Geographic Information Systems
Prof. Bharath H Aithal
Ranbir and Chitra Gupta School of Infrastruture Design and Management
Indian Institute of Technology-Kharagpur

Lecture - 60
GIS Data Standards

Hello, Namaste. Welcome back to the course on Geographic Information System. In our previous classes we learnt about what do you mean by open source the open I mean what are the different pros and cons of open source and we compared it with proprietary source, we looked at some softwares. Now this class is very specific about data standards.

What are the different standards, what are the procedures, how you represent a metadata what basically a metadata contains. So this is what we would see in this particular class.

(Refer Slide Time: 00:58)

So as I would go through the concepts, the first thing is I would look at what are data standards. What are the standardization levels? There are three levels of standardization, how we have to do standardization. All these are very well established levels. So we will look at that. Then we will look at elements of standardization. So there, there are certain ways of how each and everything has to be standardized. So we look at that.

Then how we can use the data standardization to the application level also. So that is very important when you are actually putting out your projects or you are trying to represent your data in to the scale that it is usable for a decision maker or for any of the users. Then the last part is the metadata. We would look at the metadata part as a last concept of whatever I was speaking today.

(Refer Slide Time: 01:52)

Now when we look at data standards, so the very objective of a data standard if you look at data as a concept, okay data is what you process to form an information. So in order before you get an information data has to be standardized.

Because whenever you are looking at two data that are produced if they are not compatible enough, you will not be able to understand or you will not be able to look at or analyze why that particular phenomena is happening if the two datas are producing the same variables for that particular phenomena. So the best thing is standardize the data, okay. So what basically does now I am the data producer, let us say.

So I am producing the data that is necessary for an analysis, and the end user to interpret and understand the data in the same way that I have produced. So data has to be standardized. Otherwise, for example, if I just represent any of those GIS data that I have collected without giving it a presentation or before without defining what kind of representation has to be there, then that would lead to errors or misinterpretation of that particular data.

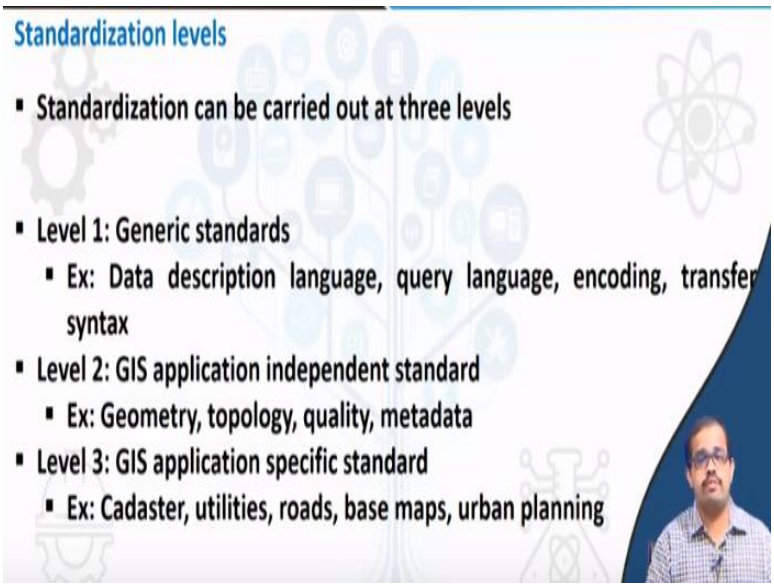
So nowadays, if you can see many of the data is put out, but the way it is interpreted is not exactly correct. So if you have to interpret it in a standard way, that is what is called data standard. That is called standardization, okay. Now when you look at standard, a standard will provide definition for a data structure. So it is a it is defining a structure of the data how the data has to be stored.

Then data content. Content in the data, what does data contain. And the rules that will, first thing is increase mutual understanding of geographic data among users. There may be different users. So the mutual understanding of each of the geographic information that is present in that particular data set will be understood by different users. Eliminate the technical problems of exchanging the data, okay.

So which means interoperability of the data will be there. So that is why the data has to standardize. That is a very important point when you are looking at standardization of data. Interoperability is the important aspect when you look at any data to be standardized. Now once you have interoperated you have created an interoperability then it is increase the integration combination interoperability in terms of data management also okay.

So otherwise your data standard would not match the manage data or the information that you have already created. So these are the three main things that you look at when you are actually standardizing the data okay. One is the data structure, data content, and interoperability.

(Refer Slide Time: 04:53)



Now so once you have done that, so there you may ask me, what are the different levels of standardization? How do you do a standardization? The first standardization level is a generic standardization which most of them without much of an issue can be easily done. That is first when I say a generic standardization, it may be just using a query language encoding or transfer when you transfer the data the syntax that you use, okay.

Then you have data description syntax, the data how the data is actually stored, the way it is stored also is one generic standardization. So data description using a data description language is also standardization. So these are generic standards, the first level of data standardization. Now in once you are, let us say that generic standardization is very important in terms when you have something that is queryable by various people who do not understand the language of GIS, okay.

So but if there are people who are mutually understand the language of GIS and understand the terms and condition then the level 2 and level 3 basic standardization are very essential, okay. When I say level 2 standardization, it means to say it is GIS application independent standard which means independent of a GIS application, the standard has to be maintained, okay.

