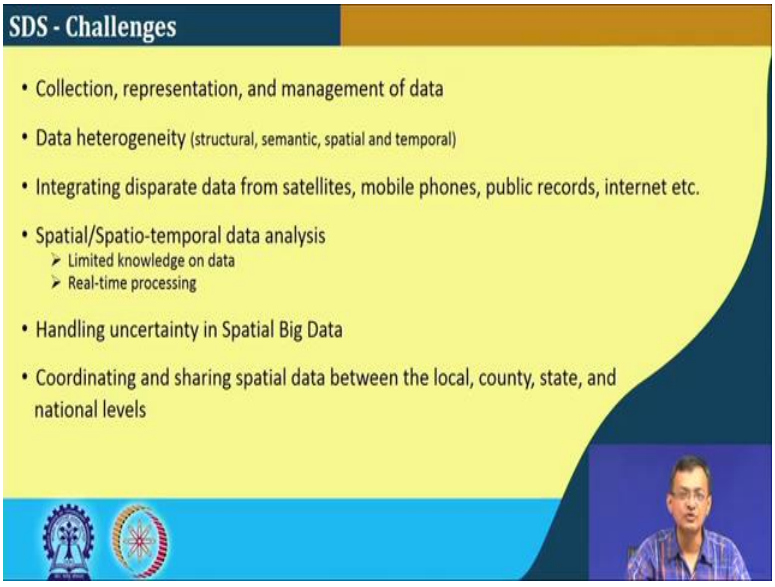
Spatial Informatics
Prof. Soumya Kumar Ghosh
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 37
SDS / Spatial Cloud / GeoViz – II

Hello let us continue our discussion on Spatial Data Science and other related technologies and in our the Spatial Informatics Course. So, if you recollect in the last lecture we discussed about this: what are the different component of spatial data science. What are the influencing factors and challenges in spatial data science and why what you need to do. Again just to reiterate that whatever we have discussed and learned in this course like starting from the modelling to spatial databases to spatial data mining's, queries, spatial networks and so to say spatial analysis or some sort of a spatial computing.

What we those are there very much there and we want to make a umbrella encompassing all these theories, models, applications. So, that it can be used for different decision making, or scientific analysis, or predictive model and type of things. So, there is the focus that how the spatial informatics helps in having better understanding of the data and analysis of the data.

(Refer Slide Time: 01:36)



So if you recollect last lecture we know that these are some of the major challenges. One we talk about collects a spatial data science challenges, SDS challenges, collection, representation and management of the data right. So, how this data need to be collected in a repetitive pattern, how they will be appropriately represented and interoperate and how to manage this data. There are several aspect of it like one data some data you are taking repeatably at what scale you want to take. What are the different scale you need to offer? How this temporal scale data will be there?

And though we not explicitly discussed, but it is important to know that in several cases in you may not be able to collect all the data related to your schema alright. Some of the data may be may not you not may not get the data about the based on your schema. So, while we model and develop the database schema it is based on the domain expertise right which takes care of all possibilities etcetera.

So, we need to have different approaches at the underlining database, underlying looking at the means storage mechanism in dictionaries, so that these are addressed. Secondly, this data heterogeneity is a major challenge like both syntactical or structural heterogeneity. Semantical by meaning by semantical heterogeneity and there are a trinity in type of spatial and temporal context right. When I say that some data you get a 1 is to 10000 scale; some data you get a some other scale right.

Some data this pixel satellite remote sensing data the pixel is 20; 20 cross 20 some maybe 5 cross 5 some maybe 23 cross 23. So, the there are issues of spatial extent also right, there are challenges of temporal extent right some are daily stored, some are weekly some maybe every 3 days, every month and type of things. So, different sort of data different scale of data, with different structure of capturing and storing the data and there are different semantics or meaning of the data.

So, this is a huge challenge right out of that we try to address some of these challenges through this. What we have looked at a spatial data infrastructure which is which able to handle this different heterogeneous data and give them in a appropriate format to the stakeholders. So, it is a somewhere a layer between the data provider and the data consumer right. So, other are integrating the separate data from satellite, mobile phone, public records, internet and type of things right.

