

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

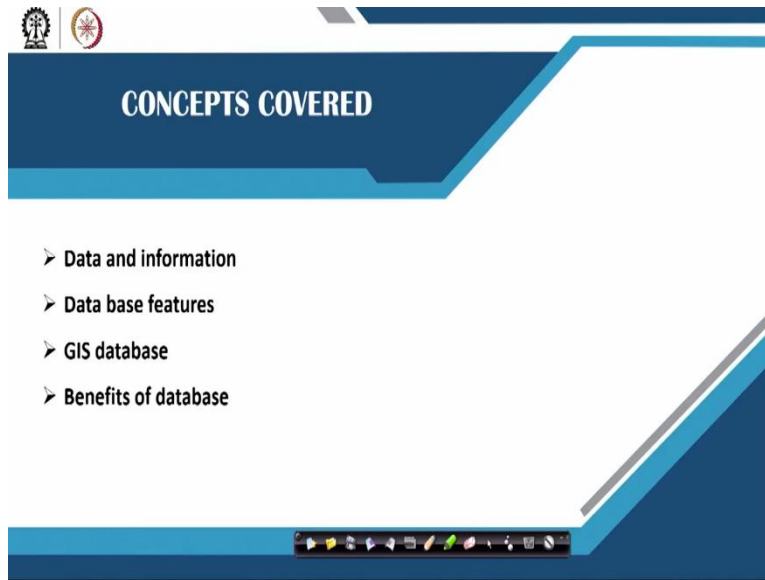
**Module No # 07**  
**Lecture No # 31**  
**Introduction to Database**

Hello Namaste and welcome to the course on geographic information systems this is about module 7. Till now when we started our journey from real world we actually started it from if we look at we started from the real world we started looking at how we collect data. How we integrate data? Now once you have the data in your hand how do we actually store it in a database? How do we retrieve in a database and what are the different database that are there?

And finally how do we process it to form information that is what we would be learning about in this particular module. And to be very specific we would learn what is a database so once the data comes in how do we store it? How do we retrieve it? And most importantly what are the different ways of storing it different database that are there and if time permits we will also look at what are the languages that are being used in order to manage this database or query this data from the database.

So this is what we would learn in this particular week and the next week. So both of this week would be packed with databases. So and probably in the last week you have also heard about QGIS a introduction about QGIS. So Prakash did speak about what is QGIS? How QGIS has evolved over the period of time and he did show you some tools that exist in quantum GIS. So if you can start looking at those tools, if you have downloaded QGIS or in the practical session they were my TAS should also show you how do you download how do you utilize the QGIS for analysis of data. But if you can download QGIS and start looking at it probably database would make more sense in terms of how it is implanted.

**(Refer Slide Time 02:29)**



So in this class we would look at the concepts for the first and very foremost concept is data. What do you mean by a data and what do you mean by information? And most of the time was most students term data and information to be the same. No data is different, information is different. So we will look at what is data? What is information? What is the difference between both of these? Then database feature.

If you want to have a database so what are its features. Then we would look at GIS geographic aspect or the geographic angle of the database. How it actually differs from a normal database. Then we would also look at if there are any advantages disadvantages of database. If there are advantages what are those advantages and how it may help us in bettering ourselves in terms of maintaining the files.

So entire instead of making hard files how do we actually maintain a database so that it would help. We most of us today have access to the at least most of them who are taking up this course I would not say all but most of them have an access to the digital access to most of the ticketing system backing system etc., So those are also have developed over the database itself the communication of data, database and as information.

So all of these put together also is a database. So in fact most of us are looking at everyday activity in terms of database itself. So let us learn some of these so the this particular session now the next 2 modules are just an basics of a database. It is targeted to those people who are

understanding GIS and those who do not understand even the basis of the database. For people who want to learn advanced data databases I would suggest them to take up the course which is completely dedicated to how the database? What is the database and how does it work.

