



Transcript provided as a free resource by



Data Mesh Radio Episode #276: Making Self-Service Actually Work Well Safely

Interview with Kate Carruthers

Listen ([link](#))

Transcript provided as a free community resource by Starburst.

To check out more Starburst-compiled resources about data mesh, please check here: <https://www.starburst.io/info/data-mesh-resource-center>

To get their Data Products for Dummies book (info gated), please see here:

<https://www.starburst.io/info/data-products-for-dummies/>

To get their Data Mesh for Dummies book (info gated), please see here:

<https://starburst.io/info/data-mesh-for-dummies/>

List of all previous episodes (including links to available transcripts) [is available here](#).

The folder of all previous transcripts [is available here](#).

0:00:00 Scott Hirleman

A written transcript of this episode is provided by Starburst. For more information, you can see the show notes.

0:00:07 Starburst

Welcome to Data Mesh Radio with your host, Scott Hirleman, sponsored by Starburst. Starburst is a single platform to help you activate all your data, no matter where it lives. Check out our new Data Products for Dummies eBook to learn more about how your organization can utilize data products. To download your free copy, head on over to starburst.io.

0:00:29 Scott Hirleman

Data Mesh Radio is provided as a free community resource by Data Mesh Understanding. It is produced and hosted by me, Scott Hirleman. I started this podcast as a place for practitioners to get useful information about data mesh. We're at over 200 episodes. I've now left DataStax, thanks for all their help in founding things, but I've left to start Data Mesh Understanding, which is also helping practitioners to get to the information needed to do data mesh well. We have free implementer introduction and roundtable programs, in addition to the more advanced yet affordable offerings, so please do get in touch if you're looking for more information on how to do, how to approach data mesh. Just check datameshunderstanding.com for more info. There's also a helpful organization of past Data Mesh Radio episodes there if you want to dig into specific topics rather than digging through 200 different episodes. So with that, let's hit the funky intro music and listen to what you'll hear about in this interview episode.

Episode 276: Making Self-Service Actually Work Well Safely. Bottom line up front,



Transcript provided as a free resource by



what are you going to hear about and learn about in this episode? I interviewed Kate Carruthers, who is the head of business intelligence at the UNSW AI Institute and the Chief ta and Insights Officer at UNSW, or the University of New South Wales. To be clear though, she was only representing her own views on the episode. UNSW is not currently implementing data mesh, but Kate and I talked about this. She's preparing to be able to implement data mesh. I think this is a great lesson in building up the capabilities to move forward towards your goals, but not rush. They're heading towards a data lakehouse architecture to start before they really want to start heading down the data mesh path. So here are some key takeaways or thoughts from Kate's point of view.

Number one, universities can teach us some really interesting perspectives on self-serve. Because universities are such complex organizations and so many departments are involved in deep investigations and research into very specific areas, they really are the only domain experts inside the organization, right? If somebody's really, really focused on one very narrow thing for their PhD or as a professor, they're the only ones that are really those domain experts. So enabling them to even own their own data can be very challenging, let alone helping them to understand how to share their data with others safely.

Number two, relatedly, each academic researcher is essentially a micro domain themselves with their own ways of working. This keeps coming up about kind of coinciding with people's ways of working while each researcher has their own. That just adds to the need to enable freedom in ways of workings, but still keep them, you know, as Kate said multiple times, keep them safe. And safety was a key theme of the general conversation. Number three, "At the end of the day, data mesh is about controlling the bits that you need to control and giving people the freedom to do what they need to do safely." Number four, "Technology is kind of the least of your problems. When it comes to data, be prepared to start with some people not even recognizing there is a problem with the current ways of working or a need to improve. Connect their pain to data immaturity to win them over. For sure, they've got some pains. Find out what those are and then connect them to the data immaturity."

Number five, the best way to win people over is show, don't tell. Show them the power of self-service instead of pitch them on it. Get a POC going and get people to tangibly see and hopefully, relatively quickly touch your self-service capabilities early. Number six, always look to anchor your data work, especially things like platform work, to a business need. How will doing the work impact the overall business, the overall organization? Why is it important to do and to do now? Number seven, when tying your data work to the overall business strategy for your organization, do not forget the people aspect. The relationships matter. Your work on the data team



Transcript provided as a free resource by



definitely isn't only about technical execution. Get that out of your head. I know it's the easiest, it's the most tangible, but get that out of your head. Number eight, potentially controversial, build a culture around data that is as focused on building human relationships as it is on building data pipelines and platforms.