So, there are there may be different sort of data some are from the satellite images, some are mobile captured through mobile, some are cloud sources, some are through social media, some are in through internet and etcetera. How to handle those data? Spatial spatiotemporal data analysis is there limited knowledge on the data right. So, you need to have a all encompassing knowledge of the data suppose satellite captures the data in a 8 band, or even my ulti hyperspectral data. In some say 100 plus bands then in order to analysis I need to know that about the knowledge.

So, there is a huge challenge of domain expertise into the things and real time processing. Some of the data needs to be processed some of the application demands the real time processing once you get the data it should be processed real time right. That means, like say a cyclone monitoring or some other type of disaster like flood, or even in some cases is with appropriate sensory things. A fire broke output which requires immediate action to be taken that regarding that overall resources how to safely evacuate people, how to encounter that things etcetera.

So, that real time processing of this huge data another challenge is handling uncertainty in spatial big data. So, as such what you have seen that these spatial data fits all the typical property of a big data analysis, or big data right of the big data. So, handling uncertainty in the spatial big data is a serious problem for that matter handling uncertainty in any sort of data is a is not trivial right so that is the another challenge.

Coordinating and sharing spatial data between local within the local area, within a county, or within a particular state and at the national level. So, it is a challenge right. So, it is coordinating sharing spatial data things are like may I say that district level data are collected say I say SDI, at the district level is DSDI district state District Spatial Data Infrastructure. Then we have a SSDI State Spatial Data Infrastructure and then NSDI.

So, this hierarchy is there this drawing this hierarchy is pretty fine, but when you get this data integrating the things like making several district to make a state and type of things. Data being collected by different people being mapped from different maps etcetera it is a herculean task right or it requires a appropriate process to handle. So that it goes on appropriately into the things right.

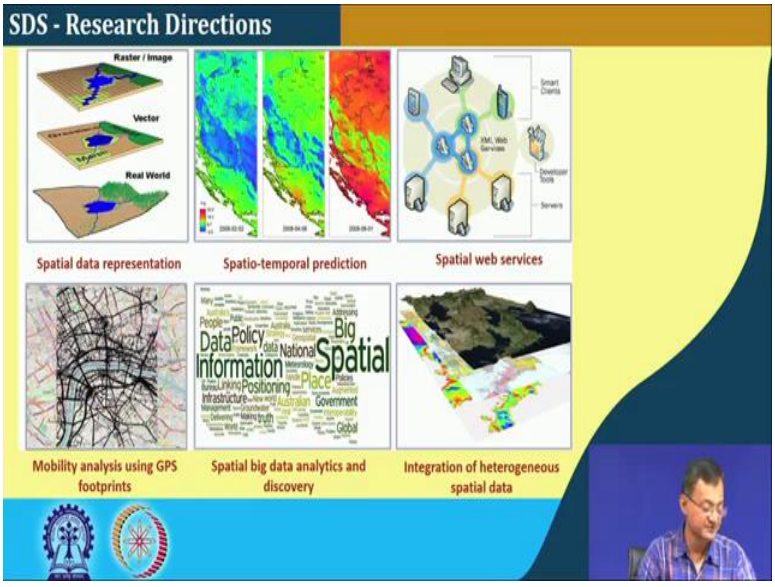
So, this coordinating and sharing is also a challenge. So, these are some of the major challenges in realization of spatial data infrastructure, or spatial data science so to say.

And SDI tries to have addressed many of them above and above and over and above what we have in SDS we need to have a processing of the data right. We need to appropriately process that data's for some to achieve some real output or useable output or product right so that is important.

So, I need to that require that may require a domain expertise to handle that alright. So, say I want to climatological data I want to predict I may require doing expertise on that things. If I want to predict something for health related things, I required some public health expertise like public health the type of thing and, or somebody who has medical practitioners right. If I require something on say pollution related things in a foreign state experience so; that means, I req there is a need to this knowledge of these experts need to be integrated into a into a form of a process into the in this whole system right.

Otherwise I cannot have much scientific or decision support things out of it right. So, there are several challenges and we will try to see as few case studies. That how these things can be addressed or how things are addressed many of the things what we did in our lab out here. There may be other advanced work in several part of this globe.

(Refer Slide Time: 09:28)



So, if we see the SDS research as we have seen spatial data representation in one part. Spatio temporal data prediction right space time data prediction, spatial web services to so that different heterogeneous data can be accessed or different application can run of the things.