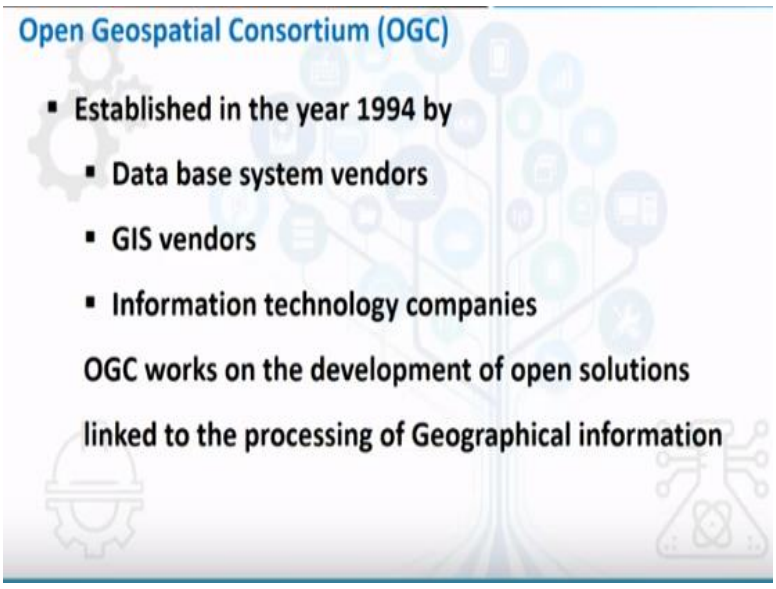
When I use a data this data should be compatible with all the GIS applications and all the applications that we may intend this data to be used. That is what is the level 2 GIS application and independent standard. For example, standardizing the geometry,

the topology, quality, metadata. So these are all comes in the level 2 standardization. Level 3 standardized data is we are very specific to what application we are using.

For example, if we are using a Cadaster for application on a Cadaster level or you application to the utilities, roads, urban planning. So all these applications have their own standards. So you as I previously said, there are certain ways of looking at the data. It is not that every data has to be looked at as a spatial resolution of 1 centimeter or let us say 1:1 ratio.

There are data, which has to be looked at a level of 1: 10,000, 1:5000 to make an effective decision making. But when you are looking at that, so you have to maintain a specific standard. So each of these applications have their own way of looking at the data. So this way, you if you understand how this how you score, choose the scale of level of collecting data and application of the data, then you would be able to easily interpret the level 3 where the GIS application specific standardization is necessary, okay.

(Refer Slide Time: 07:56)



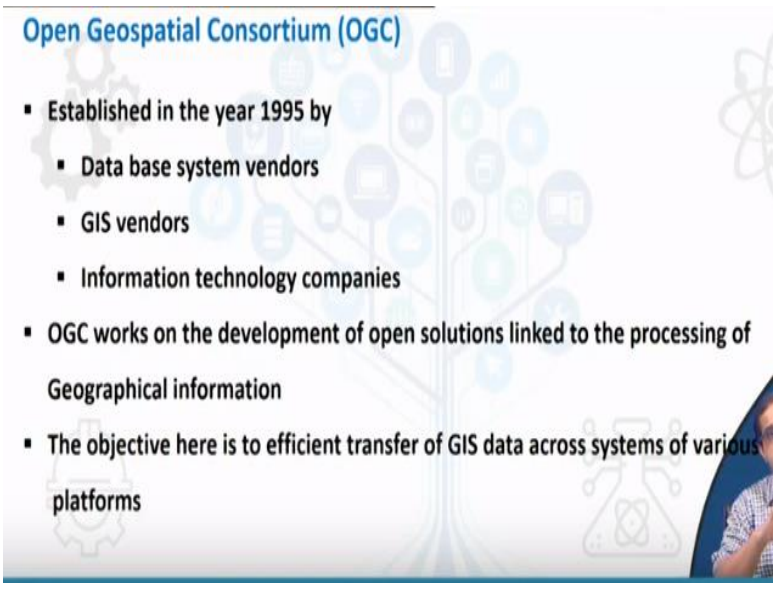
So when you look at this, there exist a consortium. This consortium is works for the development of basically an open solution, developing open solutions, which is essentially towards geography informations which means any data that has geography connected to it okay. So that is where the open source Geospatial Consortium or Open Geospatial Consortium works in a larger context.

So this was established in 1994 by the data system database system vendors, GIS vendors and information technology companies. Why is it basically necessary? It is necessary for various things. The first thing is if everyone has to understand what is a kind of a data standard, what is the kind of data needs and what are the standardization needs, then one has to have a certain common global level in which people understand what is the standard.

That is where the Open Geospatial Consortium was established. OGC standards are well framed at a level of how the data is converted to information, how the data is stored as a database, and how the data is accessed by any of the users. And this is where the open space Geospatial Consortium forms a very important aspect in any anyone's data processing.

