

Transcript of Recordkeeping Roundtable session:

Due to technological incompetence, this transcript is missing the preliminary introductions.

Acknowledgement of country: Before we begin, I would like to acknowledge the First Nations peoples of all the unceded lands where we are meeting, including the people of the Kulin nations where Monash operates, and the Gadigal people of the Eora nation on whose land I live, and pay my respects to their Elders, past present and emerging and to First Nations participants in our meeting today

Introduction to speakers conducted at the beginning:

James Lappin:

James Lappin blogs at <https://thinkingrecords.co.uk/> and recently wrote the thought provoking article 'Rival records management models in an era of partial automation' in Archival Science. James has over 20 years of experience in records management as a practitioner, consultant, researcher and trainer, working for both public sector, private sector and international organisations in the UK and Europe.

Chris Hurley:

Chris Hurley is well known in the Australian and international records and archives communities as a provocative thinker (NOT THEORIST!). While resisting labels, Chris is widely associated with records continuum theory. Now retired, Chris has held many senior roles in public and private organisations, including as The Keeper of Public Records at Public Record Office of Victoria, Chief Archivist of New Zealand and Bank Archivist at the Commonwealth Bank of Australia.

Andrew Waugh:

Andrew Waugh is a specialist in digital archives. Coming from a technology background, Andrew worked initially on the Victorian Electronic Records Strategy (better known as VERS) in the 1990s and subsequently has been involved in implementing the VERS standards, transforming the technological environment and digital capability of the Public Record Office of Victoria.

Nicola Forbes:

Nicola Forbes is the Director Information Governance for the Fair Work Ombudsman. She has worked in a number of roles as a senior level in information/records and has

led several teams managing complex programs utilising multiple technologies and grappling with the practicalities of recordkeeping in the workplace

JAMES LAPPIN

I've got three questions about our current era of records management. I think it is clear when it started. I think it's clear that when email and corporate networks and the Internet arrived in the workplace in the mid 1990s, that that started a new era of records management with new possibilities, but also things that we couldn't do that we could previously do. And I think it's also clear that when we reach a point where artificial intelligence has is so advanced, and we are so confident in it that we are willing to use artificial intelligence to just reorganize any record system into whatever order we want, that that will also be a new era of record keeping with new possibilities and new challenges.

So, what is this era that started around 1996 and that we're still living in now, but that we're aware will end and possibly relatively soon? What do we call it? I don't think we can call it the digital era, because the era after AI is all pervasive in records management will also be a digital era. And I think we need a phrase for it that kind of tells us something about the reasons why we haven't been able to do what we set out to do at the start of the era.

So at the start of the era, we as an international profession, led by yourselves in Australia, came up with an international records management standard that set out what we wanted to achieve. And what we wanted to achieve was to be able to consistently and routinely capture records into a system (or systems) that organised records by business process and that we could apply our business activity-specific records retention rules precisely to. That's what we set out to do in the first International Records Management standard, and we quite patently haven't been able to do it.

So something's wrong: either we're wrong and we haven't been doing things right for 25 years or that initial idea, that initial aim, was unfeasible for our age. And I think it's the second. I think it's not possible in our age to have a corporate wide system that

consistently captures records into a system that's organized by business activity, business process, functional analysis, business classification, whatever you want to call it.

And I think the reason why is to do with the fact that the correspondence, the information that's moving around our organizations and in between our organization and other organizations, most of that correspondence is being automatically filed. It's being automatically filed into whatever your main messaging system is. And for most of you, that's email. But if it if it was individual teams chat, it would be the same. So it's being automatically filed into a structure that is not organized by business activities, not organized by business process. It's not even a structure of our choice. It is a structure determined by the need that if there is going to be an instantaneous exchange of correspondence between two computers That there has to be some protocol behind that, whereby each computer can file that message so that the recipient or recipients can access it. So email systems file in order to communicate. They are filing systems first. They are communication tools because they can file, because they can automatically file.

So that international records management standard, and the whole DIRKS way of thinking behind it, kind of presupposed some kind of blank canvas whereby we're setting up the structure of systems and we're setting them up in a structure of our choice. But that isn't and wasn't the situation we find ourselves in our era. Correspondence, which accounts in our era, as in other eras, for the vast bulk of information that we have to manage, is automatically filed, but not into a structure of our choice, and not into a structure that supports the complex records retention rules that our organizations want to be able to apply to support the compliance needs that they have.

