Spatial Informatics
Prof. Soumya Kumar Ghosh
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 36
SDS / Spatial Cloud / GeoViz – I

Hello we will be continuing our discussion on Spatial Informatics and for this couple of lectures rather four or five lectures we will be discussing on different aspects keeping in view that whatever we have learned or discussed in the previous lectures right. So, in this final link of this particular course we will try to see that some of the case studies, what are the different application and how whatever we have looked into a learn how they come into play when we talk about when we look about specially application related to spatial informatics.

How spatial informatics helps in decision making or how informatics helps in different scientific research and type of things. So, some of the things which is coming up now, in a big way is one is Spatial Data Science in short we have written SDS spatial cloud and also though we have not discussed about the visualization this is also totally a separate ballgame altogether. But, we I thought that it will be nice to discuss little bit on geo visualization that how do you visualize the data right same data set or from the subset of the from the major set how do I visualize right based on my decision making things right.

Like if you look at the same data layers of the things like if I look at the land use, land cover or general or road network, build up area say water bodies or different type of a feature sets. Then from that different people or different decision makers are different view of it; like a say a transport engineer or transport department one should to look at more at the road network part per se taking into consideration how other parts influence the things right.

So, whereas, in case of a expert from the say water resources we want to look at that water body in a bigger way, but the input datasets may be the same right. So, maybe it is added by when we do for a particular domain related operations. So, or in other sense a different type of visualization may come into play right it may be over the maps, it may be emphasizing some of the maps to a particular region of interest, it may be a tabular data, it may be tabular plus we have map data and type of things. So, it makes a lot of it a

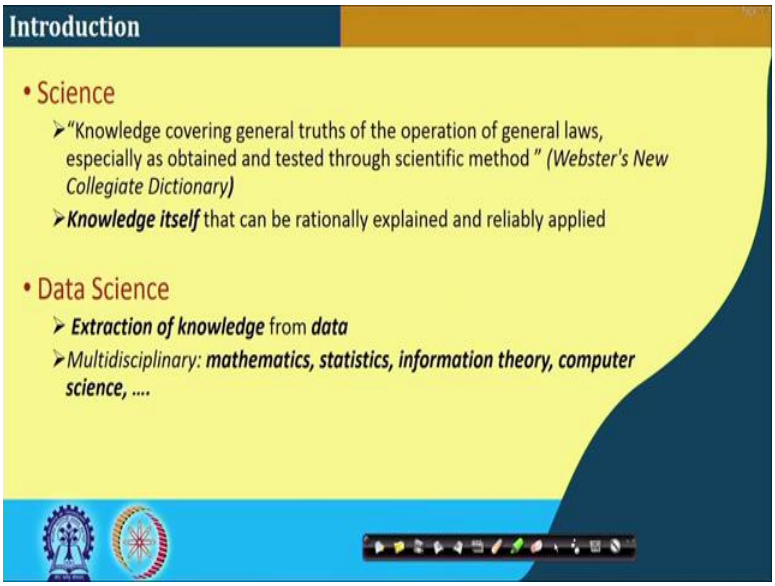
lot of analysis and computing, spatial computing to make from the raw data to a product which is usable type of things.

So, in this couple of few lectures we will try to see that how things how spatial informatics cells in this. So, we will start with a spatial data science spatial data science or data science per se people have different type of definitions. We will try to look at one way of looking at it, there may be variant of those definitions also, but I believe that these encompasses in the different aspects of the things.

(Refer Slide Time: 03:51)

So, primarily these three things we want to cover there are sub items which we should go on discussing when we will be taking up.