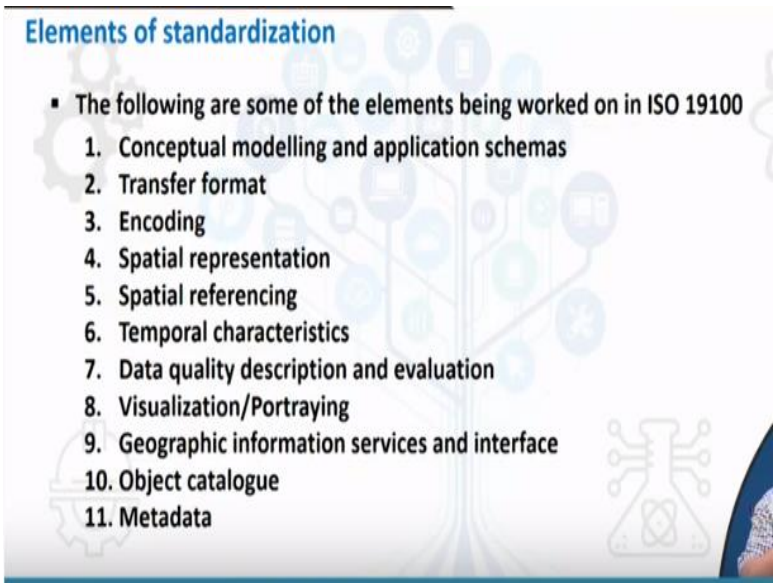
So Open Geospatial Consortium is I would speak about this in detail probably in my next lecture. But as of now OGC is one works for the development of open solutions linked to processing of geographical informations.

(Refer Slide Time: 09:47)



Now when you look at OGC consortium, as I said it was established in 1994 but came to force in 1995.

(Refer Slide Time: 09:56)



So when you look at that it has certain elements or it has listed certain elements of standardization. The first element is conceptual modeling and application schema. So when I say conceptual modeling, it gives you the concepts on how the modeling and application has to build. What kind of schema has to be used and what application it has to be used. It comes under the ISO 19100 series.

And it very clearly determine very clearly defines the schemas of how the application can be developed, data into applications and how the conceptual modeling can be done, how a model can be built basically. Then, it also defines the transferring formats, which means that how a data can be transferred from one maybe one format to the other format or one place to the other place without having any connect with any of the GIS software's in use whether it is proprietary or open source.

Then there is encoding, how the data is encoded. So when I say how the data is encoded, there are certain ways of representing a data, storing a data. So it has to be stored only in that way. So how do we do that is actually represented by in this particular elements of standardization. Then you have spatial representation. How the data is spatially represented. You cannot just represent anything as any object.

It has to be represented in a specific way. That is what we were stressing about how the vector data model is represented, how the raster data model is presented. So this kind of representation has to necessarily come out of how the data has been carried from the field or how the data is being generated and how the data has been

represented. So with that, the spatial representation also becomes extremely important.

The next thing is spatial referencing. See if I have a reference that refers to a geographical version in a different referencing system and there is someone who is referencing the same space in a different referencing system, we will not be able to match both the qualitative and quantitative aspect of both of these data. If both has to be matched, then you have to maintain the same spatial referencing system.

That is what it says with spatial referencing. Then you have temporal characteristics, which is very important in terms when you are looking at any natural disasters or any of the natural or human made effects that is happening in and around. Then you have data quality description and evaluation. This is extremely important. Data quality itself is a biggest thing that one has to look at.

Just generating a data, anyone can generate. For example, if you have just a sensor planted anywhere okay, so that can generate any amount of data that you may need. But the quality of data and the data needs is very important for any user, okay. What kind of data it is giving and how do you evaluate that data? It is not that every sensor will work all the time. I mean it is extremely accurate.

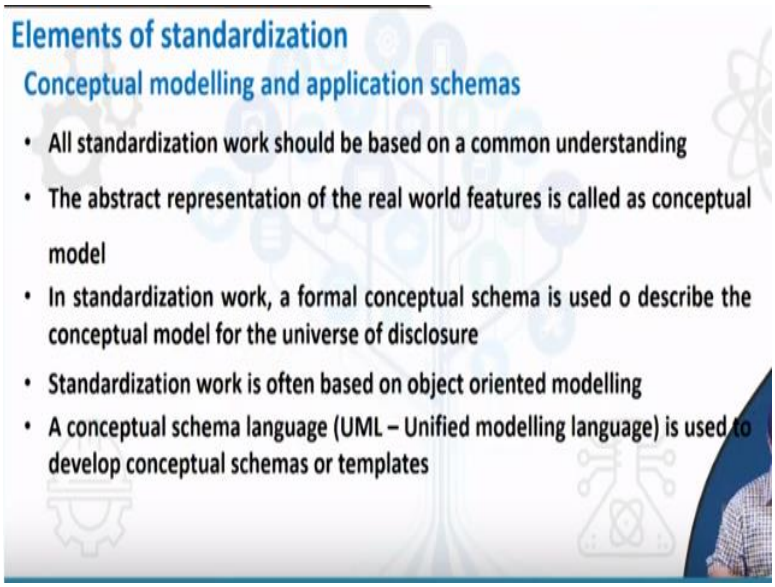
You have to evaluate a particular sensor, particular device and look at what is the error mode. Only when once you have understood the error mode then you add that error to whatever the readings it has been provided. So that error evaluation is extremely important. Whenever you are putting out your data, you have to mention what is the error that data has in itself.

So that also has to be mentioned, without which the standardization or maybe transcending of your data may not be possible at all. Then you have visualization portraying. How do you visualize the data? How do you portray your data, how do you represent your data? That is extremely important. There are certain ways of looking at it for SOI has its own way of looking at it.

And any of the USGS has its own way of looking at it. So you have different ways of looking. It is very country specific as of now. And when you look at Geographic Information Services and interface, this is a well-established standards. Whether it is when you look at services, it maybe web services, it may be desktop services. So based on services, there are certain standards.

For example, now you have web GI standards, the WMI standards etc. So all of these are actually contributing to developing a quality data, okay. Then you have object catalog and metadata. We will look at metadata in our end of this class.