So we have this disconnect in in our age, and only in our age, between how we need to organize records and how we can organize records. And that is not something our profession can remedy in this era.

I just want to really quickly whiz through the history of recordkeeping just to place ourselves in our wider history.

So it hasn't always been the case that organizations have wanted to organize records by business activity. For the first 700 years of record keeping, if I take a slightly Anglo centric or very Anglo centric perspective, organizations were so small, but they didn't need to have sophisticated records organization schemes. They didn't need to apply sophisticated records retention rules. They just kept correspondence in chronological sequence and they were quite happy with that. And there was no need for records management and there was no records management theory. There are some archivists, but no records management. But when we hit the 20th century, organizations get really big. And therefore, because they've got a lot of people in them, they create a lot of records and therefore there is a need for a more sophisticated organization of records. Therefore there was also a need to control the volume of records. So eventually we get Schellenberg and the creation of modern records management as a science, as a social science. And we get Schellenberg's idea that if we organize records by business activity, by function, we can apply the rules that relate to that business activity to records. And he said it's the most efficient form of organizing a record system. And before email, he was right.

But is it the most efficient way of organizing a record system now? I don't think it is.

Because if you were to try to organize correspondence by business activity now, what you'd effectively be doing is to ask people manually to refile correspondence that had been automatically filed by a machine, and that is both impossible and inefficient. Because obviously the machine - the email systems - file according to a set of automatic rules. They don't use machine learning. They file into a very simple structure. And because they file quickly and instantaneously, the volumes of correspondence exchanged in our era are exponentially bigger than the volume of correspondence exchanged in the previous era. Which means that we simply haven't got the human effort necessary to refile all the business correspondence that we exchange, even if even if we ignore the personal and trivial stuff.

So the whole electronic records management system movement failed because, even though we were able to come up with great structures for those systems, we weren't able to consistently file into those structures. So if we compare the previous era to our era, in the previous era, an organization could consistently file correspondence into a structure of its own choice. And it could do it by the simple expedient of intervening in

the correspondence process, by interposing registry staff who could file in between the post rooms where correspondence would arrive and the destination recipient where the correspondence was being delivered to. So in other words, if an organization was really serious about this records management, it could set up registries to file correspondence into a structure of its own choice routinely.

When email comes in, it bypasses this whole structure. Email just gets delivered straight to the individual instantaneously. And we try something which actually David Bearman, a theoretician, said was doomed to failure and he was proved right. We tried to impose electronic record systems to the side of this and ask individuals to file into them. And it didn't work. And it wouldn't matter what you put in the place of electronic records management systems. So when we then said, OK, well, we'll have something more collaborative like SharePoint. That is also distinct from the correspondence process and it is no more possible to file into, to get people to manually file into, SharePoint things that had already been automatically filed into their email system than it is to an EDRMS system or any other system.

So in our era, we are faced with a situation where we cannot optimize the structure of a record system for correspondence and at the same time have a consistent capture of that correspondence. It's not possible in our era. It might be possible in the AI era.

So we are now in a situation where we are managing a tension. And the tension comes between the quality of capturing to the system, the quality of the organizational scheme, the structure and schema by which we organize a record system, and the recordkeeping rules that we want to apply. So where we have a system like an email system, like Exchange or an individual chat system like Teams Chat, where we have a massive volume of messages stored by individual email accounts where it is not possible for us to precisely apply retention rules and access rules. We have to decide what's going to give. We can say, oh, well, we don't want this structure, we will have a separate structure. You need to move things: you need to take screenshots of your Team's chat and post them into SharePoint, or Teams Channels, or you need to drag and drop your email into SharePoint. But that won't capture that correspondence consistently. That will have an inconsistent capture of that correspondence into another record system, SharePoint or Teams Channels and therefore we can't square the circle. Yeah, so we either have to apply much simpler retention rules directly to the original

records system for correspondence, and Exchange for email in the in the Microsoft 365 world or Teams Chat for chat messages, or we apply the complex retention rules to a complex structure, knowing that we miss a large part of the business correspondence.

