Learning Analytics Tools

Professor Ramkumar Rajendran

Educational Technology

Indian Institute of Technology, Bombay

Lecture – 2.4

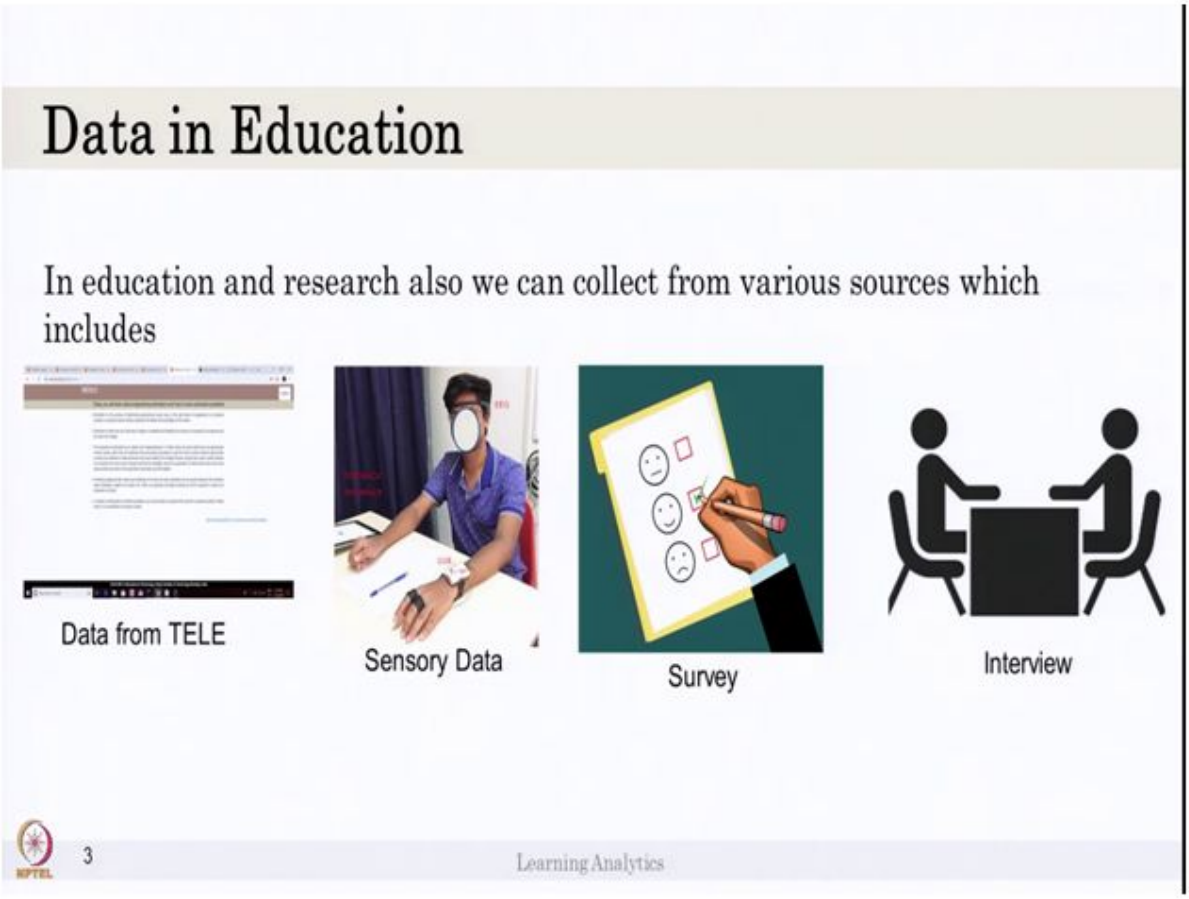
Ethics in Learning Analytics, Student Privacy

Welcome to Learning Analytics Tools course. In this lecture, we will talk about Ethics and Privacy in data collection.

(Refer Slide Time: 00:31)

In the digital age, that is Web 2.0, we have access to data and we can collect a lot of data and also we have access to a lot of data on the internet and we can use this data for analysis and to provide recommendations and everything.

(Refer Slide Time: 00:41)



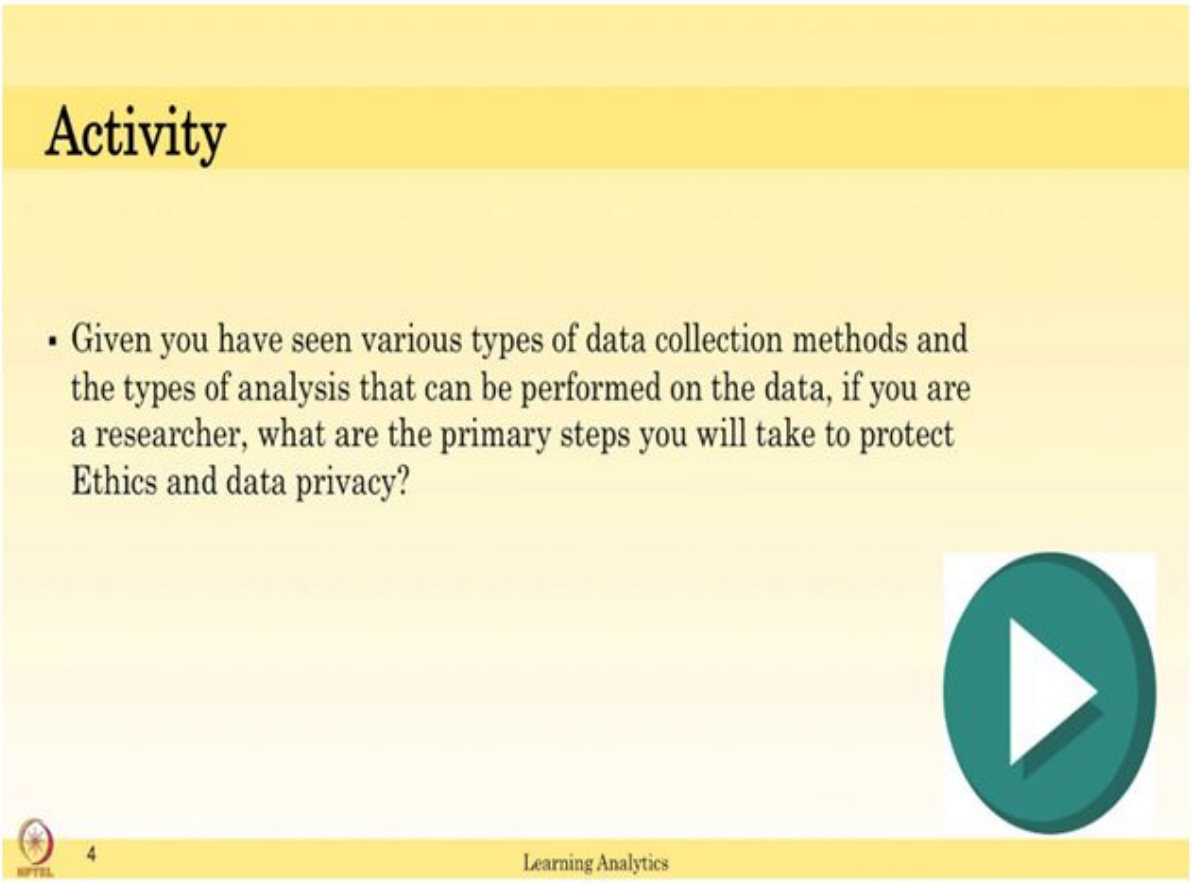
Similarly, in education, we talked that we can collect data from Technology Enhanced Learning Environments such as we discussed “MettLe” or we can collect data from sensors like a GSR or Eye Gaze detection or you can collect data from survey questionnaires.

For example, the pre-test, post-test, survey, performance survey, survey questions like, what is, how the system looks like, a lot of data you can collect from the student’s perception also. Apart

from these data, we can also conduct interviews after the interactions with the system or before the interaction or you can simply conduct interviews to understand the learning process.