(Refer Slide Time: 14:44)



Now when you are looking at elements of standardization, the very important part is how do you conceptually model it and how the application schemas are developed. So when I look at all standardization work, it has to be based on the common understanding of all of the users. It is not that one user understands one kind of standardization.

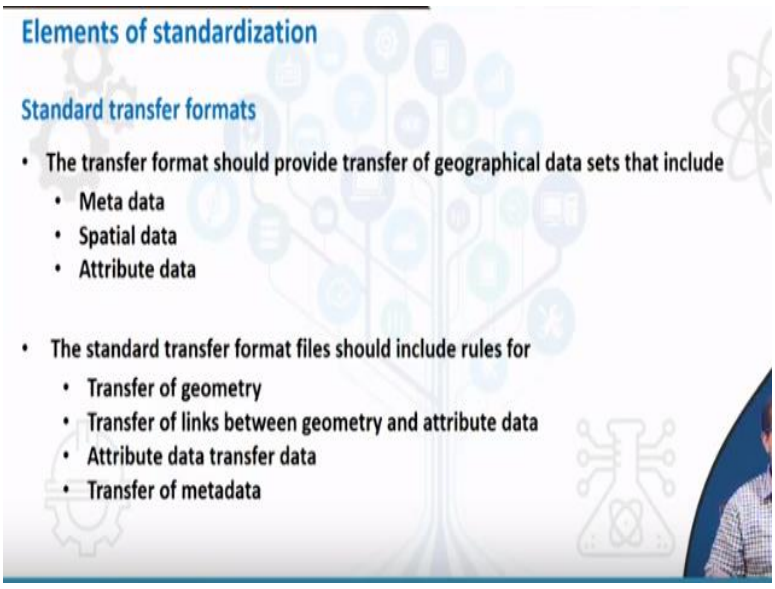
That is why I do this this way of standardization which may not be useful to any other users who are using it. It has to be completely represented in terms of common understanding. The abstract representation or the real word features is extremely important and is called as a conceptual model. The conceptual model is only an abstract. It does not really represent the real world phenomena.

In standardization work a formal conceptual schema is used to describe a conceptual model for the universe of disclosure. So always whenever please look at any model that you have, always a conceptual model as a universe of disclosure is mentioned. Then you have a standardization work, which is often based on object oriented modeling. So when you are looking at every data, every data every model, each of this model are considered to be an object.

So when you are looking at modeling and when you are looking at standardization, both of these two go hand in hand. So when you are looking at both model and standardization, each of these data has to be considered in terms of foreign object. And object oriented standardization should be done in order to have better standardization that is done over a period of time.

Then you have a conceptual schema language, normally it is a UML that is our Unified Modeling Language is used to develop a conceptual schemas or templates which is the order of the day.

(Refer Slide Time: 16:35)



Now once you look at this, there are different standard transfer formats. So when I say standard transfer formats, the very basic three parts take any file for example, let us say a shapefile. Shapefile normally has five different files along with the shapefile, right. So there is an attribute file, there is a spatial file and there is a metadata file. Without these three files, it is not possible for any geographic data sets to be transferred or the data is kind of corrupted, okay.

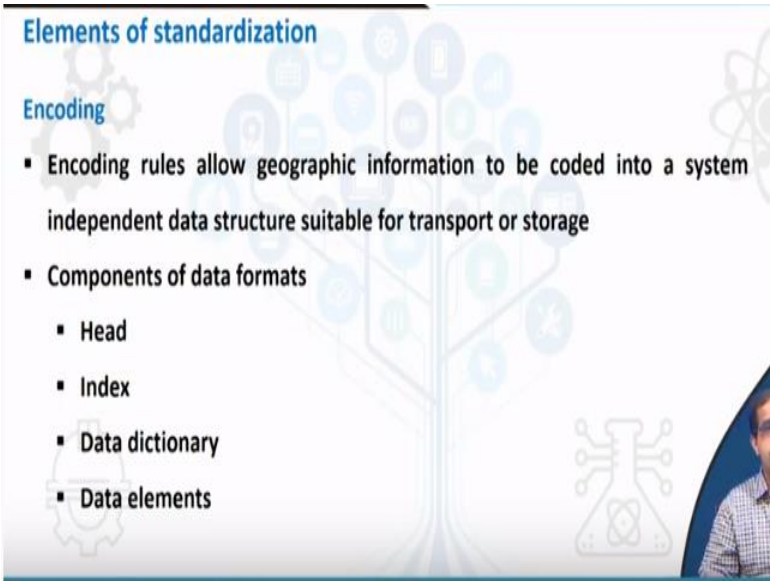
No one you cannot provide a usable data to the user without this kind of standard transfer formats, okay. So when we look at this if you take any of the geotips etc. So you have their own way of representation. There are a series of files which have different types of representation, the way the data is stored in a different files and all of these files contribute together to form a single I mean a single file.

Which is actually your, whatever the data you have created, okay. So as I said our example of shape, you have a database file, and you have one or two more files, which is actually of if without one file also you are falling short of your standard format. Then the standard transfer format files should also include rules. This is very important okay.

So when whenever you are looking at standard format other than the metadata data, other than the spatial data, other than the geometric data, other than the database that is there the very important point is you have to look at you have to include rules which actually defines how the geometry is transferred, okay. Then transfer of links between the geometry and attribute data.