CHRIS HURLEY (at 13.28)

Well, I'm not quite sure what I'm doing here. I'm told I'm to talk about theory. I don't like theorisation because it's so often contrasted with practicality. I like the disjunction between theory and application. I agree with James' history of recordkeeping, although my timing timetable will be slightly different. I go back and I say at the start of it, people made records because they needed them. They didn't make records because they were told to. And I once defined humans as recordkeeping mammals for that reason. It's something that people do. And our involvement in that, I would say up until the late 1960s, was really is tradesmen rather than as theoreticians. We had a few principles and we had textbooks and we had manuals, but we didn't yet understand the difference between theory and application.

When Sue McKemmish gave my exposition on the serious system to David Bearman, he said he was delighted, but he said it wasn't conceptual analysis. I was peeved at the time, but I think I understand the difference now. What changed? In addition to the other things that James has mentioned, the two biggies - computerization, which encompasses a lot of that, but also the disappearance of middle management. Middle management, was the people who kept the made the rules and who made sure they were enforced. So I would place the real emergence of our theory in the aftermath of that, when they no longer have the basis for good recordkeeping. And we had to think it out again from first principles. Paradoxically, that also took us back to the very beginning. And people made records when they needed them rather than when they were told to.

When I joined the bank just at the beginning of this century, I found that their corporate recordkeeping, like everybody else's, was lousy. But they had one area of activity where the recordkeeping was superb and this was in their dealings with customers. And you can understand why that is the case. But they didn't think about that as recordkeeping. I thought about it as banking. Later on, I got involved in a corporate wide email project where they've asked us, and the lawyers and other people, how long do we need to

keep emails for? And we patiently explained that you sentence by content, not by format. That led somewhere. But I actually discovered that the bank had implemented what I think James and his colleagues have called an in place approach. One area of the bank, the part that deals with stocks and shares and trading, had a real need to keep a record of exchanges, which mostly took place by email, by phone and by other communication mechanisms. And what they decided to do, again, without any recordkeeping theory to back it up, was to provide the traders with a little email system, all of their own, which nobody else could use, and they were forbidden to use the corporate wide system. And they said, we're going to keep those emails forever. Well, I don't know if they will keep them forever, but the theory, or the practice, was pretty much what James describes in and his colleagues described in their article, as the approach taken by NARA.

I thought the lesson from that was not that their approach was a good one, but that it reinforces the idea that when people need records, they make them. They find a way of making them despite the technology, not necessarily because of it. And I'm assured that the same thing applies in selected cases: in the blood bank, which keeps perfect records, I imagine land titles, births, deaths and marriages, student records at universities, and as an occasional use of trains, I would hope that the railway engineering records are pretty good. And as a devotee of that TV show, Air Crash Investigations, I'm fascinated to see each time they do an investigation, the first thing they go for is the records, not just the black boxes, but the airline records, which I assume are good records.

You could envisage a big organization like the bank separating out functions like harassment, discrimination, occupational health and safety, establishing specialized applications so that communication would begin and continue separate from the universal e-mail server. If a complaint arose, you would make that complaint in the application. You would deal with it in the application. It would be resolved in the application. Quite separate from the other activities, technological activities. And that would be a basis for in-place management of the kind that was described in the article.

I think there's a misunderstanding about the Pittsburgh requirements. The idea of functional requirements comes from systems design, but that idea is meant to analyze the needs of a particular set of circumstances to design and build on the basis of that,

and then to analyze the result. The requirements identified by Pittsburgh were not like that. They were generic, and they weren't developed with any particular application in mind. It follows from that, that the Pittsburgh requirements can't be used as a boilerplate. And as David Bearman himself once said to me, once they developed, you then have to adapt them to a particular set of circumstances and you're free to select those ones that you don't need and to discard them. That can be a little bit scary, but that's different from the way I think some of the standard making use them. What they did, and I'm not against standards, but what happened was that they amounted to a manual, a one size fits all instruction for how to produce good records. But that is a misapplication, in my view, of what the requirements were intended to do.

And the other problem with standards is you have a problem with auditing because audits, not just in our area, but in all sorts of areas, tends to audit compliance with the standard rather than outcomes. And so that's why people are dying in nursing homes from malnutrition. The nursing homes get a tick for conforming to the nutrition standards, the nutrition standards talk about the quality of the food and the preparation and the delivery. But what happens in these places, and I've had occasion to see this, is the food gets to the person and because they can't swallow anymore, and there's nobody to help them, it stays untouched and gets taken away and they die. But that's because they're auditing the performance against the standard and not the actual outcome. Anyway, that's an aside.

