

## GMT20200429-081721\_Digital-Pr\_1920x1080-converted 1.mp3

**M** [00:00:03] Ok let's continue.

**M** [00:00:11] So you've seen this picture before. We are agreeing on the question meaning we've done that and we've made the initial estimates for everyone. So now today's or this morning's task is to discuss the similarities and differences and then this afternoon we'll go on to provide final confidential estimates. I do recommend you have the spreadsheet in front of you that you filled in yesterday and whether you like to take notes on paper or directly into a random column on the spreadsheet then as we go through, you might find it easy, easier, if you make notes about the various discussion points that strike you as something you hadn't considered yesterday, so when you come back to that question, you can, ~~you can~~, take that into consideration when you when you decide if you want to change your estimates from yesterday's or whether you want them to be the same. If they're the same, we we'd ask you to write them in again that would be very helpful. So we're very happy to see the comments as well, that also helps us with our modelling to see how people are thinking about it, particularly if it pokes us into thinking away about it in a way that we hadn't considered. As, ~~as~~, Alex said, with the TNA focus, sometimes it can be easy to miss details which are important in other ~~other~~ archives.

**M** [00:01:37] So I've had a look through the range plots and I am seeing what exactly what I would expect to see for round one plots. Shall we just have a quick look at plot number one?

**H** [00:01:54] OK. It's at this point, you'll have to bear with me, so I've got work out how to put the screen on, so I'm going to stop. It might be about two minutes for me to get this working, so bear with me.

**W** [00:02:10] Can I just step in while you're doing that?

**H** [00:02:11] Yeah. Good idea.

**W** [00:02:14] I'll just say a word just now, if I may, to everyone. So I mentioned at the start yesterday that the DPC is doing an evaluation of the project, that's one of our roles here. And so not surprisingly, every good workshop has this evaluation form at the end. This therefore, being a well-run and well-organized workshop, we've got our evaluation form ready for the end of the workshop. So I will be sending that round as a Google form. I'll do it at the end because there's no point in doing it midway through, really. But I do, I really would like to lean on you, please, to just to follow up. Our experience in DPC ~~of~~ evaluation forms is that it's actually pretty hard to get people to do the online evaluation forms. It's a lot easier to circulate bits of paper. You get much more responses, even though that's a huge amount more work. So I will be sending that around.

**W** [00:03:10] It's partly about the workshop but also it's about setting ~~and~~ and measuring your sense about the outcomes of the project. Remember, the National Lottery Heritage Fund has quite specific requirements about project outcomes. So we need to evaluate against their kind of set list of project outcomes. Anyway that's going to come to you towards the end of today so watch out for that in your mailbox in due course. I think that was all I wanted to say about that just now H and A so, yeah.

**H** [00:03:51] Thank you. I will just. I've got everything ready now. Let me work out how I returned to the call. Right now share screen. Got it I think, can you all see a range graph on the screen? Yes, excellent.

**M** [00:04:24] Well done. That's good. All right.

**M** [00:04:29] I'll change the ratio between the shared screen and the the people so I can see the range graph a lot bigger. So what I think we'll do H if it's okay with you is if you will read out the question and then I'll discuss the range graph and we will facilitate the discussion. But before we do that, what I'd like to do is just to look at question one's range graph, not as a question one, but as in just an example of a range graph and pick out some of the features that we are seeing and typically do see on range graphs like this. Is that OK?

**H** [00:05:08] Yeah. Good with me.

**M** [00:05:10] Thank you. So here we have 19 experts who've given their very valuable opinions. And what one thing we notice is that some people's range graphs are much wider than others. So if you look at the blue one, number 13, number 3 rather, down the bottom here, and the green one just above it, number four, if we look at number 15 the brown one nearer the top and 17, which is the sort of pale brown one near the top as well. We see ~~that that~~ that their ranges are very, very wide. Now, during the discussion, we hope to find out whether they think that the system itself is very uncertain or whether they personally were very uncertain. If they personally were very uncertain after we'd have a discussion they may be confident to reduce the range of those graphs. [coughs] Excuse me. We see two very narrow ranges tucked down there to the left. The red one, which is Expert 2 and the blue one which is Expert 9. So these ~~these~~ experts are very confident that they know the value of this, the answer to this question. And they know it within very narrow certainty bands. They think that it's very, very close to the left there very close to zero and there's no way it could be anywhere near as big as 800, 900 or whatever. We have also Expert 11 in green, and Expert 5 in purple who have exactly the same answers. They are very relatively narrow experts and narrow uncertainty bands and they are up at the 800 give or take end of things. So there's no actual connection between Expert 11 and 5's estimates and Experts 9 and 2's estimates. [coughs] Excuse me.

**M** [00:07:06] So we'll discover during this discussion whether that's down to a misunderstanding of the question, we've often had this before, we've had one or two people completely out ~~from~~ from everybody else and they've sort of understood the question almost in reverse. So when they understood the question the same as everybody else, they revised their estimates in the other way. Or whether it's actually a genuine difference of opinion and that's fine too.

**M** [00:07:40] Notice that Expert 3 has very wide uncertainty bounds, but is the only expert to capture everybody else's views within their own. OK. So we've got the people with the very narrow uncertainty bands are very informative, but they can't all be accurate because their ~~their~~ uncertainty bands don't overlap. The people with very wide uncertainty bands are going to be accurate, they're going to capture the correct value, but they're not very informative. And then everybody else is somewhere in between. Does that make sense to everybody? Any questions in the chat?

**A** [00:08:19] You can always, if you want to just put your thumbs up and say yes, you can always use the reactions in the bottom bar. Expert 16 very kindly tested it for me this morning. And a little thumbs up will appear in the corner of your image.

**M** [00:08:42] I like that the reactions are only ever positive, you can either, you can put a thumbs up and you clap. [--] to have. So it's very kind of [--] and really positive here [--] yeah very good, we need a bit of positivity in lockdown. I find nothing more tiresome than everybody moaning about everybody else. Oh, come on, people we're in this together. Anyway, to the quest to the to the.

**H** [00:09:08] Yeah. Oh, so shall I read question one? Are you ready?

**M** [00:09:12] Yes please read question 1 yep.

### Question 1

**H** [00:09:14] So out of a thousand UK archivists, how many would you expect to say that the archive has a list of all the digital file formats collected?

**M** [00:09:28] So the Experts number 9 and 2 think that very few of them will have a list of all those, most people think it's somewhere in the middle. Experts 5 and 11 are pretty certain it's going to be fairly high, about at around 80 percent as far, maybe up and up as far as 90 percent, but not lower than 70 percent. So can we think of reasons why it could be down low with Expert 2 and Expert 9? Could it be as low as that? Could it, is it likely to be as low as a well I can't remember it will look, looks like about ten or twelve to me. Would it be down that low? You might need to unmute yourselves, actually.

**E07** [00:10:23] I would suggest not, and ~~and~~ my reasoning would be that, so we're not in this question talking about the skill of individual archivists have we're talking about their archive. So within an institution, you might have lots of archivists, many of which will have no idea how to create lists of digital files formats, but that there might be someone within their archive that does have that skill, so that's why I would put it a bit higher than that.

**E11** [00:11:00] I agree. I assume that be some duplication within archive services. I think you've get a different profile if you're talking about archive services. But given the likelihood of multiple archivists being in the same repository, that does change the likelihood in my view.

**M** [00:11:25] So any anybody who can think of any reasons why it might be low?

**E20** [00:11:30] Well, if I could pick out the word 'all'. So it says a list of all those digital file formats. Sorry, I think Expert 19 is speaking as well, should I continue?

**E19** [00:11:48] Yes, please do.

**E20** [00:11:49] [00:11:49] Sorry [name]. So I was thinking of the word all. OK, so the word all implies complete control. And that's something that everyone would aspire to but whether or not they would, they would have that is an open question to me.

**M** [00:12:11] So that's a good reason why it could be low.

**E19** [00:12:14] Yeah. I was basically going to make the same point as Expert 20. I mean, to ~~to~~ disqualify that from ~~from~~ being ticked, then you just need a single unidentified file. I know that in our legacy collections are things like we've got known unknowns and, you know, we're one of the well-equipped ones. So that's why I thought it would be quite low.

**M** [00:12:39] So your known unknowns include things like the file format.

**E19** [00:12:46] Yeah.

**M** [00:12:50] Ok that's fine. Good good.

**E07** [00:12:51] Ah so I thought, maybe I'm got it wrong, but when we were discussing the question yesterday, my interpretation that I took away was you might have a list of all the file formats, but some of them it would just say unidentified so you wouldn't actually know what those file formats are. But ~~I would still~~ I would still make that, I'd still count that as a list of of all the file formats. So I guess I'm just ~~just~~ thinking about the question in a different way.

**E09** [00:13:23] I think I was thinking of it in the same way as well, because we sort of mentioned that it wasn't necessarily knowing each PRONOM identifier but just having a list of sort of formats in a more generic way.

**E14** [00:13:45] Yeah, I was I was reading it that way as well. File formats, a list of file formats could include unidentified as as a classification.

**M** [00:13:56] So in that case, your known unknowns would actually be a classification of its own and that would count as a yes rather than a no. H can you remember, does that accord with the discussion we had, and A?

**H** [00:14:13] Let me have a look. I think that was what we, I do remember talking about known unknowns.

**E10** [00:14:26] Hello there, it's [name] here. I was just thinking about this, so I thought it might be quite likely that archives or archivists would know that they have collections, legacy collections on formats like floppy disk or CD-ROM, but they wouldn't necessarily have done the work to understand that the floppy disk contains specific file formats like JPEG or even Word documents or whatever else it would be. So I thought it was less likely that they'd know specific individual formats and more likely that they would know carrier mediums, which wasn't the question.

**M** [00:15:07] No, that's ~~that's~~ helpful, thank you. Does anybody agree with ~~that~~, is that is that likely ~~to have~~ to have carriers with file formats that you don't know on? It might vary from archive to archive, of course.

**E07** [00:15:35] I think it's possible, but also I think that most people understand that one of the key first steps of digital preservation is to get all content off portable media into a kind of a safe storage space where they can then start to do activities like file format identification. So ~~I do~~ I do recognise there will be some archives that have not kind of taken those first steps yet. But I still think that quite a lot of people would have done that. That might be one of the first things that they've done.

**A** [00:16:21] Quite a few people in the conversation that are saying they focussed on the word 'all' if that makes any, if that helps.

**M** [00:16:29] So I think we just need to have a common understanding of this. When we say all what do we mean exactly by that, do we mean that known unknowns count as a category?

**A** [00:16:45] Yeah, I guess I think that, I hadn't thought of that.

**H** [00:16:51] I think, I think we should include known unknowns.

**E10** [00:16:53] Yeah. I thought that's what we'd said in the question.

**H** [00:17:01] So we're going to include known unknowns when we we answer this again.

**M** [00:17:07] We know there is a file format, but we don't know what the file format is. That's on the list. Any ~~any~~ other comments or questions for Question 1?

**H** [00:17:28] So they've attempted to identify all files in their collection says Expert 14, that is probably a better.

**E07** [00:17:37] Yeah. It makes sense.

**M** [00:17:42] Excellent, well, last call for questions? Question 2 it is. OK.

## Question 2

**H** [00:17:50] So question 2: out of a thousand UK archivists, how many would you expect to say that they had at least some knowledge or skill to perform file format analysis of a digital accession? And I think when we were discussing this, we were talking about using a tool to identify the file.

**M** [00:18:14] Yeah, so we use the word identification, I wrote that down. We thought it might be a better word than analysis didn't we. Yeah. So we're looking here at there's a big group in the middle. So from Experts 9, 9 up to 19 and also Expert, the green one Expert 4, so there's ~~there's~~ a big range sort of in the middle there. We have Expert 8 who's got wide uncertainty bands. But notice how the distribution is very asymmetric, so it's much more likely to be high but there's a long tail to the left where it could be low. And again, we have Expert 6, similar pattern. It's more likely to be high than low, but it's possible to be quite low, very wide uncertainty bands. We have Expert 7, a very small range on the left. So. Yeah. And we have 1 and 2 completely non-overlapping, they just meet at the bottom end of the range for Expert 2 is exactly the same as the top end for Expert 1. OK. So what are the some ~~some~~ of the reasons why we might have very few saying that they had the expert, they had the knowledge to perform identification?

**E16** [00:19:53] For me, I think this was potentially about the understanding of the question. So I took it and I thought this is what I thought we discussed as well, to mean that you may not actually have the tools to do it and you may not have done it, but you had an understanding of what was available and had an understanding of how to do it.

**M** [00:20:16] Yes.

**E16** [00:20:18] Which is why I thought that actually the probability was very high. I think most archivists have an understanding of what should be done and how to do it. But they don't necessarily have the tools. So I think the response to this question is actually quite a high probability.

**M** [00:20:37] Yeah. Yeah. So you're going to close to a thousand out of a thousand.

**E14** [00:20:43] I think my experience of kind of talking to people is that most of, they kind of vaguely know it in theory but they don't have the confidence to actually try using it. So I think in ~~in~~ a survey they would probably answer quite negatively about their ability to do that.

**E10** [00:21:05] I thought there was something here about the demographic of the archivists, because I think if you're trying to look at a spread over a career, all of the archivists over a career, and I think there would be some who will be much less confident with ~~all~~ anything digital and this which is quite specific, would be even harder. Whereas if you're perhaps looking at more newly qualified archivists who have come across this topic in their training, then they, I think, feel they'd be more confident to say that they would manage it. So I was sort of assuming, trying to assume that a thousand archivists spread over that range of experience.

**M** [00:21:53] That's very clever, that's exactly the sort of thing we should expect to think about. Yes, it's me, but it's also my colleagues in other archives, some of whom are more recently trained, some of whom are more experienced, et cetera. So that's very good.

**M** [00:22:14] And so we've got some reasons why it might be quite high because we were talking about some knowledge to perform that, I think, was it this, was this the question where we said even if you didn't have the tools to do it yourself, you would know where to go or something that? You knew it could be and you knew you could buy in the expertise. Was this the question we were discussing that?

**D** [00:22:34] I think that was Question 3.

**H** [00:22:36] The next one. But I think it's a ~~similar~~, probably similar idea we can.

**E03** [00:22:40] What I noted down for this question was that we said it was going to be about identification rather than analysis. And that you might be using a tool rather than doing it manually, as it were, on Windows Explorer. That was the one thing I noticed about this this question.

**E20** [00:22:56] So for me, if I may, I pick up also [Expert 11's] comment about the demand for introductory training being very high and our experience of delivering the introductory training, we typically address that file format question at some depth, which is a weird paradox because on one hand it leads me to assume that the demand for training reveals a shortage of demand or a lack of skill. On the other hand, I'm tired of repeating myself. I've done it so often that it leads me into a kind of paradoxical position of expecting it to be both low and high at the same time. Not very helpful.

**M** [00:23:40] But it gives us reasons to think about why it would be low. And this might be to do with training or it might be to do with confidence and we are asking ~~would~~ would you expect them to say...

**E03** [00:23:51] I think that's a really critical factor, actually, because I think probably people do have more skills than they're confident about. And it's just about the practice and maybe thinking about what Expert 20 was just saying, there are probably people who are doing training courses, but you know how it is. If you don't, then go back and apply that training straight away, if you've still got factors in your institution that are sort of holding you back from taking that forward, it may be that then by the time that comes around, you think, oh gosh, do I know what I'm doing? What if it's all changed? Because I know from talking to colleagues over the years about training in digital preservation, you know, it sort of feels you're going over the same ground, that things might have moved forward a little bit and there are new things to learn, but you can sort of feel a bit like you're, it can be hard to make progress for those sort of various reasons that are to do with your institution and applying those skills. So I think looking at this again, ~~it is a~~ it's really critical about that expect to say that they had at least some skills. I think that's where it starts to pin down. And so for me, I was one of the people who did quite a wide range, but with a high point that I thought, you know, this is where I think it probably is for that reason, because there's so much variety and institutions and people's ability to apply knowledge that they've learned. But I think people do probably have more skills ~~than~~ they than they realise even just to do those sort of basic things.

**E20** [00:25:11] I agree wholeheartedly with that Expert 3, and I think also people have skills that they're not able to apply because of internal procedures associated, particularly with IT departments and they know where the tools are but they would never get to install them.