How the attribute data and the geometry are linked. Then attribute data transfer data or attribute data transfer how it is done. So that information. Then transfer of metadata. How the metadata is transferred. That is also very essentially need to be understood or need to be represented when you are representing any standard format. So now I am speaking about the format of how it is data is stored, okay.

(Refer Slide Time: 19:01)



So now in case of storing you have to encode your data. There are various ways of encoding the data. Whether you take any raster image or a vector image it has to be encoded in a certain format. So it means it entail certain components to be included in the data in order to encode it in a specific format. So when I say encoding, encoding rules allows geographic information to be coded into a system of independent data structure suitable for transport or storage, okay.

So when I say that if you see this definition has three different parts. One is rules allow geographic information to be coded which means geography information to be coded you have certain rules. Already that we have mentioned previously that all of this data that is stored has to have rules of the way it is transferred. Similarly, for encoding you will have rules.

Now this data codes into a system with an independent data structure okay, which is suitable for transport or storage. Whether it is transporting that is from one system to the other system or one user to the others user or storage from your desktop to your hard disk or a pen drive. So that is called encoding, okay. So the normally the components of this data would be a head, index, a data dictionary and data elements.

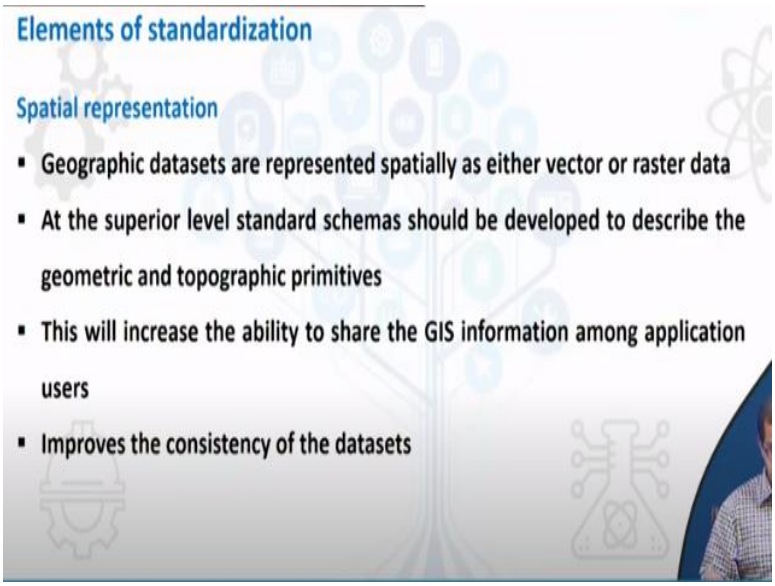
So all of these are very important in terms of having a component having any data stored as a particular structure so if it has to be usable by any of the GIS software. So when I say for example JPEG, let us take JPEG. So the first few lines of any JPEG image will define what is its total header, what kind of index it has, number of rows

and columns which means data dictionary and data elements that are mentioned there in the first few lines of your image.

So which means to say that your program should be able to read that, capture that, then it should know from where it has to start reading an image and how the image is represented. So this is how the encoding is done. So encoding is also very important and it has to be done only in the standard way.

Otherwise it may not be, if I have created the data, if I have not encoded in the standard way, if the JPEG is not saved as the JPEG standards that have been already defined, then I will not be able or any other user may not be able to open it the way the data is actually represented or collected, okay.

(Refer Slide Time: 21:46)



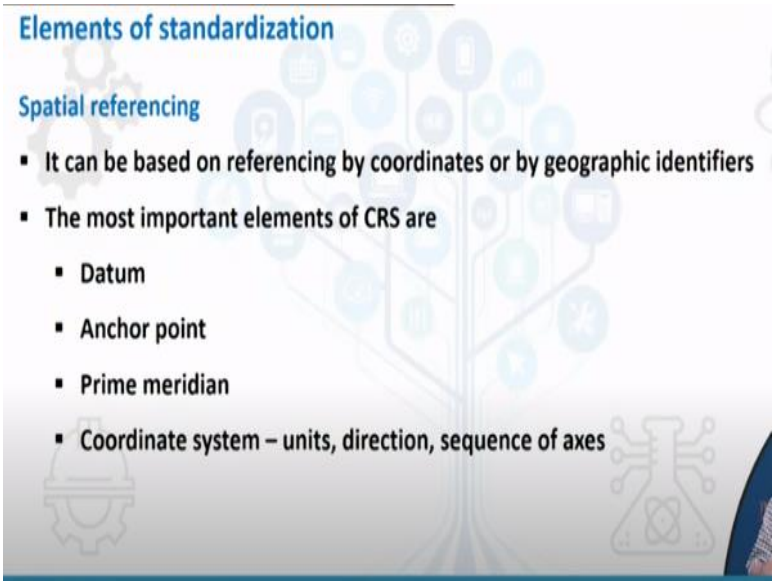
That is one part of encoding. Now the spatial representation. Whenever you are looking at a geographic data set, these are represented spatially either in a vector data model or in a raster data model. So whenever you have a vector data model, so it has its own way of looking at it and when you have a raster data model, it has pixels. The way it is stored is in a number of rows and columns.

So whereas in a vector data it is point line and polygon. So it has its own way of representation. Now when you are looking at a superior level of standard, each of the schema should be developed in terms of to describe the geometric and topographic

primitives. Which means to say that geometric representation or topographic representation, the schema has to be developed so that it represented accurately.

This what it basically does, it also helps in developing the best ability to share the GIS information among most of the application users. And most importantly, improves the consistency of the data sets, okay. So this is how this is about the spatial representation.