(Refer Slide Time: 04:01)

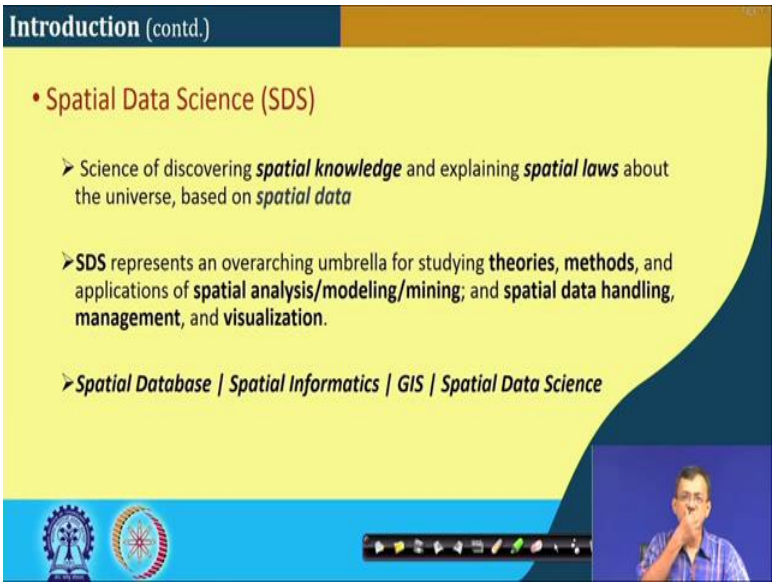


So, initially we look at the spatial data science encompassing the whole thing as a under the scope of spatial data science. Now, if we look at that definition of science per se right. So, like if you look at Webster dictionary definition knowledge covering general truths of the operational of general laws specially as obtained and tested through scientific methods right. So, this is a Webster New Collegiate Dictionary which says; so, data science per se extraction of knowledge from data right.

So, data science wants to look at the extraction of knowledge from data. So, it is a multidisciplinary definitely it encompasses typically any if we any domain of data science if we look at is maths, statistics, information theory, computer science and then maybe a lot of other things which are more on a domain specific stuff. So, when we talk about spatial data science. So, science of discovering spatial knowledge and explaining spatial laws about the universe based on spatial data.

So, it is whatever that knowledge discovery information retrieval or explaining different spatial rules and laws about a about the universe as a whole that is about any spatial dimension or a particular region of interest is based on some spatial data.

(Refer Slide Time: 05:10)



So, when we talk about spatial data there is also non-spatial or attribute data which is there right. So, spatial plus implied attribute on a spatial data when we take some decision that encompass a spatial data science.

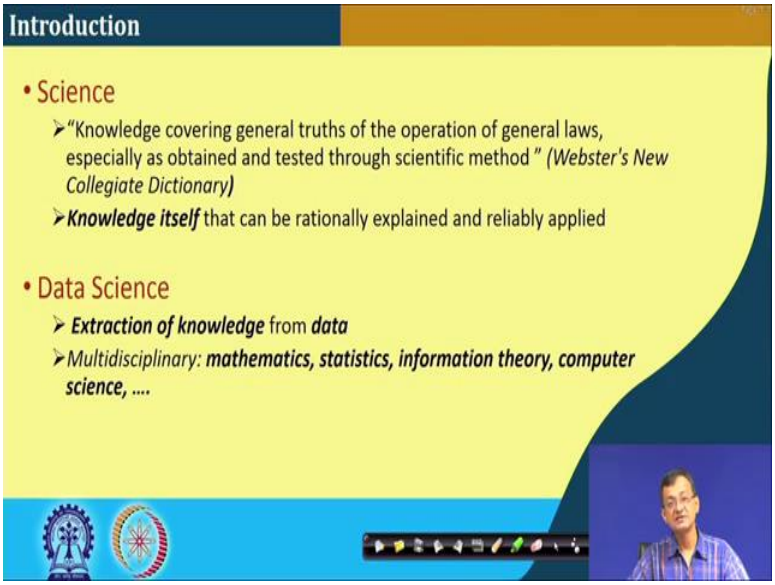
So, SDS in short we though it is not maybe even will not be very commonly used on the things, but SDS as we refer represents a overarching umbrella of studying theories, methods, applications of spatial analysis modelling, mining and spatial data handling, management and visualization. So, if you look at that overall umbrella it talks about everything right it is the theory, methods and applications; applications would have been highlighted also theory, methods and applications of spatial analysis, modelling and mining data mining or spatial data mining, spatial data modelling,

Spatial data analysis which we seen watch which have discussed in our previous lectures and spatial data handling how it can be handled or how it can be acquire interoperate type of things manage and of course, visualize the data. Though we have not implicitly talked about geovisualization, but in several for our case studies someone in some demo or some example scenarios we have seen that visualization effect of visualizing over maps, over a particular highlighting a particular context or domain and type of things.