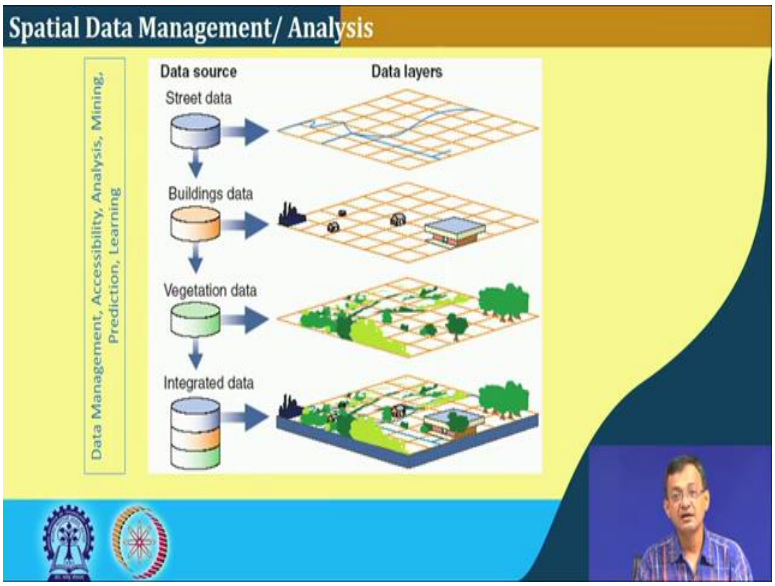
Mobility analysis uses GPS foot pints as we are discussing, it is a becoming important to look at the GPS footprint and find out the mobility analysis. It may help in different type of things that movement patterns helps us that where more resources are required like resources in terms of say public vehicles, or public transport resources in terms of say even public amenities type of things. So, the sources in terms of say security arrangement if there are things etcetera and there are different things which gives us these mobility analysis right.

Big data analytics and discovery so how to have this big data analytics and discovery of interesting patterns, or features in this particular data set is definitely there. Integration of heterogeneous spatial data it comes with the back to back with the spatial web service. But integration of this heterogeneous spatial data is a major issue. Well, there are several other important factors or important components one is visualization. How to visualize this heterogeneous data once integrated, or once have the different algo etcetera for different analysis analytics things are there.

How to visualize this data set right. So, that is also a thing other things like though that is at the background how to archive this data right. I need to archive need to looked into the data when require when requires historical data sets in a temporal scale and type of things. There are several other challenges we are not gone to the things like that same data can be supplied by more than one stakeholders right.

Like say I say that road network of Kharagpur that may be owned by the national transport agency or national transport authority. Even those type of data can be collected and shared by state transport agencies know whom to take right. So, that there are several other complicacies or challenges we need to bear. Never the less these are the major research direction what we see in today's geospatial world where computer science or computer science IT related knowledge or schooling will help us in addressing these problems.

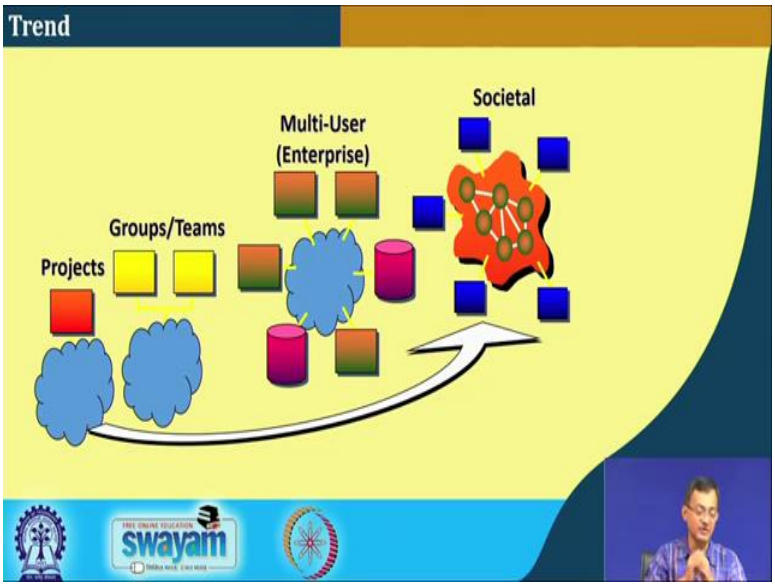
(Refer Slide Time: 12:38)



So, this is a picture already we might have seen earlier nevertheless so it is spatial data management analysis. So, what we have that is a different layers and we want to integrate or disintegrate the layers for different analysis right; so data management, accessibility, analysis, mining, prediction, learning on this data science. So, same region I have that different layers one is say vegetation layer, build up building layer, street data. There can be different other type of things even popular population map, pollution map, right drainage system.