**E03** [00:25:24] That's right. So and then by the time you can, you're not sure you've got the up to date skills anymore, so you kind of go back to square one and sign up for another DPC course.

**E20** [00:25:35] The gift that keeps on giving.

**E03** [00:25:37] Yes.

**M** [00:25:39] So we've got some comments on that. So A do you want to comment on those? I don't know what Droid is.

**A** [00:25:56] Yeah, Droid ~~is our~~ is our tool for identifying file formats and yeah, one of the problems we have is that lots of people with outsourced IT are not allowed to install something even if it's completely harmless and from a hopefully trusted source like The National Archives. So, you know, we tell people, you know, you can use this tool, it's freely available, we support it, we'll help you use it but ~~if you~~, if your IT department has a policy of charging you thousands of pounds just to install something, penetration testing, etc., etc., it means they can't do it, they just, they physically can't do it. So that's one of the problems we have around people being able to use that tool and I'm assuming that goes for other tools as well.

**E14** [00:26:43] And so as a result, they won't feel they have the confidence to run it even if they have been on a course or ~~or~~ whatever. So in a survey, if they're asked, they're unlikely to say that they would feel confident to run it.

**M** [00:26:57] ok, I think that's very helpful. I think we're coming come to the conclusion it's about the saying and there might be a lot of social or social reasons like, oh, well, that person's always better, they always do it, there might be, you know, a lack of practice back

at the organisation after having training, there might be barriers to do with IT and IT availability, which would cause people who objectively we might consider to have the skills to say that they didn't have the skills. Is that a fair summary?

**A** [00:27:30] Yeah, I think so.

**M** [00:27:32] OK, so is everybody happy to make their own notes on question 2 and then move on to question 3?

### Question 3

**H** [00:27:40] So question 3: out of a thousand UK archivists, how many would you expect to say that there was some capability within their organisation to carry out at least one of a) file format migration b) software emulation or c) data recovery from damaged or obsolete media, even in a limited way?

**M** [00:28:06] OK. So we're asking people would say they had a limited way of doing these technical things.

**M** [00:28:13] So again, we're looking at the range graph here and we're seeing that expert number 2 in the red, expert number 4 in the bright green, expert number 13 in the pink and expert number 15 in the brown are all very much on the low end with ~~relatively~~ relatively narrow bands, fairly certain that it lies in that range there. We have an expert number 5 in the purple and expert number 6 in the turquoise with moderately confident but completely non-overlapping. So they have a difference of opinion there. Most other people are somewhere in the middle. Expert 16 ~~is~~ is fairly confident. It's in the middle there somewhere. We've got Expert 17 again, exhibiting very wide uncertainty bounds but thinking is more likely to be high than low because their best estimate is in the high range.

**M** [00:29:04] So again, say the same same question really, under what circumstances would you expect that to be very high? And under what circumstances would you expect to be very low? Reasons for that.

**E16** [00:29:17] I think, again, for me, this came down to the interpretation of the word organisation. So, you know, we wouldn't necessarily and in fact, none of the organised none of the services I've worked in have had the expertise, knowledge, capability within our service. But within our wider organisation there most definitely is. Doesn't mean that I have had use of it, but I know it's there.

**E03** [00:29:55] That's how I was, that was how I was understanding this question too.

**E02** [00:30:00] Yeah, I would agree.

**E14** [00:30:02] Yeah, I think that's what we said, even the knowledge that they could buy in a service for this one was, this is the one where we mentioned that, so saying having sufficient knowledge to understand that that option was available.

**M** [00:30:20] So ~~se~~ your discussion there would suggest it should be higher, at the higher end? Most people would have somebody in their organisation or some capability within their organisation to do these things, even if it is limited, is that is those ~~those~~ your, am I understanding that correctly? Your argument is that you should be at the higher end or high end?

**E14** [00:30:44] I think it would certainly be higher than the previous one. I think the overall state of play means it probably still wouldn't be that high overall.

**M** [00:30:57] Is that because we're still asking people to say, or?

**E14** [00:31:01] Yeah, to some extent. People, again, a sort of general lack of confidence in digital skills in the sector.

**M** [00:31:09] We're not asking this time to say about themselves, but about their organisation.

**E14** [00:31:15] Yeah, true

**E13** [00:31:18] Good morning everyone, [name] here. If you can hear me, we're happy in great difficulty with our audio. I'm not sure if I'm coming in the right place or not. I was going to explain [...] figure.

**E13** [00:31:40] That is from.

**E13** [00:31:45] Experience working extensively across archivists within the [...]. The majority had to[...] virtually no IT support whatsoever. And so I think, as I like the comment that A gave earlier about capability to access tools such as Droid. Capability is next to nil. And that's why I would pick from low, low figures.

**M** [00:32:20] Thank you, that's very helpful. Any other questions or comments about why this, the answer to this question might be high or low or?

**H** [00:32:47] So I may suggest maybe people's answers might vary depending on their organisation size. So if they are an archive, which is a smaller part of a wider organisation that's very digital or they've got a very good IT support then they may be more likely to say higher. But if it is just a very small local archive or a very small organisation that doesn't do as much digital, then maybe they'll answer lower.

**M** [00:33:14] Absolutely right.

**M** [00:33:16] Any comments type typing into the spreadsheet where you send it back about, you know, whether you think your archive is typical or whether you're answering for your particular archive or whatever? Any any comments along those lines would be very helpful. So I think I interrupted somebody then.

**E07** [00:33:31] Yeah. Sorry. And I was just going to say that I think in this question, so I would suggest it would be reasonably high because file format migration is one of the options here. File format migration can be quite easy, actually. It can be hard as well, but it can also be easy. So it could just be as simple as converting a jpeg to a tiff file. It could be as simple as converting a Word document into a pdf. And I would say that a large percentage of the population would be comfortable doing that.

**M** [00:34:11] Hmm.

**E07** [00:34:12] Whether they do that with adequate documentation and with adequate QA processes as part of a kind of a more formal file migration workflow is another question,

but some forms of file migration I think are very easy and very easily within reach of anyone working in the sector.

**M** [00:34:34] Well, given that and given we're asking even in a limited way, you think that's a reason for it to be quite high.

**E07** [00:34:40] Yeah, though I think the bar is much higher for software emulation or data recovery. I think that they're a little bit more complicated.

**M** [00:34:51] Excellent. Thank you. Any more or shall we move on? Have you made enough notes as to how you might think about adjusting or otherwise, ~~your~~ your estimates?

#### **Question 4**

**H** [00:35:07] Question 4. Out of a thousand files with file formats, that are ubiquitous and/or open, how many would you expect to have the tools to render?

**M** [00:35:25] The files with file formats that are open. So we've got a cluster largely on the right here with ~~so~~ so the general opinion is it's fairly high although we do have Expert 13 who begs to differ and is quite confident that it would be no. Again, we have Expert 4 in green, Expert 16 kind of turquoise and Expert 9 in navy blue with very narrow uncertainty bands. And we have Expert 10 in the ~~in the~~ mustard yellow and Expert 15 in the brown and Expert 8 in the lighter brown. We've got very wide uncertainty bands, so it could be low. They think it could be even lower than Expert 13 thinks it could be. So again, the same question, why would it be high? Most people obviously think it would be high. And under what circumstances would it possibly be low?

**E20** [00:36:33] The part that confused me and we did discuss this yesterday, though, is again, that question of rendering and the quality of the rendering, you know. And we I think we agree to phrase along the lines of a kind of a sufficient quality for our regular use case or some such phrase. And if you put it in those terms, then I think with confidence, my confidence be fairly high at that our top end of that graph. But if you take away those, I mean, if you really want a forensic and thorough performance of a file under complex circumstances where you need to reproduce the original environment, then confidence will be incredibly low. That difference would transform my answer to the question.

**M** [00:37:18] Right, thank you.

**E13** [00:37:25] [name] here, as far as question four is concerned my reading of it was that having the.

**D** [00:37:41] Sorry, we've lost you Expert 13.

**E13** [00:37:47] a file, ok, i will start again. I found that the notion of rendering. [muffling] or open it's almost a definition to say that a file which is ubiquitous and open has readily available tools allowing you to render them. That is what an open and ubiquitous file is.

**D** [00:38:28] Yeah, I mean, we're essentially talking about things like Word and Excel and jpeg and there are probably some open formats that are a bit harder to deal with because they're not ubiquitous but a fairly simple web search should probably find you something

then you might run into some of those technical issues again about not being able to install software, but on the whole.

**A** [00:38:56] We did define the rendering as being something you're doing now, not about something you're looking into the feature about, it's about what you can do now so it should be, again, things like Word and jpeg, [muffling] proprietary alternatives.

**E13** [00:39:20] But but even with those qualifications, it's about the [lost sound] virtually always immediately [lost sound]

**A** [00:39:44] We can't hear you [name].

**E13** [00:39:49] Oh, OK. Oh, OK. Can you hear me now? Well, coming and going. It's. If you can hear me now, I think I'm going to have to abandon using audio.

**A** [00:40:05] You can dial in.

**E14** [00:40:17] I think Expert 12 was about to say something.

**E12** [00:40:19] Yeah. Yeah. Sorry. As I think I was just thinking, from, so was thinking from kind of my experience and background to kind of umm where I'm working now, kind of from a AV point of view, was that we would we would kind of need, we kind of need to have those tools to render in order, because in terms of archiving, but also access. So yeah, it was just kind of knowing what I would kind of need and so having those tools available it's it kind of I couldn't do what I would need to do without having them.

**E14** [00:41:03] I think that's true probably in your situation. But then maybe another archive that doesn't deal with like AV as much might struggle with some of the open, some of the open AV formats. Yeah, they would necessarily have the tools immediately installed and they might run into some of those, then run into some of those technical barriers trying to trying to get them there. So I think that's where some of my uncertainty was, that kind of situation where it's an open file, but perhaps not something that would be in every collection that we're running into all the time. I can see, yeah obviously in your situation where you're largely dealing with perhaps quite a narrow band of these types of format then, you're probably more likely to have those tools.

**E12** [00:41:47] Yes. Yeah.

**E07** [00:41:51] So is this the question, sort of question that you would expect us to have a really broad range of views on because we're actually talking about us like we're not talking about or broadly speaking, out of a thousand archivists we're talking about, how many would you expect to have the tools to render? So is there a sense that you don't really need a consensus on this question?

**M** [00:42:15] Absolutely. Absolutely. We're not aiming for consensus on any question. What we're doing here is just saying the reasons why it could be low and reasons why it could be high. Now, it may be that everybody answers it in the second round, exactly the same as they answered in the first round and that's absolutely fine. It's just to clarify how people are thinking. Just to clarify how people are understanding the question. And so you can make whatever notes you feel are useful to you when you go away and do the second round and then keep whichever first estimates that you want to keep and adjust those that

you want to adjust. I've got a little note that says 'a typical archive'. So are we doing a typical archive or we're doing your specific archive. Has anybody else got a note on that?

**E07** [00:43:08] I think perhaps that note was as a result of when we were discussing these questions yesterday of me saying I don't work in archive, therefore how should I answer this question. [**M**: Right.] So I think for people who do work in an archive I assume they would be answering for themselves.

**M** [00:43:28] Yes. And then you're drawing up, I think we discussed you were drawing on experience.

**E07** [00:43:33] This experience. Yeah.

**M** [00:43:34] With previous experience. And those who haven't got recent means of experience on this or general. Yeah. The general experience of what they may see as a typical archive currently. OK.

**E14** [00:43:49] And I think as kind of just discussed that means there will be potentially quite a bit of variance between answers depending how broadly people are collecting and the types of formats they normally experience.

**M** [00:44:00] Well, this is so, so far, this is the most, the question where we have most consensus.

**E07** [00:44:07] Typical.

**M** [00:44:09] Yeah, but that's fine. I mean there's still quite a lot of variance, isn't there, really. And if it could, you know, as some of our values are down at sort of seven hundred and some of them up at sort of nine hundred out of a thousand. So there is there is some some variability but we do. Yeah. The sort of the impression that the graph gives is a cluster to the right, rather than a complete spread over everything. So does everybody feel that they can have another go at question four? So we've got from [E19]: "formats tend to become ubiquitous via the universality of renderability". Well, that's a good point.

**M** [00:44:55] Any more comments on question four or shall we move on? That's a move on. We got some thumbs up here. Thank you, [E12 and E20]. I can't see everybody.

## Question 5

**H** [00:45:10] Question 5: Out of a thousand files with file formats that's are neither ubiquitous nor open, at an archive where staff have good technical skills, how many would you expect to have the tools to render?

**M** [00:45:28] And I think this was the question where we said, you know, render good enough for the user's purpose, didn't we? So, again, what pattern are we seeing here? We've got an expert number seven in the pink on the left. Very tight uncertainty bands very sure it belongs down there. Expert 10 slightly wider bands, but fairly sure it belongs up in the nine hundreds. And Expert 15, even ~~even~~ narrower uncertainty bands again up at the top end. Some very wide uncertainty bands, maybe capturing the fact that different archives are different and maybe capturing the individual expert's own uncertainty. So Expert 19 in the pink at the top and Expert 18 in the orange, very wide uncertainty bands and Expert 2 down the bottom in the red. So again, same question we've got get 15 and

10 think definitely at the top end and we've got 7 and 6 saying it's definitely down the other end, other people saying it's somewhere in the middle with various different levels of uncertainty. So again, the question when could it be very low? When could it be very high? What all the uncertainties in this question.

**E09** [00:46:56] I think I went quite high on this one because of the focussing on the good technical skills and the sort of not needing to render absolutely perfectly, just, would have thought with good technical skills, a lot of formats you could get some ~~some~~ data for.

**E03** [00:47:17] I went with a similar sort of view on it that, you know, archivists are quite resourceful and if they've got good technical skills, then you know, they will find a way in a lot of cases. Not all at all, so I didn't go super high, but that's what I was focussing on too.

**E07** [00:47:36] Yeah, I was focussing on the fact that in my previous context, I was in a big university and a university's got a huge range of software because they're supporting all different researchers and different departments. So there was always a way of not always, but quite often there's a way of finding a piece of software that might work for a particular file format.

**E10** [00:48:01] I suppose I conversely went quite low because you might have good technical skills and you might have good knowledge and understanding of what you need, but you might not actually be able to get it. So that's in the context of the capability and the capacity of the organisation to support it. So ~~se~~ then I thought, you know, perhaps in my context that that would be hard. And so the figures were lower.

**E06** [00:48:28] Yeah, just to support that point. From an organisational perspective. It didn't really doesn't really matter how good your technical skills are if the organisation that you're working for just won't let you actually access some software or hardware.

**M** [00:48:45] Yeah. So there are perfectly good and legitimate reasons why it could be low in some cases it could be high in others. So [E15] says, I took the position, that effort would focus on a small number of formats that were most frequent or common would give the best coverage of ~~learning~~, of leaving odd files or formats for another day.

**A** [00:49:26] [E21] is agreeing is saying the same yeah if there are skills and staff available with good skills, they'll find a way, consulting others online if needed.

**E06** [00:49:35] But personally, I don't think this should be about finding a way that is outside of your organisation. I know we do and don't get me wrong, it's great that we do it for the purposes of the model, which is to demonstrate where the organisation needs to support you and give you more resources to mitigate risk, We shouldn't be thinking, I dont think we should be thinking about going outside of that organisation for the tools to render. If they haven't given them to us then for the purpose of this question. We shouldn't be going outside.

**A** [00:50:14] No, that's a really good point actually for the model to do its job to prove to help archivists make the case.

**E06** [00:50:22] Yeah. Yeah. We have to remember what we want this model to do. This isn't about, you know, saying how resourceful archivists are, this is about saying what is lacking in an organisation. Where do we need the organisation to support us in order to

mitigate risk? And if they're not giving us the tools to render, despite the fact that we can find a way around that, then that's what we should be reflecting.

**A** [00:50:49] Yeah, absolutely.

**E04** [00:50:50] There's also an aspect of IT or openness of IT support and ability to download your own test versions, because you may have to try five or six different ways to get the actual results you want. I ? or a particular IT model where you can't do it yourself, an IT member staff delivering one tool to you is maybe going to be a bit of a long process.