So if theories aren't councils for perfection, what are they? Well, sometimes the theories can be used to actually improve recordkeeping. But above all else, what they do is they form our understanding and knowledge. They are how we relate to the world and they support us when we're talking with managers who actually get good recordkeeping and actually want to do something about it. And you can tell them from the others because they're the ones whose eyes don't glaze over after 30 seconds. They're also a way of enabling us to talk to our allies or people who should be our allies: the lawyers, the auditors, the risk managers, the compliance work and then the I.T. area, the data quality people.

So to finish, they're really our entry into a conversation with people, so we shouldn't abandon theory in order to be able to have that conversation. And the hopeful bit, I would think, is that conversations can change the world sometimes.

ANDREW WAUGH (23.00)

I'm trying something new, I normally free flow, but I tried to write some notes down, so if you see me glancing off a little bit, that's because I'm looking at notes. I'll start off by saying that my views on digital recordkeeping have been developed as a series of clicks. So every now and again, very rarely someone says something and it enlarges my understanding of a topic, particular recordkeeping. And I think in this context, for the purpose of this talk, my first click was at a seminar organized by PROV on investigators and records. We invited both the Victorian Ombudsman and the Auditor General to speak to us about how their investigators used records and both of them independently got up and said to us, and all the records managers in the in the large auditorium, investigators look at the records, but the smoking gun is always in the email. And that brought me up short. I thought, what are we doing wrong, that the important records are actually not in the recordkeeping system? And I've looked at many, many investigations subsequently. And they're right. They always talk about email. That's where the important stuff is.

The second click was by the Australian archivist, Adrian Cunningham, who noted that particular set of 19th century files with the best records he'd ever seen, and that was because the record, the files themselves were the unit of work. A clerk did something on the file, put their stuff on it, and they passed the work over to the next one. The way of communication was actually by the movement of the files around the organization. And that was true, I think, for many organizations even up to the 1950s and 1960s. And that gave me the answer to my question: the important records weren't in the modern recordkeeping system because the recordkeeping system wasn't where people did their work. They actually did it, they worked, in email. These days they'd probably work in Office 365 or in a client management system or some business application. What's in the formal recordkeeping system is actually selected by the staff from the bigger collection of real records. That's why the formal recordkeeping systems are not actually that useful to people like the investigators, because it's not where the records are actually located.

I had a bunch of clicks reading James's paper because there's a whole bunch of interesting stuff about history, connected to what I learnt and understood about records

as well. And I think the fundamentally the distinction between the business system where people are working and the formal record system, is why what James calls the Durante model, where you register in place records in a separate record system, often fails. Because there's no need to make the formal records complete or reliable as the day to day business is still in the original recordkeeping system. For an Australian trained on the continuum approach, what James calls the Durante model is doubly strange because records come into existence the moment they're created in the business system and to move them, not when they're registered in some other system. Business systems, consequently, are always a record system.

We now know also, that Bearman's model where you actually add functionality to the business systems doesn't work either. And the reason for that is that basically organizations don't actually develop their own business systems. What they do is they buy in commodity systems. They used to change them and modify them, create code in them. They don't anymore. It's too risky because as soon as the underlying system changes all their investment's out the window. What they do now, from our observations, is that they buy in commodity systems and they configure it, to change its behaviour, and look and feel slightly, but they can't change the actual underlying behaviour of an organization. So if we are going to say to Microsoft, put this functionality in, it's useless because unless they can fit neatly into the underlying model, they can't implement it. They can't put it in. And I think that's the problem, even though people are actually talking to Microsoft at the moment.

So that leaves James's third approach. You work with the records in the original business system. While you may not have the functionality you want in the system, they all have sufficient records functionality to be used as the business system. That's why they've been selected. Email a terrible record system from our perspective, but that doesn't stop organizations working in it and it doesn't stop investigators basing their investigations on email. And it certainly hasn't stopped court cases based on e-mail. It's a perfectly adequate record system from their perspective. Pretty much exactly what Chris was saying.