So more of that comes but this is just a basic just a give just to give you a brief flavor of how the database works. So that it is easier to understand when you are implementing it on using a GIS software.

(Refer Slide Time 04:55)

**Data and information**

- Data consists of a series of facts that can be collected, stored, processed and recorded.
- Data is a collection of values
- When data gets organized it becomes information.
- When processed it helps to draw generalized conclusions

<u>Data</u>	<u>Information</u>
Raw facts	Data with context
No context	Processed data
Numbers & text	Value based data

The slide features a background with various icons representing data and technology, including a gear, a smartphone, a laptop, a network diagram, and a molecular structure. The NPTEL logo is visible in the bottom left corner, and a navigation bar is at the bottom.

So as I said when we come to something called as a data and information. So now you go to the field you have your GPS in your hand you go to a field you collect the data. Now you come back and these will be a series of information that is collected. So maybe that you have gone around the entire city collected the footprints of the various building looked at different aspects and recorded all of those.

So data is nothing but you have collected a series of facts that can be either sorted or processed or even recorded okay. It can finally once it is process it can be even stored. So when we define data it is the one that is collected based on facts, phenomena or any of those aspects that can collected stored processed and recorded. Data is actually a collection of values it may be based on perception it may be based on various other various issues. But these are the collection of values that maybe in the form of a collected or it is from the perceived data.

But when data get organized ok and processed it becomes information. So now you have data ok let say that you have collected all the points in your GPS you have come back to your lab ok then you start working or you import all those points in into your computer system. Once you have imported all the points then you will start looking at what are the different features maybe you have a background as a google map or a or any of those reference values.

So now what you will try to do is that you will start disguising these points and say that this what I have collected this is what is a quantitative value this a qualitative value ok. Once you have collected it you will process to obtain certain information out of it maybe using a raster processing using some other raster data or satellite data or any other data ok you just get some value sort of it a qualitative quantitative information.

And you process that information now that becomes an information. Once you have process it that becomes a information. So those are the values which are stored as a information ok. So when this information is processed normally it gives a generalized conclusions. So you should be very specific so not always the information is very specific conclusion it is most probably a generalized conclusion of all the data that you have concluded.

You would not say that I have taken this particular building this particular site that particular building. So if you are pinpointing each of this information storing as a different value then it is a specific information. Otherwise it is a conclusion of all the data that you may have collected and stored in your GPS and processed in your computer system. Now when we look at data and information, data is about raw facts ok.

So they may they most probably most of the data would be a context ok. But when you look at it, it may be numbers, it may be text, it may be GPS traces, it may be your just in photograph information's or it may be your high altitude low altitude photographs that you may have taken at that particular of time. But once the data comes into your lab and comes into your system you connect the data with the context.

You will say this particular building has is somewhere next to this particular place in this particular street ok. So it has this is the latitude longitude and this is street and this is where you collected the data okay. And the data is all of this data it has seven floor it has maybe some of

those data you have collected. So it is a data bit context ok. Then when you look at information it is the one that is processed after you have got data.

So when you look at information it is most it is information it is a processed information and the generalized information generalized conclusion. Normally information's are value based data when you look at data these do not have value to it, it has to be processed to get a value. So information's are those which have value based. So that is the different between the data and information. So when you are speaking next time to anyone so be careful that how do you define a data? How do you define our information? Okay.

**(Refer Slide Time 09:54)**

The slide is titled "Data and information" and features a blue header. It contains two bulleted lists defining data and information. Below the lists is a flow diagram showing three "Data" boxes pointing to a "Processing" box, which then points to an "Information" box. The slide also includes a small video inset of a man in the bottom right corner and a navigation bar at the bottom with the NPTEL logo.