Number nine, another potentially controversial one, to share personal or sensitive information such as PII, a producer should justify why it's appropriate and a data controller should review that. Keep humans in the loop. Stop trying to make all of the governance aspect of sharing sensitive data automated. It's a recipe for getting yourself into trouble. Number 10, giving data owners, and this is what UNSW calls data controllers, but giving those data owners a say in how their data is actually used, can get them more excited to share their data. It isn't a silver bullet to data sharing incentivization, but it adds value to them. They get to say, how should this be used or how shouldn't it be used? Number 11, good conversations about access to sensitive data shouldn't be yes or no, right? It shouldn't be that simple binary, oh, we can't do this, right? They're about getting to what is acceptable and maximizing value within that framing of what is acceptable. Get people to share what they're trying to accomplish and then partner to best achieve what they're trying to accomplish.

Number 12, invest in business analysts. They are your front line to figuring out how to proceed around data and generate value around your data. You need people who can speak business and data simultaneously to drive to great outcomes and business analysts, that's kind of the exact rule, their business and their analysts. Number 13, find ways to prevent data puddles, especially places where people are copying data and then not securing it well. You know, one, you don't want to have all these copies of data for many, many different reasons, but that not securing data well is a big potential problem. Number 14, "People overestimate the power of making change really fast and underestimate the power of sustained incremental change." Number 15, somewhat relatedly, give people a mental map for change, let them understand what's changing, why, how. It removes the fear of the change and it lets them lean in. You're creating change now with and through them instead of pushing change on them.

Finally, number 16, potentially controversial, ChatGPT and other related GenAI can actually be a great benefit to education. We have to lean into it as it's not as though students won't have access to these tools in their work life and their real life post-education. So getting them to still learn in general, but leveraging these new emerging better tools is essential to their progress as they progress through their careers and lives. Okay, enough of just me. Let's hear from our awesome guest in this interview episode.

Okay, very, very excited for today's episode. I've got Kate Carruthers here, who's the



Transcript provided as a free resource by



Head of BI for UNSW AI Institute. To be clear though, she's only representing our own views. And we're going to talk about UNSW's journey towards thinking about data mesh and why they're kind of taking a stop along the way and not trying to jump right into those data mesh waters, go for the lakehouse first. We're going to talk about how do we actually think about exposing data to people without having to do the work each time? How do we get people into such a mode where they can actually fend for themselves? But how do you also limit risk? How do you think about creating that as a scalable approach instead of, again, having to do kind of one-off, one-off, one-off? Just the general concept that I talk about this with data mesh a lot, not everything should be decentralized. How do you figure out what should be and what shouldn't? And then also, Kate's got some really good perspective on how GenAI is starting to be used in universities. And so I think that's a topic that a lot of people are very interested in. But before we get to that, Kate, if you don't mind giving people a bit of an introduction to yourself, and then we can jump into the conversation at hand.

0:10:47 Kate Carruthers

Thanks very much, Scott. It's so nice to join you. So I'm the Head of Business Intelligence for the UNSW. That's the University of New South Wales AI Institute in Sydney, Australia. And I'm also the Chief Data Officer for the university. And the AI institute is all about inventing the future of AI, which is really exciting to be part of. It's a research institute. And as in my other part of my day job is Chief Data Officer of the university, which is actually managing data as an asset across the entire organization. And how I came to that was, I had a number of roles in ICT in big organizations, big global multinationals, and I did a lot of data projects.

And I suddenly realized about 10 years ago that we were about to undergo a very big digital transformation and that data would underpin it as an organization. We would need to be on top of that data if we were going to be able to ride that wave. So that's what's really been driving me for the last 10 years, is this journey to get our data into a place where we can actually undergo that kind of digital transformation.

0:12:00 Scott Hirleman

And I'd love to hear even a little bit about what is the data of a university, right? Is this just on the actual students and things like this, or are you helping the departments manage their data around their research? How does that actually come about? If we could jump into a little bit about that day-to-day job.