I think, and we think, that progress is to let go of the regret for functionality that's just not there and exploit the functionality that is in those systems. As an example of this approach, over the last year, we've done some experiments in PROV on processing

emails. And again, that echoes the ideas in James's paper. We took all of the available email for PROV staff for three years: a small organization, about 60 staff, but that was still about two hundred and fifty thousand emails. Would we do anything with it? Well, the key problem simply the size of the collection is too big. So first of all, what we did was we removed the duplicate emails. So if I sent an email to a fellow staff member, a colleague, there'd be two copies in the collection. So we get got rid of one. By doing that, we eliminated about 40 percent of the emails, getting it down to about one hundred and forty three thousand unique emails. Then we threaded the emails. So when you hit reply, or reply all, metadata is put in the email to tell the system what emails it's replying to. And so you end up with a chain and you see this if you ever use Gmail. That's what Gmail does. Threading turned out to be fantastic from our perspective as record keepers, we reduced the emails from one hundred forty three thousand to eighty thousand threads, about 32 per cent of the original collection. That's a significant reduction, and a consequent reduction in the management work. The threads are actually better than the original individual emails. Threading supplies the context of the email: what's the response to what and how have people responded to it? An email that wouldn't have been selected for preservation is often selected by threading because it's linked to a selected email. So you don't have to have every email having all the good metadata. And finally, threads match what users want, they actually want they want to find an individual email, they understand the story in which it happened. And finally, in the future for AI, threading increases the information available to the system.

But then we looked at it and said, what have we done? The first rule of archiving is to respect the original order. If we think of email, isn't the original order structured by user, then, by however the user structured their mailbox. You don't see any of that in what we've done. We've actually completely eliminated that. There are two answers to that, and both are relevant to the larger questions of handling digital records in systems like Office 365.

The first is the digital records have many original orderings. They're not like paper. Paper had one ordering. Digital records have many, many different orderings. Some older things are highly visible, like the user mailboxes, and some are implicit in the metadata and the carrying out of the work, such as the email threads. As archivists and records managers we can choose the ordering we use depending on the use of the records. If you're Melbourne Uni archives, for example, and you've got the Germaine

Greer collection, which includes her emails, it's perfectly natural to present that, and structure and handle it as a mailbox because that that helps explain who she was and how she thought. But in a government archives like PROV, well, we're not interested really in individuals and more interested in how the organization collectively constructed business. And that's why we don't actually like the Capstone approach very much, because that's all about individuals. And the threading organization emphasizes how people carried out the work.

That second answer is there's no need to choose one ordering anyway. There are many different ways of ordering emails and we can actually work to de-duplicate your management using threads and then present the original ordering again back in the emails. What we've done is to use the metadata and functionality in the original digital recordkeeping system to perform records management functions on the record, just not the way expected by classic paper-based record keeping systems. We think that the key to managing records from digital systems is to focus on outcomes and to look for mechanisms that will achieve those – not focus on how we achieve the outcomes in paper based record systems.

As an example of this approach, the Australasian Digital Recordkeeping Initiative recently published guidance for records managers on procuring software as a service system, such as Office 365. We really tried to focus on outcomes, not on traditional mechanisms. For example, we ask the records manager to consider how the system arranges and describes the records, not whether it's implemented in a business classification system.

So the key ideas I want to leave you with are:

- The best records are in recordkeeping systems in which the work was done. Moving records out of those systems into a formal record system degrades the fonds.
- The strength of the digital recordkeeping systems is in the range of metadata associated with each record. Exploit this strength to achieve record goals and think laterally about how to do this. Don't consider them as paper recordkeeping systems from the second half of the 20th century. It doesn't work.
- And finally, don't confuse aspirational records practice with actual business requirements. Whatever rudimentary record management is present in the

system may be perfectly satisfactory for the risk appetite of the business. So focus on the things that you need to do as records managers that the business is not going to think about like retention and disposal.

NICOLA FORBES (34m)

The approach that I thought I'd take when thinking about this talk, and James touched on it at the beginning of his talk, so I guess we're looping back in a sense, is to consider the post digital era. A lot of people consider we are now in the post digital era. And there are two ways that I've seen that defined. One is where the digital is so much part of our systems, processes and the way that we live, that we no longer notice it. So it becomes unnecessary to mention that digital in the sense that it's just part of what we do. The other way that I've heard it defined is where things such as record players, audio tapes, all of those technologies that you would see as being analogue technologies, become something that's retro, something that becomes cool because it's old. If we look at the eras that James has spelled out and we think about the post digital era, I'd like us all to think as professionals, as recordkeeping professionals, as archivists, as information governance people.

