Ethical Research

Welcome to this module on ethical research.

My name is Gerald Maguire, and I am the professor at KTH for Computer Communications. I have done this material together with Dr. Ellen McGee, retired professor from Long Island University.

[slide 2]

The goals of this lecture are to discuss: What is research? Why should we do or not do research? How can we conduct our research in an ethical manner? What is misconduct in the scope of research? Some questions about authorship. Who should be authors? Who shouldn't be authors? Some questions about intellectual property rights (IPR) and privacy.

[slide 3]

We begin with the first question: What is research?

Shuttleworth, in 2008 said in the broadest sense of the word, the definition of research includes any gathering of data, information, and facts for the advancement of knowledge. We note the importance there that the goal is to advance knowledge.

[slide 4]

So some important questions to ask ourselves when we think about doing research are: Why we do the research? Typically we do this because we want to achieve knowledge, we want to understand what is true, or you want to refute falsehood. Because we believe something to be false, and we want to establish that it is false.

We also have to ask ourselves: How do we do research? And we want to do research in a proper ethical fashion. And then a big question is: What to do research on. What are the topics? What are the topics not to do research on? We will investigate some of those in this module. We are explicitly leaving aside the questions of where to do research, when to do research, who does research, and how to fund this research, to your discussion sections.

[slide 5]

The "Why" of research Heidi McKee and James Porter in their book, "The ethics of internet research" said "ethical research begins with a coherent, valid, and sensible research design" and the research project needs to establish a "worthwhile purpose for the project the why the research". And so the objective is that you really want to gain new knowledge or understanding and it should be of some benefit to society, and if your research involves research subjects, then it should be of some benefit to them, and/or you should gain useful knowledge. That means you aren't doing it simply because of your

personal interest or that respect to gain from it in some way. You must begin by saying, "What's the value in this" and "Who is it valuable to?"

[slide 6]

So to conduct research ethically, we need to consider the issues about accountability, fairness, trustworthiness, and establish collaborative values.

The Swedish center for research ethics and bioethics (CODEX) has an extensive collection of information on research ethics, and you can find further information about it at their web address as indicated here (http://www.codex.vr.se/en/forskningsetik.shtml) - The information is available in English and also available in Swedish.

[slide 7]

So the first question, whenever you are considering research, is, "Do you really want to be involved?" Is it something that you would be embarrassed to put your name on or to be associated with it? If so, then you should say no. Because if you don't want to be associated with it, then you shouldn't be doing it. You need to understand what to do if you find that the research is something that you find personally repugnant. You just thinking is all wrong. You should again probably say no, I'm not going to be involved.

An how do you formulate your hypothesis testing to avoid causing harm, because of course, you don't want to intentionally cause harm to anyone or anything.

[slide 8]

Some or many of you will consider doing surveys.

One of the difficulties is that when we use computer or internet-based research and conduct information [gathering] involving human subjects, we need to consider: Do we need an ethics review board approval to do this?

If so, then we need to go and make sure that we apply and got approval - before we do the study. Now in many cases, such as in the US it is required to have an Institutional Review Board (IRB) approval if the work is going to be published and that includes your master's thesis or dissertation or because the study involves some vulnerable groups such as minors, prisoners, etc. people who wouldn't be able to say legally: "yes, I want to participate or no, I don't want to participate". And it includes a whole variety of types or means of gathering information, such as the use of questionnaires or interviews, whether you send email, or whether you mine the data from websites.

Now, Kate Kelley et al., in a book called ‘Good practice in the conduct and reporting of survey research', emphasize the importance of a single clear and explicit research question. You want to know: What is the question? If you don't have a clear formulation of the question, then it is really very hard for you to make a successful research project.

You need to think about the ethical issues of confidentiality and informed consent. Can you keep the data confidential? How can you keep the data confidential? How are you ascertaining that the people participating were really informed so that they're giving their informed consent to participate-Because if they don't know what they're agreeing to, then it is not informed consent. And, of course, you need to develop a good survey instrument-You aren't simply collecting data and analyzing it. That means the design of the survey has to be very carefully thought-out.

[slide 9]

Now the Council of American Survey Research Organizations (CASRO) has a Code of Standard and Ethics For Market, Opinion, and Social Research. It's divided into all these parts. You can read more about this. But they correspond to some of the things that we just discussed: privacy, confidentiality, want to make sure that you don't want to harass people, so, therefore, you shouldn't be asking people to participate in your survey unless they expect to be asked to participate in the survey; otherwise you're harassing them.