0:12:21 Kate Carruthers

So I always think of the university data in three realms. So the first realm is the realm of administration, what you would in, if you're in a bank, you'd call it enterprise. So the running of the place as a business. And we have the same kind of data that any



Transcript provided as a free resource by



other kind of organization has. You know, we've got finance systems, we've got student systems, we've got CRMs, we've got all of them. And the second realm of data that I think about is learning and teaching, which is kind of similar to enterprise data, but it's specifically around the teaching of students. So they're specific custom systems that are effectively our custom enterprise systems. And then the third realm is research, which is the pure research for the, you know, finding new information and discovering new things. And the way I always describe that is we're inventing the future.

So there are three different realms and sort of the first two, the administrative data and the learning and teaching data are kind of like enterprise data, but research data is kind of really wild, is there's every single kind of data that you can imagine, from we've got satellite dishes on top of our buildings where we take feeds from satellites. We do climate research, we do medical research, we do petroleum research, we do all kinds of research you can imagine as a large research intensive university. So they're the three realms of data that I always think about for us as an organization.

0:13:58 Scott Hirleman

Yeah, I think when you started to think about data mesh, was it, you know, I mean sometimes those organizations when you talk about that kind of enterprise data, the administration and the learning and the teaching, maybe there are enough domains to say, maybe we should look at this, you know, or this, or is it that the research is really pushing you into that realm of data mesh where you give people that self-service capability? What made you start to think this is something we should look to in the future if not immediately jump into right now?

0:14:32 Kate Carruthers

Well, first of all, I'll just explain, I always clump the first two, the administrative and learning and teaching into one sort of bucket because they're very similar and research into its own bucket. And there's really good reason to consider data mesh for the future of both of those, so both of those areas. And indeed, we're actually contemplating what kind of research or workbench we need to provide in the future right now. So we're starting to think about a project to conceptualize what we can do to provide a workbench for researchers. And the challenge you've got when you're providing something for researchers is they're all individuals, they're like small sharecroppers who run their own small farms. And so there's a lot of them and they have a lot of individuality and they don't want that to be crushed. And so that's a real challenge to manage that and allow them to flourish the way that they need to, but also to keep them safe.

And in the other two realms, the administration and learning and teaching, there's even more good reasons to move towards a data mesh because realistically at the



Transcript provided as a free resource by



end of the day, a data mesh is about controlling the bits that you need to control and giving people the freedom to do what they need to do safely. And I think that that's a really attractive thing for me to contemplate for both kinds, for all those kinds of data, because that's what we really want. We hire smart people and we want them to be able to do their jobs and we want them to be able to use data to drive important outcomes, but we want them to be able to do it safely. And we don't want things like data duplication. We don't want, I keep calling them data puddles. We don't want them to be creating their own little data puddles that are insecure and stuff like that. So we really want to be able to provide safe environments for people to do their jobs.

0:16:35 Scott Hirleman

I mean, and you mentioned the word safe quite a bit, which I think is an interesting approach here. One question I would have, I mean, I think on the administration and the learning and teaching side, that interoperability is far more, it's kind of more of an obvious thing as to why you would want that and all that. On the research side, are you seeing that different departments really would like to collaborate and this is what you're trying to unlock for them? You talked about kind of unlocking the future and designing for the future. Is that something where, or is that getting way ahead of ourselves of like, hey, I don't know that this team really needs to collaborate with this other completely array on the finance and...

0:17:18 Kate Carruthers

No, no. It's all about collaboration and its collaboration across the organizational silos, you know, we call them faculties and schools, but it's people doing cross-disciplinary, trans-disciplinary collaboration now. I might just tell you a story, because it was about 10 years ago when I first arrived at university and I was doing a research project with some folks in the School of Medicine. And we were doing some work with indigenous data, so our indigenous people are called Aboriginal and Torres Strait Islanders and we were doing some research about their health. And it was very sensitive data and I said, "Where are we allowed to store this? How are we supposed to look after this data?" And the everyone was going, "I don't know, I have no idea." And so we've done all this paperwork, we've done ethics approval, we had all of this stuff that we went through and at the end of the day, nobody helped us work out how to secure that data. So we knew what we were wanting to achieve from the research, we knew how to do it.

