**Module1b – What makes information identifiable? - Transcript**

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

[00:00:05] This module is going to look at what makes information identifiable.

[00:00:10] We've already used identifiable when we were talking about the common law of confidentiality, and actually the same definition that we're going to explore here is going to be used in later models, when we start looking at GDPR itself. So what does make information identifiable?

[00:00:27] Do you think that year of birth on its own is an identifier? Probably not.

[00:00:33] What about gender on its own, nothing else, just gender? Again, probably not.

[00:00:39] What about first part postcode, on its own, nothing else, just first part postcode? Well some of you are already shouting ‘probably okay’. Probably not in most cases but some postcode areas are very small. We've all been to meetings where we are the only man or the only woman. So there are situations where actually all of these parameters are identifiable on their own. Not usually, but there are some.

[00:01:05] And the same could be true for place of birth. Again, on its own with nothing else, not identifiable. You know what I'm going to do now don’t you?

[00:01:15] What happens if we stick it all together? What happens if we say we have information constituting year of birth, gender, first part postcode and place of birth? Is that identifiable do you think?

[00:01:24] Let's test it by actually filling it in with some real information. Year of birth – 1926. Gender - Female. First part of postcode - S.W. 1 W. Place of Birth – Mayfair. Some of you may already know this and shouting it out. Others it may take you a few minutes to google it but it won't take you long to determine that that information relates to the Queen.

[00:01:51] This just illustrates, I think, the difficulty in defining identifiability. It is dependent on two things: the content of the information, the sort of identifiers (weak and strong) it may contain, but also, just as importantly, the context in which that information is going to be viewed. So by that I mean what other information, for example, would you have access to? Do you have access to the Internet and all of the information available at your fingertips? As a consequence, if you do, can you put information together with the information that we're looking at, to render that information, the limited information, identifiable. So that’s why I put a picture of a jigsaw here on this slide. If you can add other information, link things together in order to identify the individual concerned, then it is identifiable. This is obviously the context again, it can be made even clearer if we're talking about something rare or unusual. It actually can be very difficult to really anonymise person level information if it contains, say, rare or unusual events.

[00:02:59] What we're talking about here is a real grey scale: from anonymous to identifiable, and where you are along the grey scale depends on both content and the context of the information.

[00:03:09] The law however is binary and therefore we have to make decisions. Is it identifiable or isn't it? We can't opt for well that kind of depends. To help us, the Information Commissioner's Office produced some guidance - the Anonymisation Code, which I should say is being rewritten currently so this is subject to change, but I suspect that change is not going to be dramatic.

[00:03:36] In previous guidance, the Information Commissioner's Office has said: yes, there is some information which is anonymous, inherently. It contains no identifiers, things like, I don't know, some batched statistics about populations is definitely anonymous, and some very limited person level data may be truly anonymous.

[00:03:57] We have another sector of information which we should class as anonymised. This is information where it may contain some weak identifiers, but we're confident that we can control the context sufficiently that identification is not reasonably likely. Now by controlling the context I mean things like limiting the access people may have to further information. So, for example, a legal agreement promising not to make any attempt to re-identify anyone. It is a question of literally or metaphorically switching off the Internet.

[00:04:40] What do you mean by not reasonably likely? It is a motivation intruder would not be reasonably likely to be able to identify anyone from the information we're talking about. A motivated intruder is not a hacker. It is someone who is slightly more motivated than others to discover the identity of someone.

[00:04:57] So for example if the information was about a celebrity, the queen or indeed any other celebrity, then there could be quite a few people who may be a little bit more motivated than most to identify whether or not this information does indeed relate to the celebrity. They are a little bit more motivated. If your information can withstand that kind of scrutiny, then you can say it is reasonably likely it’s not identifiable. It is anonymised. And as I say the best way, or the most common way, in a research context of controlling that context is through entering into confidentiality agreements. So, you can share information that you have stripped of all strong identifiers, share it with others who do not have a duty of confidence with the individuals concerned. Provided you also control the context in which you share that information by entering into a data sharing agreement with your collaborator, whereby your collaborators promise to make no attempt to re-identify the information that you're supplying them with. And then we can say that it is as good as anonymised, and there has been no breach in confidentiality.

[00:06:19] I hope that's helped you work out the answers to some of those workshop scenarios that I talked about in the first module. If you are interested in how those scenarios can be worked through, we have put up a short module, which explains how it's safe to share information in those different scenarios.

[00:06:40] That's it for this module. In the next one we're actually going to talk about the General Data Protection Regulation.