**Module 4 - Safeguards, archiving and secondary uses of personal data - Transcript**

[Spaces indicate when the slide changes, and some timings are included where these are captured by the software]

[00:00:05] We're now going to look at safeguards, archiving and secondary uses of personal data.

[00:00:10] Let's look at the principles of GDPR again, but this time with research goggles on. Research uses data in some rather peculiar ways, and at first sight there would appear to be a bit of conflict between what the principles demand and how we commonly use data in research. For example, does the principle that personal data should be collected for specified, explicit and legitimate purposes mean that if data was collected, say to manage clinical care, we can't use that for research to address a research question? Does it mean that data collected to deliver social care cannot be used to address a research question? Well thankfully no, because GDPR and the Data Protection Act together…

[00:00:58] …have ensured that if you are doing, or if we are doing research within certain limits, so with certain safeguards applied, and we’ll come on to what those safeguards are in a moment. But if you are doing research, that's a limited type of safeguarded research, research is not an incompatible purpose. So research is not incompatible with clinical care. Research is not incompatible with social care, with education, so you can use data for secondary purposes. However, remember one of the big things in GDPR is to increase the importance of transparency. So although you can use data for these secondary purposes, you have to be transparent about it.

[00:01:49] So you have to ask yourself, is research a new purpose? If the public are not expecting their data to be used to support research, they haven't been told that anywhere in any transparency information, then it's a new purpose, and you're not being transparent about the secondary use and you have to do something about that. We will come onto what that might be in a moment. Equally I think in the past it was very common that NHS organisations would have a couple of sentences in their privacy notice that said: we're a research organisation, we could use your data to do research - full stop. I think in the era of transparency that just isn't enough. It's not specific enough and we need to put a bit more meat on the bones, to give the public a bit more of an idea about what that might mean. What data are you going to use? What purposes exactly, what are you going to do with it? Rather than just a very generic statement. So I think now, NHS caregiving organisations, really ought to have words to the effect that: we do research, we may use your personal data to support that research. For example, your care team may look through your medical notes to see if it's appropriate to invite you to take part in a drug trial and they'll look through your medical history to see if you're appropriate or not before they invite you. You will not be recruited to a drug trial without being asked. They're simply looking to see if you’re appropriate or not. Something like that gives a much clearer picture of what's going on. It's not like we have anything to hide at all and therefore we should be entirely transparent, there is nothing to lose. So making sure that research is mentioned clearly and quite specifically, with a bit of an idea of context is really important now in this era of GDPR. To cut a long story short, you can use data to support your research when that data was collected for some other purpose originally. But you can only do so provided you're transparent about it. Another principle that I think often looks like it may be problematic as far as research is concerned is the principle that you can only keep personal data identifiable as long as is necessary, and that's almost like how long is a piece of string.

[00:04:10] Well thankfully again GDPR and the Data Protection Act says that if you're doing research that is safeguarded. Again, it's this special limited type of research, we’ll come on the limits in a moment. Provided you are doing that kind of research, you can keep research data identifiable indefinitely. There is no need to get rid of it. So anyone who tells you that you can't keep data any more, archiving is off, because of GDPR. No. If it conforms to the safeguards that apply to research, you can keep research data indefinitely. Finally, and this isn't necessarily just a research problem but keeping data up to date. In research we often like out of date data. We, for example, might be looking at a timeline of the treatments of hypertension. So we know that Patient B had poor blood pressure, very high blood pressure, two years ago and we followed them up until now and their blood pressure's improved. GDPR does not, and should not, demand that we delete the out of date blood pressure measurements for patient B, and it doesn't.

[00:05:19] You only need to keep data up to date when necessary. If it's not necessary, you don't need to do it. And for lots of research, it is not necessary. Having said that, in the same research study, say you wanted to call that patient be back in for another clinic visit to measure their blood pressure again, you do need to know their contact details and you do need to make sure those contact details are up to date, so you do contact the right person, living at the right address. So although you don't need to keep all research data up to date, some of it, it is necessary to keep up to date, and when it is necessary that's what you must do. Okay, so most of the incompatibilities that people immediately start to suggest with GDPR, it's not true, provided your research conforms as to, as I say, to these specific safeguards. So let's talk about what they are.