(Refer Slide Time: 22:59)



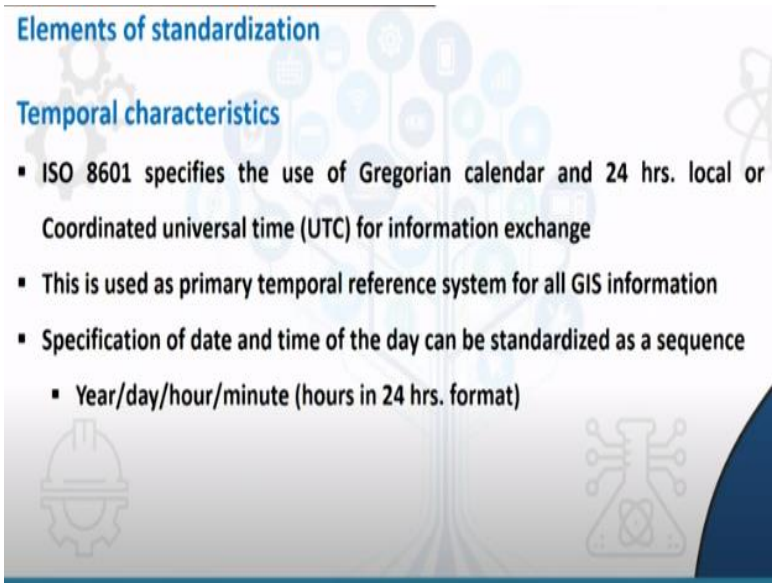
Then there is spatial referencing. Spatial referencing, we have already spoken about this in our maybe in the fifth week of the course. So when whenever you are looking at spatial referencing please keep in mind do not use different systems of referencing because you will not be able to match the data that you have already generated. It has to be in the same referencing system so that it is much easier for a interpreter to interpret.

Also to understand, also to compare, also to generate the different models that may be required for any kind of analysis or addition supports. So when you are looking at this the five most elements of or four most elements of the CRS are you have to look at the datum, you have to look at the anchor point, you have to look at the prime meridian and the coordinate system that is with units, direction, sequence of axes.

So all of these come under that part. So this has to match otherwise it will not be if it is not provided in the standard format, you will not be able to have the data

transferred to anyone. If someone has generated a particular way of spatial referencing it means to say that that entails that other all the users should know the or should be able to access that kind of spatial referencing. Otherwise, you will not be able to open that data at all.

(Refer Slide Time: 24:22)



Or your data may be misrepresented for somewhere else. For example, if it is created in India, it may go to eastern coast of Africa or any other place. So that so you will not be able to find an exact locations on that surface. So match it properly, then it will be much easier for any kind of analysis. Then the temporal characteristics.

So when we look at temporal characteristics, we refer to the ISO 8601 which actually specifies the use of a particular calendar and a 24 hours local or coordinate universe time for information exchange, okay. This is a primary, temporal reference system for all GIS information system. So whenever you are using this, you are supposed to use and coordinated universal time which is UTC for any kind of temporal characteristics.

So specification of date and time of the day can be standardized as a sequence which is as represented here that is your day, hour, minute. So not in the regular style of your date, month, and year. It has to be in this format which is year, day, hour and minute and in 24 hours format.

(Refer Slide Time: 25:37)

Then when you are looking at data quality and quality evaluation. At a very superior level the objective of standardization is to maintain the data quality and handling the data quality with respect to any of the GIS data. So quality statements should consist of following elements. One is what is its positional accuracy, okay. What is its attribute accuracy?

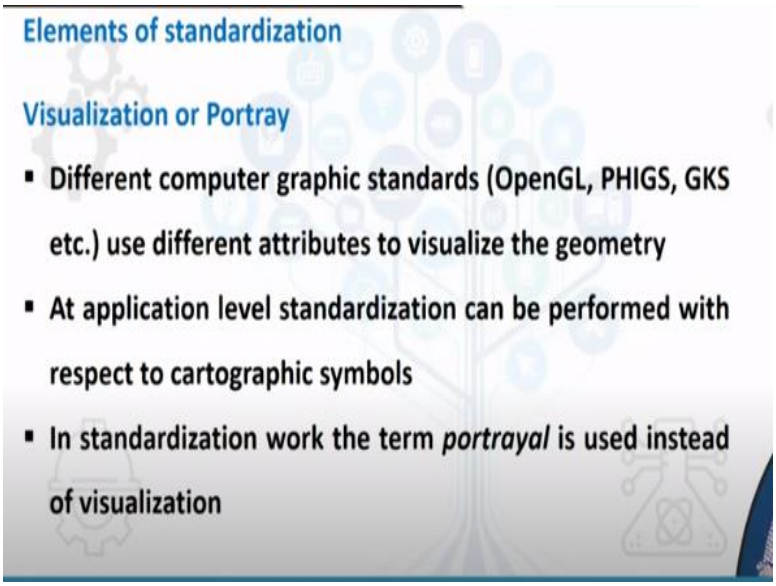
So nowadays what we see in many of the data is the positional accuracy is okay but when you are looking at attribute accuracy is what the concern is. So attribute accuracy should be very clearly understood by the user. Also it has to be understood by the data developer and has to be mentioned, so that the user knows that the attribute accuracy is that he or she can then define what attributes he or she has to collect.