**Data and information**

- Data can be any character, number, images, words, text
- Data is based on records and observations

---

- Information is on analysis of the data
- Information is organized and processes

Diagram: Data → Processing → Information

So to go ahead with the same thing data can be any character, number, images, words so it can be anything. It can be based on records and observations also. Many a times most of the data's collected based on records based on various information from the governmental records. And certain observations there may be hundreds of observations that would have people would have done over the period of time and recorded. That can be converted into a digital form that becomes a data. So this is what we mean by a data.

So when you look at the information is an analysis of whatever the data has come. Information is organized and processed. So information is always organized ok and this and entire has certain process. So when we give a generalized structure for this you have huge number of data here ok.

We have lot of data and when we process it, it forms an information this is generalized form of how we say how a data is converted into a information.

**(Refer Slide Time 11:05)**

The slide features a blue header with the word 'Database' in white. Below the header, there are two bullet points: 'A database is defined as the organized collection of data, stored and accessed from computer system. A database can also be on physical records' and 'An efficient computer database requires proper design and modelling techniques'. In the center, there is a graphic of a blue database cylinder with the words 'Data Base' written on it. The background is white with various icons like a gear, a tree, and a person. At the bottom, there is a black bar with the NPTEL logo and a navigation toolbar.

Now when we look at a database now we have understood data next is the base. Which means to say that it is a storage system where for that particular data. So when we look at database, database is an organized collection of data stored and accessed from a computer system. A database can also be on a physical record. So that also can be a database for example the file for example there are a attendance systems in many schools or in many institution where physically the attendance is mark that is also a database ok.

That is a physical records that have been maintained over a period of time. If the same attendance system is maybe put it maybe put it may have been stored in the form of a thumbprint. When, many of institutions now have sensing system where thumbprints are used to give attendance so that there is no inconsistency in the attendance system. So if that is the case then you have a digital form of the database. So you can have a physical form you can have a form that is stored in the system also both are called as a database ok.

And when you look at computer database these are more efficient and if they have to be efficient it has to be properly designed ok. So if has to be properly designed then the correct modeling techniques have to be involved in terms of designing of that particular database ok. So when I

say a database when I so now when I say a database it is where your data is stored to be processed that is nothing but a database as simple as that ok.

So you have your attendance system where to add a particular teacher or tutor marks your attendance and that particular attendance may be processed at end of your examination or when you write your exams to see how much is your total attendance whether you have a shortage or not. So that is nothing but a database ok.

**(Refer Slide Time 13:15)**

**Database system**

- Database systems are designed to store large bodies of information. Its an organized body that stores related information
- In addition, it must provide for the safety of information stored, in the face of system crashes or attempts of unauthorized access
- Database and database systems have become an essential component in everyday life. In the course of day most of us encounter several activities that involve some interaction with database

The slide features a blue header, a white background with faint icons, and a small video inset of a man in the bottom right corner. A navigation bar with various icons is visible at the bottom.

Now well if the shortage information is given how much the percentage of shortage you have then it means to say that it is nothing but information that is processed and given from that particular database. Now when we look at the database system so when I say system it is software, hardware etc., It is a entire system where the database comes in is processed is then delivered. So that is nothing but a system ok.

So when I look at the database system these are normally designed to store large bodies of information. It is an organized body that stores related information's ok when I be very I am very specific here stores related information not junk ok. In in addition it must provide for the safety of information stored this is a very essential in today's context. So safety of the information that is stored is essentially very integral part of database development.

If you are if you are able to give a lot of security aspect to your database then probably your database would be more secure in terms of illegal surfing that may have happened on your database. So in the face of a system crashes or attempts to unauthorized access you may have to save it. So you may have multiple locations of looking at it are those most of the database now nowadays are stored in in cloud connected systems.

So that any even in the physical systems that are that the user uses if it is if it actually shuts down or under repair or system crashes. So you will have always have you will not have any issues when you are handling that systems or always a backup of that database is made always and now may be a backup on any of the cloud systems are made. So when you look at database in a database system so database is where it is stored it is a way it has all the algorithms it has the way it has to be stored.

