**Module 2 – GDPR, Personal Data and Data Controllers - Transcript**

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

[00:00:07] In this module we're finally going to start talking about GDPR, and in particular going to visit two definitions: personal data and data controller.

[00:00:18] So far in the previous modules we spoke about information and what makes this confidential. But the reality is, as researchers we accrue vast amounts of information, electronically or file it away in some paper filing system. In other words we collect large amounts of data. We do not live in a world where we can just accrue data without people knowing what we've got on them. We do have a right to know who's holding our data what are they doing with it, is it accurate, what kind of decisions are they making on the basis of it. We have a right to correct it if it's not accurate. We have a right to know that it's safe from hackers and from accidental disposal.

[00:01:03] And this is because we live in a world with data protection law. GDPR is the latest iteration of data protection law applied across Europe. It places additional requirements on those who collect, hold and use personal data, additional requirements, over and above the common law. GDPR has no impact on common law. The common law rumbles along as it always has done. GDPR places additional requirements on top of the common law. As I said if you collect, hold and use personal data. Now Personal data we're going to define in a few minutes, it's not the same thing as confidential information. There are similarities but there are also differences. I cannot talk about GDPR in a research context without mentioning also the Data Protection Act 2018. European government derogated aspects of GDPR to member states to legislate specifically around the details of how GDPR would be applied, for example to cover research, other areas also. So whenever we start talking about GDPR and its application to research, we also have to bear in mind the Data Protection Act 2018 here in the UK. Throughout the rest of the modules I'm going to keep using GDPR, the acronym, I don't just mean when I say that, the General Data Protection Regulation, I'm probably also incorporating details taken from the Data Protection Act 2018. But I'm not going to really worry too much precisely which detail sits in which piece of legislation. I'm sure you've all got better things to worry about. Suffice to say, data protection or GDPR and the Data Protection Act is very much a corporate responsibility. It is the data controller who is responsible for compliance. And that's why I've spent a little bit of time towards the end of this module talking about who the data controller is likely to be in a research context.

[00:03:04] Let's start by looking at what is personal data. First of all I've already alluded to this. It's data, structured information that could be electronically structured or structured in some kind of manual filing system. It also has to relate to or be about a living person. Once you are dead, your data is no longer considered personal and therefore is no longer covered by GDPR. But whilst you're alive your data may well be considered personal, and therefore covered by GDPR. So it has to be about or relate to a living person. And finally, and possibly the most difficult part of the definition, it‘s identifiable. That is, identifiable on its own or identifiable when in combination with other information that you are likely to have access to. Now we've talked about identify ability in previous modules. If you're not clear on what makes information and data identifiable, I would revisit that module just to get it really clear in your head. I'm going to do a quick summary here. Again you might want to spend a little bit more time thinking about it, in which case please go back to the previous module (Module 1b).

[00:04:11] Identifiability, as I've said before depends on two things: the content of the dataset and the context in which that data is viewed.

[00:04:22] The test of whether it is identifiable or not is whether or not a motivated intruder would be reasonably likely to be able to identify an individual, that is taking into, again, into account both the content of the data and the context in which it is being viewed.

[00:04:37] And as we've said before, you can control the context of how data is viewed by placing agreements, legal agreements, in place when data is moved around. So if you share data with a collaborator, you can strip all the obvious direct identifiers out of it, you share the data with your collaborator and provided you control the context the recipient views that data through the use of a confidentiality agreement, for example, then you can say that they are holding anonymised data, because it is not reasonably likely that they would, or a motivated intruder, would be able to identify individuals from the data you’ve just sent. It’s common practice in a research context to pseudomymise data, that is to strip the data we're interested in, the health data, from all the identifiers. We keep the identifiers separate in a locked filing cabinet. We often have a cipher which will link those identifiers through a code to the gubbins, the actual research data we hold on a computer system somewhere. Common practice, good practice, and we'll see why in later modules, you should continue to do that, it's an important process under GDPR. But the one thing it does not do, is render the data you're holding not personal. The fact that you have access to, you wouldn't go near it I know, but the fact you do have access to the filing cabinet with the cipher and all the direct identifiers in it....

[00:06:08] ...means that pseudonymised data is likely to be personal. Again this does come with the proviso that the Information Commissioner's Office is revisiting their definitions of anonymised, pseudonymised and identifiable data. But at the moment, without that new code, we're going to work with the old code, the anonymisation code, and that is how things were, subject to change.

[00:06:35] OK so that's personal data, but GDPR doesn't just talk about personal data, there's also a subset of personal data called special categories of data. For those data protection anoraks out there, this is roughly equivalent to what we used to call sensitive personal data in the old Data Protection Act. All health data, if it's personal, is also likely to be classed as a special category of personal data.