[00:06:11] First of all you must implement additional, technical and organisational measures. That is over and above the sort of usual I.T. security stuff that your organisation will have in place anyway. Don't panic, if you are conforming to good practice and your research conforms to good practice, the chances are you already have appropriate technical and organisational measures in place. But we are going to visit these in a few moments, so hold your horses for a minute we'll get to it. The second safeguard is that your research must not cause any substantial damage or distress; and the final safeguard is that your research processing will not lead you to make any significant decisions about the participant, so that is the research processing of the data will not lead to any significant decisions being made. [00:07:04] Now we think it is true that for the vast majority of research you will not be making any significant decisions about the participant. But if you feel you are, then there is an alternative, which is your research is approved by an ethics committee. So it's either or. Either you make no significant decisions or you have ethics committee approval. That is not to say that you don't bother getting ethics committee approval at any other time, there are lots of reasons why you would get ethics committee approval, not to mention it being NHS policy and that the law demands it in other places. But as far as GDPR is concerned and the application of safeguards to your research that is the time you need ethics approval to conform to this final safeguard. So if all three of these safeguards are in place, you can keep data indefinitely and your research will not be considered incompatible, although you do need still to be transparent. So as a consequence these are called safeguards.

[00:08:08] There are other safeguards I should say before I move on. These are not the only safeguards referred to in GDPR, but they are the safeguards which you have to have in place, as I say, in order to keep data indefinitely, in order for research not to be considered incompatible.

[00:08:23] So let's look at what technical and organisational measures actually means. In a nutshell really it covers two things. Two things that the chances are you are probably already doing. The first is pseudonymising. We've already talked about pseudonymising a little bit when we were talking about personal data, we know that pseudonymising does not render data no longer personal. But one thing pseudonymising does do is it helps manage the risk of disclosure. By pseudonymising, what I mean is you take your research data and you remove all the real world identifiers from that research data, you squirrel those identifiers away in a locked filing cabinet. You have a cipher which links real world identifiers with your research data, which is probably on an Excel spreadsheet or on a computer somewhere. You limit the number of people who have access to the cipher. This is really good at limiting disclosure because you limit the number of people who have access to anything identifiable. And if anyone were to break into your office or to hack into your computer, it's going to take quite a bit to be able to put the two bits of the jigsaw together to help identify anyone. So it really does limit the chances of disclosure. However, by pseudonymising you do introduce a bit of a risk yourself to the research.

[00:09:48] Mistakes can be made easily when you start applying a kind of randomised cipher to data, and as a consequence data can start to get muddled up and be attributed to the wrong individual, the wrong coded individual. So although pseudonymising helps manage the risk of disclosure, that is balanced with the threats it may present to the delivery of your research, and really you should try and limit the threats to research delivery through pseudonymising: by ensuring you have a standardised method to code and pseudonymise; train people in how to do it. They may know how to pseudonymise on another study, they've done it many years, many many times before, but how do they know how to do it now, how to apply this system? You audit the process to make sure that mistakes are not being made, you keep the process simple and if you really can't limit the risk sufficiently then maybe you have to start relying on less robust codes, and not having it quite as heavily anonymised as de-identified as it might be.

[00:11:00] So that's one of the safeguards: pseudonymising, comes at his own risk. The other safeguard is data minimising: you only collect the data you need; you minimise the number of participants you recruit; you minimise the amount of data you collect per participant; and you minimise the degree of sensitivity of that information.