But the database system is a collection of how this database is organized. And when we look at it data and database system these have become an essential component in everyday life. When you look at for example we encounter several activities when we look at interaction of the database very common activity that you may see is for example most of us have looked at Amazon, would looked at Flipkart.

So when you are actually surfing all those items on this Amazon, Flipkarts, Smart deal, Snap deal all the stuff. So what are you trying to do is that you are actually accessing the database to look at what are the items that are there for sale ok. So different vendors so now you have different vendors each of these vendors have different items for sale. And each of these items have their own way of representations ok. Own constrains, own way of representation, own policy etc.,

So each of these become a database ok so data in a database. So now you have several data's in the database now you select whichever you need you then connected to the database which wherever the Amazon server is located. So you talk with the database and say you want to buy this and that once you have connected this it ask for your address that may be stored again in one of the server of Amazon ok or one of the cloud systems of Amazon.



So now once that is done then it redirects you to your payment gateway which maybe some of the backing servers where again your database is stored. So when you are giving your particular UPI address or your net banking value it is going into that particular server accessing that database only the things that are visible to that particular server is accessed and when once you give authorization it deducts your whatever the value that has to be deducted.

So this is how the chain that normally works. So that is what you are trying to do in everyday life or even if you are going and trying to look at your mini passbook in an ATM and you are trying to take a printout ok. You can take a printout of whatever is last 10 transactions. So what are you trying to do? You are actually trying to access the database collect the information of all the transaction that has happened in last maybe in last 10 transactions.

That is nothing but you are querying in the database you have stored some information bank has collected data about you and has stored certain information there based on your transaction and you are accessing that database in order to get that information that is exactly what the database system does a database does and how the information will be ok. Most of us are involved in this digital era in terms of daily interactions with the database and database system.

**(Refer Slide Time 18:29)**

**Database system examples**

- Transactions in banks
- Online railway/airline/hotel Reservation
- Computerized library catalog to search for a Bibliographic item
- Institute/college ERP
- Online examinations and results

And if you want to look at very good examples of the database system banking system is a very good example of a database system you could even see that there is a lot of issues that is happening sometimes a their system is down there is a upgradation or some hack someone has an

unauthorized access to that system. There is a online railway now if you look at the railway system with the digital network that has improved and the digitization that has improved.

You can see the previous the railway platform today's railway platform is much better in terms of user friendliness and how you actually start looking a booking a particular ticket on that particular IRCTC website. So similarly different website given have been given an access in order to book that particular railway ticket it means to say that there are multiple instances which are actually looking at that particular database.

So this is a wonderful thing of how database can help multiplicity of user in the they will normally do not have any counter number of users who can get into the database but they look at the security of information the amount of information that has to be leveraged and amount of information that is necessary for particular operation to happen. Similarly when you look at hotel when you look at airlines so all of these have a certain ways of interaction of database there is very well known system.

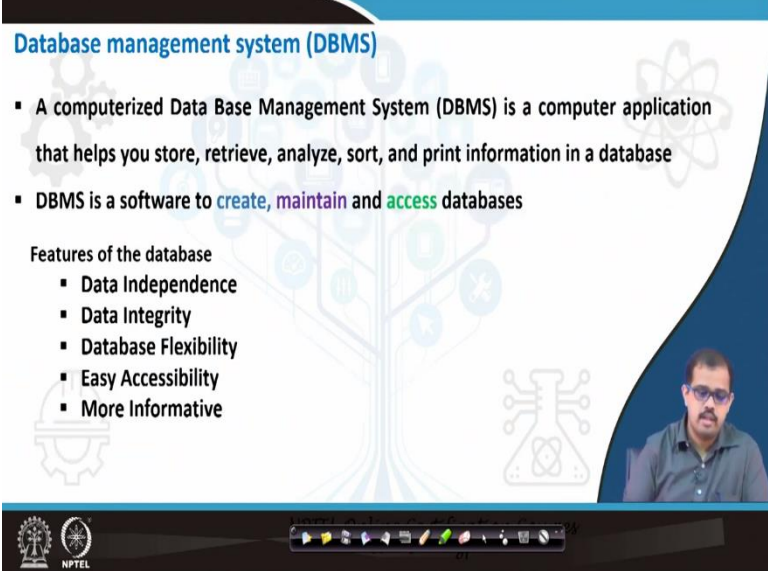
If you look at some of the libraries of your schools which are completely computerized where you go you look at your title of the book and where the shelf that is actually are kept. So that is one way of accessing your accessing a database of that library of that institution. Then some institute have ERP and probably in some days IIT Kharagpur is going to develop ERP for the most of the institute across India or a platform that is generalized platform for most of the institute.