So, if we try to see that it is a some sort of a overarching or what we say there is a lot of a communality or that of interaction between spatial database, spatial informatics GIS spatial data science right. So, to say and of course, along with that these things are there

that maths statistics, information theory and computer science different algorithms and different methodologies and type of things right.

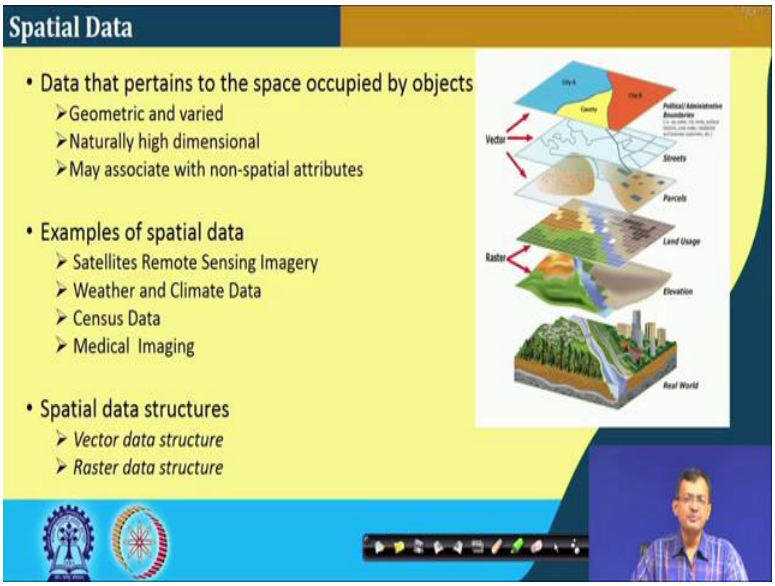
(Refer Slide Time: 07:48)



So, it is overarching sitting using this we if we again put on the basic data science things which comes to a spatial data science. It is a pretty interesting and in today's world it is been extensively used for any decision making specially for say development planning, methodal operations say natural resource management for that matter say forest management or say water management and water resource management and type of things or all things.

Even to the extent of extensively used these days in a transportation where we have spatiotemporal data of these moving objects like vehicles and human being and type of things and try to see that how this overall transportation challenges can be addressed right. So, it is not only natural resources, it is several kind of resources which are being addressed in this spatial data science.

(Refer Slide Time: 09:04)



So, we do we have discussed at length, but I thought one or two slides to again come into this page that where have it means in space with that particular discussion. So, when we talk about spatial data that pertains to the space occupied by objects right geometric and different varied geometric naturally high dimensional may associate with non spatial attributes like in this particular layer taken from some internet resources.

So, we have different type of things right those this is maybe the political or administrative boundary right. There are streets there are parcels right land parcels there are land uses or land use land cover there is elevation map like how the elevation things are there right how it is how far this based on the things what we say how is the terrain means height wise right. So, there is elevation map and if we clubbed together or we can if we extract from the things this is our real world scenario may be right.

There can be lot of other things like there may be something or build up regions where the these are the some build up regions there may be typically water resource and type of things like I can have specifically for a hydrological map and type of things right. Nevertheless what we see that thus a particular real world can have a different layers another interesting thing what we can see that some of the layers what we say vector right and some of those layers which have raster layers right.