**E09** [00:51:20] I think as well, in terms of highlighting the risk one of the key things for the question for me was the good technical skills. If that wasn't in the question, then I would have put a lot, lot lower. So not so much.

**M** [00:51:38] So I think we've got an agreement. Correct me if I'm wrong, Alex and Hannah, that we we're thinking in terms of high technical skills, that might not be enough because you're ~~you're~~ focussing on what ~~what~~ tools you're given within the organisation.

**A** [00:51:53] Yeah, yeah, that sound a sensible approach.

**M** [00:51:57] Just about the wonderful resourcefulness of going elsewhere or getting your friend to do it or whatever.

**H** [00:52:04] Can I ask, did anyone consider the ideas of file format migration or software emulation in this question? So even if IT said or you couldn't download one tool considering technical skills is including things like that?

**E14** [00:52:21] Yeah, but you still you'd still need the support to get those things up and running. So unless you've actually got that environment already available to you, then you can't do it. So, I mean. You know, we're lucky in our context at The National Archives, we've got a separate lab network and so on that we've got admin control over that we can install pretty much what we want on, other organizations, even if they've got someone with a high technical skills if they don't have that, that, support then they're not going to be able to do that still.

**E04** [00:52:56] Yeah. I mean, with the thing about a spread of tools as well as maybe possibly one particular closed or maybe even an open version of software, you may have to have five or six tests each one against the other and see if there's any variation in the output. Yeah, I think just one selected item wouldn't be very useful at all.

**E20** [00:53:30] And this is sort of talking me down, actually, I'm moving my scores lower than where I think on this one.

**M** [00:53:41] OK. Are we we all done? So I think we we think that actually there's a lot of different experiences depending or whether you've got the the support or otherwise and it's not the key thing is not necessarily good technical skills that's not the barrier. Excellent.

## Question 6

**H** [00:53:59] So Question 6. Slight change of theme here. Out of 1000 born digital files, how many would you expect to have the content metadata that meets an archive's requirements?

**M** [00:54:18] So question 6 again, we've got some different ideas. So you've got Expert 13 at top thinks there's no uncertainty at all. It's definitely by looks of it, all of them. Expert 2 thinks it's most likely to be very, very high, not all of them. Expert 8 is on the other end, thinks it's very low. And Expert 18 in the orange at the top is also on the low side, but with wider uncertainty bands of similar size uncertainty bands on Expert 10 in the mustard on the right and Expert 7 in the pink on the right. Most of other people reasonably wide uncertainty bands somewhere in the middle. Maybe a group more in the six hundred to eight and another group of the two and four hundreds. So, again, reasons why it might be high, reasons why it might be low the reasons why there might be a lot of uncertainty.

**E04** [00:55:17] I'm looking at a scenario where this would be an archive receiving a collection from very variable institutions or bodies. So some may be very adept to putting together an excellent metadata list, whereas others may just basically dump stuff on the hard drive and send it off to you. So really, the scenario we're looking at, it could be as broad as that. And we are talking about receiving metadata from an institution rather than creation.

**M** [00:55:45] Yes. That's a good reason why there might be a lot of variability.

**E04** [00:55:50] Yeah.

**E14** [00:55:52] Yeah, and it's you know, we're looking at this from our own context at TNA, we've got fairly, we don't actually expect institut[ions], you know, we don't expect to be provided necessarily with that much content metadata, just the file name and last modified date and things like that, but sufficient. Others may be trying to kind of hold out for more as to expecting to do more actual description on files and things. So this one depends a bit, you know, if you're entering from the point of view of your own institution's expectations, there'll be quite a wide variety on this one.

**E04** [00:56:31] It could be as well the time period of the materials that you're receiving, obviously, as it goes back further in time the people involved, to the origin point has an effect on the amount of metadata that may be collected. I suspect there tends to be an improvement in the metadata as it comes towards recent times, whereas maybe with earlier collections not so well. So there's gonna be some missing items back from the day.

**M** [00:57:00] Comments in the chat saying uncertainty over interpretation of archive's requirements.

**E09** [00:57:10] I sort of focussed on one of the things we mentioned in our discussion is the accuracy, because I'm sort of looking at it from The National Archives perspective, where it's fairly high because we have quite relaxed minimum metadata standards, but I wouldn't be 100 percent sure that all of that file metadata, especially the dates, is completely accurate.

**E10** [00:57:39] I was just thinking, I was just thinking of taking this broader context and that this is not a unique challenge to digital. And I don't know whether that matters or not. I just did my experience. We try not to content digital objects different from any other archives that come into our care. So the collection of the content metadata or [...] at that point of accession is something we would do regardless of format.

**E19** [00:58:26] I was just going to say that at The National Archives, I think our bar is actually very low and we set it very low because of the difficulty in getting, you know, reasonable descriptive metadata. In an ideal world, we would get that descriptive metadata on a file level. So I would perhaps suggest that what we're collecting right now isn't adequate.

**M** [00:59:00] So that was a comment about the the the archive's requirement.

**E19** [00:59:08] Yep. So I put it over on the low end because I think for digital, because of issues like scale, it's very difficult to get that contextual descriptive information that would be adequate I can't remember the actual exact wording of it, that meets the archive's requirements yeah.

**M** [00:59:28] [00:59:28] **OK.** [0.0s] Any other any other points on this one? Or shall we move on now.

### Question 7 and 8

**H** [00:59:41] So question 7, which is identical to question 6, except it's talking about digitised files.

**M** [00:59:51] And we have again, we have the spread. We have Expert 4 in the bright green, very narrow and very low, expert 5 in the purple very narrow and very high, and 14 also in purple, slightly pinkier purple, very, very narrow and very high. And then Expert 9, it could be anywhere, pretty much. Lots of uncertainty there. Expert 17 and Expert 19 quite uncertain. Expert 15 as well. So yeah. We're not looking for consensus. We don't have much of a consensus here. We do have some people who are very sure it's high, some people who are very sure it's low and some people who are not very sure at all. So, again, reasons why it might be uncertain, reasons why it could be high. Reasons why it could be low.

**E14** [01:00:42] I think one issue here is that the idea of a digitised record versus a surrogate is perhaps a bit specific to TNA. And then the context in which we're creating those, it's very often in order to make sure we can give kind of more granular access to paper records that would otherwise be closed. So there's a lot of making sure that we know the content metadata so that then it is often kind of personnel files and things like that. So making sure we know the name and date of birth and things like that so we can make accurate closure decisions. So typically in these cases, which, don't happen and we're not creating hundreds of these digit[ised]. Well, we might have thousands of files, but we're not dealing with many series where this has actually happened. So there's quite a specific effort on these to gather the content metadata, in order, otherwise there'll be no point doing the digitisation anyway. We wouldn't gain anything from having it as a digitised record versus taking the paper.

**E20** [01:01:52] So I took out I took out, ended up taking a slightly odd view on this one and relating the digitise to the surrogates. I couldn't do a Question 7 without also thinking about Question 8 and the data appear and I ended up thinking to myself, digitise content. Presumably is a broader category than surrogate, and the surrogate content is probably something which you have digitised yourself. So there is in Question 8 imply that you have a control over the the workflow of digitisation, whereas in question 7 by implication you don't have control of the workflow of digitisation. Now, obviously, if you've got control of the digitisation of workflow, you're in a much stronger possession in terms of getting the

correct content metadata. Whereas if you're accepting material which has been digitised by others under conditions which you may not be able to control, then, you know, your ability too can guarantee or warranty that content metadata is different. In doing that, I have written or read a whole or extra into the question perhaps then that I should have done. So I thought I wanted to share that with you.

**E14** [01:02:59] Yeah, that's an interesting one, because actually if the original originator of the records has actually digitised them kind of in course of business and then passed them to us then we treat those as those born digital in terms of how they're catalogued and so on. Whereas the digitised record is a very specific instance where we would still have a lot of input into the ~~into the~~ process and control over what would be delivered to us. And so, yeah, as I say, this is perhaps a bit of difficulty around the very specific definition of a digitised record versus a surrogate.

**E20** [01:03:42] Yes, so I'm thinking about. I mean, I'm thinking of my own example, you know, my own experience of just receiving, just routinely receiving, you know, for example, expenses claims. Where people have scanned the receipts, you know, you can no control whatsoever as to how that goes. Whereas obviously, if it's been digitised in a much more controlled, managed project, big investment thing, you know, to try and clear boxed files or to move things into deep storage, well, that's a whole different proposition.

**M** [01:04:12] Would we generally agree that if you're digitising yourself and then you really ought to have control over whether or not the content of the data is good enough, whereas if you receive it, then that's a lot more variable.

**E07** [01:04:26] Yeah, I was thinking about this in very much the same way as this Expert 20. Around. I couldn't think about this question without thinking about who is doing the digitisation, therefore who had control over how it was done. But I know you know the scenarios where a depositor might bring you that thing to digitise, but then they want to keep the original.

**E07** [01:04:47] So you're still controlling that digitisation process as the archive. But they're taking away the original again. So you still, that's still at a digitised record, but yet you're doing that work on making sure that you've got what you need.

**E07** [01:05:02] It's ~~it's~~ really, um, it's really variable depending on the circumstance I think, isn't it?

**M** [01:05:08] Do you have a handle on the proportions where that might be the case, though. Is it is it mostly that they've been digitised elsewhere or on those occasionally you get to do the digitisation or is it mostly you do the digitisation and they take the originals back again and does that make a difference to the number you give.

**E07** [01:05:30] Yes, it's probably mostly that other people are doing it. But then you might have some input into that into that process regarding guidelines, and good practice that you can pass on around what metadata and file formats you require.

**M** [01:05:51] And would that vary by archive?

**E07** [01:05:55] Probably, yes, a little bit more variability. Yeah.

**E20** [01:06:01] There's also something in here, sorry to labour the point again, there's also something in here about time. So if this is material which had been digitised as part of there was a digitisation project, say, in the late 1990s. I would have a very different view of it if it was something which been digitised say in the last five years. So it. Well, I think as with all these, it depends.

**M** [01:06:27] And it's good that we have this discussion. And we do work out why "it depends". And so we can have we can have this level of uncertainty, but we're not thinking of just one situation here.

**M** [01:06:37] And yes, as well the discussion is as valuable as in a different way as the numbers themselves. So understanding why, or under what circumstances it could be high, low or whatever. You've done the digitisation yourself, even though you don't hold the original, you'd expect it to be more close to Question 8 where you still hold to the original. You would have no control over the metadata that was generated in the process. You might have an archive where you've got rules around what you will and won't accept as digitised.

**M** [01:07:10] And then you might have, you know, you might just accept anything that's given to you with the hope and the guideline that actually it might be done in a good way. So what proportion of your collections would fall into each bucket might influence your answer.

**E07** [01:07:26] That's a really good point about the time frame saying, as in the previous archive I worked in, we had loads of digital surrogates that we'd digitised ourselves 5 or 10 years ago that we couldn't make any sense of because they'd been digitised, ad hoc, in different ways by different people, for different purposes.

**E07** [01:07:48] So in the end, it actually becomes cheaper, I suppose, in the end, to just get rid of them if you've got the originals, and redigitise if you need. So yeah, it really depends on the time frame as well.

**E10** [01:08:06] Just to add that yesterday when we were talking about the difference between digitisation and surrogate, I kind of I might have got my ~~my~~ head in the wrong place because I wrote down that digitisation was the idea of replacing the originals. I was thinking of like a scan and destroy. Yeah. So if the idea was that scan and destroyed, I thought your ~~your~~ methodology would have to be spot on. Because you getting rid of the original and access to it? So my scores for this question were actually higher than my scores for surrogates, which I was thinking of kind of access and use value and there being lots of variables in it.

**E14** [01:08:46] So that's certainly how we're generally thinking about it. But I see I kind of take the point that if people have just standing course of business and you're thinking about that kind of thing where you can't necessarily specify actually what they've ~~they've~~ captured with it and it's possibly a bit lower or a bit more variable at least. But yeah, I would I would certainly see it's because it's being, the original is being destroyed and there's more likely to be control over the process.

**M** [01:09:17] So there obviously a link between 7 and 8, and when you're thinking about the numbers you want to give, then do keep those in mind. You don't have to think they are completely independent questions. You might want to have some consistency between what proportions you think would be in each in each category under each different

circumstance. So are we done with question seven? I think we've covered a lot of question 8. Anyone want to say anything more about Question seven or eight? We are seeing in question 8 the grouping tends to be more to the right, but it tends to be sort of, yes, if we're doing it ourselves, we ought to have some metadata or if we hold the original, we have the capacity. Lots of ideas going on there.

**E14** [01:10:02] I think this one is, again, definitely one, where the age of the digitisation comes in, because I know we we've got quite a lot of stuff that's either been done by licensing partners, so although they ~~have may~~, may have done some additional description on it, we don't necessarily have access to that because it's been digitised by the commercial partners. So they're using that, as that's the kind of intellectual property on their websites. So we might just have a whole bunch of images still just digitised at file/folder level. And they may have created indexing and things, but we don't have access to that or we have all the stuff that was, it was digitised in-house. We have big sets of masters, but then the actual further itemisation and stuff of that, and what's on our online access system because things were. The way ~~that was~~ the re-cataloguing was done and then re-shoots and things to replace poor images weren't incorporated, we can't, it's very hard for us to match up the itemised version that's ~~that's~~ available to the public actually, to match that against our masters that we hold is pretty difficult.

**M** [01:11:26] Anymore or are you alright to go on to Question 9?

## Question 9

**H** [01:11:30] OK, so Question 9. It was: out of a thousand UK archivists, how many would you expect to say that ~~they digi?~~ their, sorry that their, catalogue management system met the needs of their organisation.

**M** [01:11:52] I've got a note, "metadata", alongside this. So, can you remind what we were talking about?

**E07** [01:12:01] I think this was a descriptive metadata about the digital collections. What would be held in the catalogue management system? Yes.

**A** [01:12:13] Right. I think you made the distinction yesterday, E07, this was just about the information about the record, whereas question 10 is talking about systems that hold the actual files themselves.

**E06** [01:12:24] Yeah, I think that's right.

**M** [01:12:26] So, yes, I wrote metadata against this one and objects against the other one so we were making that distinction, that's right. So what we've got we got expert 19, an expert 10 think it's very low with very narrow uncertainty.

**M** [01:12:40] Expert 4 much wider uncertainty but thinks it is probably high. Pretty much everybody else in the middle, though, Eexperts 1 and 2 don't overlap, just meet in the middle. So, reasons why it might be uncertain, reasons why it might be high and reasons why it might be low?

**E20** [01:13:03] This may be one from my own perspective, which is clearly unusual here, but I've never had an archivist come to me and say that we think that our catalogue system's great and it meets all their needs. It's just never happened in my career, it could

just be that people only complain, that's also possible. But that's what leads me to that by relatively low numbers.

**E14** [01:13:29] Yeah, I certainly didn't think this would be towards the high end because you don't have to look around, you know, Twitter or Archives-NRA very long to see that most people aren't particularly happy with the catalogue systems.

**E12** [01:13:45] I went the opposite. Again, based on my experience that I've been quite happy to have. And yeah. So I again, just based it on my experience.

**E20** [01:13:57] [Expert 12], I can't tell you how happy that makes me.

**E12** [01:14:02] I'll tweet it after this.

**E07** [01:14:07] Yeah, I would say the people that complain about their catalogue systems are more vocal perhaps, and the people bigging up their catalogue systems in my in my last job we implemented a new cataloguing system and we were really happy with it. We did talk about that a lot. So sorry you didn't get that memo, [Expert 20]. I should have shouted louder, but I still think that even in that context, there were lots of people within the organisation that would still complain about it, because no system is perfect, you know. But cataloguing system is so, that your requirements are so great that you're not going to find a system that's going to do everything you needed to do absolutely perfectly, in an incredibly user-friendly way, it's just too big an ask. So everyone's always gonna have some ~~some~~ gripes and grumbles, even if it even if overall they're pretty happy.

**E07** [01:15:01] So, yeah, I think I went somewhere in the middle on this one.

**E03** [01:15:08] I was trying to distinguish the sort of the overall aims and objectives of the cataloguing system as being good enough against the sort of functionality that individuals doing specific tasks within the system might find not to their satisfaction. So, I felt there was that two pronged element to it, that even though people might complain, they might complain about the system being slow or not quite being user friendly. But that doesn't mean that the overall function of the cataloguing system in identifying archives within the repository isn't met.