You can see there is an this image is showing the and ERP system of IIT Kharagpur which again a entire thing is a database. It has student information, it has faculty information, it has information about different aspects of a particular department you have certain accesses given to certain faculty, certain accesses given to students, there is certain access only given to administrators. So this also is evolving over a period of time and giving and helping it to become more of a paperless transaction over a period of time.

So well this is one very good example that even IIT Kharagpur is having a very efficient system in terms of ERP and other one other very good example as nowadays the online exam has come up and the results that show up. So these are also a good example of how the data is processed as

a database and how do you have the information in terms of results that could have been processed.

**(Refer Slide Time 21:43)**



**Database management system (DBMS)**

- A computerized Data Base Management System (DBMS) is a computer application that helps you store, retrieve, analyze, sort, and print information in a database
- DBMS is a software to create, maintain and access databases

Features of the database

- Data Independence
- Data Integrity
- Database Flexibility
- Easy Accessibility
- More Informative

The slide features a blue header, a white background with faint technical icons, and a video inset of a man with glasses in the bottom right corner. The NPTEL logo is visible in the bottom left corner.

Once we have understood what is a database? What is a database system? Now database management system so you have to manage that database right. So in order to manage that database it is a computerized management system when I say database management system data management system ok. So it is an application that helps you to store, retrieve analyze, sort, and print information in a database.

So it can be also called as a software or it can be called as an application to create maintain and access a database ok. So that is called that is nothing but a database management system. Now data is different, database is different, database system is different and database management system itself is different. So it is an application so you should understand all of this concept and then only you will be able to handle most of your data in terms of attribute data that you have for any of those calibration that you want to make.

Then you have a features of databases. When you look at features of database the very important features is data independence. So when you look at the entire database data independence will become a very important part of any database. So most data that is stored on data are independent in terms of the way they are stored and data integrity. So that is another aspect when you look at the integrity of any data is very essential in terms of maintaining the entire database.

So data integrity becomes a very important feature of the database then you have flexibility then easy accessibility. So whatever the example that I give are very good example of easily acceptable. So where once you are connected to that database through your 4G network you have easy accessibility of whatever the information that you need ok. If you need something administrative information definitely you have to go the bank.

But until unless you are at your own access level so you can get all the information that is necessary which is easily available and most and very important aspect to user is these are very information in terms of information that is required. The big informative database system that you can always most of us always look at is the Wikipedia ok. If you look at the Wikipedia part of it these are huge number of users who actually maintain the entire database who edit regularly the database who add information to the database ok.

So that is accessed by us as a processed information. So that information is then can be used by any people to understand what is that particular concept. So a very good example of database system ok.