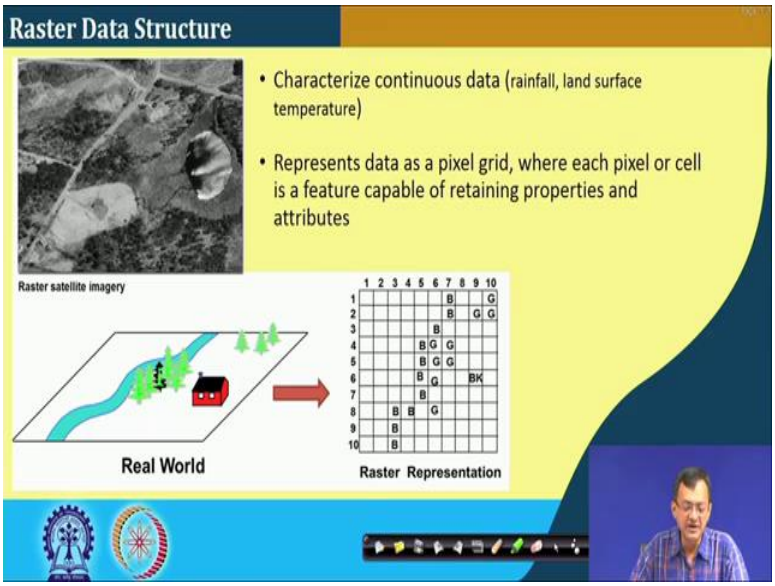
So, like last couple of days we couple of lectures when we discussed about remote sensing data set etcetera they are primarily raster datasets right captured from the from

satellite or airborne any objects like aircraft or drone or something like that or some high rise things right. So, examples of spatial data in satellite remote sensing imagery weather and climate data right which are being recorded by different sensors census data right census data which are being surveyed and collected by the things that also have a spatial context right

So, this population map etcetera the spatial context even to the extent we can go to medical imaging right. So, it is going from the geospatial to say medical imaging. So, it is if I consider my human body as a reference point then it has a different spatial context right. So, starting from blood flow to different organ placements and other things etcetera not distribution etcetera all are medical. So, we are not going to the medical imaging part per se in this particular course, but that also people consider as it can be considered as a over the space which is human body as the reference.

So, and we have seen that two major data structure which helps us in managing these datasets if no one is raster another is vector data sets right.

(Refer Slide Time: 12:24)

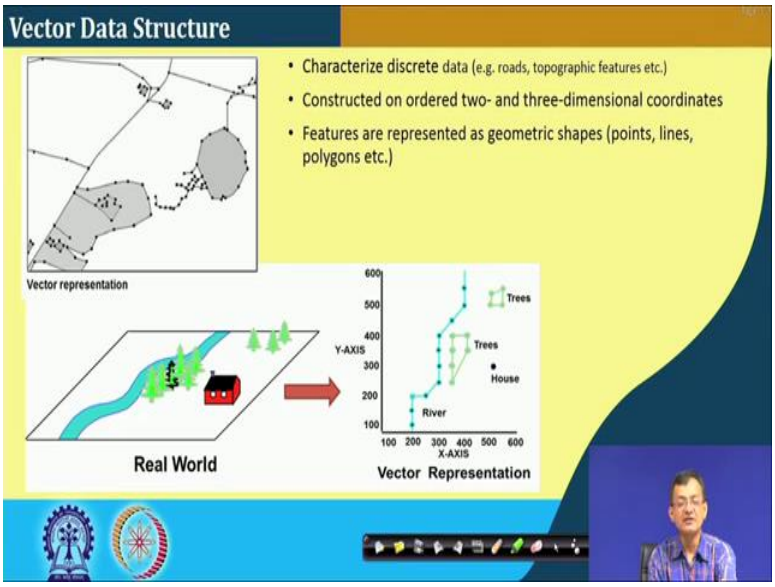


So, just as we already have seen, but like in this case of a raster data science characterized by the continuous data rainfall, land surface or temperature or like anything which are continuous in design or more specifically captured from a from different sensory or mostly satellite image like that this is a raster satellite imagery data of a particular region.