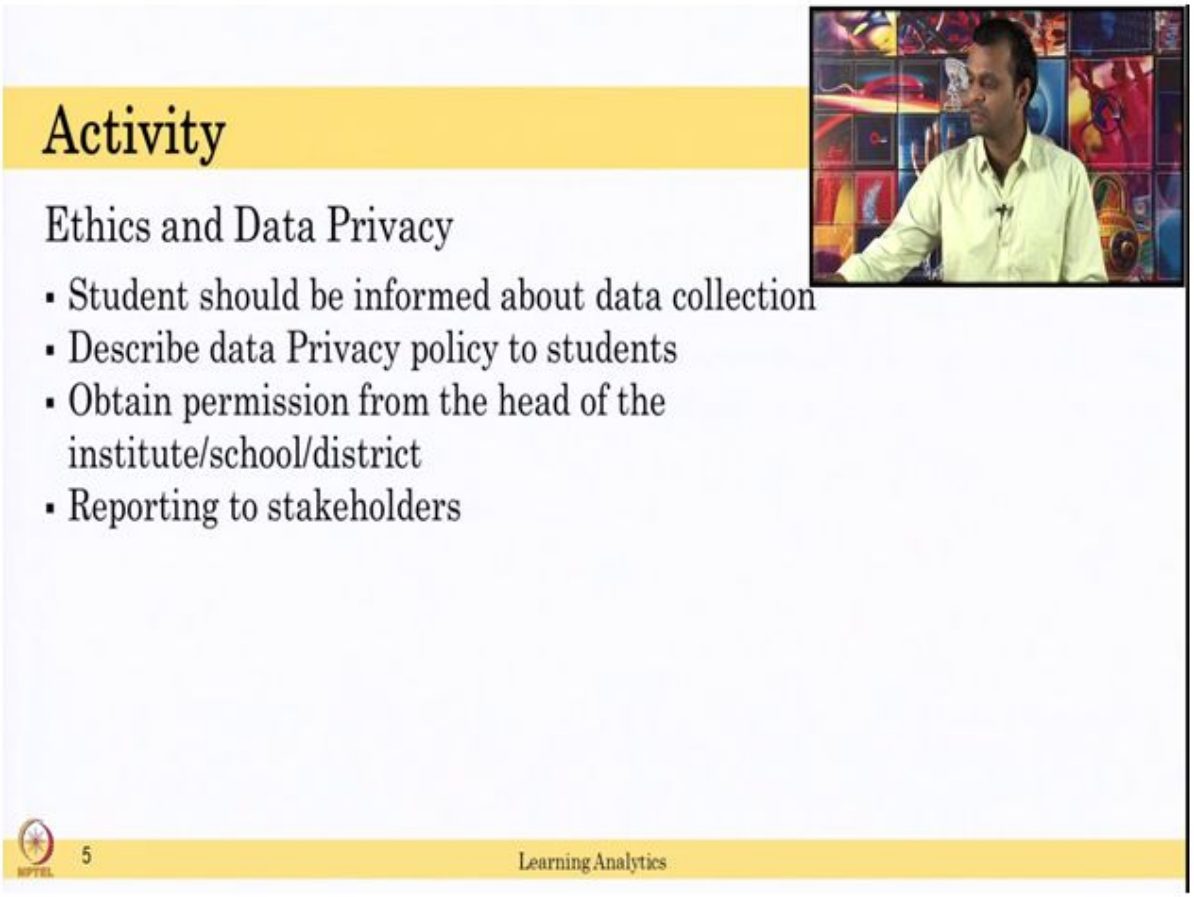
These interviews can be recorded and can be manually transcribed and coded and used for the data analysis. So, you know, you can collect lot of data from different channels and different sources. So that you have seen the various type of data collection methods and type of analysis that can be performed on the data.

(Refer Slide Time: 01:50)



Consider you are a researcher, you want to start collecting data from a particular learning environment. What are the primary steps you will take to protect the ethics and data privacy of the learner? What are the ethics in education? How do you protect the data privacy of the learner? Please pause this video, write down your answers and resume the video to continue.

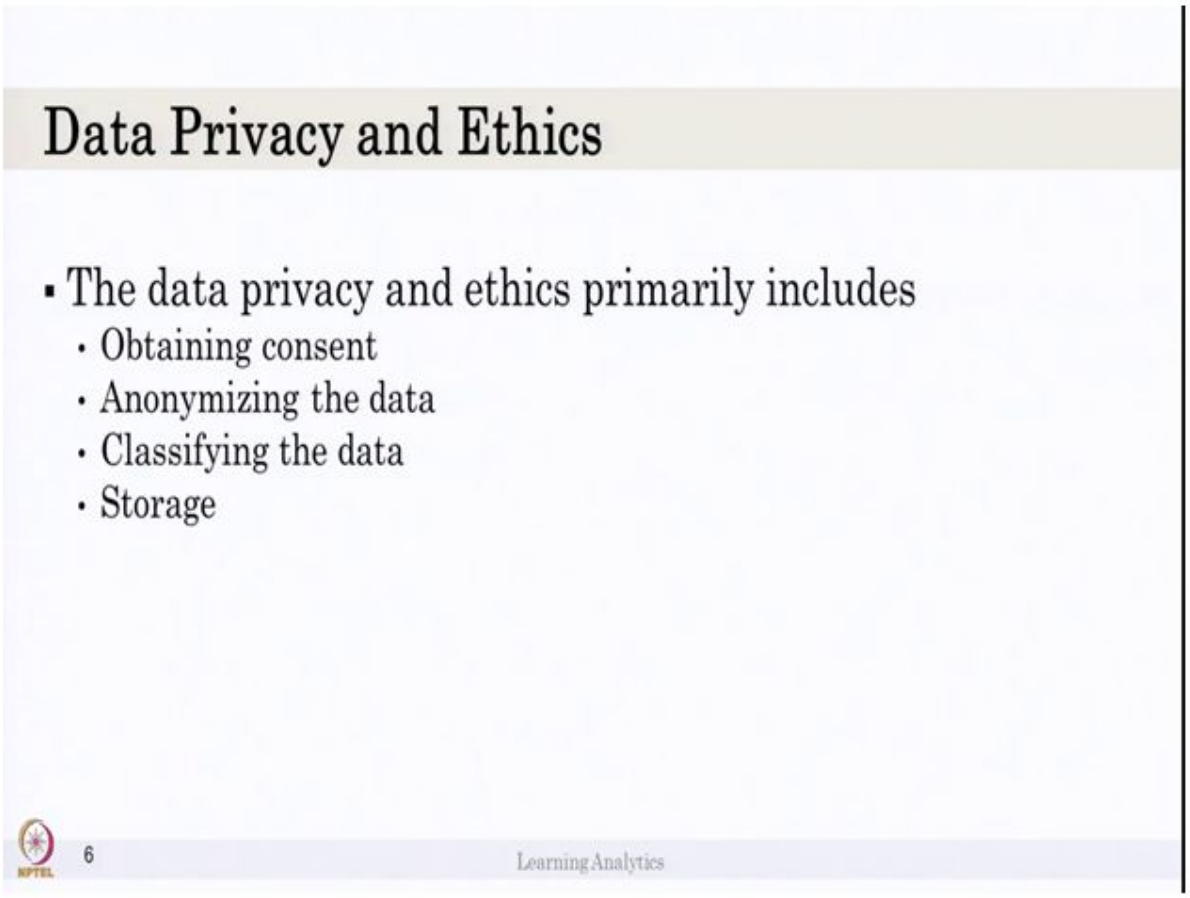
(Refer Slide Time: 02:13)



Ethics and data privacy, which means students should be informed about data collection. Like you should inform students that you are collecting the data. Also, you have to describe that data privacy policies to students like who has access to data, why do you keep the data, all this information. Also, you have to get permission from the head of the institute, school, district or the teacher based on your data collection, based on your research question, your study.

Also, you need to know what data will be reported to stakeholders like if you want to report all the data to all the stakeholders in the education setting, you have to think about that.

(Refer Slide Time: 02:54)



In general, data privacy and ethics primarily includes these four things. One is, obtaining consent. The consent form is as I mentioned previously, you have to inform the participant. So, you have to get the consent from the participant that they are ready to give their data for research purpose or we can use the data for research purpose.

So, the participant has to provide his/her consent. And you have to anonymize the data. It is highly recommended that immediately after collecting the data from the research study, please remove all the identifiable information and anonymize it. You have to keep the master key of the student name and that map key(but that should only be with you), kept highly secure. You have to use only the anonymized data for the search purpose and share this with your collaborators.

Also, you have to classify the data into identifiable and non-identifiable. For example, I mentioned that you can collect the student's facial expressions using a web camera, right? You cannot anonymize that. If you anonymize the student's facial expression with the mask or

something, you will not able to understand the student's facial expressions. So, you have to classify the data such as anonymize-able or identifiable and non-identifiable.