**(Refer Slide Time 24:33)**

**Database management system (DBMS)**

Advantages of the database

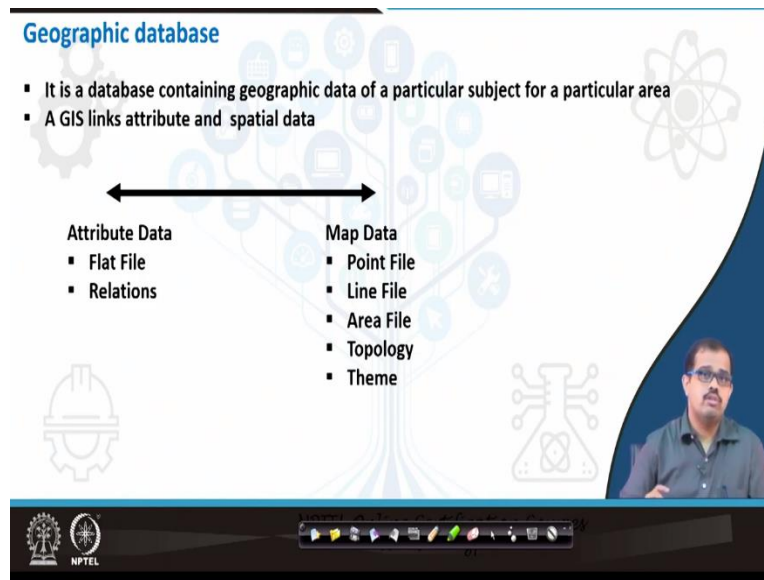
- Non-Redundant
- Data Independent
- Program Usage
- Data Inter-Relationships

The slide includes a central graphic of a blue cylinder labeled 'Data Base' and a background of various icons representing data and technology. A small video inset in the bottom right corner shows a man speaking.

So when you look at a database management system we have certain advantage of when we look at database also. one is it is non-redundant, data independent they are not worried they are not dependent on what kind of data it is. Then program usage that is also quite essential it is

advantageous in terms of different programs in usage. Then inter relationships is what matters you when there are a lot of interrelationship among the data, among the database and among various sets of database so it is much easier to handle in terms of relations that exist across the data sets.

**(Refer Slide Time 25:11)**



So when we look at database and geographical context so both of these are quite related in terms of how they are being processed. When you look at that database it can contain a geographic data of a particular subject of a particular area. So when you look at the difference between the database and the geographic data for example if we look at this specific thing they are quite related in term of how they are stored.

But when you have data when you have GIS it actually links whatever the attribute information you have with the spatial information ok. Spatial information is stored in different models so you have different models and connected attribute data. So when you look at attribute data these are normally flat files or files that may have numbers or just information quantitative qualitative data which are there. So these are just attribute information's. Then these attribute data can be related in form of a relations equal to greater than less than or it is related to something else.

So these are the two things that attribute data may have. But when you look at map data that is in the spatial form has point line polygon it may have a topology that is built in already, it may have schematic information that has been built or processed over a period of time. So combining both

of these and putting it on the database is nothing but geographic database or GIS database. So connecting a geographic information with the attribute information all the keys that are there in the entire database is nothing but you are using GIS database ok.

We will learn we will also use we will create ourselves a particular vector layer then we look at how we can manipulate that then how can we store it as a database. Then how will we manipulate that particular database. How will you use attribute data to quantify certain things like area etc., Then how do we actually show that attribute data as a factors in terms of relating to your unground situation. So all of these will look at it in the practical class wherein we would do an hands on each of this aspects.

**(Refer Slide Time 27:35)**

**Advantage of database over files**

- Avoids redundancy and duplication
- Reduces data maintenance costs
- Faster for large datasets
- Applications are separated from the data
- Applications persist over time
- Support multiple concurrent applications
- Better data sharing
- Security and standards can be defined and enforced

And when we will look at advantages of database over files the very important advantage is to it avoid redundancy ok. Let me go to the next slide or maybe next set of slides which actually tell you what kind of redundancy may happen ok. But as of now let us think that it avoids redundancy and duplication in terms of for example if you have a file for a particular purpose now you keep on if there if someone does not know and creates another file in the same context so that becomes a duplication of the same set of files if the papers are different it is redundant file but if the papers are same then it becomes a duplicate file.