So, in the real world it may be like this right. So, whereas, when we give a raster it is a grid representation right. So, this was suppose this is a real world the same data has been taken as a satellite image and one we take bring when we bring the image down to the earth surface and storing as the image form it is basically a m cross n matrix like in this case it is a ten cross ten matrix, where it is have different wave value or d n value as we have discussed right.

So, indicating like in this case some values which are b for blue which indicates water body green for greenery which indicates something that vegetation and there are building etcetera which is are by the BK and type of things. So, that means, I have a graph paper type of things and color accordingly and try to represent the things. So, this is a raster image right it is a continuous things like that.

(Refer Slide Time: 13:57)



On the other end we have seen that a vector image. So, same data instead of the satellite if I have a vector image then I have this sort of thing characterized by discrete data row topographic feature etcetera constructed on order two or three dimensional coordinates.

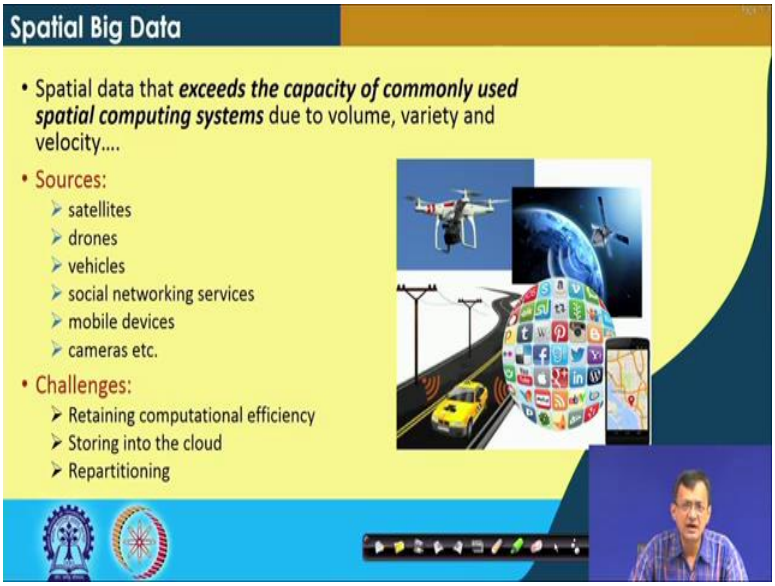
So, typically two dimensional coordinates if we are only looking at the what is the in a Euclidean plane or in a 2D surface how things are there. If I want to make it in a 3D surface that three dimensional space of there with the height information x y and h; so, in the 2D we have only x y information x y and h information then I can have a 3D space

right constructed on a order 2D or 3D coordinate systems features are represented as geometric shapes.

So, as we have seen the basic three geometry point, line, polygons or what we say point, polyline and polygon right. So, this thing we can represent a any map data alright. So, this can be another way of representing data by vector. So, it or it depends on that what sort of applications you are having how the data has to storage how the data is captured and what level of granularity or clarity you are looking for based on that we can have raster data set or vector data set right.

There is a way to convert from one to another and there are obviously, there are issues of what we say generalization or there are errors and type of things, but there was a different factors. So, we can have two type of representation raster or vector representation.

(Refer Slide Time: 15:50)



Another thing which are these days becoming a major challenge is spatial big data right. If the data is huge volume same thing, same area of same region of interested data are collected by different type of a sensory equipments or different type of mechanisms which are not necessarily under single command and control, that means, they are independently taking the data and the any decision maker or a public in general, citizen in general or users have a may take the data from different sources right.

So, first of all this data size is becoming a huge problem because it is a very repeated tip and data being captured at a very high frequency. So, spatial data that extend the capacity of commonly used spatial computing system due to volume, variety and velocity. So, these are as we are talking about this spatial big data. So, our normal computing mechanisms, our normal computing algorithms etcetera are becoming major problem right.