What are the privacy laws and regulations that you need to follow and be extremely careful in any research involving children, young people, or vulnerable populations. And there are special considerations for online and mobile research, which means that you're conducting this research communicating with people via their cellphones. You also have to think if you're conducting research for someone else, how you meet the responsibility for those people, and of course, how you meet your responsibility to the public.

And you have to think about your relationship to your subcontractors - people that you might be hiring either to collect the data by acting interviewers or subcontractors analyzing the data.

[slide 10]

Similarly, the European Society for Opinion and Market Research (ESOMAR) has a set of guidelines, and you can read about those. But you will notice they cover basically the same set of points.

[slide 11]

So whether you're in North America or whether you're in Europe, there are good guidelines available if you're going to conduct this type of research.

Other issues that you're concerned about during your data collection are you need to assess the risk for psychological, physical, social, economic, or even legal risk for someone participating in your survey. This means you shouldn't be asking people questions whose answer would cause any of these types of risks.

So you can't be asking someone: Do you take illegal drugs? Well, of course, the difficulty is were they to answer that someone might be able to track the answer back [to them], and then they might face legal repercussions being charged and imprisoned.

You have to be very careful about special populations, the so-called vulnerable populations, people who are mentally incompetent, minors, or prisoners because they don't have the ability themselves to be able to give informed consent. And you also have to think about if the date is considered sensitive. Typically today, we think of that in terms of medical data or political data or sexual data.

CODEX has a nice description of handling personal information on the web indicated on the slide (http://www.codex.vr.se/en/manniska3.shtml). And later in another module, we will talk about some of the issues about handling personal data in terms of the context in Sweden involving a law whose initials are PuL (Swedish: personuppgiftslagen)

[slide 12]

So with regard to data collection, we said that it requires an affirmed, informed consent - someone has to actively say "yes", and they have to be informed, and they have to understand what it is they are saying yes to, and they have to give their consent or say "no I'm not going to consent" - in which case they can't be part of your study.

Now Heidi McKee and [James E.] Porter in their book "The ethics of Internet research: a rhetorical, case-based process" have a lot of useful guidance when it comes to collecting data in chat rooms, discussion forums, or blogs, social networks in massively multiplayer online games, etc. It's important to understand that just because the data is available and might be read by many, many people that doesn't mean that you can simply use it.

It's also important to remember that you should never collect data that you don't actually need. So if you don't need it, don't collect it!

[slide 13]

There are also ethical issues in data analysis.

For example, the simple problem of how do you protect anonymity. And one of the difficulties is it turns out it rather hard to actually provide anonymity.

In survey research, lots of investigators disassociate the names from the survey responses whenever they do the coding and recording, so the result is they only have a participant's number five six seven two. Only they know the mapping between five six seven two and the actual person's name. [Hence] no one else can see that information.

Well, there may be some limitations on that there may be others who actually do need to see that information. How is it kept? How do you maintain it? And in qualitative research, it is very common to use pseudonyms. Because of course, if you're talking with the person, you don't want to reveal their identity.

Other questions are how long should you keep the data and even the very, very important question of who actually owns the data. And it might not be who you think!

[slide 14]

Now the common problem from society's point of view is scientific misconduct. It is the violation of the standard codes of scholarly conduct or ethical behavior in professional, scientific research.

Now, Peggy Fischer, in the Office of Inspector General of the U. S. National Science Foundation, has described a set of policies where it carefully defines: What research misconduct is. It's the fabrication, falsification, or plagiarism in proposing, performing, or reviewing research or in reporting the research results. And the report carefully describes what fabrication is, and we all understand what that is - it is that you made up the results. Either you made up the results - to begin with. Or you made them up when you reported them, or you made them up when you recorded them. Aha, the user is not here, so I will guess that they would have answered this way. That is fabrication. Falsification is manipulating the research materials, equipment, or the process or changing or omitting data or results such that the research is not an accurate representation of the research represented in the record. Aha, that measure is not what I expect - so that I won't write it down. That is also fabrication.

Plagiarism is the inappropriate taking of another person's ideas, processes, results, or words without giving them the appropriate credit. And the policy defines the policy also defines "research" and "research record" - what are these? And it suggests that countries need to develop systems to deal with misconduct. In Sweden, the National Medical Research Council committee to deal with scientific misconduct was already established in 1997.