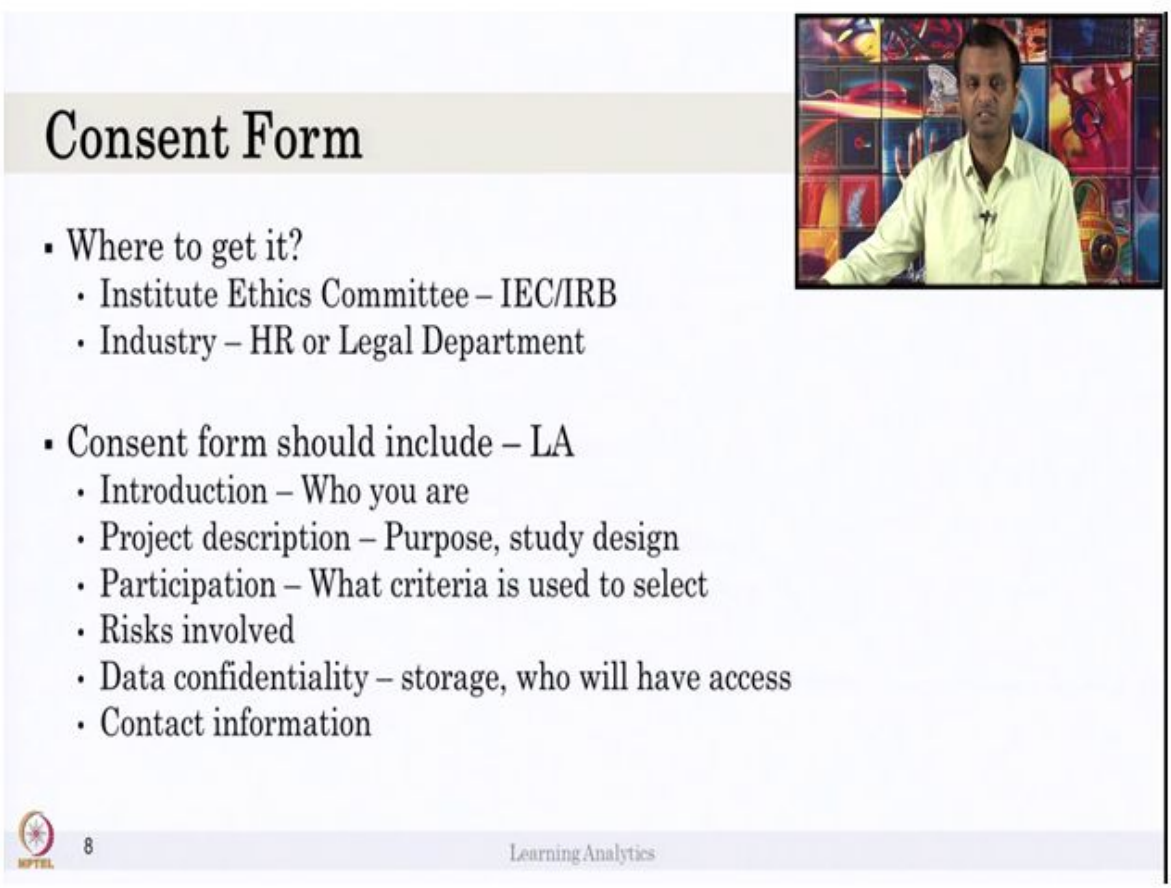
Also, you have to classify the data into videos, surveys, different types of classification can be done. And the last thing is storage.

- Where do you keep the data stored?
- Is the data storage spaces secured, free from hacking?
- Is data in a properly secured place and the key is protected with some software?

Everything has to be checked. So, let us talk about the consent form. So, we had to get this consent form. So, usually in academic institutes, there will be an institute ethics committee (IEC)

or Institute Research Board(IRB).

(Refer Slide Time: 04:52)



They will issue the consent form. If an institute does not have one. I strongly recommend talk to your institute administrators and create one committee. Committee of researchers who already have experience in conducting research and come up with a set of rules so that anybody can apply to the institutional ethics committee to get the consent form. Usually, the institute which has Humanity Science definitely will have it; also the institute which has animal study or bio or the human studies definitely will have the institute ethics committee.

An industry HR and legal department very well know, what to do because they do a lot of data sharing with the clients, collect data and share data with researchers, so they have a huge department, legal department, HR department with a big disclamation form and everything. So they know what to write, so talk to them and get the consent form ready for your research.

So what are the things that should be in the consent form? I am going to talk about the consent form for the learning analytics purpose. This is not applicable for conducting a study in a biological sample or conducting the study in humanities and science. This is basically for education setup.

The consent form first introduces who you are, that is your name and what are you are doing and why you are collecting the data. And the purpose of this project study design, you are collecting data for what purpose and what is the study design; like how long the students has to interact with the system? Is there pre-test, is there post-test? Is there any other information the student should know about the study to participate because the student should know what are they willing to participate and why do you select a particular type of student?

For example, if you want to know about student's background knowledge who are joining in MCA, you might be selecting only the first-year student of MCA to know what is their background knowledge. That is a very good time to collect data. So now you are, based on your research question you already filtered a particular type of student or a particular group of students. So that participation criteria should be explained in the consent form.

And what is the risks involved in participating in this study or not participating in the study? For example, the risk involved in participating in the research study you are conducting might be they are losing a lot of time because they have to spend another one hour, two hours, three hours

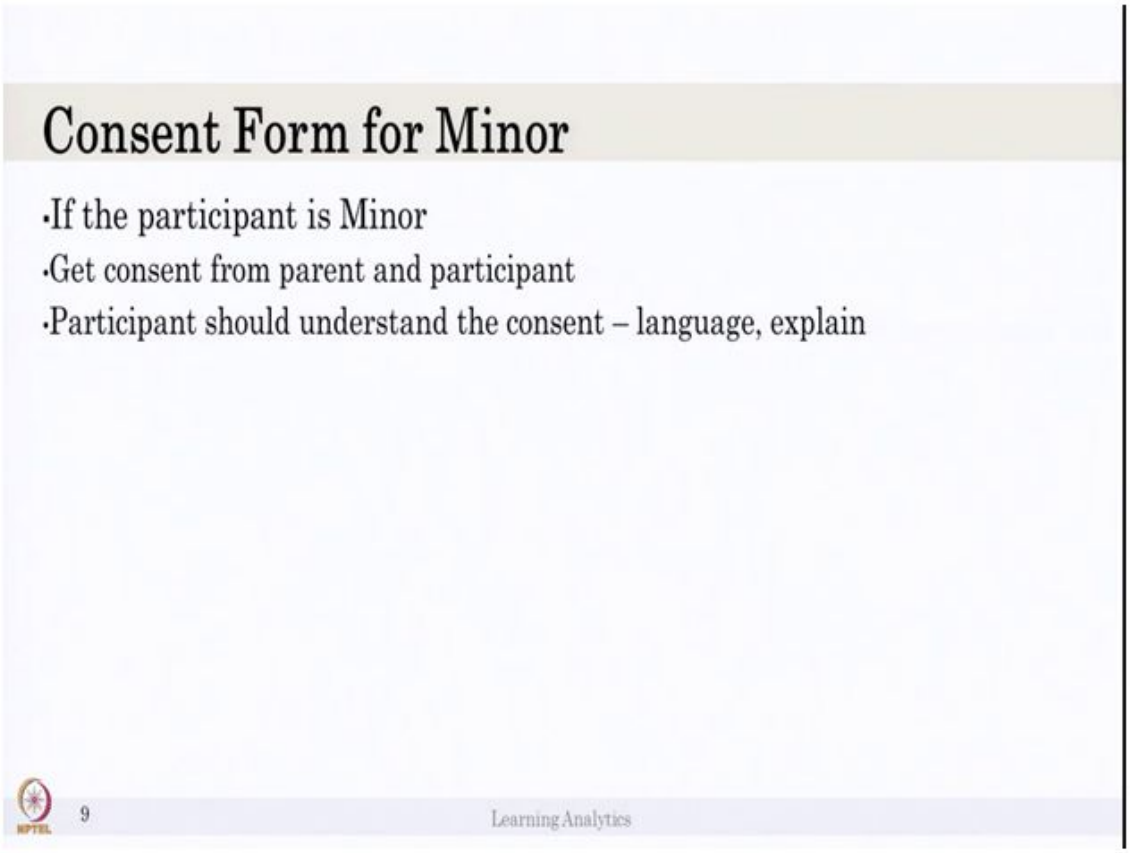
based on your research study design. Or the performance in this particular study might be used for the grading that may be a risk. This is a risk maybe for participating in your study.

Also, the students who are not participating in the study, who are not interested to give their data to you should not lose something their peers are learning. For example, the students who are participating in the study, you might be introducing a high-end simulator to teach a database course; so they understand everything well.

But the students who are not interested to participate should be given an equal amount of opportunity. You should not force them by saying that, only students who participate will get the particular content about the database, you will not learn it; that is not correct. This involves the teacher also. The teacher can say that I will make sure that all the students get all the content equally while participating in the course.

And importantly data confidentiality. Like where you will store data and who will have access to the data. That should be informed in the consent form. Students should know who will have access to his or her data so that they can provide consent to you. And the contact information. It is not your contact information, the contact information of institute ethics committee or the legal person in the industry so that the participant can contact them to report about your study, is there some problem in your study they can report or they can withdraw their study data something like that; they should have contact information.