The image that I have in my mind is taken from science fiction movies, so you'll have to bear with me. But we're leaving one ship, and we're going into the airlock, and we're going to come out on the other side in another ship. What are we going to take into that airlock with us? So what are the tools that we need to take from the digital era into the post digital era? What is going to be useful to us? What theories are useful? What standards are useful? What ways of practice are useful? To me, I feel as a profession, we're standing in that airlock at the moment because the change in the digital economy has been so rapid, and continues to be so rapid, and has so many impacts on society. What we see as the digital era, which was about conversion of format, digitization activities, building repositories, working at how to deal with interoperability, all of those things. What do we need to take into that post digital era and what defines that post digital era? And if we look at other disciplines, like sociology, for example, they're all grappling with this as well, but they're using particular theories (and I'm going to use the word theories there) and keeping those theories, and keeping those principles, and

keeping what's important to them as a discipline and as a profession, and defining those in order to look at how they will deal with the post digital era.

And to look at it in more detail, I thought I talk a little bit about what I see the post digital era being and to use some examples. So before my current job, I worked for an organization that was given billions of dollars to build integrated transport systems for a state, and it was the biggest spend ever given in Australian history. Now, that was such an eye opener to me, because certainly I was now dealing with technologies such as automated vehicles. Dealing with trains, for example, that create and generate their own data. They use that data to look at passengers, to look at safety, to look at running the vehicles. It is all built up in the vehicle itself and the systems that support the vehicles. And it makes you think, well, how do you transfer custody of that information and data? Can you transfer custody? The data and information was held by a Spanish company in this case. And it was very proprietary around that vehicle. Another, area where you can see it, is if you look at Twitter. So it's not the statement. Let's take Trump, for example. It's not the information statement that's important. It's also what's behind that. So it's the likes, the dislikes, who he's friends with, what happened, what people said. And I think that's part of the post digital era. And it's that definition by big tech. They define what becomes a record. They create other information that becomes a record. The record takes on new meanings. That becomes something that can be used to test customer tastes, can be used to define individuals.

How do we grapple with those sorts of issues, is my question. And with Office 365, you see it with the debates coming around recently around the data it was keeping on individuals in the workplace, which Microsoft has now announced it will no longer do. But when I first started using it, for example, it tracked when I used Teams, when I used emails, what other apps I used. It gave me suggestions - would you like to up your Teams use, it showed when I logged on, when I logged off. And so the record itself is not actually the content that I'm putting out by that chat or that tweet or that email. There is actually underlying data which sits under it, which itself is a record of something else. So some other aspects that I thought of around what I consider to be the post digital world, is that concept of intimacy and digital intimacy - what the data knows about you and how it can predict what you do. How do we grapple with those things? What are the principles and what are the theories that we're packing into our airlock which will enable us to grapple with that post digital world?

And I think that the speakers today have really helped with that. And I think there's a need to be positive. I think the technological determinism of something like the tech giants, which you see with Twitter, you see with Office 365, you see with a lot of products, makes people despondent in our profession because you think this has already been determined for us. We're no longer those intermediaries which say what a record will be, and how it will be captured, and what will be kept. We're sort of become like, I guess what everyone feels, where's our agency. Are we just part of this tech universe where the tech giants determine what is kept and how long? An example of that, I guess if you look at the way that Office 365 sets up what emails are kept, and particularly their rulings around seniority, and being able to set certain rules - certain rules for junior staff, different rules for senior staff. Well, why are they defining that? And what is the meaning of that in a world which has become increasingly collaborative? Where the concept of seniority is still there, but it's changing.

How do we have those discussions with technology companies on those issues? What is the language that we use to have those conversations? How do we make ourselves heard? And I think for a long time it was you need to talk to your manager or above. You need to make yourself heard at the executive level. And in a sense, how easy is that? How easy was it then when that was all we needed to do? How do we have those conversations with someone like Microsoft? How do we have those conversations with all of the other technologies that are coming out?

CASSIE FINDLAY AND KATE CUMMING COORDINATING QUESTIONS (42m)

Cassie: I've been just really gripped by all of those talks, really so thought provoking and chiming with a lot of my experience in workplaces over the last five years. So thank you so much, everybody. I guess one sort of one question that occurred to me, and I think we touched on this a little bit, you know, in both government and the private sector, everything is 'as a service'. That is how things are progressing. In what ways (and I might start with James) would you say should recordkeepers consider that in devising recordkeeping strategies - the everything 'as a service', whether it be software platform infrastructure?