[slide 15]

Daniele Fanelli, in a report called "How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data" looked at the results of research involving survey data. And it has some really rather shocking results because collecting this data together 1.97% nearly two percent of scientists admitted that they had fabricated, falsified, or modified data or results at least once and that's a rather severe form of misconduct and up to 33.7% admitted to other questionable research practices. So nearly a third of the people admitted to knowing about this kind of misconduct or carrying out the misconduct. And in his surveys regarding the behavior of their colleagues, the admission rates were 14.12% for falsification and up to 72% for questionable research practices. This means that they believe that the majority of their colleagues were involved in these questionable research practices.

What are some of these questionable research practices? They include things like dropping data points because of gut feelings or changing the design or methodology results of the study in response to the pressures from the funding source. Aha! If you will say that our product has these good properties and you ignore these bad properties, then we will fund your next research project. However, that is inappropriate. Note that this particular paper by Fanelli is based upon samples of 21 surveys, included in the systematic review and 18 in the meta-analysis - where they took data that had already been analyzed to look for this misconduct.

[slide 16]

Plagiarism, as you are all aware, is unacceptable at KTH.

Plagiarism consists of not just copying people's words or ideas or processes, but it also includes stealing their code or putting people down as authors who didn't actually do the work. All of this is Involved under the umbrella of plagiarism. And most publications have guidelines about who can be an author and who cannot be an author. You need to follow these guidelines.

[slide 17]

So, for example, the authors must have made a significant contribution either to the research conception or the design or the conduct of the research or its analysis and/or the interpretation of the scientific work. So it's perfectly permissible to have an author whose sole function was they design the experiment; hence, they should receive credit for that. And you might have another person who only was involved in the analysis, and it's perfectly okay for them to be an author of the paper. They needn't have done all of these things, but they must have done one of them. Therefore gift authorship, including someone on a paper who isn't actually contributing in any of the above ways, is unethical.

There is also a question about ghost authorship. What happens when someone other than one the named authors actually writes the paper?

And this is supposed to be relatively common in the pharmaceutical field where people hire ghostwriters to do the writing of the paper before them. And there's a very, very hazy line between how far this writer can go when they have crossed the line over into really being an author and not simply being an editor. And you need to think about that. It's also important that you only publisher data once.

And if you're in a university, as you are here, your faculty adviser should discuss how publications will be organized, who will get credit. It's helpful as a student if you got this in writing when you're going to work on your doctoral thesis because, in many cases, you're going to be collaboratively writing papers with your adviser and with other students.

[slide 18]

It's also important to avoid any conflict of interest why because you want to make sure that someone who looks at their score will say I can believe these results.

So if there was a conflict of interest which might even be financially beneficial when these reports come out this particular way, then the stock that I own this company is going to go up in value, so I'm going to leave out these are the values that I'm going to go through the good values and could change all - so that I'm going to benefit is again inappropriate.

[slide 19]

There's a whole variety of intellectual property, and it has various forms of protection.

Some of the products of the human mind are protected by copyrights. We all know that today under the Berne convention copyright extends from the time that you write it (you don't have to file it, you don't have to register it) happens as soon as you write it or you draw it in the case of drawings or carve it in the case of carvings, etc.

There are also patents. To get patent protection, you have to file an application with the patent office. And depending on what country you're in, you can patent a whole variety of things. In some cases, you can patent processes. And in other countries, you can't. And you need to explore that in terms of where you believe that you want this protection. There is also trademark protection to protect brand names and product symbols. I should mention that there's even patent protection on design, and many people file design patents. Yes, you can protect this design of these rounded corners because that's an important part of the look of our product.

There are a lot of international agreements regarding intellectual property law, and you can read about them. The primary organizations involved in intellectual property protection are the World Intellectual Property Organization (WIPO) and the U. N. Educational Scientific and Cultural Organization (UNESCO) and the UN Conference on Trade and Development (UNCTAD)

[slide 20]

Now one question that you might run into today that is a very popular social issue. It is: What about hacking? Is it okay if you hack into someone's computer system? Or is that stealing? But I only looked and peaked. I didn't actually take anything Is that ethical will or not? And you need to be asking yourself all the time when you're carrying out actions: Was this ethical or not? Do I want to be associated with that or not?

[slide 21]

Another common set of problems concerns privacy, and today, of course, we have entities like Google's Gmail where they process the email. They extract out of that information to build a model you that they use to decide how to target advertising to you. Is this violating your privacy or not? Just because no human has looked at it, does that mean it is private? Or is this building a model of you? What happens with that model? Who owns that model? So it's a complex problem.

Today we have laws about workplace spying that limit your employer even though you're using a telephone for which your employer might be paying. Are they allowed to listen to all of the conversations? Are they allowed to know who's called you and you have called?