(Refer Slide Time: 08:56)

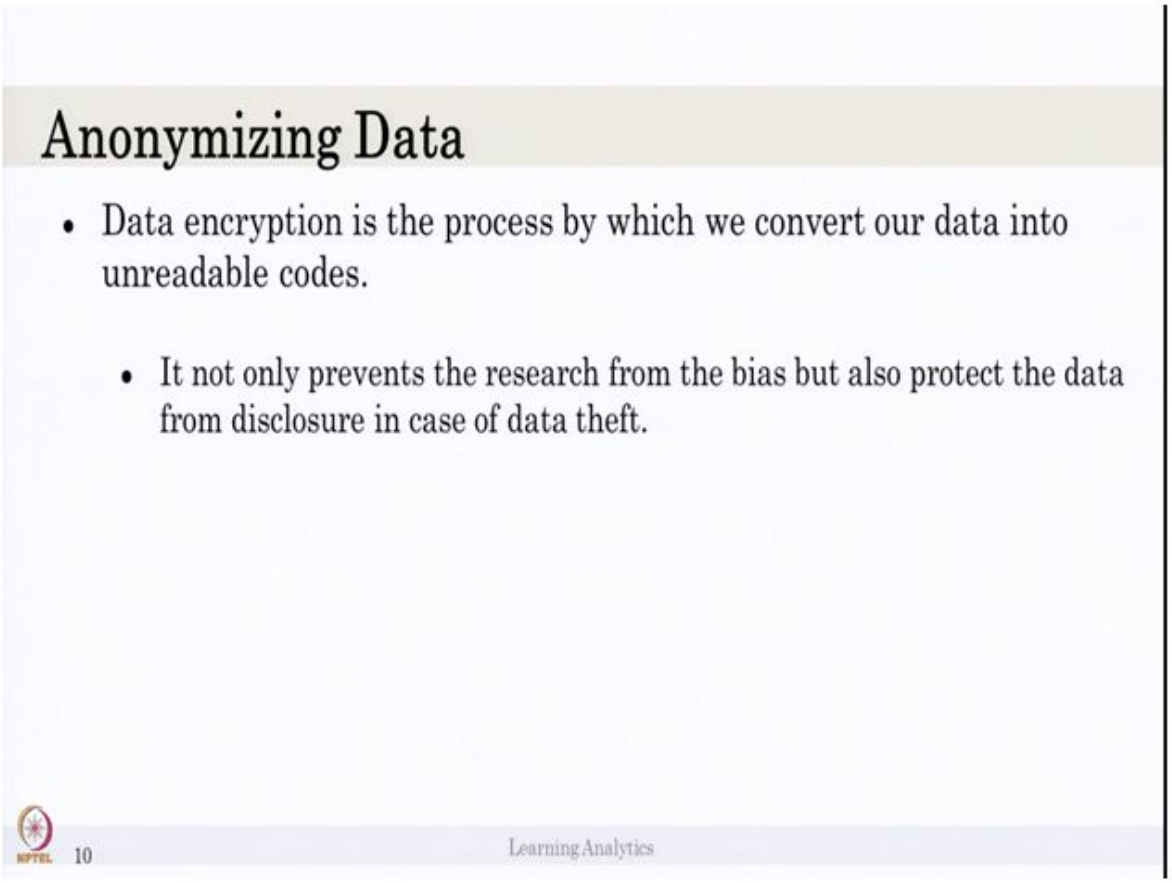


So consent form for the minor. If the participant is minor, get the consent from the parent, also the participant. For example, if the participant is Class Sixth student, you need to write a consent form in the language they understand. You should not use too much of technical words which they do not understand.

What if the participant is Class Two students, say 6 years old or 7 years old? So you cannot write the consent form which they can read and understand. There the teacher might need to explain the consent form to the student, also the parents. So they will understand your study design and they provide consent saying that I am interested to participate in your study and you can collect my data. That is one thing.

What if the participants are kids, infants or toddlers? You cannot conduct a study without the parents or guardian. That is very important. You have to get consent, also they(parents) should be with you when you are conducting a study. And the second thing is anonymizing data. Like you have to anonymize the data as I mentioned.

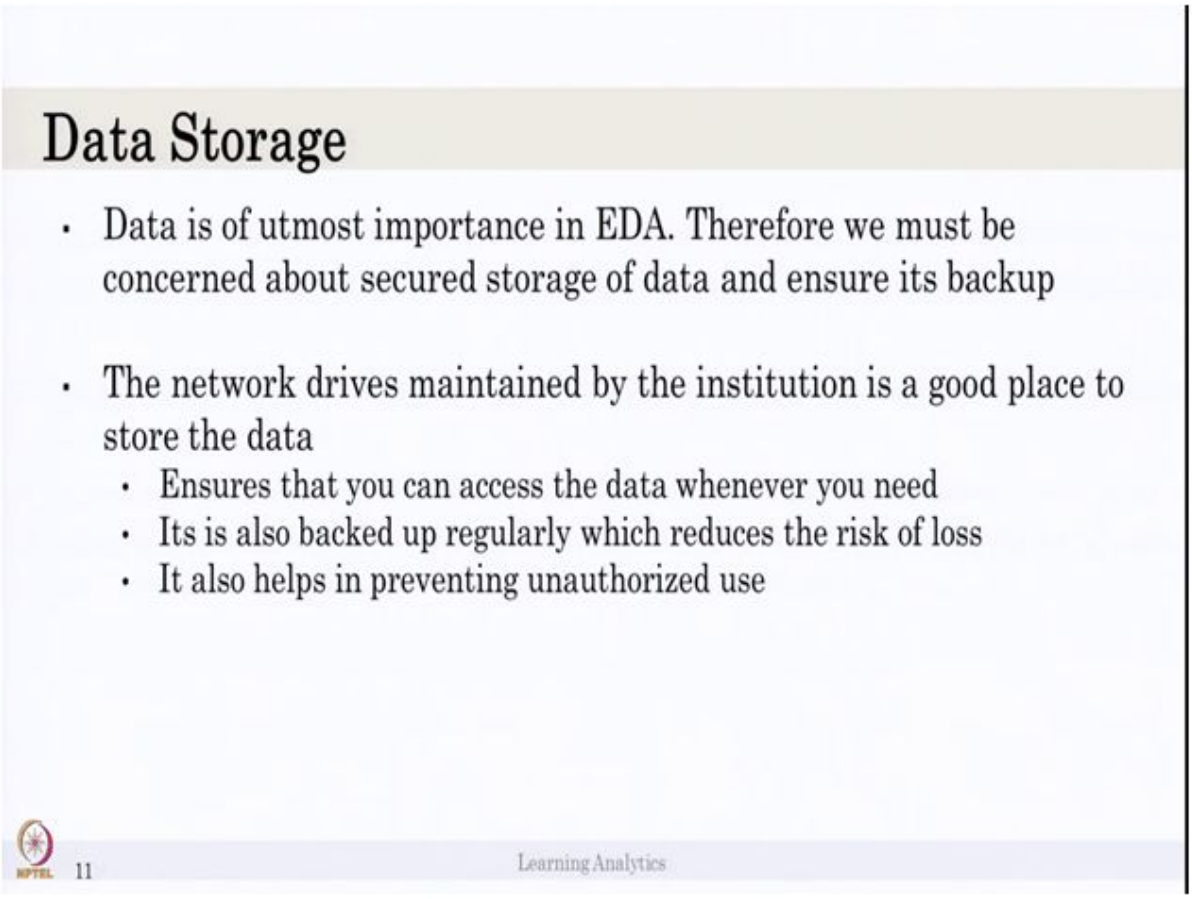
(Refer Slide Time: 10:09)



This data should be anonymized so that nobody is able to identify which user data it is. Before anonymizing, I recommend you to encrypt the data to convert all the names, e-mail ID, phone numbers to unreadable codes, de-identified data. If you are using e-mail, mobile number all information has to be converted into some unreadable code, some unique code or something like that, all user ID.

And they also encrypt the data, encrypted data actually helps researchers from the bias; also, if the data is lost, encrypted data cannot be accessed without a particular key or a password. I am saying encrypt the data with a particular key. Also, it is highly recommended to keep all your data, not just research data, your personal sensitive data in an encrypted format so that if it gets lost not everybody will have access to the data.

(Refer Slide Time: 11:15)



Coming to data storage, data is very-very important. Data is essential in educational data analytics or in data analytics. Therefore, we should be concerned about the security of data, therefore, we must store the data in a secured place and also make sure that data is backed up frequently. The network drives maintained by the institute is a good place to store the data. Why? Because you can access the data wherever and whenever you need.

For example, you need a data when you are in the campus, you can go to the particular network and collect data or if you are out of campus, you need that data you can remotely connect to the server, encrypted space then collect the data or use the data. Also, it is backed up regularly. If you keep data in the server maintained by the institute, institute usually backs up the data once in a week or once in two weeks so that loss of data is less because the data is backed up regularly.

