explain it in my further slides but for it is quite dynamic real world is quite dynamic. So it is hardly very difficult for us to understand or put out in the form of a data model. So that is how your data model has to be quite efficient in handling the real world data. So when we look at entity has type classification entity has attribute classification entity has relationships.

So type of entity the type of attributes that it can have and relationships between these attributes define your entire database so this is what everyone has to remember always entity is dependent on type classification attributes and relationships.

(Refer Slide Time: 22:19)



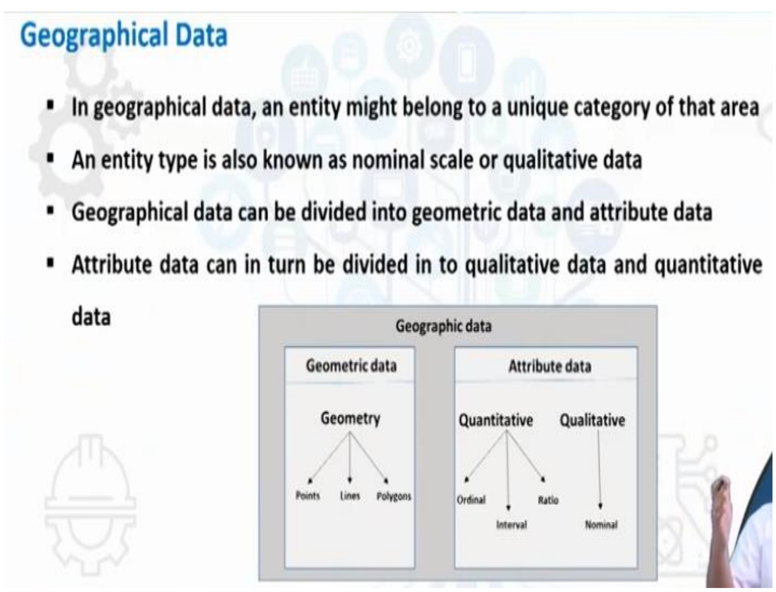
Now when you are entity at type classification the concepts of entity types assumes that uniform phenomena can be classified as such and it need not be different phenomena. Some entities may need to be categorized for example road when you are looking at a road has a class, road is entire road is an entity okay. But when you are dividing it you need to have a whether it is a NH road whether it is SH road, ODR MDR whatever kind of roads these are.

So this is how you actually classified the roads so entity is then classified based on this each entity must be uniquely defined to previewed and ambiguity you cannot have something that is NH1 and NH1A so that kind of I am just using an example but that kind of a ambiguity should not be there whereas if it is NH1, NH2, NH3 it is very well taken because it is easier for your database to understand these are different representation of the same entity or classes of representation of the same entity.

So for example if you have house that must be defined in such a way that it is either a detached house at number 1 UNESCO mark is classified as under house not under industrial building. If for example there is let us say we are looking at academic campus so there is a house of a faculty that is there in the academic campus. So when you representing this house this has to be under house not under industrial building okay.

So and it is and detached house at number 1 UNESCO mark that is the way you have to represent if you just say a detached house and you do not give a reference to that it is an this not the house and a industrial building then you are going to mess up the entire database. So it would give wrong results for different kinds of analysis so it has to be completed classified properly.

(Refer Slide Time: 24:30)

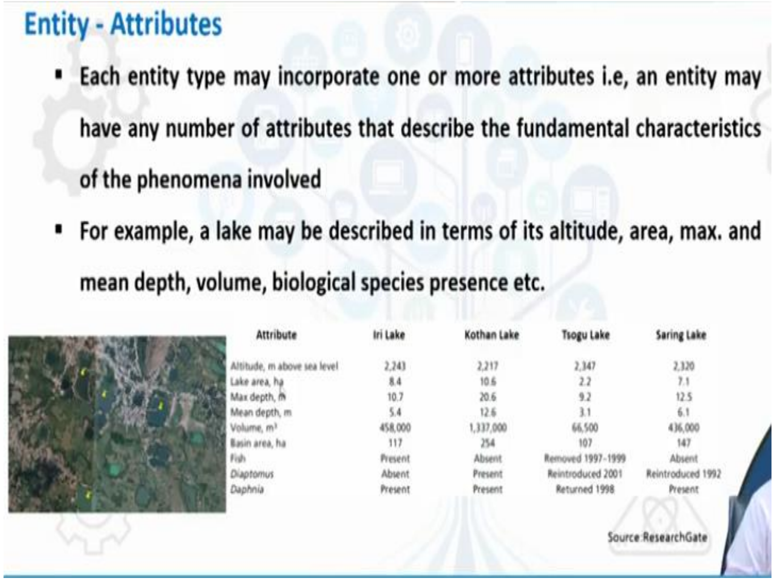


Then we look at geographical data and an entry might be along the unique category of that particular area or it can be at dependent category. So when we are looking at entity type is also can be nominal scale or a qualitative data so you have an nominal data and a qualitative data. So but geographical data can be divide based can be see there is a difference so now what I am speaking is about attributes and a entity.