**E20** [01:15:53] That's entirely right. And it occurs to me that if my grim reconnaissance turned out to be correct, then no one would ever pass archive service accreditation, which is, you know, not the truth. So I think good enough is probably enough.

**M** [01:16:09] And wouldn't it also depend on the sort of range of your archive? So if you've got ~~some very~~, if your archive for a very specific thing, you can you can pick your management system, which works very well for that.

**M** [01:16:23] That wouldn't be great for somebody doing something else. So if you're handling files that are a little bit out of your usual range, then the system becomes more clunky to use. Is that something that might happen?

**E06** [01:16:36] It's more about the size of collections, the size of collections has a huge impact on how adequate you think your cataloguing system is. I mean, as [Expert 20] said, no, no archive or very few archivists are going to be 100 percent happy with ~~with~~ the catalogue system because at the end of the day, every user ultimately wants a Google type experience and they want to know the exact contents of every single page of a file, be

that physical or digital, that will pinpoint the exact word that they're looking for. And, you know, we ~~we~~ don't have that. That's ~~that's~~ not how our catalogue systems.

**E14** [01:17:21] Yeah, but we're talking about archivists here. It's the archivist.

**E06** [01:17:25] Yeah. So I think archivists such are generally always going to be a little bit kind of reserved in in coming down and saying, oh yeah, it's great because we can always see how there could be improvements. But that perhaps ultimately, if pushed, if if you said to them. Yeah. But you know does it. Does it meet the needs of the organisation in that generally you can respond to an enquiry and ~~a~~ or a user can identify something that they are interested in, most would probably say yes. So that's why I think it would be really middle ground, perhaps a little bit higher than middle ground, but you are never going to get it overwhelmingly near the top or overwhelmingly near the bottom, really, I think we're quite pragmatic about it, ultimately. We just like to grumble about it.

**M** [01:18:25] [Expert 11] has put in the chat, "but it's also a fact that many know there isn't funding to improve. So they make the best of what they have. And being vocal is the mechanism to encourage the players to make changes."

**E14** [01:18:43] Yeah. That's probably fair.

**M** [01:18:47] Well, then all done with question nine.

### Question 10

**H** [01:18:53] Ok, question 10, its very similar, very similar. So out of a thousand UK archivists, how many would you expect to say that their digital asset management system met the needs of their organisation? And we sent a definition yesterday to clarify. But I think we should talk it through as well.

**E14** [01:19:14] Yeah, I think broadly this is quite similar to the, to the stuff we've been discussing about the previous one, but my feeling is it would generally be lower than the previous one, people would be even more unhappy with where they are with this side of things than the previous question.

**M** [01:19:31] That's what we're seeing. Generally there's a shift to the left compared to Question 9.

**M** [01:19:47] Any further comments?

**E07** [01:19:49] I guess I was thinking on this one that there's so many people that would say, well, I don't have a digital asset management system, so how would they score on this one? I need to check back on the definition that you circulated [H]. It's to see how to get to this one next time.

**E14** [01:20:12] I think that's what you know, if you haven't got anything, you're not going to say that you, you've got no system to meet the needs. So it's definitely going to be it's gonna be low.

**A** [01:20:21] We did define a DAM as as processes as well as a system. So if you've got robust workflows to manage and record all the information you need.

**E14** [01:20:35] Again it depends if you've got already small collection just having a network share and a spreadsheet that tells you where things are, might you might feel is sufficient for you now. But you kind of, on the other hand, you know that's not going to last you forever, you still might not feel it's really enough.

**E07** [01:20:51] Yeah, I've just checked the definition now and that makes sense.

**M** [01:21:00] Good. So I'm conscious that we got, is it 10:40 for our break? And it's now 10:48 now.

**A** [01:21:04] Yes. Shall we have a break?

**A** [01:21:09] And we said, we said we'd be back? I'll just check the timings. We said be back at 11, 11.

**A (plus interjections from H)** [01:21:18] Yep, yep, yep. I said 11:05, so between 11 and 5 past. Well we'll come back. If that's all right with everyone.

**Various** [01:21:27] Yeah. Thank you, ok. See you then. Bye

[break]

**H** [01:21:54] OK. So, question 11. Which is, let's get it up, oh I've temporarily lost it.

**A** [01:22:13] I've got it, if you want.

**H** [01:22:15] Can you read this one, please, Alex, I'll get it up in a minute.

## **Question 11**

**A** [01:22:16] Out of 1000 files with insufficient content metadata, at an archive where there is sufficient information management. How many files would you expect to be able to identify. And by identify we mean knowing what they are and where they are from.

**M** [01:22:39] So what do we see on the range graph, we see Expert 3 pretty much things that 0 with almost no uncertainty and Expert 4 things is pretty much all of them with no uncertainty. Expert 14 is again, along with Expert 3, very, very hard on the left and ditto Expert 16. We've got slightly less certain, but definitely at the low end, we've got Expert 6 and Expert 1. On the other hand, we have up with Expert 4, we have a lot of people plus tending to cluster to the right with Expert 8 saying there's a lot of uncertainty, Expert 15 and Expert 18 saying there's a lot of uncertainty. Expert 18 tending to think it's likely to be higher, but it could possibly be low. Same with Expert 8 there could be reasons why it could be low. So again, reasons why it could be low. Reasons why it could be higher. Reasons why it could be very uncertain?

**E13** [01:23:49] Can I bleep in to here, it's [name] here, to talk about my high there. And, and, that is because I was strongly influenced by the phrase where there is sufficient information management, because if there is sufficient information management, then you know what they are and where they're from.

**M** [01:24:12] Yeah.

**E09** [01:24:18] I was the same, I think. I was thinking on TNA experience where we might not have detailed content information, but we would generally know where material come from and have some understanding.

**A** [01:24:33] Would anybody be willing to say why they gave a low I gave low answer to that and presumably with a different interpretation of the **question?**

**E20** [01:24:40] Yeah, I'd be happy to do that. I read that differently. So I read sufficient information management with emphasis on the management. So there is a capability with the information management rather than the existence of the information. If that makes sense. So that I say pulled me down. Just a slight, slight variation of the wording that I think could make quite a difference.

**A** [01:25:04] Ok.

**E10** [01:25:05] I was wondering whether the thousand files came from the same source or different sources because if there was. If there was if it came from the same source and you didn't have sufficient information management to know what they were or where they've come from, we wouldn't be able to identify any of them. Whereas if it was different, a spread of different sources, that you might have better information management for some than for others.

**E14** [01:25:31] Yes. Interesting point. Yeah. Oh, yeah.

**M** [01:25:42] Alex, Hannah, do you want to clarify the question in any way, we, I think I mentioned, we're thinking about a thousand files selected at random or coming from the same place probably.

**H** [01:25:54] Yes. So I think we're thinking a thousand random files and the definition of information management, which you should have on your sheets, was internal systems and support for coherent information management and documentation of preservation actions. So I think we were thinking more about having the systems in place. [...] this question's about, you might have really good information management, but you didn't have, poor content metadata. Maybe we need to, yeah, interpretation of whether that means because we have good information management, we have sourced information or we can still identify it.

**E14** [01:26:35] Yeah, you might not be able to say much about individual files, but you've probably got information as to where the accession came from. Although, again, there's probably a bit of a, I think I had a spread, and was a bit lower because, you know, it depends how old some of this is. So you might not, you know, if it's a disk that was found in the middle of a paper file later on you might not have the kind of full traceability.

**H** [01:27:06] So [E20] saying about you could have very good information management process and way of organising things, but you just keep poor information because we have insufficient content metadata.

**E20** [01:27:18] Yeah. Or it may be completely wrong.

**H** [01:27:25] I think that's a fair interpretation, but maybe people do think that. Interpret information management differently.

**M** [01:27:38] [Expert 22]'s made a good point as well. "All of those reasons are why my spectrum was wide: the high, because the archivists are great problem solvers, but low because you may know where the accession came from, but not what the individual files are," is that fair?

**A** [01:28:01] Yeah. Surprisingly hard, isn't that?

**E20** [01:28:11] I mean, I certainly know of examples in my own previous work in ADS where we had databases with, which we could process and look after, and we had very good processes for looking after that material. But we didn't have to look up codes for the databases and it was for all intents and purposes, useless. So very good information management, really rubbish information.

**A** [01:28:40] And that is, that it is, common. You know, you're getting, you're getting legacy data. I've just done a project where I've tried to reconstitute a database from the information we have about it. And yeah, you can collect as much information as you like. But if you don't have the information to start with it isn't hugely helpful.

**E07** [01:28:59] But the word, "identify" in this question is this, we're not talking about file identification here, are we?

**A** [01:29:09] No, we're not, we're defining as knowing what the material is and where it's from. You know, can you locate the file? Is it sufficiently described for you to know that this is what you want? Can you understand its context within the archive. Can you find other versions of it? Can you find the provenance, that kind of information? So it's less about technical characteristics than the kind of descriptive access.

**E14** [01:29:40] Yes, guess some of that is actually related more to the content metadata, the context and the sufficient description. So, yeah, I think that does go towards [Expert 20]'s point, actually, that you might have good information management processes, but if you just don't have the information to start with, then tough you might be able to to reconstitute some of that as effectively, as [Expert 22] has said, by, you know, doing a lot of hunting around and what you were just mentioning as well, Alex. But it is going to take a lot of effort and.

**E13** [01:30:19] Hi. Can I suggest that there's something of a paradox emerging where if we contrast the generally high response regarding the satisfaction of archivists with information management, with their cataloguing systems in Question 9, that that if, if the answer to all of that is yes, it's great, then why would you go on the low side in question 11?

**E14** [01:31:03] We know lots of archives have backlogs, they might have a catalogue system that's perfectly capable of dealing with the content metadata, but for whatever reason they haven't actually been able to do the description.

**E13** [01:31:16] It's a very broad definition of meeting the needs of the organisation. Thank you.

**A** [01:31:30] Sorry what was that? No, thought somebody said something.

**M** [01:31:37] Anymore comments or questions, has everybody had the chance to make the notes they want to make on Question 11. Yep.

## Question 12

**H** [01:31:45] Good. So question twelve so slightly changing a bit. So out of a thousand files as an archive where staff have good technical skills, how many files would you expect to have sufficient technical metadata?

**M** [01:32:06] So we're clustering at the high end here. By the by the looks of, Expert 19 has popped it right in the middle with no uncertainty. Expert 7 begs to differ. And Expert 18 strongly begs to differ. Expert 2 thinks it's very high and there's almost no uncertainty, no uncertainty on that. So, again, we may have less to discuss here, but all the reasons why it could be very low are the reasons why the uncertainty could be very high over?

**E13** [01:32:40] Hi, [Name] here, can I come in on, on, the reason that it might be very high on this turns on our conversation yesterday regarding technical metadata and where, where, for example, one is not doing file migration or anything like that. Then basically the files are static digital objects and the technical metadata is just what the technical metadata is. So this stuff just sits there.

**A** [01:33:22] OK assuming, assuming, one that the staff had not done a great deal to the files.

**E13** [01:33:32] As regards technical metadata. If one is, is, leaving the file in place, preserving the, preserving the original bit stream, then the there is no impact at all on technical metadata.

**E06** [01:33:52] Is, sorry, is this not the question where we decided that it wasn't actually about the technical metadata existing because as [Expert 13] says, it just exists. It's more about the technical metadata having been extracted, somehow and put in a different format.

**E14** [01:34:13] Yes, I think things about having it usable to, to, an extent. So it's, it's, there implicitly. But unless you have, I mean, here we are talking about a situation with good technical skills, though. So we would certainly, we would, we would expect that these perhaps all people who would know what they're doing and could could get what they needed.

**E06** [01:34:30] Yeah. So that this is actually. How many files would you expect to have sufficient technical metadata extracted from the file. Or. Yeah, I. Yeah. No. Logged. Recorded. Logged. Yeah.

**E14** [01:34:45] Or even as they would have the skills if they needed to answer a particular question. They've got good technical skills so they would not know how to go about dealing with that. Yeah.

**E06** [01:34:55] Which is why I said it could be as high as a thousand, but I actually came down a little bit lower. No more from a time point of view than anything.

**E13** [01:35:06] So, so, does this mean that when we're saying about how many files would you expect to have sufficient technical metadata? We're actually not asking a question about the files. We're asking about, how, how much of that technical metadata could be extracted and exploited?

**E14** [01:35:31] To an extent, I think, yeah. Because, as you say, it's always there implicitly, but that's not actually very, it's not a very useful thing to say necessarily, unless you can access that, then it's, it's not sufficient to use what you do with it if you try to answer a specific question or.

**E13** [01:35:57] Because well, I would say that if we're asking about how many. How many archivists or how many archives do we think could extract technical data from a thousand files. That's an entirely different question. And it would need to be re restated or reinterpreted.

**E07** [01:36:27] And I think it's not just about extracting, is it? So the definition of technical metadata is not just, it doesn't just constitute running DROID or JHOVE or another tool and extracting something. It might involve other things that might involve a bit more human intervention. I'm speaking to the depositor and establishing what operating system they were using and things like that.

**E13** [01:37:00] Yes. So. So the point is we're not talking about files here, because that's, that's the bit I find misleading.

**E14** [01:37:10] Well, you still it's still on a per part per file basis. You know, is it about to file or not. Some of it. Yes. Is it implicitly there as part of the file. But unless you've kind of actualised that in some way, it's not actually. Well, that's not helpful to say it's there implicitly, you can't actually make any use of that.

**E13** [01:37:34] OK. I'm satisfied that the conversation is it is prompting me to completely reinterpret the question. That's ok, that's fine.

**A** [01:37:46] That's it. That's exactly what this discussion is about. The answer that we can refine for the next round.

**E14** [01:37:55] Right. OK.

**M** [01:37:57] So, [E11], and [E09] and [E06], have been making some comments about tools and time.

**A** [01:38:04] Mm hmm. Yeah. U. Yeah. You might be able to install some relevant tools, but then you might not be able to because of your IT situation.

**A** [01:38:16] And it takes time. Yes.

**M** [01:38:20] OK. We all got sufficient notes on that question. We can move on.

### **Question 13**

**H** [01:38:25] So question. 13 is very, very similar. It's identical. Apart from this time as an archive which has staff have poor technical skills.

**M** [01:38:38] So any comments?

**M** [01:38:41] How much difference does that make?

**M** [01:38:45] We've seen a general shift towards the left in the in the ranges. Again, more experts still think it's very high with no uncertainty about the general shift is slightly to the left.

**E13** [01:39:00] I put my hand up on that one. I used the same interpretation as previously, so I can now, adjust that.

**M** [01:39:13] So, no problem. About the differences poor skills makes. OK.

**M** [01:39:22] OK, let's move on.

#### **Question 14**

**H** [01:39:28] Question 14 out of a thousand hard drive disks kept in a monitored commercial environment. How many drives would you expect to fail in any 12 month period?

**M** [01:39:45] So generally we've got a low estimate. People are thinking that that's not very common thing, though. Experts 11 and 8 and 2 think that actually there's quite a lot of variability in that. So generally speaking, the opinion most wasn't that much variability. But a few people thought there was more variability than others. So you're right. When isn't it?

**E20** [01:40:17] Yeah, for for me, this depends very much on things like the manufacturer and the particular technology which is in use are very significant differences. And we see significant failures. For example, for most recently is it Seagate is producing shingle-based drives which are feeling all over the place. We're looking at failure rates up to 20 to 30 percent in that case, within a year. But you know, that's not the mainstream, thank goodness.

**E21?** [01:40:49] Mine was quite wide here, but that was based on lack of specific knowledge. So.

**A** [01:41:00] And everybody remembered, I'm sure that we were talking about not a commercial environment, as in kind of an Iron Mountain storage facility or Google AWS but in an archival environment and fail as in just stops working.