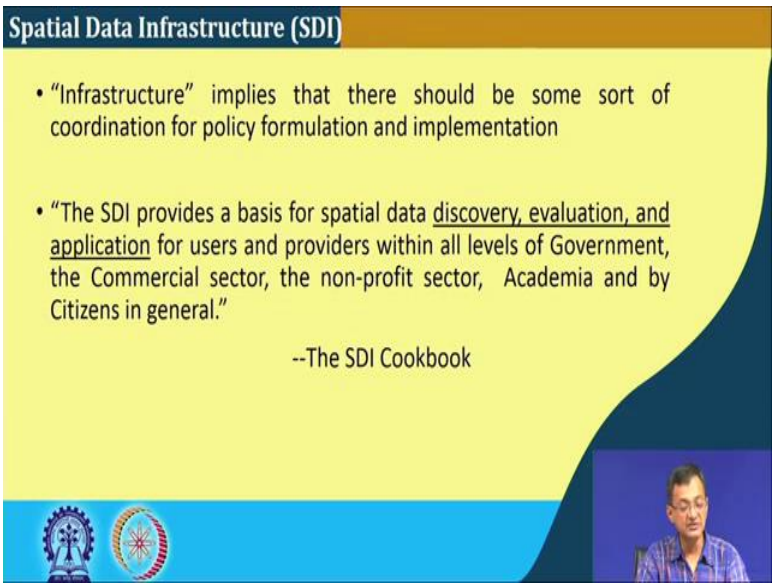
So, there can be different type of aspects into the things build up areas and no building data is there. So, these are different layers of information of the same region of interest right. So, once I have the same region of interest. So, what we can analyze the things based on one or multiple layers? So, it is a multi layer analysis, or multi layer modelling type of things.

(Refer Slide Time: 13:49)



And, also we have seen that the trained as from which as in the project mode has now graduated to some societal mode that is more of a cloud related type of a cloud type of platform it has graduated right. So, it has grown from the projects to group teams to multi user, to society at large. So, it is some sort of a spatial cloud where can hook into the things and get it the work.

(Refer Slide Time: 14:18)

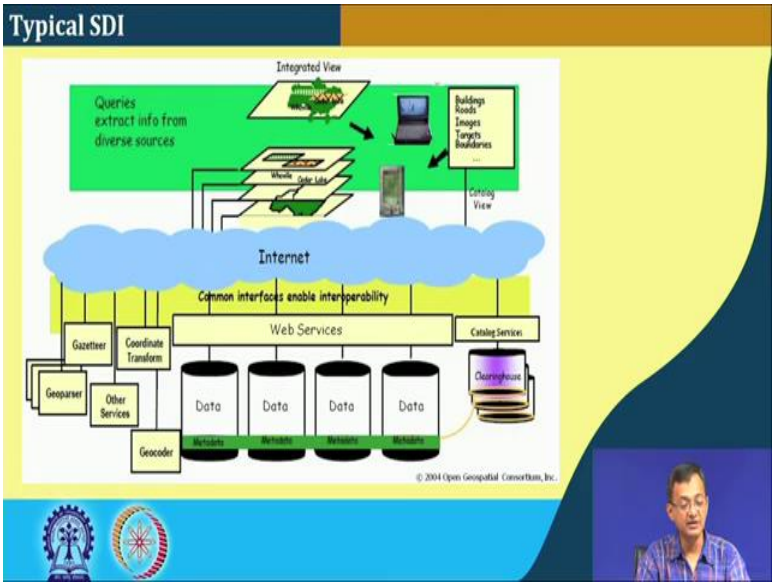


This is a definition already we have seen in our earlier or at the beginning lectures when we discussed our spatial web services. So, what we were is what the things need to be

developed is a spatial data infrastructure. So, it that implies that there should be some sort of coordination for policy and formulas and employ a policy formulation and implementation.