James: Can I knock it back to you, Cassie. I mean, what do you think?

Cassie: So in my experience, the lawyers just get involved and there's a lot of contract stuff - there's a lot of contract negotiations and then there's a lot of discussions about what is going to happen to our data at the end of the contract - if we're lucky. Sometimes those conversations don't happen, but that sort of skews the whole recordkeeping conversation so that the business finds their requirements and if records and information professionals are brought in, it's typical to say, OK, at the end of the contract, just delete all of our data and or give us a copy of our data. And then there's sort of a bit of a disconnect with how organizations have been recapturing that information, if at all, at the end of the engagement. So I don't know. That's my experience.

James: I worked on a complex construction project once and I realized that this construction project, which was using massive data sets, really complex data sets. At the end of the construction project, it would have all of these data sets that it would want to hand over to someone or it would want to keep for an accountability period. And yet it wouldn't want the applications because it's not building anything anymore. So then you have this problem of what do you do when you want the data, but you don't want the application. And the data is super complex. It's not just documents with a bit of metadata. It's a complex dataset.

Cassie: Yeah, and I think also there's this question, as Nicola alluded to, that you are kind of then being defined by that provider's structures and functionalities. And is that a problem or not as far as the records that are produced?

James: I mean, as a profession, in terms of strategy, in an ideal world, we'd be able to move data from any application to, a generic, perhaps digital preservation system where we could just say, OK, we don't want that contract with that supplier anymore. We do want the data. We import it into our digital preservation system. And then when we get a replacement system, we might move into the application. But that requires such a complex digital archiving skills, doesn't it?

Cassie: Yeah absolutely, and those are those are the kinds of problems that digital archivists and preservationists are facing. And I think there's an awful lot that perhaps this new era can learn from the lessons of digital archiving and preservation. This

department that had this ancient system which has been decommissioned and now we have to ingest as an entirety that recordkeeping system. How do we go about doing it? How do we what what's the toolset we need? etc etc. And I know there's lots of knowledge from Andrew who's with this stuff. So but I don't want to I don't want to take up all the time because I think there's probably a bunch of other questions. And this this is Nicola's stuff, too. So I want to say something.

Kate: Andrew, a couple of questions about the tool that you are using for your threading technique?

Andrew: We wrote it ourselves – well I wrote it actually. Actually, it turns out to be quite the threading is quite a complex process because there is the standard, of course. But not all the vendors, and I'm looking at you Microsoft, actually follow the standard. So it's complex. It works. It works very well. And it's interesting, Cassie was talking about digitally archiving, because one of the things we did was we actually turned the threads into records and so we could put them into our digital archive. We haven't done that yet. We structured it. And one of the things that occurred to me at the end of it was one of the reasons we were thinking in terms of that structure, was that the techniques we used were that we can actually ingest structured objects into our digital archive as tree structures, which threads failing to naturally. We had the tools that we could use to do that. So it allowed us to think in certain ways. The threading was very interesting, a lot of fun, and it was surprising just how powerful it was when you looked at the threads. All sorts of funny little things came out. So, for example, one of the things that was a surprising number of what I called singletons. So emails that were sent or received from the organization but never got any responses. And what that meant often, was that it was a boring email. It was like the someone was sending you a conference invitation or sort of seminar or it was the administratrivia - are we going to have a staff meeting or something like that? The other interesting thing with threads was that often very long trips were very boring. The longest thread we had was 37, a depth of 37, a huge wide spread that was all about purchasing office chairs. So, you know, it's interesting to hear what threads actually meant and how you could deal with it. One of the really, really interesting things about threads was we looked at it and eventually we worked out that it looked exactly the same as top numbering. And if there are archivists around, that's a classic 19th century correspondence management technique before vertical files you top numbered things. And it was really like, oh, wow, just back to a future.

Cassie: the next question is, and this is a sort of a provocative conversation starter, which I think Carla in the chat, asked. How can Microsoft accommodate all of the international variety of recordkeeping rules and requirements?