So now when you are looking at the entity types can be only nominal and quantities whereas when you have a geographical information along with the entity then becomes a geometric data or an attribute data. So you can have two types of data attached to an entity with a geometric data and attribute data. So that is how an entire data model is defined then attribute data can have a qualitative data and a quantitative data.

So when we look at this entire scape you have an entity you have entity can be represented in two types of data's or can be connected to the two types of values or classes it can be geometric data and attribute data. So then attribute data can be both qualitative and the quantitative data as per the requirement okay.

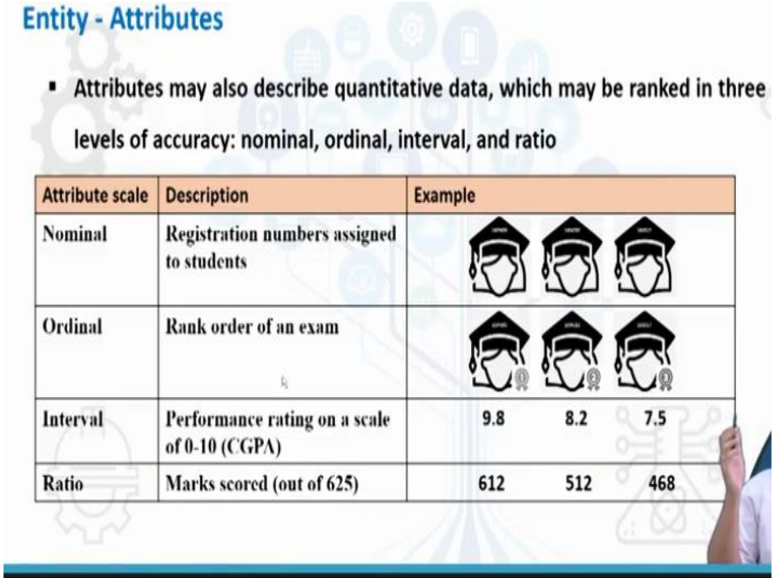
(Refer Slide Time: 25:54)



So when we look at the attributes for example when I have considered here is you have a particular remotely sensed image this is of certain number of lakes. So when you look at it these attribute data can have all qualitative and quantitative data for example here the attribute as altitude lake area maximum dead the fish if it is present or not whether what is the volume of an area the basin area in Hectare's.

So all of these have quantitative whereas these are qualitative data so both of this can be represented so each of this attributes is then linked to an entity would have ID based on all of these data can be easily extracted whenever necessary.

(Refer Slide Time: 26:46)

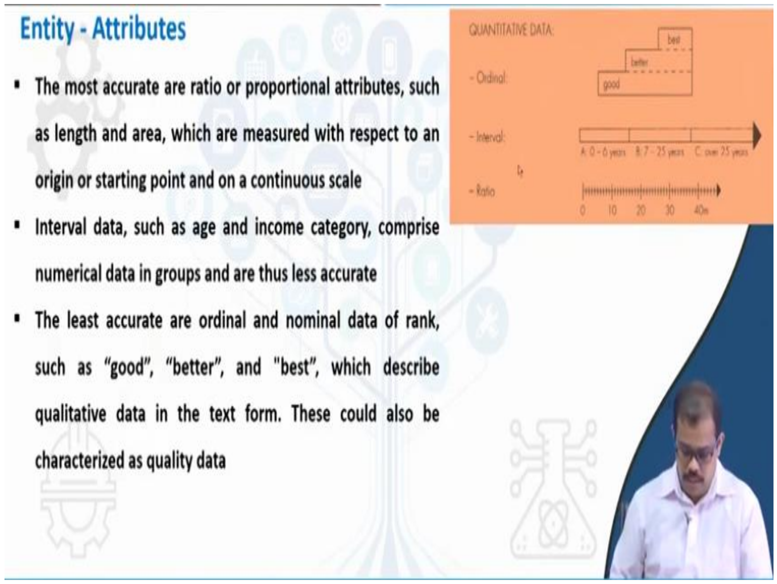


Attribute scale	Description	Example
Nominal	Registration numbers assigned to students	
Ordinal	Rank order of an exam	
Interval	Performance rating on a scale of 0-10 (CGPA)	9.8 8.2 7.5
Ratio	Marks scored (out of 625)	612 512 468

So when we are looking at attributes as I said can have a quantitative data so it is based on kind of accuracy that you need you have nominal, ordinal, interval and ratio scale the way the attribute scale is represented as something like this. For example most of you know what do you mean by nominal scale. So similarly we have an ordinal scale where you rank the order whether you are students whatever are the data then you have the internal scale where you represent intervals and you have a ratio scale where you are representing a ratio.