The SDI provides a basis for spatial data discovery, evaluation and applications for users and providers within all levels of government, commercial, non-profit, academy and by citizen atlas. So, that everybody participate in to the process either contributing to the system, or getting information, or knowledge out of the system or both alright.

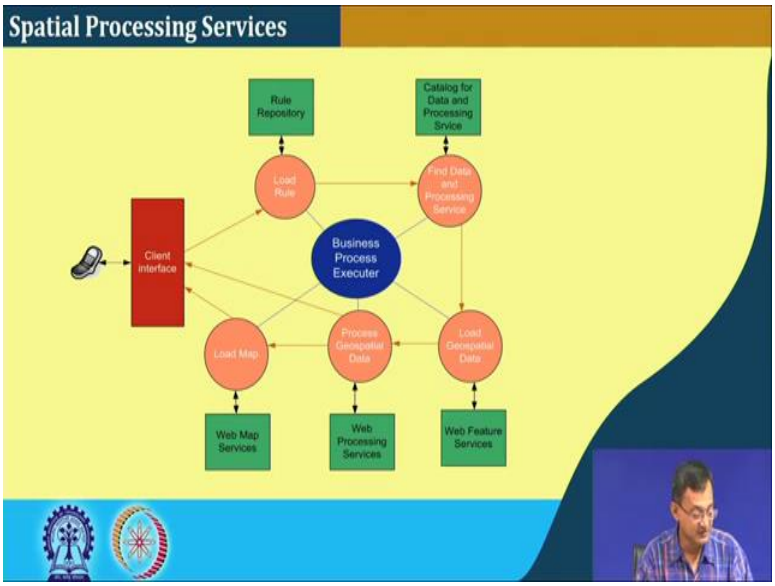
(Refer Slide Time: 15:12)



So, this type of things where the infrastructure gives this is the typical view. So, we have different data sets in between there is the web services there, spatial web services there and there are internet by which things are being accessed. So, it can be different layers or information or analysis of the some data and etcetera right. And also as we have seen earlier there can be different other type of site like one it coordinate transformation. If you have one set of coordinate to another set of coordinated like say you have lat long.

Then go to WGS 84 or something like that and then you need to prepare a coordinate transformation there is a gazette or to have this a registry services and other things are there. And, there are different type of other services are there right and over that what we have has different application, or interfaces, or analytics tools which are know about the things.

(Refer Slide Time: 16:12)

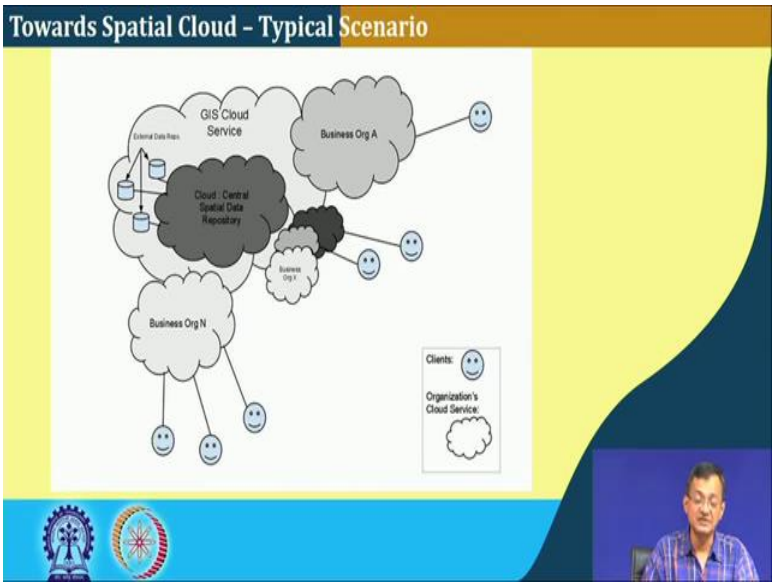


So, a spatial processing service like what we want to process is basically what we have the client interface. And, there are there are several type of services map services, processing services, web feature services. There are other like registry services and based and there can be a rule repository. Like there are different set of predefined rule, or rule set or rule or policy set for particular set of data.