**E14** [01:41:19] I think looking at [Expert 21's] point, yeah, I mean here we're talking about just a range of discs. So, you know, they they could be quite a wide range of ages and and so on. And then we go on to the next question, why we're specifically asking about basically pretty much, you know, as soon as you turn it on, then do you start to see more failures oh, yeah, a different number of failures in that instance. Once you once once you've got it, perhaps it's like the old thing about life expectancy is actually a lot. The thing that changes your overall life expectancy over time was just the huge kind of childhood mortality, people could still live to 70, 80. The bias of all that the that the deaths that a very young age brought, brought, the overall life expectancy down.

**E20** [01:42:06] I mean, there's a subtle question here, which is actually, I think, relevant for archivists. Maybe I'll just pose it, without necessarily deflecting the conversation. So disks don't appear to fail because oftentimes they have self-healing capabilities. So there's some repair mechanisms on the disk which is identifying broken components, clusters, whatever and responding accordingly. But of course, in a fundamental sense that's sort of copy and repair mechanism, it's not something you would do with a paper collection. I really wouldn't

just think oh well actually about fine we'll just have we'll just repair that straightaway. So in a sense the disk has failed, but it has failed and it's repaired itself, and we just move on, you know, and I'm not sure we're very good at getting reports back from how systems repair themselves. But I mean, that's a deviation.

**E20** [01:42:59] Yeah, I think here we're talking about, you know, the drive has completely died and has had, the actual drive has had to be replaced rather than those kind of some sort of read error or whatever, that's, that's been recovered from, or parity error.

**M** [01:43:14] So there are good reasons why there could be a lot of uncertainty here in terms of different types of different failure rates. You don't know at this particular point how old any of these are. So some of these might be at the end of their natural life like a whole you've got a thousand of these distributed over the usual replacement interval.

**E13** [01:43:37] Hi. Can I follow up on E20's comment? I was confused by both questions 14 and 15, and the idea of disk drive failure because it all disk drives are starting to fail from the point of manufacture.

**E14** [01:44:03] But they don't fail, to lose, or they don't lose the data, particularly in a monitored commercial environment

**E14** [01:44:15] The hard drives fail and I swapped out and everybody carries on in complete satisfaction that the data is safe. So what's the, what's the, thinking behind asking the question?

**H** [01:44:37] I think with thinking about the durability of that medium, so you might not lose the data altogether because you might have other copies. They'll be replaceable. But this will, sort of that's all you had and you didn't do any replacing, but you monitored the commercial environment was more about the fact that you were looking for when things fail and maybe you've got more stable conditions than it being on a shelf.

**E13** [01:45:02] But the, the, question specifically refers to monitored commercial environments where of course, in modern storage systems, the data is spread widely and single disk drive failure would never result in any data loss.

**E13** [01:45:30] I can't really see what we're getting at here.

**E13** [01:45:35] A domestic environment is is completely different.

**E14** [01:45:40] I think that's why it does specify commercial. So, you know, we're talking about a bit more, not not, necessarily a full on data centre, but we're talking about disks that are in in servers in that kind of thing.

**E14** [01:45:56] No, just I. We're not talking. We're not talking about data loss, necessarily we're talking about drive failing.

**E13** [01:46:04] The additional point here is that monitored commercial environments would be using enterprise certified drives, which are a completely different kettle of fish to a domestic drive.

**E14** [01:46:18] That's not necessarily true. Actually, some of the data centres have decided that actually the extra cost to the Enterprise Drive is not worth it.

**E13** [01:46:31] Indeed. So they just rush around replacing cheap and cheerful drives. Yeah. But but the the whole environment is set up to not lose data.

**E14** [01:46:44] Yeah. But again, we're not talking about data here. We're talking about actual rates of the hardware failing.

**E13** [01:46:51] Okay.

**H** [01:46:58] OK.

**M** [01:47:03] They're just, a couple of people said, archivists may not have a good handle on this, but do your best. That's what we're asking people. We're asking you for your best opinion. And hopefully the discussion will have helped a little bit. Should we move on?

### Question 15

**H** [01:47:15] Yes. Question 15 is a very, very similar. But now we're talking about hard drive disks failing within their first twelve months of use. So young ones rather than a mix of ages. How you might expect that to be different?

**M** [01:47:31] Generally we people think it's lower. Although there are some people who think in the first 12 months the uncertainty is much higher. And I think this probably goes to the the idea that, you know, if something breaks, it either breaks after 10 years or it breaks in the first six weeks because it wasn't manufactured properly, whatever. Any comments about that? Any questions about that? Where is, where is, the uncertainty here. Experts 18, 12 and 2, thought it was a fair bit of uncertainty.

**M** [01:48:24] So we also have Expert 2 who thinks that it couldn't be plausibly less than 100, but we'll find out to four thousand. So 10 percent would definitely fail. Pretty much definitely fail. You would be very much a winner if we got less than 10 percent for you in the first twelve months. Is that reasonable or is or are the reasons why that might be the case?

**M** [01:48:59] nope, nobody wants to discuss it. Nobody has any comments or questions. We can move on. OK.

**H** [01:49:04] And so I wonder if there were actual figures on this and this could inform us more than our guesses? So I guess a reminder that in all of these questions there will be some which are seed variables or calibration questions as well. Well, we want your just your honest answers and your impression on your knowledge on these.

**E14** [01:49:40] So in these, these, we're now on essentially two questions that are basically the same as the last two, we're just talking about a different storage technology. A lot of a lot of what has been said would still apply to this. It's just the kind of inherent reliability of each storage type.

**E20** [01:50:00] And the nature of the failures isn't it.

**E14** [01:50:03] Potentially.

**H** [01:50:04] Good question 16 talking about NAND solid state drives and fail in any twelve month period.

**E20** [01:50:12] I mean, for me that the risk here isn't the likelihood it's the kind of impact and sense that we might experience is that they typically have a lesser, less likelihood to fail. But when they do fail, they fail spectacularly. If you see what I mean, rather than simply losing a few sectors or bits, I.

**M** [01:50:47] So, generally people think it's less likely than the, than the the hard drive disks.

**M** [01:50:55] Is that fair? Expert 1, 6 and 10 have got an opinion which says there is more variety has more uncertainty than others. The reason why there might be more uncertainty here?

**E14** [01:51:14] I guess it's still a newer technology than harddrives overall. We don't have as long a track record, although as was mentioned on the previous one, there are some slightly newer technology changes in technology coming in in harddrives as well that potentially affects the variability as well as perhaps those become more common than.

**E21** [01:51:37] Think again, mine was a fairly cautious estimate because of inexpert knowledge.

**E19** [01:51:49] Sorry. Just coming into this late. I had to be on a different call. Just an observation on flash media, if it's not been made already, that unlike harddrives flash, it a very, very definitive rewrite cycle, at which point they will fail regardless, which can be a particular problem when you've got say an array of flash drives that there's a risk that they might all fail at once, which is really, really problematic.

**E14** [01:52:19] Or the recent, there was a recent issue where it was actually the firmware that was an issue you that drives automatically failed after reaching, reaching, a certain number of hours of use, which if you'd switched all your drives on together was gonna be really problematic. Slightly aside from the main point to this question.

**M** [01:52:41] Are we all happy with that? OK, let's move on. Are you there Hannah?.

## **Question 17**

**H** [01:53:02] Sorry I was talking on mute. Again, a thousand NAND solid state drives how many would experience a persistent read error within a twelve month period.

**M** [01:53:20] The same kind of drives, different kind of error. And again, generally the view is that it's fairly small. Expert 12, a lot of uncertainty with that's their uncertainty or they think that that drives themselves are uncertain. I'm not sure. 17, 16 and 11 also got a decent amount of uncertainty that could be going up to 400 or 500. 17 and 16 don't think it's likely to be less than about a hundred. Any thoughts where the uncertainty comes from, that hasn't been already mentioned.

**H** [01:54:15] OK, maybe. Question 18.

**M** Yeah, lovely.

## Question 18

**H** [01:54:21] So out of a thousand media of type A, which is a less stable. How many would you expect to reach their end of their life within 12 months? So talking about fails of a type A, which is a broader category of less stable media.

**M** [01:54:45] So what do we see, we see a lot of people thinking generally clustering to the left. You think it's going to be low, but then there's a significant portion, maybe about a third of the experts think that actually it's going to be bigger than that. I'm having a great deal of uncertainty. So where is the uncertainty coming from? Is it the system? Is it. The experts? Under what circumstances could it be on the low end? When could it be at the higher end? When could it be as high? Got three people as high as 800 plausibly. When could it be that high?

**E14** [01:55:15] I think, you know, we are now talking more the case of stuff that maybe has just been sitting on a repository shelf. Now you want to try and access it and kind of oh, actually. In a way we're really pushing the limits on this storage device now, as opposed to the previous, previously we were talking about that commercial environment where there is monitoring and so on. Now we're just going to pick this thing up that we've had for goodness knows how long. What what's the likelihood it's actually towards the end of its life?

**M** [01:55:49] Would, would that make it higher? Would that make it more uncertain?

**E14** [01:55:54] Probably a bit higher. And and a bit more uncertain.

**M** [01:55:58] Thank you.

**E7** [01:56:01] Do we, do we, mean with this question? Twelve months, so. So if one of the examples of less stable media is a floppy disk.

**E7** [01:56:09] So 12months from the time that floppy disk was ripped written. Or twelve months from now.

**E7** [01:56:19] Because obviously, the floppy disk. Well, people don't write floppy disks anymore. So all the all the floppy disks that we have like have already been sitting for for quite a while.

**E7** [01:56:32] So how are we thinking about this question?

**M** [01:56:36] So we're thinking about the next twelve months.

**M** [01:56:38] Alex, Hannah.

**H** Yes, I think we're talking about now if, if, we've got this media, I'm going to mix of ages would it fail now.

**E20** [01:56:49] I think that's a very good question. And the clarification there, Hannah retained me to to increase my estimates. You know, I think a larger number than maybe I put in my first effort.

**E7** [01:57:04] Yeah, I think so, too.

**E20** [01:57:06] I'm looking at my own answers and also the graphs thinking actually, this should be a lot higher. Just intuitively. Well, I'm not entirely sure what I was thinking.

**M** [01:57:16] [laughter] That's the point of the discussion right there.

**E7** [01:57:19] But saying that so, you know, I've dealt with five and a quarter inch floppy disks from the 80s. And it's, you know, say out of 20 of them, 19 of them, would still be working. However many years on? 30, 40 years on. So you're kind of in some of these things are actually quite robust in, you know, they survive against the odds.

**E7** [01:57:47] I mean, we know we know that they're vulnerable and we know that ultimately we want to get this stuff off portable media and put it in a safe storage system where we can manage it and monitor it.

**E7** [01:57:57] But it does surprise me, actually, how you know, how these things do continue to tend to last.

**E20** [01:58:04] That's right.

**E14** [01:58:05] So long as no one's stuck it on a fridge door with a magnet

**E7** All depends on how it's been treated. Yeah.

**E20** [01:58:12] Also, the phrase end of life is an interesting one in this context, isn't it? Because of course, this reached the end of his life. It's more than exceeded the end of its life. It doesn't mean to say it's not still usable. If you see what I mean, it's like food that's been on the shelf at the supermarket too long. You know, its passed its best by date. But, you know,.

**E14** [01:58:33] I think we because we if we had got a rule at the end, we've got definitions.

**E14** [01:58:40] So storage life the length of time for which the physical storage device is expected to store the digital objects bit stream, but then the end of storage life can be described as the point at which you can no longer store or retrieve data due to the hardware, due to hardware defects or malfunction or say there's there is a little bit of perhaps contradiction almost in that there's the expected time versus when it really does become completely dead as a dodo.

**M** [01:59:14] OK, we all done on question 18.

### Question 19

**H** [01:59:20] Question 19 is very similar, but now we're talking about media of type B, which we're saying is more stable thinking perhaps more about LTO tapes or CD ROM disks. How many of those you'd expect to reach the end of their life within 12 months. So again a mix [audio lost].

**M** [01:59:43] And predictably we're seeing a clustering to the left more strongly than with err, than with the others. Still, some people think it couldn't be lower than 400 and it might be as many as 800. So under so under what circumstances would it be very surprising to be as low as 400 out of a, so 40 percent?

**M** [02:00:33] No. No. No questions, no comments, no discussion, are we all done with that?

**H** [02:00:41] OK. Question 20.

## **Question 20**

**H** [02:00:46] So this time there is a flood at your storage location and the archive has inadequate mitigations to protect against this flood. Out of a thousand media of type A (so the less stable) how many would you expect to be destroyed?

**M** [02:01:05] So we've got Expert 16 thinking it could be anything from 0 to, to, a thousand, to all of them, with a median of about 400, lots of people with very wide uncertainty bands here. But Expert 18 thinks, it'll be a few with very narrow and certainty bands. And it's what, 19 things would be very high with no uncertainty bands. So fair to say some disagreement here. A cluster to the right, a cluster to left. So under what circumstances could we have a high number and under what circumstances could we have a low number? And where is the uncertainty coming from?

**E20** [02:01:43] So can I ask a quick question here, which will help clarify for me again? Apologies if we covered this already, but in this sense, what did we mean by destroyed? Okay, I'm assuming that means that the data can't be read or at least the process to dry or to recover the material would be so lengthy and so expensive as to be practically unaffordable. But that's my reading of that. But of course, I still think the drive will survive. You know it won't want dissolve in the water. But it's just, you know, the informational content is beyond use.

**M** [02:02:19] Is that fair?

**E14** [02:02:21] Yeah.

**H** [02:02:23] Yeah, I think it's. I view two different words to the previous two because it's not. I wanted to make clear that it's died because of the flood. And damage from the flood. Not because it's come to its natural end of life, but I think it's the same idea of, it's gone. It's not usable.

**E20** [02:02:43] Exactly. So it's gone beyond use or it's economically impossible, (D a write-off), actively to afford to recover it. I mean, not to have quite high for me because the costs involved in that sort of recovery, especially if it's like magnetic media potentially, quite, quite high.

[02:03:02] Hey, you know.

**E13** [02:03:04] Can I leap in here something which may or may not be germane, but having experienced some of this, it all depends. For example. What the contamination in the water is. Is it salt water? And exactly what the media is this that's we're, that we're talking about.

**E14** [02:03:30] I think that there is there is quite a lot of uncertainty because actually things like flash drives can be surprisingly resistant to water, you know, people or compact flash cards and things. You know, people have dropped cameras in the water. They've been

pulled out later, even after having gone through a seal I think in one instance, some people have managed to, you know, just pop the card into a card reader and, and, get the photos off it and reunite the photos with the person who lost the camera. So although the media might be somewhat unstable, it can be surprising. Some of them could be surprisingly resistant to water.

**E13** [02:04:02] Indeed, appropriate conservation techniques can work wonders.

**E13** [02:04:07] But the risk could be could be total.

**E14** [02:04:13] Yes, I think there's this potentially quite there is potentially quite a wide range on this because there are quite a few uncertainties as to the exact mixture of media.

**E14** [02:04:22] And as you say, you know, again, I suppose that, you know, an optical drive optical CD Rom or something might be a fine if it's just fairly clean water, if it's been you know, there's actually lots of silt in there, and so it will get quite abraded and things as well.

**E7** [02:04:39] Yeah, I was thinking. So I was trying to picture this flood in my mind and I was trying to work out whether the like, the floppy disks were actually swimming in the water or whether they were were like, you know, in a box, in a box higher up and whether they'd just been dripped on. So it really depended on what that scenario looked like in my mind's eye. And also having been at an archive where we had a flood in a strong room. It was in a media, like a, audio visual media store, not digital as such, but that the damage depends on where the items are and if the strong room is full, you know, to be some items that more damaged than this and there are some directly under a drip or a leak or nearer the floor and some are higher up. And so I found it really hard to answer this one. And I put a really broad range for all of those reasons because without being able to picture what this looks like in my mind's eye, I couldn't really then consider what, what, damage would have been done.

**?** [02:05:44] All right. Yeah, I did. I did the same as Jen, in terms of different formats and also locations within, within, the storage location. And that kind of factored in a bit more uncertainty as well.

**M** [02:06:00] So we got two different sources of uncertainty, have they or have they not been deluged, and they have been deluged have they survive if they haven't. What's the nature of the flood effect and have they survived that, that's very good.