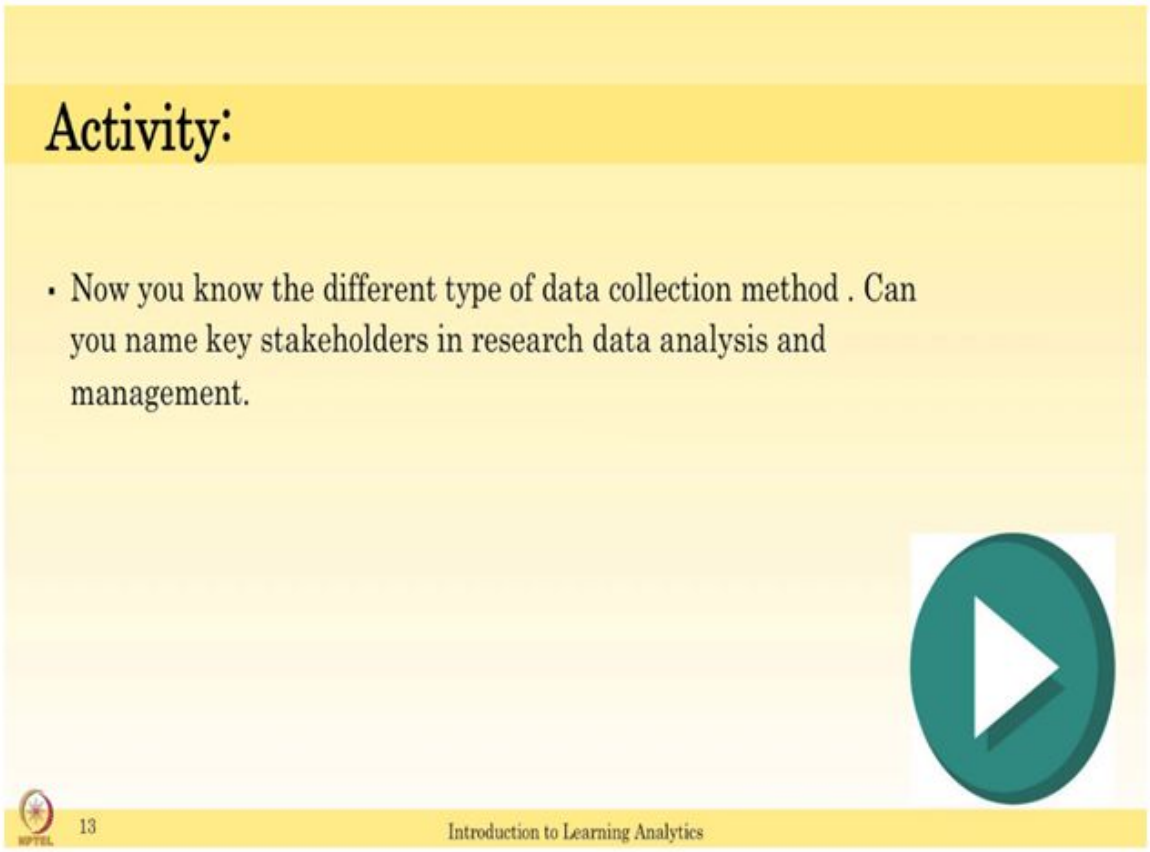
Also, it prevents unauthorized use. Only the users who have access to the institute network and the permission to access the data will be able to go and access it, so the third party, the members

who are not part of the institute cannot even access data. Even if the members who have access to the institute network will not be able to access your data because your data is backed up in your secured drive.

(Refer Slide Time: 12:36)

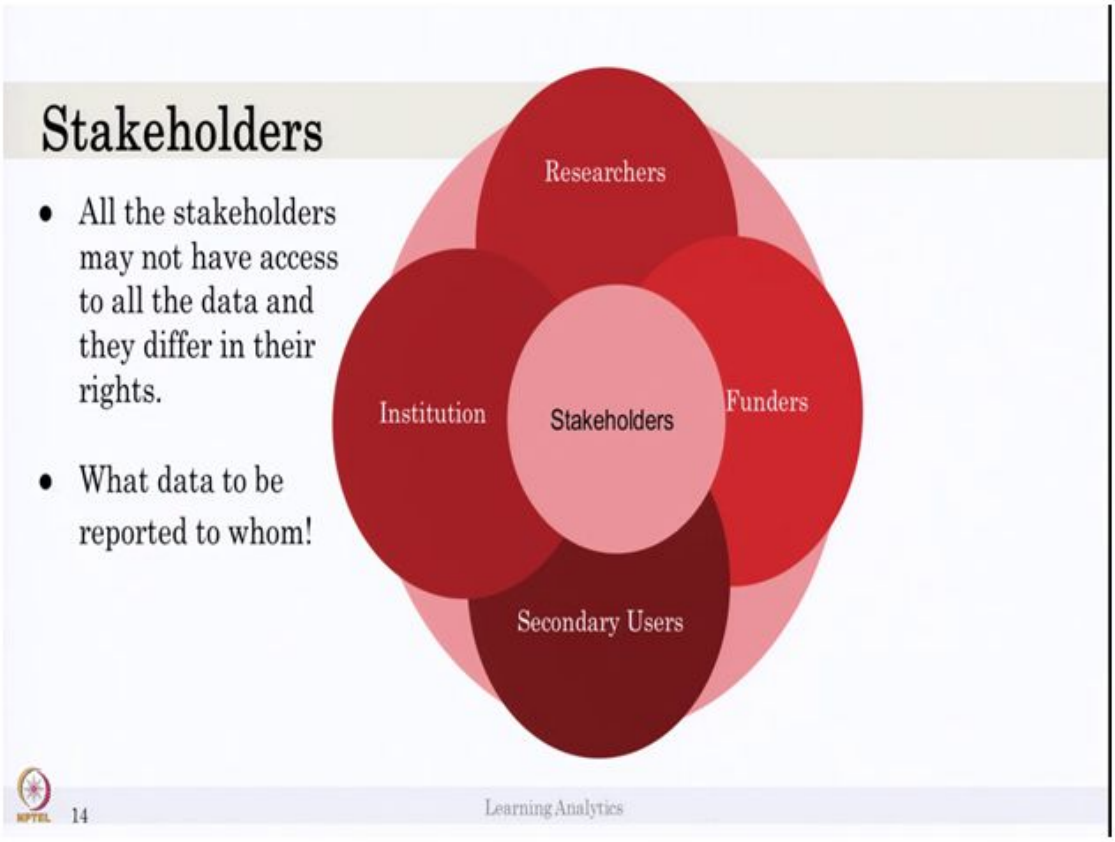
So what is data security? We talk about data storage and everything. Simply put, data security means ensuring that your research data is kept safe from corruption or it has limited access to limited people. This means backup regularly and keeping this data in a server which has less virus or has antivirus software, secures it and very few people have access.

(Refer Slide Time: 13:02)



Now that you have seen different type of data collection methods, can you name the key stakeholders in research data analysis and management? We have talked about stakeholders in Week 1, think about who are the stakeholders, what kind of access they have on the data. Please pause this video and write down your answers, resume the video to continue. Please recollect who are the stakeholders and what kind of access they have on the data.

(Refer Slide Time: 13:30)



So there are four, stakeholders like researchers, founders, or secondary users, or institutions when it comes to data collection and data analysis. So these are the stakeholders. All the stakeholders may not have the same type of access on the data you collect. For example, each one has different types of data access. So what data to be reported to whom, that is the main question.

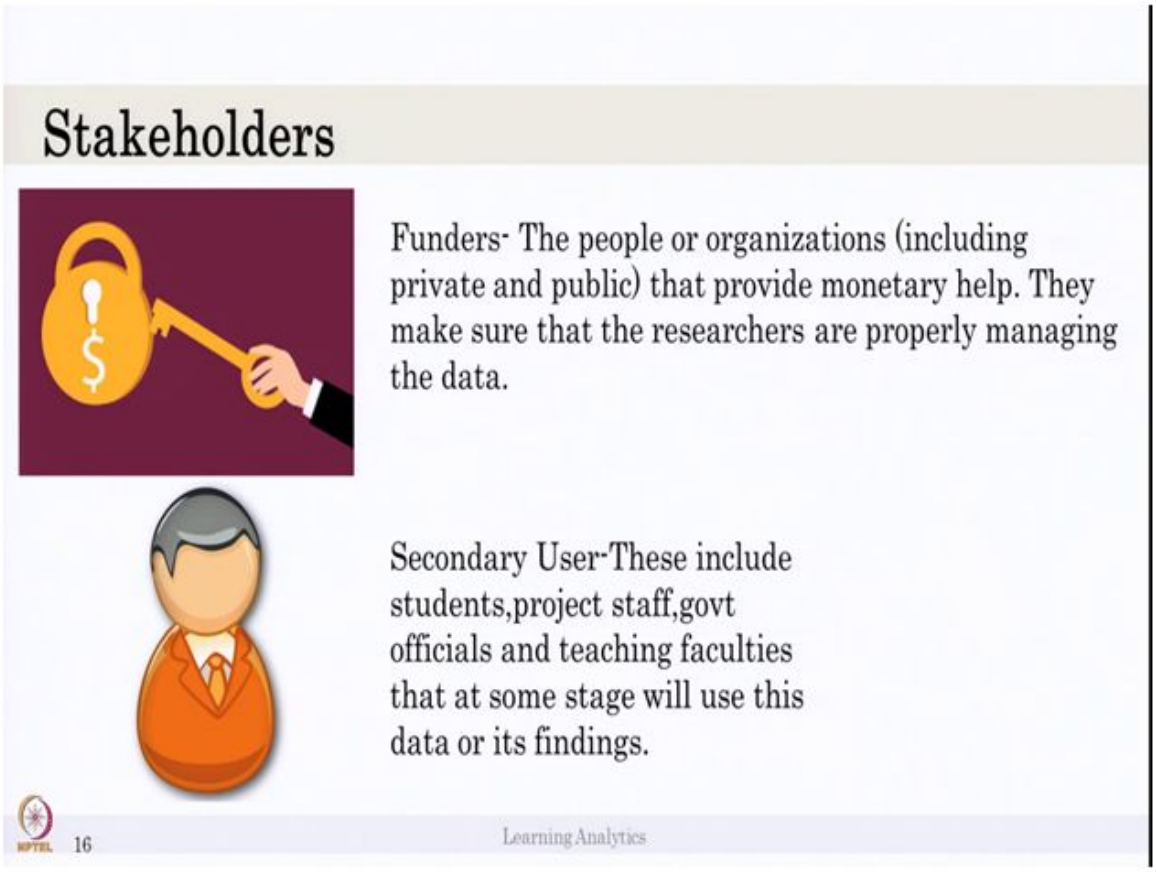
(Refer Slide Time: 13:55)

For example, if you are a researcher you are the one who is running and designing the study and collecting data, you might have all access to data and you might know what data to collect, how to analyze the data and you can think about what conclusions can be drawn from this data.