So, it is based on that that different type of application may come into play. So, in between the core of the thing is the business process executor; that means, say I am going to use the spatial data. So, I have a typical business process in mind or I have a thing like as at the example now and then we pull out like I want to make a 4 lane node to a 6 in 6 N node.

So, my objective is to make it 6 N node I want to find out what are the say cut and field whether it is going over any residential area, or it is how much it is agricultural area, how much it is barren area, whether it is going over the educational area and type of things and I want to have a estimate of that thing. So, that is my objective so to say right. So, this is my business process for my particular working of my things.

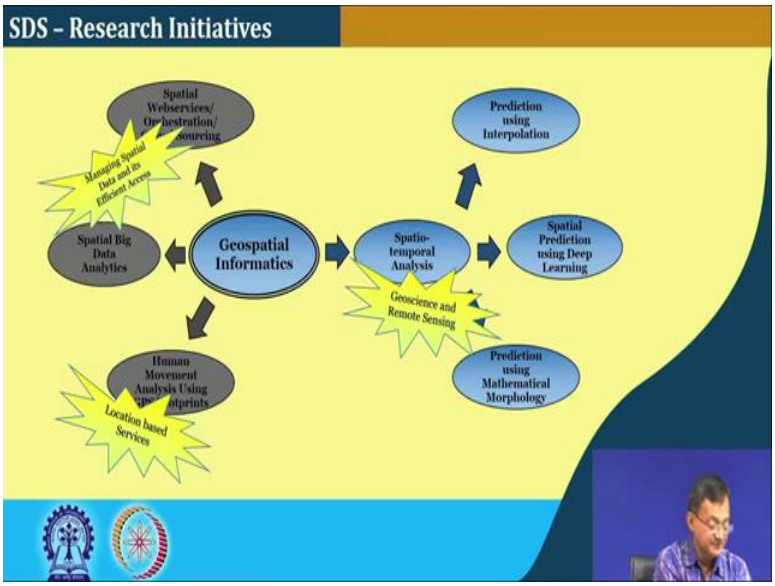
(Refer Slide Time: 17:55)



So, to satisfy these business processes whether I have all these components to the things, and later on we will discuss a little in detail in spatial cloud. But what we see here is that a instead of having a client server type of things from the from SDI point of being, or web services we my whether it is possible to have a spatial cloud right. So, where I can basically hook into the cloud and request my queries and get the results right, or requests give my business rule and get the results.

So, those type of things like in this case there are two business organization A and B several actually A to N and they give different type of services alright. So, whether I can if there is a query cloud that basic cloud acting as a mediator can get the results out of these things.

(Refer Slide Time: 18:57)



Now, little we will go little on some demonstrably things to see that how the spatial data are things can be utilized right. So, in this case is the mostly we will be using the work done by several scholar here in Department of Computer Science and Engineering at IIT Kharagpur on the spatial data science lab. Like so some already graduated some are working on it. So, we will it will give us give you a idea that how what are the different how this computer science want to play a role. And what are the different challenges there are interesting problems which can be addressed or solved.

So, if you look at so we have spatial informatics and which lead to spatio temporal analysis or spatio temporal analysis. So, spatial informatics may have spatial wave surfaces orchestration services cloud sourcing and type of things which is more of building a set up or infrastructure so spatial web services right. Other aspect is spatial big data analytics right, so that is also doing some analysis on a big data on hue large scale spatial data sets right; other interesting area where is human movement analysis using GPS footprint right.

So, whether it is or move I should say instead of human the movement analysis from GPS footprint. Like the movement of a human being a movement of a particular car and a vehicle etcetera. The basic hinges or basic motivation behind that people move with the intent right unless is gone in a different psychological state. So, to say what we say that they move on a particular intent.

So, whether it is moving by individually by walking or by some transport system personal or things they have a underlining intent to do that. So, whether I care whether it is possible from the GPS footprint to capture the intent of that particular person. Actually instead of individually individual we try to look at it in a group like student, UG student huge is PG students, faculty, staff, guests that type of category. Whether there is the underlining patterning that if there is an underlining pattern that may help us in framing different policies based on government, private etcetera right.