Or how to improve that particular database so that his or her analysis goes on in a perfect way. Then you have temporal accuracy. Temporal accuracy when we are looking at it is essentially necessary in so that all your data are standardized in a particular way or it has the particular repetitive cycle. Otherwise, the temporal accuracy if it is missing, you cannot provide a satisfactory interpretation to how the phenomena is actually changing.

Then the local accuracy. Yes, it is very essential. Completeness of data. So completeness of data is very important. The quality statement without completeness of data is essentially said that the data lacks the qualities that are required. So

completeness is very important when you are putting out the elements of standardization.

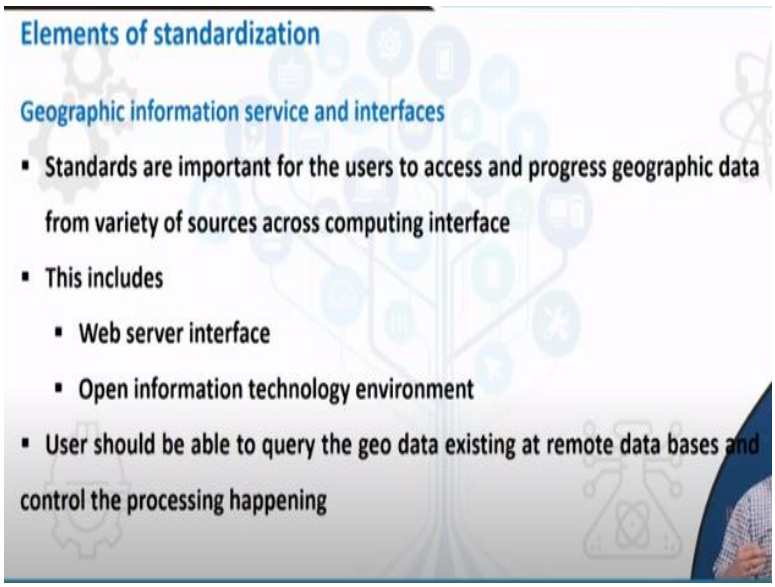
(Refer Slide Time: 27:16)



So then the visualization. How do you visualize? How do you use certain computer graphic standards. Whether it has OpenGL, whether it is GKS. So use of different attributes to visualize the geometry. How do you use it? Whether it has standard attributes or whether it is just used for the sake of using. So at application level standardization it can be performed with respect of cartographic symbols, okay.

We have discussed what is cartographic symbols. Based on those symbols, we use application level standardization. And standardization work, the term portrayal is used instead of visualization normally, okay. So visualization is, is the word, is a common word. Or portrayal is the one that is normally used in terms of standardization.

(Refer Slide Time: 28:03)



And when you look at Geographic Information Services and Interface standards are important for users to access and progress geographic data from variety of sources across computing interfaces. This includes web server interface, open Information Technology environment. So when whenever these things are included, the user should be able to query the geo data existing at a remote databases.

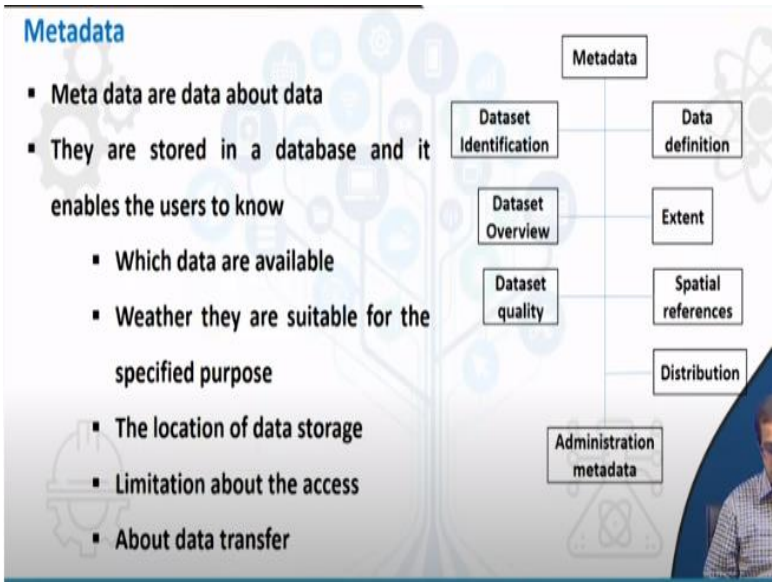
And control the process happening. So that is very important. It is not that just you put in graphs and show the web server interface. But it is very essential that user will be able to access or query that particular geo data and you and maybe use for his or her analysis. So that is very important.

(Refer Slide Time: 28:48)

That is, where the web GIS is today turning into more of more user friendly in terms of how the data is actually represented to the user. Now there are different themes that needs to be standardization at application levels for example, the buildings, the way it is represented, the transport, the utilities, the land use, administrative units, water.

So all of these have certain ways of representation that is already standardized. So we have to just follow that up, so that it is represented properly. Normally survey of India whatever the kind of representation is used, so we normally follow that kind of standard. So heights, control points, annotations, text, styles, trees. So all of these has to be basically used theme wise for standardization.