If you are part of an institution, these educational institutions have an ethics committee, so they have their own set of rules about how to use the data and what data to be published in the media and outside world. Also, the institute might provide data to researchers sometimes.

For example, if you are a researcher who gets data from institute's LMS, you do not have all the data, the institute inquires "what kind of data you want access to?". So that you can make some recommendations. So sometimes institute will also act as a source of data and also institute will act as a receiver of the data you analyze.

(Refer Slide Time: 14:50)

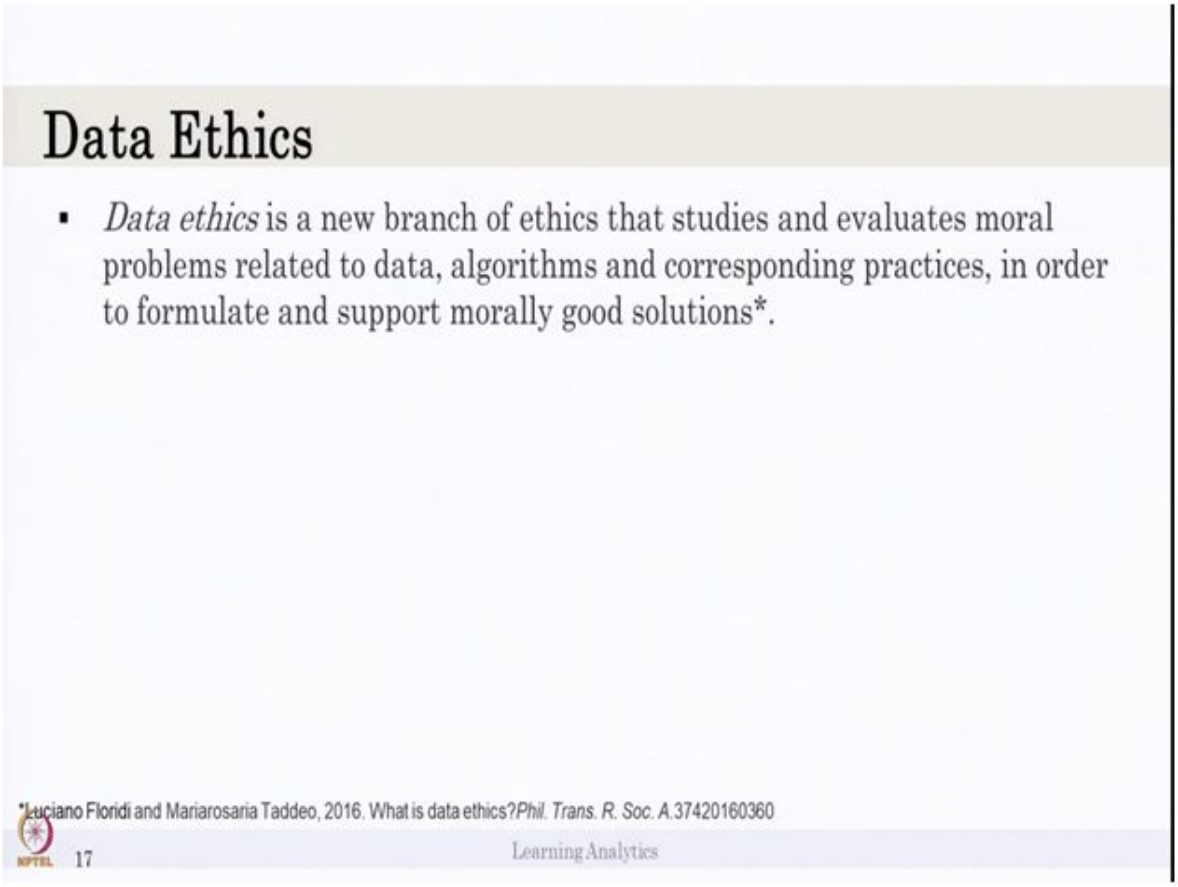


If you are a funder i.e. you funded this particular project, it is very important. They have their own strict declaration statements saying that what kind of data should be shared, and what data should be used for publications. Most of the time, most companies do not allow you to publish the study you conducted using their data. So you have to be very careful when you are running a study with funded research studies.

Secondary users are the ones who were actually reading the research paper you published, or the staffs in the institute, or the government officials you are reporting a data, they do not have access to all the data collected, they do not even know what kind of data you collected. But they know what kind of analysis, you did in abstract level, what data you have collected and what are the influences; that is enough for them.

So you should be careful about what data to be given to secondary users. Especially please do not give any user-related information to secondary users compared to all other stakeholders.

(Refer Slide Time: 15:56)



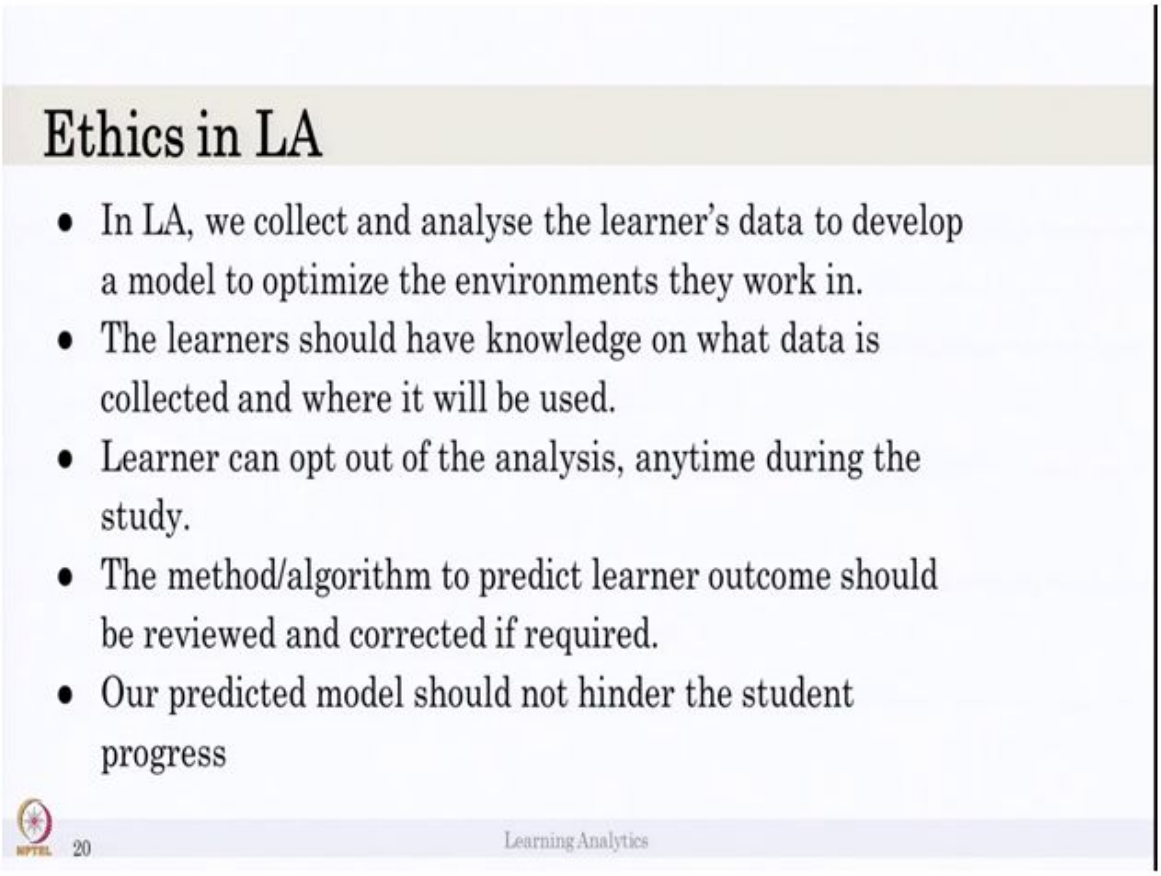
So what is data ethics? We talked about data privacy, how to keep the data secure, safe. What is ethics? Data ethics is a new branch of ethics that studies and evaluates moral problems related to data, the data from data collection, data generation, data analysis, all these things; and algorithms such as ML algorithms, AI and what other algorithms you use and the practices like why do you use this algorithm? To invent something new, in order to formulate and support morally good solutions.