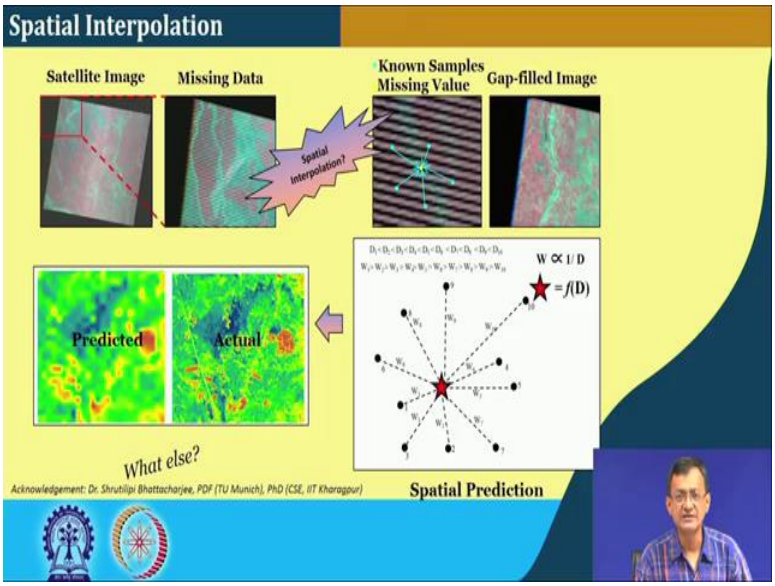
So, this is a very interesting work for it is useful for transportation engineering or transportation transport management and type of things also for taking predictive model that how things the how the growth rate and type of things will go up. So, on the other hand on the right hand side what we see this spatio temporal analysis one is prediction you doing interpolation this is one important factor. Because there is missing data is a major challenge in the digital world right.

So, data here are being collected by different sources starting from satellite to ground survey or using mobile survey. So, different sources so it is difficult, it is important to have interpolation interpolate this missing data or finding of the things, so this is one aspects. Spatial prediction using different learning techniques is another challenge that whether we can do spatial prediction using different learning to me and the spatial prediction using mathematical morphology or type of things.

So, one is that a temporal analysis mostly spatio or spatio temporal analysis which helps in prediction and for forecasting of some activity or events into things right or some mechanisms. So, we can say that these sort of things is managing spatial data and is FEC and access is mostly this who are around. How this spatial data can be properly managed and accessed into the image by access for different analysis purpose whereas, human movement analysis using footprint is more of a more towards location based services right.

So, the services which is dependent defined or can be used for location based activities right. So, these are LBS type of things. So, whereas, this prediction and morphology etcetera these are mostly over around geo science and remote sensing activity or on primarily these spatial temperature spatio temporal analysis.

(Refer Slide Time: 24:49)

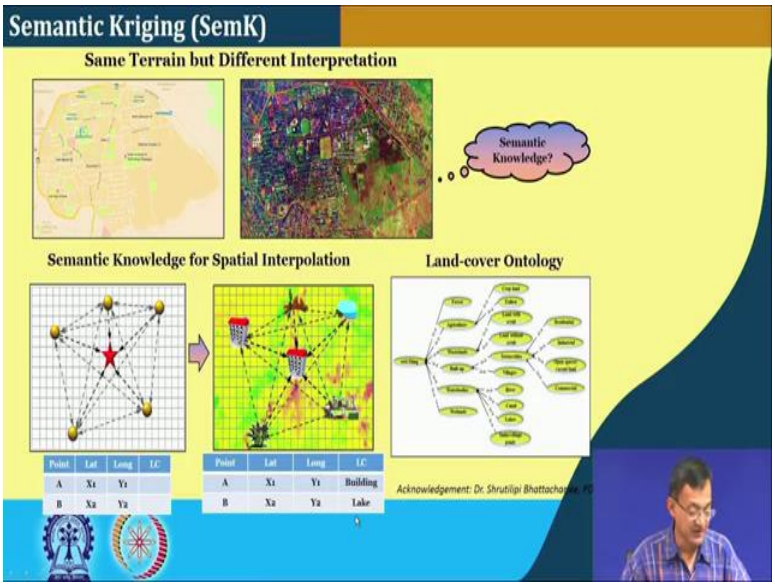


So, with that let us see one use case where things are like we have two satellite, a satellite image which there are several missing data. So, we do a particular type of interpolation or we can do interpolation like this like I have several data within the means participating data from where samples are there and then based on that the prediction is there known sample missing value gap filled by images.