A lot of questions, particularly here in Sweden, about surveillance cameras in public places, So in the case of KTH, we are a no camera zone. The result is, in Stockholm county, if you want to have a camera in a public place, you have to have a permit for it. And they will question why you want to have this because they want to protect the privacy of other people who might come into this public space.

There are lots of questions about GPS inside cars and trucks because now we can potentially have the ability to know exactly where the vehicle has been and when.

And for example, in New York City, not only were they tracking the vehicle, but they were keeping track of what other vehicles were near that vehicle using cameras mounted in public places reading the license plates and correlating this information together.

Most of you are aware that your cell phone is constantly being tracked while it's on – in order to provide you with better service. And there's a law about that information. In both Sweden and Finland, it requires a court order to be able to access that log, unless you give written permission to another entity to have access to that data.

There's also a lot of questions about collecting data while you're visiting a website, which websites you visited, which links you clicked on, how long you looked at certain parts of the website, etc.

[slide 22]

A very important thing to understand is what you can't promise as a researcher. You cannot promise that no one outside your research group will ever have access to the material or information you collect in your study. And the reason is there many means [&] many situations in which access to that material _is_ justified and may be necessary. So, for example, if their charges about scientific misconduct. If there are charges about plagiarism. If there are charges or questions about authorship. Did you really carry that out? Were these really the results that you found. So it's very, very, very important for you not to promise something that you cannot deliver.

[slide 23]

Now, another area this very much attracting a lot of attention today is data mining.

The goal of data mining is that you want to take these very large amounts of data that appear in databases. And you want to extract the _implicit_ patterns.

However, there are problems about what privacy is there regarding extracting these patterns. Now it's fairly clear in many countries, such as Sweden, what is _explicitly_ protected in the database, what is confidential – for example, data regarding individuals' medical, financial, or academic records. And what happens when we exchange data between our across bases. But there is today, really no legal or normative protections, once we start manipulating that data.

Where it is the implicit connections in the data, it wasn't explicitly there; it might not have been even understood that there were connections that exist in the data when the data was collected. So what is the obligation about privacy afterward. What happens if the data was collected from non-confidential sources, where people knew that the data wasn't going to be confidential and now because we can infer patterns across it, you might be able to derive very sensitive information even from this non-confidential

data. And what happens when the data is not exchanged between databases but is all accumulated from within one database?

So these are some of the areas that you may want to read more about if you're going to be doing data mining.

[slide 24]

Today, the Swedish Research Council and many other funding sources are promoting openness. They want transparency. They're encouraging data re-use. And part of the reason is that of course, they don't want to have to pay twice to collect the same data. And therefore, there's lots and lots of pressure on researchers today to make their data available.

But at the same time, we have to make sure that we protect personal integrity. In Sweden, there is a very nice article written by Zoran Slavnic, discussing the differences between the UK and Sweden with regard to how in this qualitative data collected - for instance, from surveys - be preserved and be re-used while providing and meeting the requirements of the law. So, for example in Sweden, it may be necessary to store the data and archive it, even though for anyone else to access this data might it might actually be illegal - one still has an obligation to preserve it because at some point it might be legal access to access it.

So CODEX has a set of rules and guidelines for research that you should take a look at with regard to this issue.

[slide 25]

As I mentioned, there's also this question about data ownership.

Now according to the Swedish Guide for Good Research Practice, in any type of research, the collected material is not the private property of the researchers or some research group.

It isn't just something they can do with whatever they wish. It must be stored and archived according to the general regulations issued by various authorities, primarily the Swedish National Archive. That means again; you need to make sure that you think about the fact that you may have an obligation to archive this data. It's not your data. So, for example, Cecilia Björkdahl, in a handbook developed at the Karolinska Institute -called ‘Research documentation at Karolinska Institutet: A handbook' says ultimately the Karolinska Institute is responsible for all research conducted at the university and is the legal owner of the raw and primary data.

So even though in Sweden there's this thing called the "lärare undantag" or teacher exception, which gives [teachers/]researchers at Swedish universities and higher education facilities the right to their own results. It doesn't give them ownership of their data.

The researcher’s right to his or her own results and thoughts are, of course, important for intellectual property issues so that we can write papers, we can file patents, etc.

The actual raw data belongs to the employer.

[slide 26]

There are a variety of other references that you can read for this. There is a summary at the end of this module, so there is a list of resources. I encourage you to go and read them.

I hope you've enjoyed this module, and I hope that you're going to conduct your research in an ethical manner.