So, the practice you have to do in data collection, in choosing the right algorithm and also doing some research is defined in this data ethics. So when you come to ethics in LA, these are the main points. In LA, we collect and analyze the learner's data to develop a model, to optimize the environment they work in. The purpose is to improve the student's knowledge to help the

students to achieve something in that particular environment. So the learner should have knowledge of what data is collected and where it will be used. It is very-very important, the learner should know what data is collected (i.e. given in the consent form).

Another important thing is learner can opt-out of analysis anytime during the study. This is very important, although they might give consent to be part of your study. After some time they can say I want to get out of the study or after giving the data within a time they just can say, no, I want to be opt-out of the study, please delete all my data. You should obey that, you should delete all the data. So the learner has the freedom to opt-out of study anytime during the study.

(Refer Slide Time: 17:43)



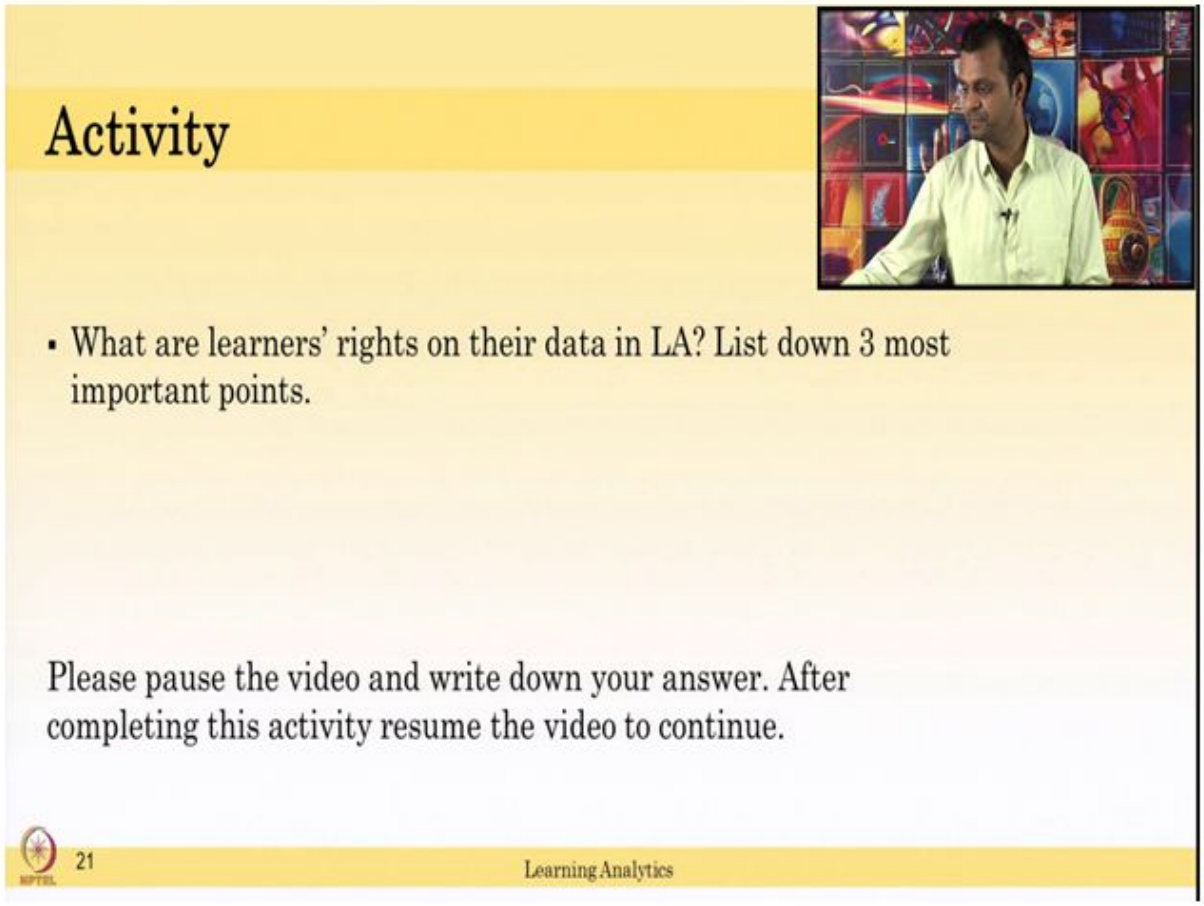
The method or algorithm used to predict the learner outcome should be reviewed and corrected if required. This means you must be thinking a student is confused and predicted students is confused based on the interaction or behaviour, something like that. But the outcome should be

reviewed, the student might say no, I was not confused, I was actively engaging I understood everything.

So, the students will be able to give feedback on what is the state you predicted so that your method/algorithm should be able to review that feedback from the students and correct it, if required. Also, a predicted model should not hinder the student's progress, when you say that a student is having problem in solving this particular problem, particular complex problem and you might think, Okay, let us give some easier solution or easier method to solve it.

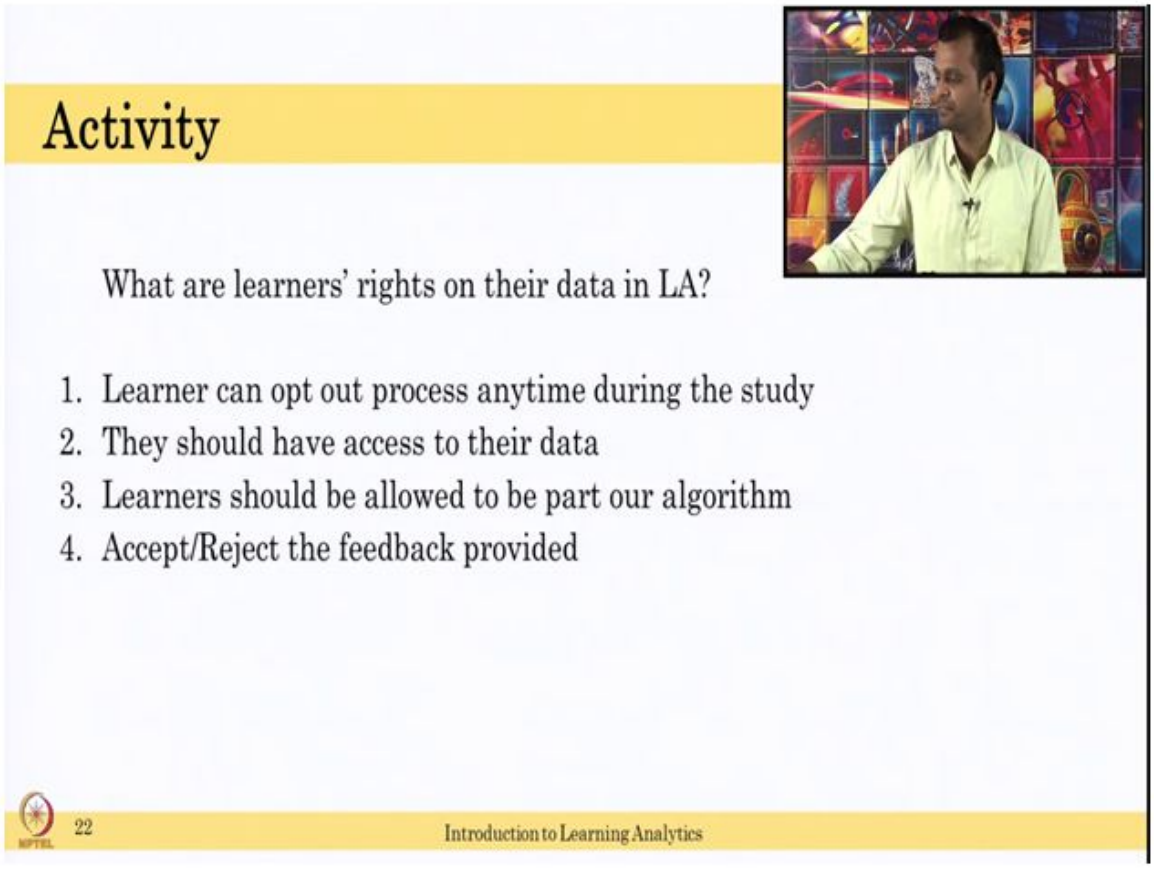
But student might say no, I do not want the easier path I want to solve the problem the way I want to solve. So give an option, whenever you recommend something, give the option to students, the student can accept or not accept. So students should have an option to accept the recommendations you provided. Which means you should not hinder their progress, you should not give too many feedbacks, too many interventions in a lecture.

(Refer Slide Time: 19:13)



So, from the previous slide, what are the learner's rights on data in LA? List down three important points. This answer can be directly from the previous slides. Please pause this video and write down your answers, after completing this activity resume the video to continue.