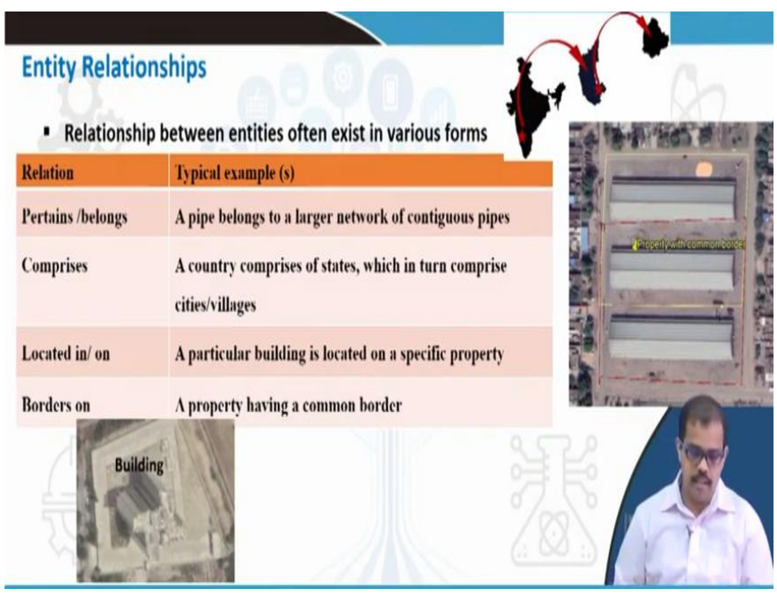
For example a marks how much marks out of 1000 that student has scored in a particular semester or interval scale i can define it in the form of CGPA what is CPG of a student on what scale okay. So they this is how you represent an attributes okay so this how an entity is linked to attributes in a form of a scale.

(Refer Slide Time: 27:49)



Now when we are looking at attributes again so when I did speak about the previous scale so this is how the attributes can be represented from a nominal to the ratio's scale so the first level of accuracy we can see is a nominal scale whereas the ratio is the best kind of accuracy you can always receive when you have a scale that is better okay. So that is what I have explained in this particular slide wherein it is have you an example of how the ordinal scale interval scale and the ratio scale works and how an entity can be represented or in the form of attributes and it is connected.

(Refer Slide Time: 28:35)



Now as I said relations so for example I have a building here so number of pipes that are actually connected to that building and then it has certain number of pipes that is connecting to the main sewerage. So that is saying the property of that particular building that is the relationship where a pipe belong to so that is the relationship. So these belongs to is a relationship or it pertain to it is a relationship then you have if you look at a country it compresses of states number of states that this particular country as the population of the states.

So that comprising is again a relationship okay so how do you represent this relationship will look at it but as of now the relationship are those which are representing those date attribute data and are connecting and giving a meaningful information. So that is what you mean by a entity relationship.

(Refer Slide Time: 29:32)

So in a summary of to summarize we have looked at real world model translating into a data model. So you have a real world you have a data world in between that you have a data model okay the way it is represented that data model then converted into a data base any query on the database to analyze and represent it as a real can be represented in a form of a maps or graphs or any kind of statistical analysis.

So I have also spoke about what is the flow from a real world to the data how the data model has been built then you have entities in the real world model that has types classification attributes and relationship. So what are different type classification different attributes different relationship in a model. Now you have a geographical data when I say geographical data always has geometric data plus an attribute data.

So when I am representing it an in a data model entity is will comprised of will need not always but will have geometric data plus attribute data when it is actually representing the real world okay. Now normal database the entity will always be connected to an attribute data. But when you looking at spatial part of it is always connected with a geometric value plus the attribute data. Then you have attribute scale of measurement so you have nominal scale, ordinal scale, interval and ratio.

So when you look at all of these scales so it is the way you have you represent your data so accuracy of this data depends on the scale. Entity relationships actually if I have to say for example you are connecting particular attribute to the other finally to the entity so you say it belongs to it comprises of it locates. So all of these are relations okay in for example if you trying to put out an a area of an building so it says this area of a building is belongs to this particular entity so that is nothing but your relations.

So this finally connected to that then in the next class we will look at how we can explore different data model how it works and we will also look at each of these data models in a more holistic way. Thank you very much let us meet in the next class.