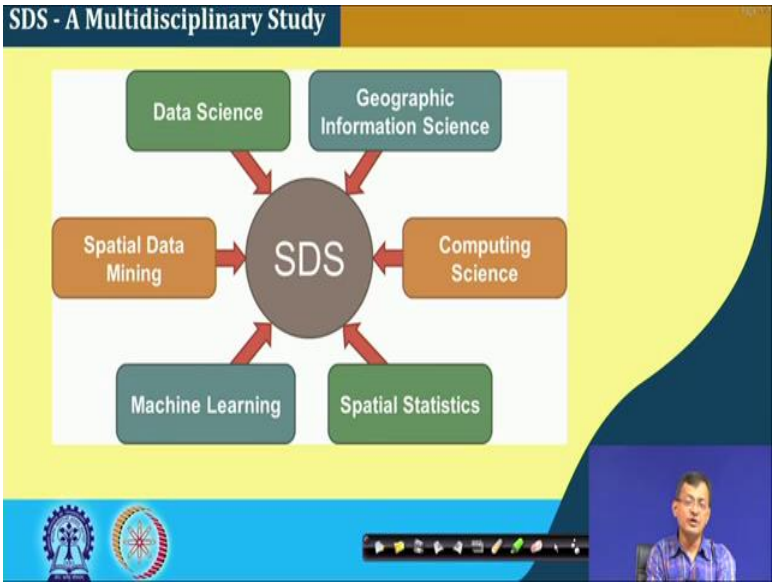
Like if you see like even our standard or well known school days processes of say doing a matrix inversion or matrix multiplication or some sort of a this because here we if the data size becoming big right. So, if I have a say of a particular image data satellite and data mosaic together to find out that particular region of interest and if it is say 10000 cross 10000 or 50000 cross 50000 pixels then stunn running this standard algorithm over our standard computing facility will go for a spin.

Most of the cases data will not be in the main memory right and running all those algorithms is a tedious or not only tedious it is sometimes impossible. So, it is both is a huge volume it is a huge variety of data like some are collected by satellite images, some are collected by drone some are connected by land survey and type of things. So, it is a huge variety of the data and coming in a very frequency is pretty high. So, it is coming in a very high frequently it is yeah. So, it is a become a volume variety and velocity becomes a major challenge.

And there can be several sources as we have seen satellite imagery, drone, vehicle vehicles, social networking services, mobile devices of individuals or cameras and which are kept captured by different roadside surveillance system etcetera. So, it is a mammoth kit it is storing is a major problem analyzing is more of a difficult problem right and whenever you try to do some sort of a mining to find out a pattern it is a hell of a job right it is a something. So, it is a ideal candidate for a big data analytics where we can have spatial hadoop and type of things which can help us in doing such sort of data analysis.

Challenges retaining computational efficiency is a major challenge storing into cloud may be a solution that were to store this data and whether repartitioning of the data etcetera are may be required or process things are challenges.

(Refer Slide Time: 19:39)



So, even if we come back to again to our spatial data science what we started with or tried to discuss to it. So, it is a truly a multi disciplinary there definitely geographic information system and other sources of spatial data are there right.

That is there along with that we require computer science for different methodologies and processes algorithms to develop for a particular domain related things right. So, it may be a very trivial algorithm or it may be a more complicated to find out that to address this problem. So, there are on the one is geographic another is all the properties of spatial data science is there it should be the spatial data science basically inherit all the properties, features, challenges of this data science spatial data sense; challenges of our standard data science problems.

So, spatial statistics is another major input for any type of analysis, machine learning techniques, AI based techniques and other things are becoming extremely popular to do any meaningful operations and of course, as we have seen the spatial data mining aspects. So, if we see starting from say data science to geographic information system, computing science, spatial statistics, machine learning, spatial data mining all those things are become major components of spatial data science.

So, in some problem we may require we may not require all of them we may require some of them and type of things. And, if you recollect we talked about spatial data infrastructure and that how the web services can help in interpretability they play major

role not only that even if you want to solve a problem that modelling the data modelling and process modelling for that particular domain is also important. Suppose you want to address the problem of say pollution in a city right.