**E20** [02:06:16] So I was also thinking about the potential of a catastrophic, catastrophic electrical failure that results from a flood and live store. I mean, a repository. So, you know, you can't you shouldn't just switch the repository on and off. You need to shut it down carefully so that there are no processes, reading and writing files at the time. So if you got a bunch of hard desks all up and spinning or in some sort of managed storage environment, if there's a sudden electrical failure and that all goes suddenly, it all goes pear shaped well the likelihood is you're going to crash the disks and that would be potentially catastrophic. And that's happened, again, a case in Canada recently where that's occurred.

**E20** [02:06:59] I'm not helping.

**M** [02:07:01] It's important to identify all the potential sources of uncertainty so that experts can then say, you know, I think the uncertainty. I mean, maybe maybe you might be sitting there thinking, I think uncertainty to be wider. Or somebody might be thinking, well, actually, I think, yes, I've given too much weight to that. That's the whole purpose of the discussion. So that's very helpful. Thank you. Any other sources of uncertainty with the flood?

**E20** [02:07:25] Can I. Yeah, can I comment on that? One more, sorry to be on at on this one. But one other thing that does occur to me is that collectively we're very poor at providing recovery advice. There's not a body of practice like archivists don't have typically a kind of first aid for digital on hand for when there's a fire or a flood. That's a real gap I think. Not something that'll be need very often, but when we do need it, we really need it. So it's almost a separate comment. I think we should be sitting down and writing a guide for archivists to recover from floods, fire, dust, whatever.

**M** [02:08:03] Bit like, we don't want to wait until we've got a pandemic, to decide what to do with a pandemic.

**E20** [02:08:07] One might possibly say so. Yes. Martine

**M** [02:08:11] Okay. So, yeah, that's it. That's a lot of uncertainty in question 20. That's excellent. And so that sounds that feels to me like that's genuine uncertainty in the system rather than experts not being certain themselves. That's great. And so shall we move on? Is there more questions?

**H** [02:08:31] So Question 21, again, was very similar. But now we're talking.

## Question 21

**E14** [02:08:35] There is still I think E18 asked us a bit about this yesterday and we sort of said, well, no, don't worry too much about the costings, whereas here we've said a bit more.

**E14** [02:08:44] Actually, some things will be so, you know, it would be physically possible given enough of time and resource to recover stuff. But practically speaking, we're not going to bother. So I think that, you know, there is probably again, there's some uncertainty within that. Well, we. Yes. Something's given absolutely unlimited resource you would you would be able to recover. But the practical, things are such that actually, yeah, even even just a kind of one time access to pull it off. You might be able for some things, but for other things it's just going to be out of reach. So I think I'm slightly changing my mind from what we told her yesterday. Sorry, E18.

**M** [02:09:28] Would that vary between archives. Would some archives have more, err, more budget, but that kind of thing and some have less.

**E14** [02:09:37] Yeah, almost certainly.

**M** [02:09:39] Another, another source of uncertainty then about exactly which of all archives kinds of things happen that.

**E20** [02:09:45] Very much so and the perception of the value of the data. You know, you could do really complex forensics, but the economics are insane.

**E14** [02:09:55] Yeah, and to some extent it depends as your comment there it's, it's, the kind of value of the data. If you're actually in it and in any case where you're then getting a major public inquiry, in effect, there's probably gonna be more effort put into trying to recover data as well than if it's in a community archive trying to keep records of purely local interest, you know.

**H** [02:10:25] I think there's also been a bit of discussion on the chat about the extent of are we talking that because we had this on a higher shelf, it's not got wet at all or it just got kind of drip off, it's been fully submerged. So it's the question states inadequate mitigation. So if it had fully adequate mitigations then we're saying it's not going to get wet at all. But does anyone have any comments about whether they answered it or if they interpret it as fully submerged in water or just maybe a bit wet around the edges.

**E20** [02:10:56] So it doesn't say a burst pipe in the question. It say's flood, I'm with Noah on the ark with this one, I'm expecting that to be a flood. You know, I'm expecting it to be fully submerged. That was my interpretation anyway.

**E19** [02:11:11] Well, you know, I was envisioning kind of assume that type flooding scenario.

**A** [02:11:20] You're all very cheerful, optimists one and all.

**M** [02:11:23] Its just how you define flood, isn't it?

? [02:11:27] You know,.

? [02:11:27] It's pretty serious.

? [02:11:29] Generally, yes.

**M** [02:11:30] It Is very serious. And I think it's a worldwide flood up to where, however my feet it was then. Yes, these will be serious. So are we done then with the with the better good mitigations question?

**H** [02:11:51] OK. So.

**H** [02:11:55] I think now we're ready to move onto question 22 if everyone else is. Yeah, I'm sure. Question 21, we've covered 20.

## **Question 22**

**H** [02:12:04] So out of a thousand files, all stored on a storage medium of type A so less stable as an archive where staff have good technical skills. How many files would you expect the bit stream to be inaccessible due to obsolescence?

**M** [02:12:22] And so we're thinking, generally speaking, most people's best estimates are sort of around 500 or less. So Expert 15 thinks the uncertainty is very narrow there. Expert 9 thinks it's a bit wider, but would have a similar lowest plausible and best estimate. And ditto Expert 5. We've got Experts 16 and 19 and 14 with a lot more uncertainty and with a higher best estimate for that. So what, what, are the sources of uncertainty here? When

could it be as low as Expert 15 suggests it might be? When could it be as high as, the, as Expert 6 suggests it might be?

**E14** [02:13:15] Even with good technical skills, some things are just dead. So.

**E14** [02:13:20] You know, you're just not gonna be able to get it off for this really obscure format, media formats, that you're just not going to be able to get the parts and things. It can be fairly high. But equally, you can be pretty lucky and we'll you get it, you know, if you can get the stuff, get things going, you're going to recover a pretty good proportion of it probably.

**M** [02:13:42] Yep

**E07** [02:13:48] The uncertainty. It does again, very much depend on the type, so this is quite a few different types of media represented in that definition. So there might be differences depending on what type. Also how old it is and how it's been stored in the past.

**E13** [02:14:18] Hi, can I leap in here. The question asks about the bit stream. Are we able to recover the bit stream off the media? Not, are we able to do anything sensible with the bit stream?

**E13** [02:14:32] So I, I was thinking more about hardware and media obsolescence than I was about any kind of software or format obsolescence.

**E14** [02:14:43] I think we're thinking more on the hardware side here.

**E13** [02:14:47] In which case.

**E13** [02:14:52] Are you are you sticking with your previous answer?

**M** [02:15:05] One or two people in the chat saying, but perhaps thinking in a slightly different way now.

**E07** [02:15:13] I was thinking I was definitely thinking about storage media, but meaning formats around floppy disk, SD drive, CD rom disks, USB flash drive being different media formats.

**E14** Yeah

**?** [02:15:35] I think for me, again, this depends on the capacity or capability of the organisation because our organisation perhaps struggles to download current, software or you know, and work with current hardware. So I feel I would have no support in looking at obsolete hardware or software.

**E14** [02:16:02] So that maybe increases the range, but again. But if we are we are assuming in this in twenty two, we're assuming good technical skills.

**E07** [02:16:10] Some of these like, flash drives, memory sticks aren't they.

**E14** [02:16:16] Yeah,.

**E07** [02:16:17] Memory sticks and CD roms and I mean, I know computers generally aren't made with CD ROMs anymore. CD ROM media's

**E14** [02:16:28] But you can still obtain an usb

**E07** [02:16:30] Yeah,.

**E14** [02:16:32] CD or DVD ROM drive pretty easily.

? [02:16:34] Well, you can if your organisation lets you.

**E14** [02:16:37] True.

? [02:16:38] If it doesn't and you can't plug it into your network. Yeah. What do you do. And then that's, that's, the bit about, you know, the purpose of the risk model isn't it. Yeah.

**E19** [02:16:47] I think, I think, those, those four or five examples on the supplementary document aren't particularly, I mean, that is that those are some of the more historically ubiquitous, media formats that we're talking about, say a DD tape or say, a zip, a zip drive storage medium that would be a lot a lot greater risk of media, media obsolescence.

**E14** [02:17:18] So I think there is necessarily a fair amount of variability in this one.

**H** [02:17:29] Anymore comments on question 22 or move to question 23?

### **Question 23**

**H** [02:17:35] So 23 is identical apart from the fact that you now have poor technical skills, so in Question 22, though, it was less stable. You had good technical skills. Now it's less stable when you don't have like a technical skills.

**M** [02:17:56] Guess the barrier is about money and permissions to the right tools that still apply. Whether you've got good or poor skills.

**E14** [02:18:08] Yeah, but then even if you, now, even if you had those even if those barriers weren't there, then you're, you're, going to be less able to do anything.

**M** [02:18:18] Yes, we see an increase in the spread of variability here. So we've got some. It could be very low and I'm quite sure about that in Expert 4 and Expert 16. But a lot of, most people, seem to think that there's a wide variety, there's quite a bit of uncertainty in the system here.

**M** [02:18:52] Any other thoughts?

**H** [02:19:02] 25, again shuffled over compared to 24

**M** [02:19:15] Can you read the questions for us?

### **Question 24 and 25**

**H** [02:19:17] So question 25. So out of a thousand files stored in a storage media of type B more stable at an archive where staff have poor technical skills. How many files would you expect bitstream to be inaccessible due to obsolescence?

**M** [02:19:35] So did we, sorry, did we skip over 24?

**H** [02:19:38] Yes, sorry. 24 was identical to the previous one. Well, I'll read 24.

**H** [02:19:45] 24 was out of a thousand files stored in a storage media of type B, so more stable, at an archive where staff have good technical skills. How many files would you expect bitstream to be inaccessible due to obsolescence?

**H** [02:20:00] so, like 22, but now with a more stable media,.

**M** [02:20:03] Yep.

**M** [02:20:04] Twenty three and twenty five, a similar. But just change in the media.

**H** Yeah.

**M** [02:20:11] So, in all these, they're all very closely related. Just in one of them, we're changing the capability of, the technical skills, of staff.

**M** [02:20:21] And then in any other case where we're changing the stability of the medium. So we'd like. Yeah, those four things. Anybody got any anything different to add?

**M** [02:20:35] We talked about some of the reasons for the variability.

**E07** [02:20:45] I thought, well, although type B medium is more stable, in some senses I thought that the media formats that are mentioned in the definition or perhaps harder to access for an archivist.

**E07** [02:21:04] So I think a lot of people can deal with USB memory sticks and CD ROMs, but perhaps not so much so LTO tapes, and bluray discs. There might be a bit, like, people might not have the necessary equipment to deal with those. So it might be more stable, but perhaps a little bit harder for people to get the information off, perhaps. I could, I could, be totally wrong there.

**E07** [02:21:36] But just dealing from just speaking from personal experience, I've got data on floppy disks and CD and memory sticks, but I've never got data off LTO tape or a Bluray disc. It may be just that no one's ever given me one and given me that challenge.

**M** [02:21:57] Anybody got any pointers? Is it more difficult?

**E14** [02:22:03] Perhaps that even increases even more the difference between the case with good technical skills and bad technical skills.

**M** [02:22:30] Well, Expert 20's made a comment in the chat about storage technology being a bit of a red herring.

**E13** [02:22:41] I agree with that.

**E14** [02:22:43] I think it is probably, it is probably true, but then hopefully that's something that the model will actually help us demonstrate in a, from a, quantitative standpoint. Yeah,.

**E20** [02:22:54] Yeah.

**E20** [02:22:54] I mean, I don't mind answering questions around it, if you know what I mean. I think you asked the right questions. When I go. But I think for the record, I think the issues, the real challenges we face are, I think, about meaning making, you know, about the context, the authenticity of the records, about the documentation, about who's taken policy lead, who's funded to do it. Some of the, some of the issues around, you know, floppy disks or optical disks. I mean, they are real, but they're they're there. Yeah, but I think that they are mainstream challenges.

**E14** [02:23:24] Yeah. But I think that's that's something we are in effect, you know, to an extent that's something we are hoping that the model will answer for us and demonstrate that actually we don't need to worry so much about side, so that it is those other things that actually outweigh the risks from, from, storage failures and things. But at the moment, it's hard to prove, that it's easy, you know, it is easy to kind of fall down that rabbit hole when you, if you do start modelling risks and those are the things that have perhaps been more amenable to modelling. So people have tried that.

**E20** [02:23:55] Yeah. Absolutely agree. Every time we open the bit list for nominations, everyone's got lots of floppy disk, you know, obscure variations, which they love to share with us, and it's true. It's interesting, a challenge, but it does lead us don't as you protect a particular type of analysis or particular hope.

**E14** [02:24:11] Hopefully we actually get now. This gives us a more holistic picture where we can actually kind of demonstrate that to some extent.

**E20** [02:24:17] Yeah. Good, good, good, good.

**H** [02:24:19] And I just welcome Expert 15 to the call, I think, who's joined us. And I know we're currently on quite discussing question 25 and looking at the range graphs. I'm so glad you've been able to join us. And perhaps at lunch in the break, I can give you a catch up on what you might missed earlier.

**M** [02:24:39] Ok, are we ready to move on.

**H** Yeah. So question 26.

## **Question 26**

**H** [02:24:42] Out of a thousand UK archivists, how many would you expect to say that they had at least some knowledge or skill to be able to generate a checksum of a digital file?

**M** [02:24:57] And so we've got quite a spread of views on this one. Again, we talked earlier about. Would they say that? Would they be, could be hand held through the process or not? Would they say it? Expert 15 is convinced that very few would say that and there's very little uncertainty on that.

**M** [02:25:15] Most people got a moderate amount of uncertainty on that. Some people think it might be a lot higher. Any particular comments that we haven't covered before about this particular question? Checksums.

**E20** [02:25:36] I look back on my own answer there and think I've given that quite a high range. And I think as I reflect on it I'm tempted to move that back down.

**M** [02:25:45] Yeah

? [02:25:57] I just thought about newly qualified archivists in the spectrum, so I would expect people who had recently qualified and were new to the profession would be like, yeah, no problem at all. Whereas people who are more established in the things had been there for ages might actually not be so confident. So one was fairly wide. Because or.

**M** [02:26:25] And do you have a feeling about the sort of population of archivists, whether they are sort of evenly distributed between those sorts of across this sort of experience range, or is it a profession where you get a lot of people who've been there a long time.

? [02:26:44] No, I think I've sort of worked on the basis that actually we've got quite a lot of new trainees and they're qualifying year in, year out from lots of different universities. What was sort of I suppose I was thinking of is an even distribution system. Yeah. Conscious that newer professionals may be more savvy. I don't know how accurate that that is, but it informed my thinking.

**M** [02:27:18] Any other comments, so shall we move on.

**M** [02:27:24] OK, shall we move on then, 27.

## **Question 27**

**H** [02:27:29] Question 27, out of a thousand UK archivists. How many would you expect to say that their IT provider supports the requirements of their organisation's archival activities to a large extent, or a very great extent.

**M** [02:27:47] Generally, they're pretty pleased.

**M** [02:27:52] So there's a lot of variation here. Expert 5 and Expert 12 think that it would be quite low with a good deal of certainty. Most of the rest of people seem to think it's somewhere around the middle, maybe about half each. Expert 4 thought it was a bit higher than that. We've heard a few comments about IT providers and support.

**E14** [02:28:18] I think in light of the rest of the discussions and the fact it actually says, you know, a large extent or very great extent, I think maybe I've been a bit generous on my initial estimate.

**E20** [02:28:26] You know, I'm going to go with this one again and say I haven't, ever stood in a lunch queue at an archives conference and heard somebody really seeing how great their IT support is. Oh, maybe it's because I attract this. But it does seem to me to be a persistent problem a persistent and universal message I hear, of people feeling really frustrated with their IT support.

**E01** [02:28:52] Well, I actually think in local government it's particularly difficult. And so I think the more local government people who've got in this, the lower it will be.

**E20** [02:29:06] I think that's true. I think that's true. I mean, I have some local government experience as well, and it was, we were outsourced and had to use an outsourced IT provider by going through the council. It was it was like a hall of mirrors. It was impossible to try and figure out who you were supposed to speak to and why. I mean, I'm sort of baring my soul here, so I'll be quiet. But it was incredibly hard.