(Refer Slide Time: 29:37)



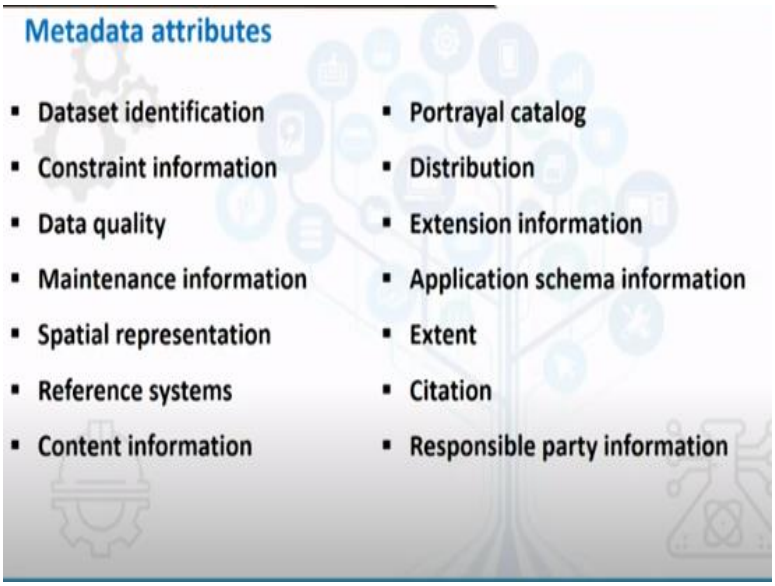
And the last concept that we would understand is the metadata. When I say metadata, these are the data about data. So we have already discussed this. But when you are looking at data about data which means to say that it is actually describing about the who is that data, okay. It gives you every representation.

For example, starting from data set identification to data definitions to data overview to extend the quality, the spatial references, the distribution, the administration of this metadata all of this is handled in a metadata okay. So it is actually stored in a database and enables. It is one of the very important concepts of your data storage, okay without which your data is useless.

It cannot give any information to the user or user will not be able to extract meaningful information from your data. So whenever you are creating data, metadata is extremely important. So what does it basically say? Which data is available, whether they are suitable for a specific purpose. There are data which are created for a general purpose, there are data which are very specific to specific purpose.

So you have to mention that kind of purposes. Then the location of the data storage, how the data is stored, where the data is stored, then the limitation about access if any, okay about data transfer, or what whatever the kind of information that is there about the data transfer also has to be mentioned in the metadata.

(Refer Slide Time: 31:05)



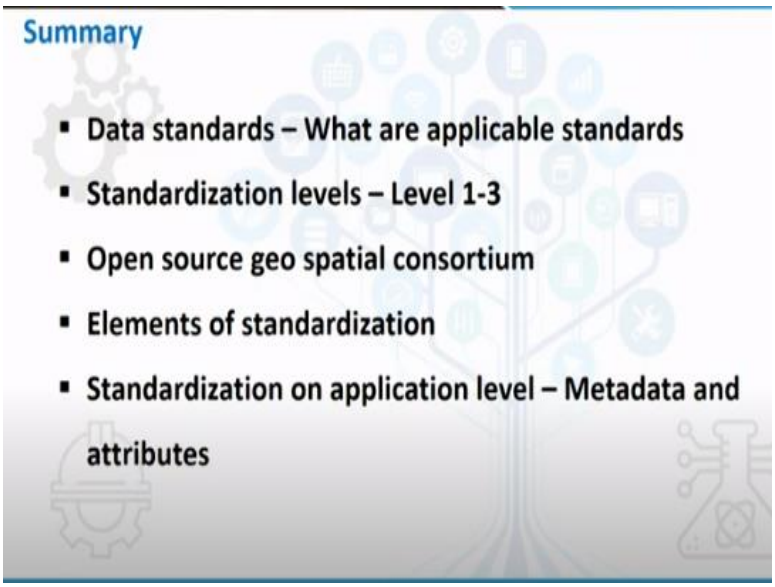
So when you are looking at metadata it has certain attributes. Now when I say attributes these are the information that a metadata data normally has to be accompanied with. For example, the data set identification. What kind of data set is there and how the data set is represented. If there is constraint, then you have to have a constraint information. Then the quality of data.

So as I said previously the quality of data is extremely important when you are actually representing a data. So quality, its coordinate system, its identification and the spatial representation this is the overall thing that basically it is that is necessary. Other than this, there are certain things like the content information. What kind of content that this particular spatial data has. Then the portrayal catalog, how it is visualized?

What are the different tools for visualization. Then distribution of how it can be distributed. Then extension information, application schema, that is very important. Otherwise we will not be able to put out the data as easy just as it is. Then extent, what is the extent of that particular thing. So you have to look at the metadata to understand the extents also. From where to where this metadata, this particular data is being represented.

The citation. So who has created this particular metadata and what does it represent. Then the responsible party. If you have any questions to that particular party, so who is the responsible party that has to be very clearly mentioned.

(Refer Slide Time: 32:37)



So this is about the metadata. So today we did understand what are the data standards, what are the applicable standards that are there. Then we looked at standardization levels that is from level 1 to 3 how it is actually standardized. Then we looked at open source geospatial consortium. We will look more into OCC standards, OGC and OGC standards in my next lecture. Then we looked at elements of standardized.

Then standardization at application level that is using metadata and attributes. So we have actually covered how the data standards or the data has to be standardized in order to put it out across to different users. In the next class, we will look at more on OGC. How OGC is organized, what are the different ways of representation. So all of the things of OGC we will look at in the next class.

And similarly, we have something called as NSDI. So government of India also has initiated the standardization of data so that data is easily accessible to everyone, shareable to everyone, and most importantly, usable by everyone. So we will look at how the NSDI also has evolved, what are the different terms that it also refers to. So we will look at that also in my next set of lectures. Till then, have a nice day. Thank you very much.