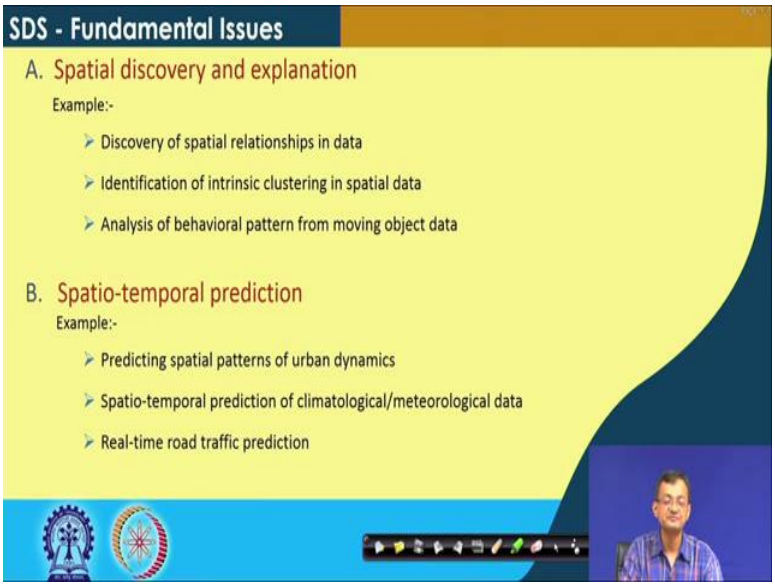
How this pollution changes over the whole city region or part of the city region, how it behaves over weekdays and weekends, how it behaves over seasons and even to one on one is looking to study for that or the pollution control board studies wants to study on that. And, then want to see that whether there is a there can be a short term long term focus that what could be the pollution level on one side another thing that whether there can be preventive mechanisms to handle this pollution levels right.

So, or reduce the pollution levels. So, in order to do that what I first I need we need is to have a appropriate data model of this pollution things right and what are these what are these standard way of modelling name and type of things right. So, as many of you have done as UML modelling and type of things that can be a one thing there are several tools available which helps in this.

So, in other sense I need to find out that what are the different major classes their attributes how they interact and type of thing. So, that looking at the in food collected data set some effective control or effective management of that particular domain can be done. So, if you look at that when we though we are talking about very high level spatial data science all these ground level things starting from spatial databases to spatial queries to how to model this data once I can model then the data to the database schema is will come into play then interoperating between different sources of that data right.

A data coming from satellite image some collected by aircrafts and collected by drone we will not sit over one another just like that. So, need to be pre process and interoperate like that right. So, those things are pretty important to realize this at the those things plays the important role the backend what we try to discussed and have a overview of the what are the different process in our previous lectures.

(Refer Slide Time: 24:11)



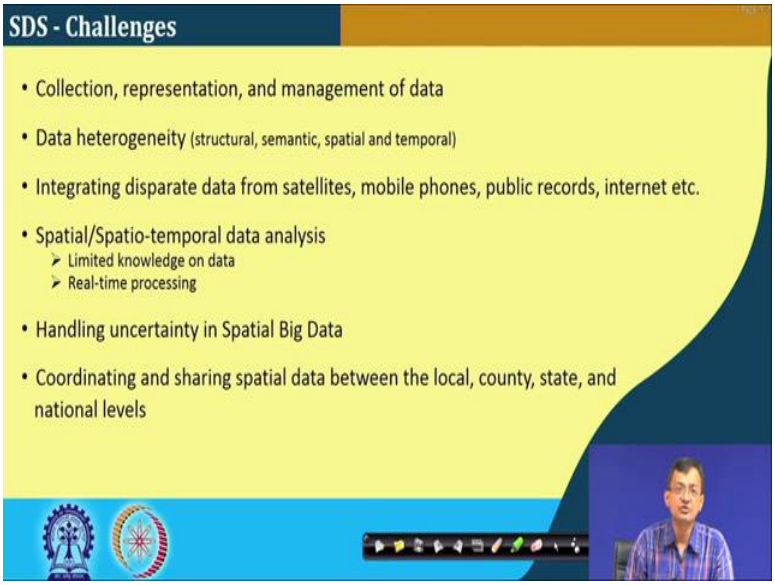
So, if you look at the fundamental issues or fundamental challenges in spatial data science is that, what is spatial discovery explanation like discovery of spatial relationship in data. So, how do I discover the how this spatial relationship or discovering the spatial relationship and explaining, that why it is like that.