[00:06:59] This would include certain genetic data and I'm going to talk about genetic data in the next slide because there's been a lot of gossip and rumour about the status of genetic data in the land of GDPR. Suffice to say, that most people who conduct health research will be collecting, holding and using, in other words processing, special categories of personal data, not just processing personal data.

[00:07:26] Ok, genetic information. Genetic data is specifically mentioned in GDPR. And I think this has got a lot of people worrying and getting anxious over whether genetic data is somehow different from other data in its classification as identifiable or not. Well it isn't. It's just the same as any other piece of data. So some genetic information may well be identifiable, but it isn't always, it depends on the content of that genetic data, and importantly the context in which that data, the genetic data, is viewed. And in particular in terms of genetic data, that context is all about what other information are you reasonably likely to have access to. If you are reasonably likely to have access to information that may help you identify an individual from a sequence of DNA, then you would say that sequence of DNA is identifiable. If you are not reasonably likely, then that sequence of DNA is not likely to be classed as identifiable. So in other words you apply exactly the same definition, it's content and context, to determine whether or not DNA sequences are identifiable or not, and therefore whether they are classed as special categories of personal data or not. You should also be aware that the generation of DNA data is governed through the Human Tissue Act in the UK. The holding of bodily material with the intent of analysing DNA is covered under Section 45 of the act. What GDPR covers is what you do with that DNA information once you have it. But the generation of it from tissue is governed elsewhere in the Human Tissue Act. That's personal data and special categories of data.

[00:09:22] Next one. What is what is a data controller? A data controller would normally be an organisation in the UK, and it's the organisation that is legally responsible for the compliance with GDPR. You don't want to be a data controller as an individual, trust me. The data controller is the organisation that decides what data is going to be collected, what it is going to be used for, how it's going to be handled, what technical and organisational measures you’re going to implement to try and keep it safe. And they're also responsible for ensuring that they're being transparent about the data that they hold and use as an organisation.

[00:10:01] For research who do we think would normally take that role? If you think about it, we already have an organisation that does a lot of these things, and that's the sponsor. We suspect that for most research, medical research, health research, it will be the sponsor who will be acting as data controller, and therefore there may well be instances where you have joint or co-sponsors. You may have joint or co data controllers.

[00:10:29] Data controllers also have to decide what their lawful basis is for holding personal data and what additional conditions have to be applied for them to hold special categories of personal data. We're going to talk about this in the next module, so don't worry about that just now. But they take on very specific legal accountabilities for compliance.

[00:10:51] GDPR talks not only about data controllers but also processors. Data processors are organisations that work under the instruction of a data controller. So you could imagine in a research context, you have a data controller, a sponsor, who asks various data processors, research sites, to collect data for them, to complete CRFs, to pseudonymise data and then to transfer that pseudonymised data to a central place for statistical analysis. Pseudonymised data, hopefully is being transferred with an agreement in place also, therefore we can ensure the data can be classed truly as no longer personal, once it's held by that central laboratory. The data controller, therefore, in that kind of scenario may actually not directly hold any personal data themselves. Personal data is only physically being handled by the data processors, with data controllers possibly receiving only anonymised data.

[00:11:51] As I said before, under GDPR, being a data controller is a corporate responsibility, we would expect controllers and processors to be organizations here in the UK. Organisations will have governance structures and processes in place that oversee activities going on within their organisation. They will have management processes in place to manage corporate risk, and data protection risk is part of that corporate risk. And part of this management of risk will include completing privacy impact assessments for certain activities, where data protection risk is significantly high, or high enough. For some research some of you may be told or asked to complete a privacy impact assessment because research is a slightly odd, if you like, use of data, or can involve slightly odd uses of data. If you are asked to conduct a privacy impact assessment, don't try and do it on your own, you should speak to your data protection officer, you should probably, failing that, or all together with your research governance office. These are the people who really do understand the nitty gritty, technical details of what data protection processes have been put in place by your organisation. You also ought to find out when your organisation is going to require you to complete a privacy impact assessment and how to go about doing it. So after this module, I would encourage all of you, if you don't already know, to find out who your DPO is and how you contact them; where is your research governance office, and how you get in touch with them; when will you need to complete a privacy impact assessment, and how would you go about doing it, how does your organisation expect you to go about doing it.

[00:13:39] Now to help you work through some of these principles that we've just talked about, again, we've produced real life examples, which we would encourage you to have a look at. And try and identify when personal data is likely to be processed in the scenarios, when and where throughout the scenario personal data is likely to be used, and to try and identify who the data controller is likely to be, or the data controllers, because in some of the scenarios there may be more than one data controller.

[00:14:06] Again, to help you through the workshop, and to get to the right kind of answers and to get you asking the right kind of questions, it can be really helpful to draw a data flow diagram. To really understand where data is coming from; what's it going to be linked to; who's keeping it; at what point is it pseudonymised or anonymised; who's analysing it; where's it going to be stored; etc. So to plot out the flow of data around the scenario can really help you come to the correct conclusions.

Thank you.