[00:11:21] As I say, I suspect you already do most of these things already, and actually often compliance with, particularly, data minimisation is overseen by other organisations. Your funder will usually only fund you to collect the data that's actually necessary to answer the research question. They don't want you to be recruiting many many more participants than you actually need. Ethics committees will have checked that you are recruiting the right number of people, that you're collecting the right kinds of data to address the question. So there are many different checks and balances already in place where you can demonstrate that you are indeed collecting, holding and using a minimum amount of data.

[00:12:03] Don't be alarmed. I think there is this worry that data minimisation means the end of big data studies. Not at all. It just means you have to justify the data you hold, and certain methodologies and certain statistical analyses do require masses and masses of data. Well if that's what you need, that's what you can collect, hold and use, but you just need to justify the volume of data that you are processing. That's all GDPR asks. It does not ask you only to hold a small amount, it's to hold the minimum you need.

[00:12:43] So, as I say, those are the safeguards. Hopefully now you understand the technical and organisational measures and I suspect, as I say, most of you will already have the technical and organisational measures in place:- No substantial damage or distress (I would assume that's not that likely any way). And no significant decisions, or if you believe you are making significant decisions then your research has been approved by an ethics committee as an alternative. As I say, there are others.

[00:13:13] Some of the others are wrapped up actually in your lawful basis. So if you rely, which we are encouraging and suggesting we should, rely on task in the public interest or legitimate interest as your lawful basis, there are, and with additional conditions of research in place, there are additional public interest tests that you have to apply, and you have to demonstrate and comply with. Your Data Protection Officer will be worrying over things like this. It's not really, I don't think, up to you as a researcher to worry too much about the public interest tests, but be aware that your Data Protection Officer may be asking you to demonstrate that you have got, for example, ethics approval or funding from a public body, or whatever, but that's why it's so to make sure that you can fulfill these public interest tests, and therefore rely on the lawful basis that we've talked about earlier. If we put all of this together this really does constitute the minimum safeguards you have to have in place, if you're doing research and you want access to the research exemptions and the kind of research understandings that have been engineered into GDPR. You can do research without those, but you do not have access to any of the exemptions or any of the understandings.

[00:14:34] So, just to kind of really reiterate I guess - Storing data: GDPR does allow researchers to hold research data indefinitely, provided, as I say, those safeguards are in place. I think sometimes there's a bit of a muddle when we talk about storing research data and archiving records. Research data, you know as researchers, it's very hard won and you will not be deleting research data unless you absolutely have to - and you do not have to. Storing research records: things like participant information sheets; copies of the protocol; amendments; approvals; all of this stuff. These records also should be kept for a prescribed period of time. The MRC suggests that for clinical studies this period of time should be 20 years. It’s very much a risk-based decision. So what are the risks of throwing it away, versus what are the risks of keeping it? If you look on our resources page, again on our website, you'll find a document that helps you go through some of the decision-making around how long you should keep research records for. I think now that we're in an era of quality assurance, if you have appropriate quality assurance processes in place, they may well mitigate the need for some of this really long term storage of research records. But it's really important, as a researcher, when you apply for funding that you do identify the costs potential costs of research archiving upfront, and you ask your funder to provide money to cover it. Archiving is not cheap. You also need to think how you're going to manage long term archiving. Who is going to be the contact, who knows what's being archived where, and again talk within your organisation, your Data Protection Officer, research governance office. These people can provide a lot of input into providing the right arrangements and getting the right managerial oversight on the archiving of research records.

[00:16:35] So finally, after you listen to this, I think I would encourage you to sit down and have a think about what safeguards you think you already have in place. Start to think about what approvals you have in place, ethics approval, funding, etc. How would you justify the amount of personal data you're collecting, holding and using? Has anyone else, like a funder, also said: yes that looks like the right amount of data to us. Do you use pseudonymisation, and if you do, have you thought about what risks pseudonymisation itself is introducing to the delivery of your research outcomes. Are you checking for mistakes? Do you follow a standard process? Is there something you could do to improve the process of pseudonymising to make sure that it really is robust. Not only does it limit the risk of disclosure, but it is also a robust process in terms of delivering research outcomes.