And when you reduce the data maintenance cost that is very important in terms of maintaining you data then faster for very large datasets if you have extreme with today's computing

capability it is you know when you have a good database design in good modeling system then it is extremely faster applications are separated from the data very different application are different and data is different, a database is different.

So application persist overtime so it there is no constraint of time then you have a support of multiple concurrent application. So you that is we look at how the multiple applications can run over a period of time. Then you have better data sharing ok so you can share your data in a better way in terms of access of data and sharing of data databases provide extreme flexibility security and standards can be defined and enforced.

As I said in today's context security and standard security is a extremely important aspects of sharing a data. So then now the concept of standard has also evolved over a period of time. Standard is extremely important in order to have the data for example the way first of all the data is saved. When you are saving it in in terms of a particular file format so those file format have also a standard way of recording that data.

So these are normally under the convention of OGS open geospatial standards. So open geospatial consortium standards based on these standards these data has to be stored ok. So if we look at these standards there it has evolved over a period of time and India is actually also looking at how we have maintain enforce this standard in terms of the development of the spatial data.

And also when we look at the interoperability of all your data all your tools that you have developed. So they need to be interoperable in all the environments in all different forms. So to make it interoperability you should need interoperable you need certain standards. So these are those standards that have also been evolved over a period of time. So both of these go hand in hand when you look at both security and standard. If standard is maintained there is a some part of security aspect is also taken care of in terms of development of that particular database.

**(Refer Slide Time 30:55)**

**Disadvantage of database over files**

- Expense
- Complexity
- Performance – especially complex data types
- Integration with other systems can be difficult

And when we look at always any system has certain disadvantages when we look at a disadvantage of a database it is expensive that is the maintenance expense. So if you are looking at a long term it may not be expensive but if you are looking at a short term measure database may be quite expensive and complexity involved. Only on until unless you thoroughly about how you maintain a database. What kind of database it is?

How it have also evolves of period of time? What is the data that is fed in? How you are connecting it with relationship what kind of relationship Etc., If you are not able to understand as a computer engineering so then the complexity is quite high. Once you have understood the entire complex database then it may be a just child's play. But until unless it is understood it is highly complex.

Then performance especially the complex data types handling the complex data type is extremely I mean it is a huge task. When it comes to spatial data it is really task that is the order of the day. So if you are very good at handling complex data in a very big complex database then probably that is the best thing that you can do in terms of GIS. Then integration with other systems can be difficult in many times.

So that is why we are now the standards have been developed if it is evolving over it periods it is not quite developed but it is evolving over the period of time at maybe in some point of time



most of the integration can be done from any system to other any other system. So it is evolving but it needs some time to evolve to get to that extent.

**(Refer Slide Time 32:44)**

The slide features a central graphic of a tree with various icons (gears, a smartphone, a laptop, a server, a globe, a magnifying glass, a gear, a network diagram, and a chemical structure) on its branches. The word 'Summary' is written in blue at the top left. Below it is a bulleted list:

- Data v/s information
- Database
- Data base systems
- GIS database
- Advantages and disadvantages of database

In the bottom right corner, there is a small video inset showing a man with glasses and a dark shirt. At the bottom of the slide, there are logos for NPTEL and a set of navigation icons.

So in summary we looked at what is a data? What is information? So we looked at data was this information. We look or we also looked at what is a database? So we started with data then looked at the database then we looked at what is a database system then once you have looked at database we have also looked at database management system ok. So then we covered how a GIS and the database matches each other.

How a geographical input and a database management system are can be inter linked. When I say geographical system it is point line polygon topology or any of those feature and when you look at database it is exposed of attribute data. How the attribute data is interacting with each other through relations. So when we combine both of this these forms at the geographical database. And then we looked at what are the advantages and disadvantages.

So when we come in to the next class probably we will look at more into how a database management system works and what are the different databases? What kind of relationships are there? What are different forms of storing your data and how we access this particular data? So let us meet in the next session next lecture thank you very much.