**E12** [02:29:29] I think I queried. I think when I came to this question again, there was an thoughts in my mind around the word supports, I think, and defining what that actually meant in terms of, supporting versus providing. So if, if, your own department was providing the requirements, but then IT could in a way support you with that. It was yeah. I kind of went back and forth with this one, I must admit, quite a bit. Based on that word supports I, I, kind of. Yeah. Wasn't entirely sure what that meant. Was it that they supported you as a department going ahead with what you wanted to do and that was all fine. Or was it that they were actually providing it for you. Or yeah that and I probably should've clarified. So apologies.

**E14** [02:30:24] I think yeah. Having a supportive, you know, relationship, I guess to some extent. And actually it says it does need to be to a large extent or very large extent, which I think was what I hadn't completely.

**E14** [02:30:37] Taken into account. Yeah. You know, we're actually talking about someone, you know, pretty satisfied customers here, which I think on reflection, I've been a bit generous as to how I interpreted.

**E13** [02:30:50] Hi, can I just draw, just talk about the phrase that influenced me the most in that. The question asks about their support from their IT provider, but not in respect of, if you like, the background day-to-day operational support, but their organisation's support in respect of the organisation's archival activities. And I would have to say that my experience working across the, the, local authority sector is that this is so close to zero that you have to say zero.

**M** [02:31:30] Yes, fair enough.

**E07** [02:31:34] I was I was wondering about what was meant by IT provider, so whether that was your, your, in-house IT team that supports and controls IT across your organisation or whether it could be an external company that helps you with digital preservation like Archivum, for example, who could be hosting your storage in your digital preservation solution for you? Would you consider them to be. Your IT provider?

**E14** [02:32:04] I guess, in that context. Yes.

**E07** [02:32:09] Because that kind of question, depending on if you were just talking about your ,your, in-house IT sections.

**E14** [02:32:19] But then I suppose you would also need, you know, some IT departments would probably not even want you having that kind of outsourced arrangement either. So you that could be. There's some tensions there as well.

**E07** [02:32:37] Definitely, yeah.

**E01** [02:32:41] I was also thinking about the broad range of activities that the archives service do. So in general, we've got our online catalogue. We've got our cataloguing management system. We communicate with, through, emails. So, you know, in all of that, that I felt that probably our IT provider did support our activities to a large extent, even though I would maintain that they abjectly failed to support our digital preservation activities. So, sort of, the broad range of the question meant I perhaps answered it more favourably than I might have done if it was more specific.

**M** [02:33:33] Right. Are we ready to move on?

## **Question 28**

**H** [02:33:39] Question 28.

**H** [02:33:41] Out of a thousand UK organisations which experienced a cyber security break or attack, how many would you expect to have been experienced viruses, spyware or malware, including ransomware attacks? I think when we discussed we were defining this as a malicious attack.

**M** [02:34:14] Any thoughts? We've got the range graphs, we've got Expert 8 who thinks it's very low and very certain, but most people are more or less in the middle. Expert 10's a bit more optimistic, larger numbers. Expert 18, same.

**M** [02:34:36] So the uncertainty here is this personal uncertainty or is it uncertainty of the system? Will it depend on the business or the organisation?

**?** [02:35:08] Can you hear me? What I was a bit uncertain about here, and I may have missed the discussion yesterday, I'm not sure. I was unclear as to what the outcome of the break or attack was. I was sort of thinking, well, there's pretty high chance that things are going to experience it, but.

**M** [02:35:30] Yeah.

**?** [02:35:31] What, what, is the question about the experience of the attack or the outcome, and I was more focussing on the experience.

**M** [02:35:42] Are you able to clarify, Alex or Hannah? So we took today. We actually have, for example, a virus that destroyed something or just a virus that successfully got into the system.

**H** [02:35:54] So we're talking about off of all the UK organisations which have experienced some form of cyber attack. How many experienced a virus or malware. So it's not about the outcome necessarily. It's how many that maybe would identify that they've had a cyber breach would also say and they've had a cyber breach. That was a virus or spyware or malware.

**E19** [02:36:19] Yeah, I mean, from my perspective, I think that the most common type of attack is a DDoS attack, which doesn't fall into those categories, that would lead me to suggest that the response to this is quite low. That said, I think it's, it's, such a wide ranging field that I couldn't give any, happy degree of certainty here.

**E14** [02:36:41] Yeah. On the other hand, it's unlikely that a DDoS would actually leads to a breach.

**A** [02:36:45] Sorry, what's a DDoS?

**E14** [02:36:47] Distributed Denial of Service attack. So it'll take down your Web site. But, you know, going actually lose any customer, client data or whatever.

**E20** [02:36:57] Yeah, but if the definition that a cyber, cybersecurity attack, which a DDoS such as.

**E14** [02:37:03] Yes, it says break and break in or attack, doesn't it? True.

**E14** [02:37:07] But then a lot of those are driven by malware, once all the bots, the doing those things are underlying result of having bots install stuff by malware on other systems.

**E07** [02:37:28] I think my, I had quite a lot of uncertainty basically because I don't know this field very well and I think I would it would help to know what other types of cyber security attack there are.

**E07** [02:37:40] So you've mentioned the denial of service potentially, but I'm just seeing that list of viruses, spyware, malware, ransomware.

**E07** [02:37:50] I immediately think, well, what what else is there? That's covered quite a lot. So knowing what they were, all the other things are, then I could probably make a kind of a more educated guess perhaps.

**E20** [02:38:01] Yeah, the other things you're looking at DDoS. You are look at things like phishing, spear phishing, those kind of attacks, also. Which could lead to injection of, of, things like malware. It is quite a broad, category, range of things, to include there.

**M** [02:38:30] Any other thoughts or questions, clarifications? OK, shall we move on then?

## **Question 29**

**H** [02:38:42] Question 28. No, 29 apologies. So question 29: out of a thousand global data breaches, how many would you expect to be due to system glitches?

**M** [02:38:59] So, again, we have this general clustering to the left here with Experts 18 and 14 and 9 and 5 being very certain it's quite low and then a middle group. And then there's Expert 6 who think this is quite high. So what might be the reasons when it would be low, the reasons when it would be high and the reasons for the uncertainty?

**E19** [02:39:31] Again, from my perspective, systems rarely glitch, the vast amounts of data breaches are generally human error from my observation and experience.

**M** [02:39:44] That would make it very low then.

**E19** [02:39:48] Yeah.

**E13** [02:39:50] Hi. Can I just ask for clarification on the metric here? When, when, you talk about a thousand data breaches, are you meaning a thousand events regardless of the

size of the data breach? Or are you talking about the number of breached data records? For example, the Sheffield breach, which has just been reported, is 8.6 million records. Does that count as one or does that count as 8.6 million?

**H** [02:40:23] That counts as one. So we talk about events.

**E13** [02:40:28] Thank you.

**M** [02:40:37] Any other comments or questions?

**E01** [02:40:42] It's [name] here. I put it really low because from my experience of dealing with data breaches in the county council, 99 percent of them are human error, not anything else. So. And then if you then put all of the IT ones in a different category, then the amount that are glitches as opposed to those others is even smaller. So that's why I've come very far on the low end of this one.

**M** [02:41:08] Very good reasoning. Has everybody had all the discussion that they need to make the relevant assessments for their second round?

### **Question 30**

**H** [02:41:20] Question 30: out of a thousand UK archivists, how many would say that their digital collections were fixity checked at regular intervals?

**M** [02:41:36] We talked last, yesterday, about that they had logged that the check had been done and they know they were explicit rather than just automatic, didn't we?

**E14** [02:41:44] Yeah.

**M** [02:41:47] So again, we're back to would they say this? Would they know this? What have we got here. Regular intervals. Some people thinking, yes, it's pretty high, 80, 90 percent. Quite a few people seem to think it's quite low, though. Expert 5 thinks it's very low. And Expert 13 thinks it's very low. Expert 18 thinks it's a bit of a range, but is tending to be on the low side as well. Same for Expert 9 and then more people in the middle.

**M** [02:42:22] So what are the reasons why people would think would know their fixity checks were done at regular intervals and why would they not know what they would know if they were not done at regular intervals? What's the difference there?

**E14** [02:42:40] I think this is one that people know it's, something you know, it's one of the fairly early things on the NDSA levels and things, so people will have awareness that it's something that should be done. But it's probably still not actually happening in that many cases.

**M** [02:43:00] Right ok, so the spread we're seeing is different archives maybe and the uncertainty

**E14** [02:43:04] and a bit of confidence, you know, how confident, even if the archivists know that's one of the things, they might depending how directly involved they are with the digital side of things, they might not be fully aware as to whether it is actually happening on their own collections and things like that.

**E13** [02:43:20] It's [E13] here.

**E13** [02:43:22] I'm really interested in anyone's reasoning who's gone to the high side of all this. Having just done a major piece of work attempting to induce the main outsource providers of digital preservation to do fixity checking regularly, or report it regularly, and to introduce fixity checking within our own storage provider and there, there, is no ability to do it.

**M** [02:43:57] Right.

**E13** [02:43:58] So I'm really interested in anyone who's gone high. If they can explain to me why?

**E14** [02:44:07] Expert 9 said he's got it the wrong way round.

**E20** [02:44:12] I think it does refer to the regular thing, doesn't it? Mean as a frequent as a regular, it can be once every five years. The Olympic Games is regular.

**E13** [02:44:22] Yes, that's that's very good provision of fixity checking on ingest, the principal tools like Bagit sort of hand that to you on a plate.

**E13** [02:44:37] But from then on, in my experience, there's nothing.

**E07** [02:44:44] I think the systems, like the digital preservation systems do it, don't they? Preservica and Archivematica and Rosetta and all the rest? [4.0s]

**E13** [02:44:55] I'd love to share any experience you've got of them reporting it to, to demonstrate that it's regular.

**E07** [02:45:02] I've only got experience with Archivematia but I know that that's yeah. I've managed to get that working. I haven't worked with the other systems. But I'd be surprised if they didn't know everything they preserved.

**E14** [02:45:16] Preservica you can see if the job is running and you, and any, failures will be reported to you.

**E20** [02:45:22] That's right. And we've we've written up a kind of assessment. The British Library's online authenticity checker. And I think they've reported something like five failures over something like 1.8 billion files. I mean, ridiculous. It's one of those weird moments where they get very, very excited about the five files. But actually, in the context of the massive collection that they had and I think they run that annually, and I think we even had an argument with them about how long they needed to keep the check in the logs for, because they didn't know how long we needed to keep the checks in the logs for the logs themselves are enormous. But yeah, yeah, there's some there's some experience of this [Expert 13]. I mean, I think you're onto something. I don't think it does happen regularly, but I think the systems do permit it.

**E13** [02:46:10] OK. Thank you.

**M** [02:46:12] There must be some other reason why it's not being done. Is this, if you've got a system and your system can do it?

**E20** [02:46:16] I think its because people don't have systems by and large.

**E13** [02:46:23] It's hugely resource intensive.

**E20** [02:46:27] Yeah, that's a good point as well. I mean, it's also fairly electricity, what's the right word? You know, consumes a lot of power, doesn't it? To actually do that over a whole system. It's not environmentally sustainable, really.

**M** [02:46:46] OK, fair enough. Good, that's. I think I've understood a lot more about that now. So that's very helpful. Any any other questions or comments that might help people to fix on their values for the second round?

**M** [02:46:59] OK.

### Question 31

**H** [02:47:01] OK. Question 30 say this is where we had a change of question, which, apologies, came quite late in the day. I think nearly 5 o'clock. I'm aware that some people might have answered Question 31 as it was written on the answer sheet. And the original elicitation sheets are not updated as done in the e-mail that were sent about 5:00. So I don't know how we manage if there are different results here.

[02:47:28] Well, could we if you whichever you'd done, if you could make a note of it in the spreadsheet before you send it back to us today. That would be very helpful.

**H** [02:47:37] And I'm going to say I'm going to send round at lunchtime, a new answer sheet for today's results. And it will have the new wordings of questions 31 and 32 and 33 in. And I'll also send an updated version of the elicitation question sheet. In fact I think I can do both of those now, because I've got it sitting in my drafts in my inbox. So then hopefully when we are answering later you will all have the latest version.

**M** [02:48:02] What is the question? The Question 31 you sent around at 5:00.

**H** [02:48:08] So the question, does one that was sent around finally was out of a thousand files. How many would you expect to have become corrupted during transfer from a depositor to an archive?

**M** [02:48:23] OK, so the split we see here, maybe the split between two questions. But generally the clustering is to the left. People think that will be relatively low, expert number 14 has a lot of uncertainty about this or thinks that the system is very uncertain. Ditto Expert 10, to a lesser extent Expert 2.

**M** [02:48:49] So any any comments or remarks about how you came to your decision making on that, that corruption.

Not E14 [02:49:00] For me. So I think I'm probably expert 14 there. I think there are just so many variables that could affect this. And, you know, if you're if you're transferring all a thousand once there is, there is, a possibility that it would be all of them. There is a possibility. But none of them. The likelihood is that it's going to be

Not E14 [02:49:30] a very small amount, if any, but it depends on. Yeah. Depends on the variables.

**M** [02:49:35] Good. Good. Thank you.

**M** [02:49:50] The comment by Expert 9, I would say the corruption of files would be low, but corruption metadata could be high depending on processes. Expert 11 has a question about how you would tell relies on depositor checksum and also transfer process.

**E14** [02:50:05] That's. Yeah, I mean, that's partly what we're trying to get at. So, there is a risk if you're not having checks on the generator beforehand and that corruption could be happening and you never know about it. If you were actually getting people to generate checksums before transmitting it, then then then you know, there's still some rate at which is happening.

**E14** [02:50:24] But yes, you can correct it, but that tends to suggest that if you don't have those checksums in place, then there's some underlying rate that's happening and you're ending up getting corrupt files in the archive.

**M** [02:50:43] Any more comments or questions?

**M** [02:50:50] Shall we move on?

### **Question 32**

**H** [02:50:52] Question 32 and again, this was another one where it changed. I appreciate some people might have answered the old 32, but the new 32 is out of a thousand files at an archive where you have poor system security. How many would you expect to have become corrupt and identified this given you had to checksum to compare to. So knowing that something has changed where you have poor system security.

**M** [02:51:22] Poor system, security and again, we're thinking in general it's going to be very low.

**M** [02:51:27] People on the high side, might have answered a different question and Expert 12, very uncertain thinks there's a big, big bit of uncertainty in the system or is uncertain themselves.

**M** [02:51:40] Similar Expert 16. Expert 17 thinks it's quite high. So any any comments about that, any any sort of thing?

**E14** [02:51:51] I think I was. I've got quite, quite a broad range. And I think it's partly exactly what the cause of the corruption is. If it's someone actually, you know, a kind of, the insider threat, they might deliberately, trash a whole load of files because, you know, they've just been sacked and had nice high level of admin access and decide they going to get there their revenge before they actually are thrown out of the systems. So, you know, in that situation, you could have a very high number of files that would be corrupted. You might then be able to identify it because you've got the checksum. But you know that that how bad that, the kind of ongoing effects that damages then depend on, on, other factors. But, you know, it could be you might get a few files damaged by accident when someone accidentally deletes a directory because they've got, you know, they probably shouldn't have permissions to actually be able to do that. But, you know, that's where the poor system security comes in. But then you go up to the scale where you're talking about a

deliberate act by someone who's been annoyed by something and wants to get their own back on the organisation. So there's, there is, necessarily quite a wide range, I think.

**E19** [02:53:08] Yeah, I think there were the questions. I mean, I did it based on the revised question. But I think the way it's phrased, you've got to two factors is pulling in different directions here. So lack of system security suggests that you might be more susceptible to some of the things corrupting your data, which would move the number higher, but then your ability to identify that would move the number back lower again.

**E14** [02:53:33] So for 32, is he you know, you've got the checksum. So you think you can identify it and then we will come onto 33 in a minute, which tries to look at a slightly different aspect of it.

**M** [02:53:54] So should we move on to 33, any more comments or questions for 32?

**M** [02:53:59] Expert 11 says, I wasn't sure what the full implications of poor systems security might be? Are we took me just left and not managed.

**H** [02:54:10] There could be a range, couldn't there, of what, what that means. So it could be left and not managed or it could just be they, they, haven't been really rigorous in this loopholes.