So, this is the predicted and this is the actual image we are doing it right. So, here what we see the spatial prediction is a function of or inversely things with a given a can be predicted with a function D, where the weight h is given by 1 by D. So, higher the distance lower the influence right. So, like if you look see remember the first Doppler's first law of geography which says that every objects influences every other objects only nearer object influences much more than the distance objects right.

So that means, keeping that that more distance is the thing. Like if I want to find out the temperature the it is influence mostly which are the nearby points less by the outer points right. So, these are the things which are there.

(Refer Slide Time: 26:39)



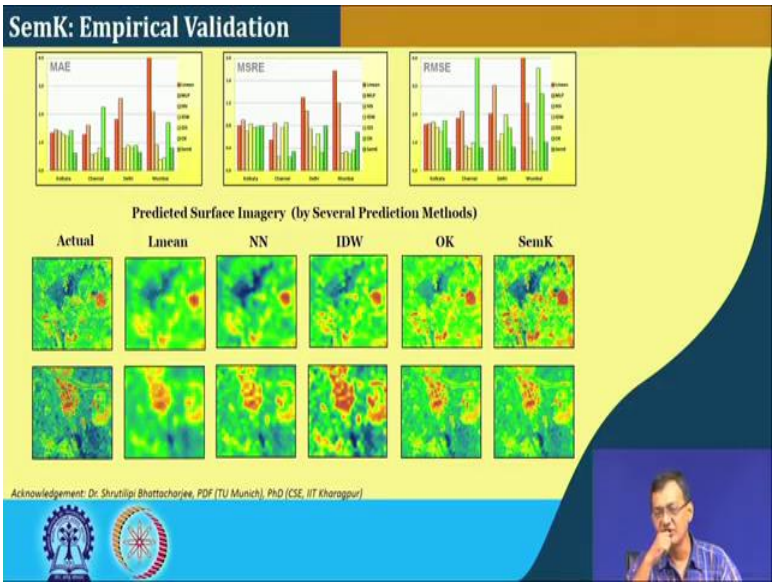
Now, here anything else so it is only the distance or something else is there. So, if I say that these are the I want to predict this value is missing. I want to so this is a typically IIT Kharagpur region right and what semantic knowledge I can having. Let us take an example suppose I want to predict this location of a this value of this. If these are different temperature values of different things then using these inverse law I can say that more is the distance less is the weightage, less is the influence.

Now, I can have this semantic knowledge for spatial prediction with that. Like the influence created by a water body may be different for to say if I consider my temperature is my the value you want to predict. So, to influence of a water body may be different from the influence by a say building or a influence of a by a affect alright. So that means, only making a inverse distance may not be the be correcting so that is definitely there.

So, whether I can consider this influence of other things right or what we say semantically how they are close to each other may have a lot of influencing factor here. So, how to quantify the semantic knowledge one is that whether we can have a ontology. So, it gives that there are different components forest, agriculture, waste land, build up and water body. And, then I can have the distance semantic distance using this tree alright; using the semantic land use land cover ontology right.

This is from a domain expert it has to come we and so I think this is something which is published by either ISRO or DST which have been taken by used in this work. So, here if you see that point A X 1, Y 1 lat long and here this basically that particular location the object is a building objects is a lake. So, their influences will be there different.

(Refer Slide Time: 29:18)



So, this semantic kriging, kriging is a very popular technique. Semantic kriging has been proposed and used in this case when it has been it can be shown that it has a better prediction; prediction than your the different type of other techniques like lumen, nearest neighbour, IWD ordinary kriging so semantic things.

So, this is a work of a part of a PhD work of one of my scholar mentioned here for our PhD thesis. So, part of a PhD work right, so it has been so see so what we can see. So, we have the data may be satellite data, so missing data or we want to predict. So, base that we used, we use the semantic knowledge to predict the things right and we could see that there are we get lot of interesting results out of it.

So, we will sees few more use cases which will help us in looking at these problems of a likes how these spatial data science can help us in different things. Here like here we have a semantic kriging for prediction we will see other things. So, in our subsequent lecture with this let us conclude our discussion today.

Thank you.