Andrew: The answer is, they won't. They'll provide a product and it's up to you as an organization, whether you buy it or not. And the answer at the moment is that organizations around the world think it's a great product. And the problem is and why we're having one of the reasons we're having the conversation now, is that then the problem comes down to people like us. And they say, at best they say to us, how do we do this in our environment. And we go, oh, we can't. And the businesses say, oh, OK, we just do it anyway. With the exception of certain things that are very, very highly regulated. And Chris talked about bank finance, for example. And I think it was Nicola who talked about transportation records, particularly aircraft - things like that are very, very highly regulated. And so they start off very strongly. But for everybody else, I think that what will happen is that they'll just ignore them and you know it. If shit hits the fan, then the courts will say, that's OK.

Cassie: Follow up question for me, when there are higher risk factors with the associated with the recordkeeping such as aviation, even university student registry systems. There is this closer connection between the work and the record that you talking about, Andrew. And then when sort of general administration, we rely on something that is less rigorous. Is that is that the natural way of things? Is that just how things should fall out about change?

Chris: So I didn't hear what Cassie just said, I'm sorry

Cassie: I was just saying that points have been made by many speakers this evening about how the connection between the business, the work and the record is typically closer. And in fact, it is the same thing, in business in which the risk levels and the requirements are more obvious. So whether it be aviation repair systems, I think you make a point that that sort of nexus. But then when that the work becomes more generalized, the connections is lessened. Is that just the natural way of things? Is that kind of recordkeeping correcting itself in some way? I don't know. You know what I mean?

Nicola: I think we also need to think about that the post digital area is also that lack of transparency. So the business, or the user, doesn't actually know what's going on behind the scenes. So an assumption that they're making a decision to make that, they know how long things are kept anymore, is actually not the case. So how many people here know what the retention rules are built into Microsoft Office 365. And if we don't know that, can you expect the business to know that? There's no transparency as to how the system and the configuration of the system is built? And so that question around transparency and visibility around how information is managed becomes important? Or I think it is anyway.

James: And it's worse in Office 365 as well, because the way that Microsoft Office 365 decides which retention rule applies is really messy and subject to some triage rules, which it's really hard to know if you if you labelled an individual item, whether that's actually going to be the rule that will apply to that individual item because other people could label it. And it's got the default rule from its from its aggregation. So it's a really messy way that they've got to apply the retention rules. So I completely agree with Nicola. Individual end users, especially in an environment like email and Teams chat where they've got something else at stake, which is their personal information as well. How do we how do we make sure that they're aware of what retention rule is by default will be applied? And how do we get at their contribution to their input. Maybe Andrew's come up with this, because obviously you've intervened in your email system and you've done your threading. How does the individual react to it in the archive?

Andrew: Well, that's very interesting because we had to be very, very careful with it because this is only experimental. And we actually did quite a bit of work with the organization to say, look, we are looking at this stuff and we are looking at your e-mails and this is what we're doing. And we had to be very sensitive to the fact that this was personal information and the staff were actually surprisingly good about it. They were quite relaxed about it. And I think one of the reasons for that was we were actually very upfront about what we were doing. We weren't going to be looking at individual emails and we were looking at things in bulk or looking at things that subject lines. But we had to be very careful about it. We had to be very sensitive. One of the curious things about it is that I think that that's one of the reasons why AIs and emails, and AI and all of these systems, are actually going to be really hard to build. Because the thing you need to

build AI on is a training data set and what training data you're going to use. I mean, you're going to use perhaps PROV training data. Well, are we going to make that data available to some random organization that look at the stuff? Hmm. Looking in the broader government community and commercial viability. Well, we've got the Department of Premier and Cabinet, we'll just pass over all of their emails to a private supplier. I don't think so. It's actually getting really hard to get generic training data for all of this sort of stuff because the information is so sensitive.

Barbara: Discussion is the tip of the iceberg. You can hear that we could go on for many a long time and I'm very sorry to have to say that we're five minutes over an hour already. So I guess I'm forced because of the pressures of time to bring us to an end. But we do have a recording of the session and we will also save the Chat. I think the Recordkeeping Roundtable will probably blog on this issue. And I so I hope that this is a spark for a conversation. And James's article was a spark for us. So thank you, James, for doing such a wonderful job in creating an environment which we all would like to speak to. And I also really want to express my thanks to Chris, Andrew and Nicola, who were cajoled into doing this. I really appreciate it. I'm sure as our audience does, too. Behind the scenes, we had Cassie and Kate, Vivien and Gillian was an early provider as well. So thank you so much to all of the people involved. And unfortunately, I'm going to have to close it off there.