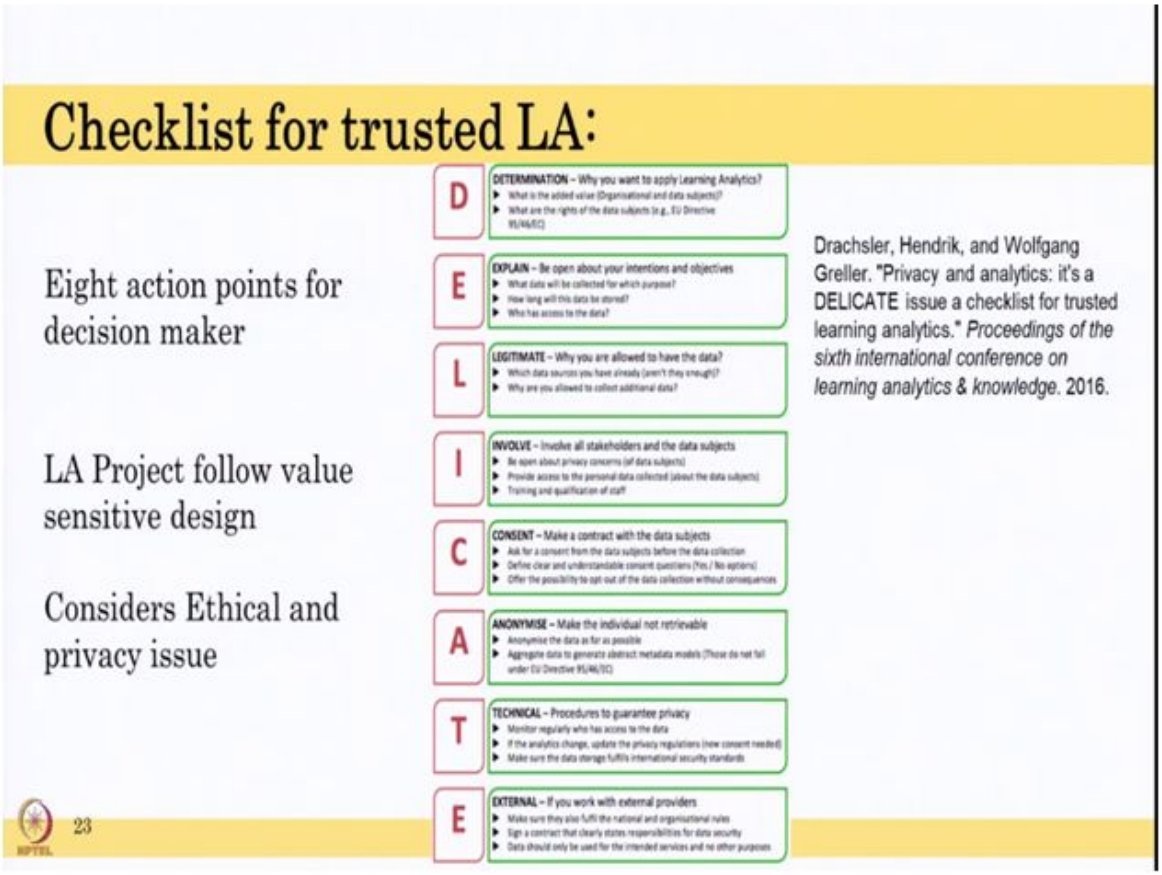
(Refer Slide Time: 19:30)



What are the learner's rights on their data in LA? So the learner's right is learner can opt-out of progress anytime during the study. They should have access to the data. The learner can ask, after completing the study, can I will look at my data, what data is collected. You may not need to provide all the biosensor data and everything you provided, but you can give an abstract view to them.

The learner should be allowed to be part of the algorithm. Which means learner should able to give feedback on the predictions you made and that feedback should be used to correct your algorithm. Also, learner should accept or reject the feedback provided. The learner should have the option so when you give feedback or some new recommendation, do not force it on the student, they should have an option to accept or reject your feedback provided.

(Refer Slide Time: 20:22)



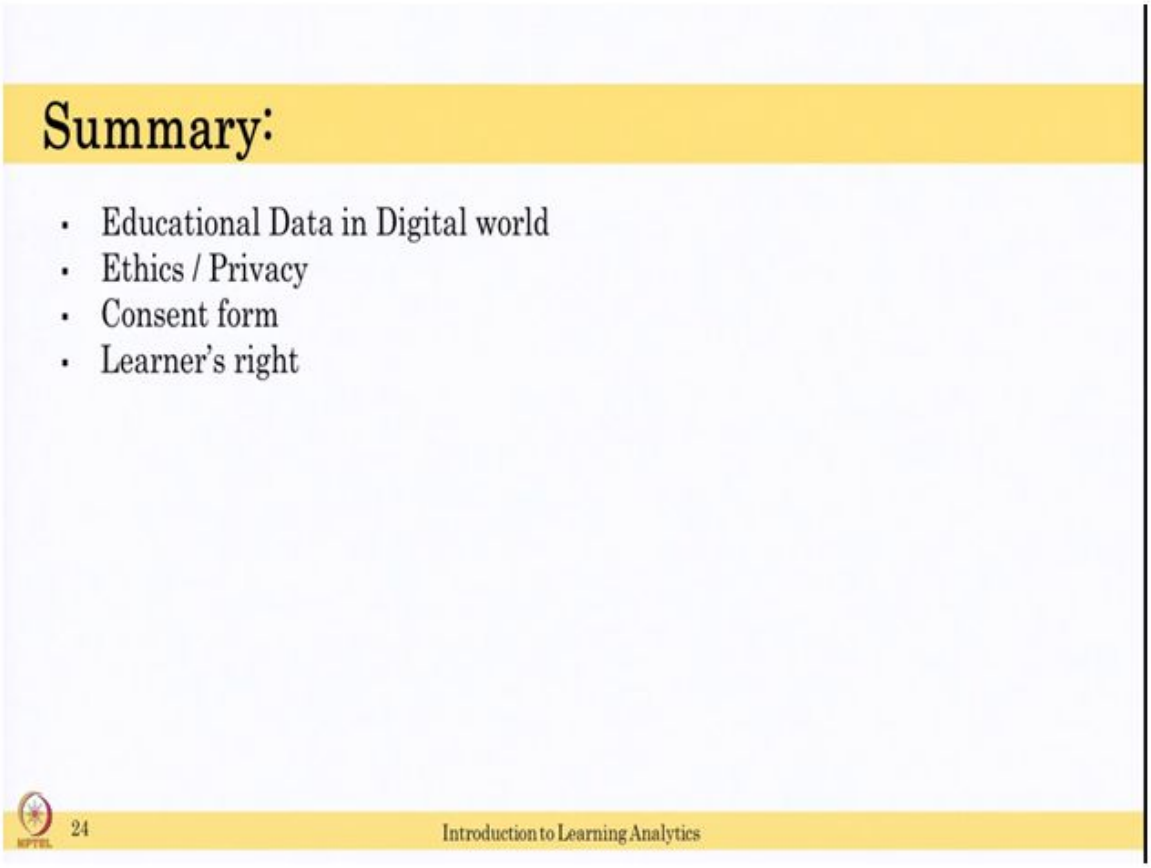
So, in short, these are the checklist for trusted LA, eight action points for decision-maker. It talks about determination, like why you want to apply data analytics and explain it to the students and what data we collected, where it is stored, everything should be explained in the consent form. Be legitimate, that is why you are allowed to collect data, why you are the one who wants to collect data of some group of people about their personal information, why you are allowed to do it, be legitimate.

Involve stakeholders and inform them what data is collected and what data can be informed to them, let them know what the process is. Also, make sure you get the consent form from subjects that is the participants. Anonymize the data after collecting, backup of all this information has to be there. If you want to collaborate externally, share your data with the other researchers in other Institute or other countries, make sure you follow the other institute's principle, their set of rules.

If you have to forward the data to collaborate in another country, make sure that the treaty between your country and other country is neutral and you can share the data between these two users without any problem. In general, in education data set we do not face these issues, mainly in defence or military documents, they have a very strong treaty not to share data with anyone else. But in education setups, yes, we can share the data with another researcher from other countries or the continent.

So please consider ethical and privacy issues when you are collecting data and when you go out and plan to collect data in your environment like classroom or MOOC or intelligent tutoring systems, and this is from the paper which is given here.

(Refer Slide Time: 22:25)



So, in this lecture, we talked about educational data in the digital world, like what is the data collected in the digital world, also in educational settings, that is data can be collected from TELE or different sensors, all these information. And we talked about ethics and privacy, how to secure the data and what are the ethics should be followed when you collect data, and what is a consent form, and what should be in the consent form, how to get a consent form, and how to get a consent form for minors. Also, we talked about learner's rights the very important. The learner has a right to opt-out of study anytime during your study.

(Refer Slide Time: 23:07)

So in summary, this is the from LAC 2016 paper, they talked about “word cloud” of all this content in ethics and privacy, they said the ownership and control is more important.

- Who has ownership on the data?
- Who has the control of data?

- Is data secure ?
- Does the learner has the ownership and control of their own data?

Thank you.