So, discovery of spatial relationship in data identification of intrinsic clustering in spatial data analysis of behavioural pattern from the moving object data right. So, there are different example scenario which require spatial discovery and explanation. There are issues of spatial spatio temporal prediction right predicting spatial pattern in urban dynamics like what will be the congestion level, where there will be a more demand for taxies or where which route of a particular bus service have a more passenger or which route is not having sufficient passenger, which requires more.

So, this different type of a decision making for different urban dynamics things can be looked into by spatio analyzing by spatio temporal data set, spatio temporal prediction of climatological and meteorological data sets right and real time road traffic prediction and these are some of the things which are which are the spatio temporal prediction. Not only that what we as we see that climatological prediction, methodological prediction forecasting on say different agricultural products like what will be the expected level of production of paddy weed etcetera in some part of the country or whole country those are

all predicting or forecasting type of mechanisms which are there right. So, these need to be looked into when we talk about these data sets.

(Refer Slide Time: 26:22)



Collection representation; so, what we see SDS challenges collection representation and management of data right data heterogeneity is another major challenge structural, semantic, spatio your spatio and temporal type of things. So, there can be structural challenges semantics and this integrating this separate data from satellite, mobile phone, public records, internet etcetera.

How to integrate or interpret as you as if you recollect that during our when you talk discussed about a web services spatial web services and type of things this is a challenge spatial and spatio temporal data analysis right. So, limited knowledge on data real time processing is still a challenge because these datasets are require a lot of dominics parts analysis to look into the things unlike our traditional data sets where also you really require domain expertise; but, that not that these extent where things are there.

Handling uncertainty in the big spatial big data is a challenge like how to handle uncertainty otherwise means we cannot have quality output and type of things. Coordinating and serious sharing spatial data between local country county state and national levels is a challenge right how challenge need to be addressed right how to interrupt a rate and type of things I we may collect data store and services.

How somebody will consume the data on and or not only for visualization or human consumption it can be used for another process at the user end to work on this data and provided definitely the data permission and other things access permission etcetera are there.

(Refer Slide Time: 28:15)

So, if we look at data science there are different research directions there can be few more, but what I see in that these are the spatial specially the different research junction one is spatial data representation right. How do I represent the data or how do I efficiently store the data. So, that it can be easily searchable easily I can extract information from that things right. So, this is a major aspect. Another aspect is spatio temporal data prediction right.

So, if I have different sets of data how good I can predict the next event or next data set right. It can be starting from criminological prediction say crop acreage or crop production prediction it can be even prediction on the based of say predicting whether this is accident prone or know attack. So, important type of things spatial web service we have discussed in detail also we look into spatial cloud in subsequent lectures mobility analysis using GPS footprint is becoming a big uses like from the GPS footprint.

We will discuss some of the things using GPS footprint big spatial data analytics and discovery right big spatial data brings a lot of other challenges like big data challenges and along with the spatial context. So, how do I, how it can be an analysis and discovery

things are there and of course, the core of the things still lies in the integration of heterogeneous spatial data which will help in a uniform axis and different decision making things right.

So, today's class what we have tried to see that whatever we have discussed how we can put them in to have a in near broadest perspective of spatial data science which are something umbrella which encompassing all those aspects right rather in the subsequent classes a few lectures we will see little deep into this different aspects right. And, try to make a somewhat we say meaningful uses or what we how this spatial informatics per se can be used for different decision making, different prediction and different scientific research and type of things.

Thank you.