**M** [02:54:24] Does that help?

### **Question 33**

**H** [02:54:32] Question 33. Again it was an updated one and David was just starting to come on to it. So out of a thousand files at an archive, again, where you have poor system security, how many would you expect to have been corrupted and not been able to identify this despite having a checksum? So you've got a checksum to compare to. But you haven't.

**H** [02:54:57] There's been corruption and you haven't been able to tell.

**E14** [02:55:01] Could potentially because someone's actually been able to change the recorded checksum as well. So, you now try and verify. And it looks like the checksums are matching. But because someone had really high level access, they've actually been able to change the recorded checksum.

**E13** [02:55:17] Hi.

**E13** [02:55:19] My understanding of the question was that the management information system was such that you had the independently recorded message digest so that possibility wasn't in bound. We had a conversation about this yesterday.

**M** [02:55:41] Thank you.

**H** [02:55:48] So I guess the question is, even if we have a good information management system. If you have poor system security, could that compromise the information management system and the information that contains

**H** [02:56:02] therein. Can it?.

**E13** [02:56:07] Well, if you if you don't have a good security over your, keeping your logs, of the checksum, so-called, then all bets are off.

**A** [02:56:31] Hold on, didn't, questions as 33. We didn't. Did we take this out completely in our final set?

**E14** [02:56:39] No, we did have one in there and.

**A** [02:56:42] We did.

**A** [02:56:42] Sorry. Thank you

**E13** [02:56:51] I think the gist of Question 33, as I read it was that you've got your system of maintaining the checksums. What, what, is your belief about the level of protection that provides you?

**E13** [02:57:10] And it's more or less absolute. Because that's, that's, its job.

**E14** [02:57:17] Yeah, but I think, you know, certainly when we were discussing this, I was thinking of the case where actually someone's managed to, to, have sufficient access to actually maybe change, change, some of those checksums.

**E13** [02:57:33] OK.

**M** [02:57:34] Fair enough. All right. Is that helpful? Any more questions? Are they any more discussions on that one?

**H** [02:57:48] OK, so thank you. So the ones I've sent through should all have the updated question, as we've just discussed, o, and so and I've highlighted it to make clear that it's not the same as the first one.

**H** [02:58:01] OK, great. Thank you. Question 34.

### **Question 34**

**H** [02:58:06] So this is back to normal. No more changes now. Out of a thousand files which were all deposited with a checksum at an archive where there is sufficient information, information management, but poor system security. How many files can you expect to provide the assurance that the bitstream is identical to when it was added to the archive?

**M** [02:58:33] So what do we see here, we see Expert 6 being very sure it's very low. And also Expert 5.

**E19** [02:58:43] Sorry that wasn't the question, the question that Hannah read out then wasn't the question that I've answered and that's in the spreadsheet.

**H** [02:58:52] Say guys, maybe I've put them out of sync.

**E14** [02:58:58] Yeah. I lost track, which one are we on now?

**H** [02:58:59] Sorry, I've put the one I'm looking at I've shuffled. Right, that one's gone. So if I read this one, is this correct? So out of a thousand files with their bitstreams stored and accessible, but where you cannot guarantee their integrity. How many files would you expect to be bit preserved?

**M** [02:59:19] is that the right one?

**H** [02:59:21] Great. Sorry about that.

**M** [02:59:22] That's right. OK. So again, we're still seeing that Experts 5 and 6 are thinking it's low, Expert 1 thinking it's low, Expert 11 is very sure that it's high. People are tending on the whole to go on the higher side. So why are the reasons it could be as low as, as low as zero or as low as, very low. Why? Why would it be higher and why would there be a huge uncertainty in the system here? Experts 13 and 16 and 18 are showing quite high uncertainty here.

**E14** [02:59:59] I think you have a reasonable chance that it will be bit preserved. You just can't actually demonstrate it.

**E07** [03:00:09] Yeah. And so lots of things. I mean, if we were just talking about corruption of files, I would say the chance is quite low, but it does depend on the system security that's there and how many people have got access to the files.

**E07** [03:00:26] I mean, I think the greatest threat here is, is people going in and changing them, whether accidentally or deliberately or maliciously or just without any knowledge of what they're doing. And that's probably the biggest threat. So it does depend on, you know, whether that's locked down or not, how many people have access.

**E13** [03:00:49] The reason that I've gone very low on this is that for me, bit preservation entails preserving the integrity. So you either you know that you've got bit preservation because you've got to checksums or you don't know. And if you don't know, you cannot guarantee their integrity because you've got no basis upon which to found that belief.

**M** [03:01:26] Anyone want to make another comment.

**E20** [03:01:32] I think I got myself in my usual muddle between integrity and authenticity, and obviously that changes the question again one more time. But I should reflect on that and think about integrity more accurately than authenticity.

**E07** [03:01:57] I've gone, on reflection, I've gone quite high on this one and quite definite, quite confident, but now I'm thinking I should be a bit less confident. I think maybe I still stay, I think being high is good because of what I said earlier about it's not a great chance that things will be corrupted, a lot of things will sit on a file store quite happily and they'll retain their integrity.

**E07** [03:02:24] But I think I need to add in a bit more uncertainty because of the uncertainty about the access conditions, and who can, who has edit, edit rights

**E07** [03:02:33] on some of these files, for instance.

**E13** [03:02:34] The, the mental test that I use on this is what happens when you're standing in court and being questioned by some quite good barristers.

**E13** [03:02:51] Are you able to guarantee the integrity of [...]

**E13** [03:02:56] You're producing an evidence.

**E14** [03:02:57] The question says we cannot guarantee the integrity.

**E14** [03:03:01] But, you know, there's still a reasonable chance that actually some of those things have been bit preserved, they haven't changed, but we can't guarantee, you know, we can't guarantee it, or necessarily demonstrate it.

**E13** [03:03:09] That is exactly the point. And that is why in my mind, bit preservation entails being able to guarantee the integrity, which is right.

**E14** [03:03:20] It's not how we've defined it. I think.

**E13** [03:03:23] Well it's how I'm reading it.

**M** [03:03:26] Are there any clarifications you could offer because I think Expert 18 has a similar, similar thought.

**E20** [03:03:33] Yeah. I see where you're going [Expert 13] with the kind of the lawyers, court, and all that, but that's a very unusual use case, you know.

**E13** [03:03:40] Not for local authorities it isn't.

**E20** [03:03:45] Oh, I am not sure that's true either. I mean, depends on the nature of the local authority holdings.

**E20** [03:03:52] I mean, if it's an oral history recording and the local history centre, who is going to challenge that, you know.

**E14** [03:04:01] Yeah. And if you look at the you know, the definition is given with the questions, say a term used it in a very basic level of preservation of digital resource as it was submitted. So, you know, the bit stream is unchanged. Yes. We can't guarantee it. But for a reasonable number of files sitting on the drive they won't actually have changed.

**E13** [03:04:20] But you don't know which one has changed.

**E14** [03:04:24] Yeah, but that doesn't matter for this question.

**E14** [03:04:30] We're just saying that from some sample, actually quite, quite, a high number will be successfully sat there.

**E13** [03:04:37] Oh, I don't deny. Yeah.

**E14** [03:04:39] And that's what the question is asking because it specifically says we can't guarantee the integrity, but the way we've defined the terms for the purposes of the question.

**E14** [03:04:51] You might disagree with those definitions, but that's how you know, that's the basis from which we're trying to do that in this case.

**E13** [03:04:59] OK. I'll think about it. Thank you.

**M** [03:05:02] Let's move on.

**H** [03:05:08] Question. Question 34?

**H** [03:05:12] 35,

**M** we're doing really well here.

### **Question 35**

**H** [03:05:16] So out of a thousand files, which are bit preserved, at an archive that has full access to rendering tools, but where the files have insufficient technical metadata, how many files would you expect to be able to render?

**M** [03:05:35] And so, again, we have we have a split here. We have Expert 11, and Expert 13, thinking it's quite high and quite certain. Expert 8 and 12 and 19 think he is quite certain and quite low and Expert 2, is also quite certain and quite high. And then pretty much everyone else has got a very wide uncertainty on that. So what's the nature of the uncertainty here? When could it be low and when could it be high?

**E20** [03:06:10] Think, I am struggling with what rendering means in this context, right? So I could imagine being in a position where you have a file, a large number of files which are preserved in the archive and they can be rendered, but if insufficient technical metadata it's, your rendering things on screen, but it's effectively gibberish. It comes back to those databases. I was describing what we said we could render them, but it didn't mean anything because we had lost the look up tables you know. So the loss of technical information, the technical metadata at that point makes the whole exercise slightly useless. So we have bit preserved. But we're waiting for an Act of God to be able to interpret in a meaningful way. So that pushed me right down to the bottom in terms of kind of usefulness.

**E19** [03:07:00] I would argue that, that those kind of use cases, though, are quite rare.

**E19** [03:07:06] So out of a thousand files, you know most of the time with the rendering tool we should be fine, you don't need that additional technical metadata. You don't need to know what video codec, for example, is in use. If you just open it with a standard media player using for an MP4 then it's probably gonna open.

**E07** [03:07:27] Yeah, I would wonder Expert 20, whether, so the keys to a database in order to turn that data into information that's usable. I'd put that in the category of documentation rather than technical metadata.

**E20** [03:07:41] I think it's hard to argue, that's a fair point.

**E20** [03:07:47] I mean, the encryption keys to the, the, codec off of it for the video file that you mentioned, Expert 19. I mean, that's a good point. But sometimes there is limited, you know, there's restrictions and digital rights management on video files. And if you can'tm if you don't have that, then we're back to square one. But it's a moot point as to where that goes into the discussion, isn't it?

**E14** [03:08:17] Do we need again to revisit how we've actually defined those terms.

**E13** [03:08:24] I have to say that I was quite happy with the way the question was phrased, and it seemed to me to be self-evident from the question that you would expect to be able to at least render it to some extent because you've got full access to rendering tools.

**E20** [03:08:47] Just to read the question again, back then. So Expert 13 that, with your interpretation, would mean that would be a relatively small number. You'd be coming out on the left hand side of that chart, is that correct? I am interpreting.

**E14** [03:08:58] No, That would mean you can render most things so,

**E14** [03:09:03] So you should be up towards the right hand end.

**E07** [03:09:09] The question here is, is whether you've got the tools, you've got the files, you haven't got the metadata, so it's your ability to decide which tool you're going to use, for which file without having the, the, necessary information.

**E07** [03:09:25] Is that right?

**E13** [03:09:27] Yeah, but the technical metadata is there.

**E13** [03:09:32] And you've got full access to tools, so I don't see where the problem is.

**E07** [03:09:39] Yeah, suppose, so you can double click on a file and hope it opens in the right bit of software.

**E07** [03:09:44] But sometimes there's a bit more groundwork that you need to do when you're faced with a file. So you double click on it and Windows doesn't know what to do with it. So you then have to work out which tool you're going to try and open it in.

**E13** [03:09:56] I'd never be using Windows though.

**E07** Whichever operating system you were using.

**E20** [03:10:02] Yeah. You need to make some sort of file association for just double clicking to work.

**E20** [03:10:08] To go back to that though for a moment.

**E20** [03:10:09] There's insufficient technical metadata.

**E20** [03:10:13] Didn't you just say that you had the technical metadata [Expert 13]?

**E20** [03:10:16] I think we're maybe going back to the idea that actually it's fairly implicit. It is there even if you then have to do some hunting.

**H** [03:10:22] But then it was about whether you've made that log of the technical metadata, is it having access to a DROID report, having inspected that video file to see how many different audio streams there actually are, I quite like [name] example of that is you might

you know, it's a video, you open it but unless you've got that sufficient technical metadata, you might not know that you should be hearing sound and you're not.

**E14** [03:10:46] Yeah, but then is that you know, again, is that fairly, is that a rare use case, so yes, that will affect some things. But for a lot stuff, like a whole bunch of Word documents or whatever, you're still be able to open it fine. But yes, there are those, those, more edge case scenarios where you do run into some issues so it's not going to be everything but a fair bit of the proportion's fine.

**E04** [03:11:06] There's far more variation with AV kind of models,

**E04** [03:11:10] complex objects. I think when you're talking about documents and whatever, it's probably a simpler case.

**E14** [03:11:17] Yeah, so partly it depends.

**E04** [03:11:20] There's any number of variations in the encoding from MP4 to MP4 whatever the format may be. So you want to be able to have as much technical metadata at hand as possible to know that what you're receiving at one end is what was actually within the file that you are receiving all the information that's in there.

**E14** [03:11:39] Yeah, So it depends a bit on your what are your thousand files, if you're. [Expert 4 Exactly]. And if you were the BFI and it's all video, you might run into more issues than if it's us at TNA where the vast majority is kind of JPEG2000 and office document type files.

**E04** [03:11:55] You know we have some examples with older audio encodings, video encodings aren't so.

**E14** [03:12:00] That's that some examples.

**E14** [03:12:02] So it's not you know, we're still where we would expect to be able to do probably the majority. There's some range of uncertainty. Yeah, but probably won't be able to do absolutely everything, but it's going to be towards the higher end.

**E04** [03:12:13] And yeah, it's just a cautious approach that if we just open in a video file inside a generic player, it may or may not, depending on the behaviour of that particular bit of software, announce to you the fact that there is an audio stream, but it can't read it. A video stream that it can't handle. So it's also with what meaningful information to get back. From your particular tools that you have on hand.

**E14** [03:12:39] Yeah, so that I think, you know, explains why we certainly wouldn't expect it to be, you know, for around 1000 files You couldn't be sure you'd get everything where you haven't made some effort to actually extract and, get some of that technical metadata into a usable form rather than saying, yes, it's, it's, there implicitly in the thing. But if you know, if you if you don't, as you say, if you just open it with a tool that would appear to be a perfectly reasonable tool to use on it, then you may be missing something without knowing it. So actually, your, your rendering is not good.

**E04** [03:13:12] Yes. With things like 3D formats as well, you get a wide spread of different behaviours amongst different software.

**E04** [03:13:19] When interest took off, it's.

**E13** [03:13:22] Whilst not disagreeing with the thrust of the of the debate the argument there.

**E13** [03:13:30] Is that conversation not more about the quality of the rendering tool?

**E14** [03:13:38] It might be a perfectly good tool, but just not for the thing you've actually, you know, for the thing you've immediately given it and without having that technical understanding, you don't know that. So it's a perfectly good file for an MP4 with virtually all codecs. But for one particular codec, it's not the one to use. So, is that a good tool or a bad tool?

**E04** [03:14:00] Yeah. VLC player may well play, may well play the file back with no reports of any particular details. But unless you go in and you look at the various streams that are available from there, you may just think it's your average stereo

**E04** [03:14:15] video file. Nothing unusual about it. When there may actually be underlying audio streams and commentary, whatever.

**M** [03:14:28] Sounds to me like its good reasons why the uncertainty bands might be quite right on this.

**H** [03:14:38] So, Expert 9's question about he assumed that the technical metadata could be extracted before rendering. Are we assuming you wouldn't then, I think. Yeah. I think that's we're saying you haven't extracted that technical metadata. If you could, then you would have the sufficient technical metadata. But here, you're trying to do it. Maybe a bit blind, so to speak.

**E07** [03:15:02] I think there's also some files for which you wouldn't you wouldn't really be able to extract any technical metadata. And I'm thinking of like my Wordstar files from a floppy disks, that DROID couldn't read. Well, it didn't give me a file identification for them. And there was a lot of hunting around to find the right tool to actually try and open those files based on the fact I didn't know what they were. But I got there in the end. So you don't always need the technical metadata in order to sort of end up being able to open the files in the right tool.

**A** [03:15:42] OK. So, I am conscious of the time, It is just ticking to 1, I think now. If we're happy to end the discussion around Question 35 there, I suggest we go to lunch and continue with the remaining nine questions?

**A** [03:15:59] We've got 44, in total haven't we?

**H** [03:16:02] Yep.

**H** [03:16:03] Question 36 and 38 are quite similar and we got through these sections fairly swiftly last time so hopefully it won't take too much longer after lunch.

**A** [03:16:15] Yes I think it took about an hour yesterday after lunch, so [Expert 2]'s hungry so that's not good. Okay. So we'll see you all back here at 2 o'clock. Yeah.