But one thing we didn't have as a research group, you know, a bunch of doctors, a bunch of PhDs and post-docs, we didn't know where we could store that data safely. We didn't know how we could de-identify it in an adequate way. And there was a whole lot of questions we had. And that's what really started me on my journey into becoming the chief data officer for the university.

0:18:42 Scott Hirleman



Transcript provided as a free resource by



That's great. And I think a lot of what you're talking about is the something that I'm seeing some people skip over, which is jobs to be done. Like, what are we actually trying to accomplish? Not just, wouldn't it be great if we could have all this stuff completely unlocked versus hey, let's talk about what this could unlock for us of the, exactly what you're talking about with doing this stuff safely. Oh, it gives us the ability to open up all of these completely new vistas instead of just like, hey, we wanna be data-driven, right? It's hey, we're actually trying to accomplish these X number of tasks or whatever, and that this is that approach. I'd love to hear why you decided to, I fully agree with your decision to do this 'cause I tell people, don't jump into data mesh too early, but why you decided to go in the data lakehouse route before you really started to go, okay, we're going full data mesh.

0:19:45 Kate Carruthers

Well, at the end of the day, it's really about the technology is kind of the least of your problems. More often you've got the real problem about the people that you're dealing with not having a mental map of where you're trying to go on this journey. They don't even know there's a journey when you start out. So you need to actually get the people, the process and the technology lined up. So we started our journey back in 2018 when I had this vision for a serverless data platform. And I mentioned it to a few people around the university and they were like, I don't know what one of those is. And I was like, so I had to build them a proof of concept that would show them, and the thing that I wanted to demonstrate them was the speed.

So we've had a legacy data warehouse for more than 20 years. So we're very early adopters of data warehousing technology. It's been an on-prem data warehouse. But I could see that we would get a lot more flexibility if we migrated to the cloud. And so the first step was getting people to buy into the idea that we should move to the cloud. And the big thing that was troubling all my business colleagues was the fact that it took six months to onboard new data resource and build the first report. So I was able to demonstrate to them, in three weeks we were able to onboard some new data sources, build some new dashboards and have them ready to demo within three weeks, which blew them away. And so they were able to say, oh yeah, I can see the benefits we will get from this.

So cloud wasn't the destination. Cloud was, how do we deliver some business benefits to you? So my team with the ball there, they were slow because the technology was slow, but if I tried to lead with I wanna do the technology thing, it wouldn't have mattered. So it was a matter of finding the business benefits. So we migrated to the cloud in 2018 and it was a really good project and roll on a couple of years, in 2021, 2020/2021, we were doing our first machine learning proof of concept. And one of the problems that we found was that we had to keep pulling the data out of the data lake, out of the data warehouse. So to run the machine learning, we had



Transcript provided as a free resource by



to pull the data, we put the data in and we'd have to pull the data out again to run our machine learning pipelines.

And it was like, we need a new architecture, we need a more, I described it to my team when I told them to go and find us a modern architecture is, data warehouse is last century's technology. What's the 21st century's technology for us that will enable us to be more efficient and will allow us to drive from our same pipeline? I want one pipeline that can drive our BI analytics and reporting, our AI and ML and our services because increasingly our data is being used to drive low-code services. And so, that was the mission that I gave the team to go and work out what a 21st century architecture that would enable that would look like. And so they came up with that. They went around, they did their research, they came back and pitched me the data lakehouse which I'd already kind of identified as the way we should probably go in the future. So we did it and we're just finalizing that migration right now. And it's really streamlined our operations. It's kind of happened under the covers. So we haven't really articulated what we've done to the business. It's kind of a plumbing thing from our perspective because they're still getting all their reports, their reports still work the same, they're all happy. But now we can drive our major ML pipelines and AI pipelines as well as our low-code apps, which are increasingly putting demands on us.

0:23:51 Scott Hirleman

That makes sense. I think I've talked about this a little bit with my unicorn farts theory, which is kind of starting to fade a little bit away, but I've talked about if you're gonna go and talk to the business people and you're gonna talk all about data mesh, you might as well just call it unicorn farts because they don't care. They don't care. We keep trying to take them on the sausage factory tour and we're dragging them, offer the sausage factory tour. If somebody really wants to know the plumbing and the how, great, like they wanna lean in, they wanna understand, great, but they wanna know like, what is this enabled? What can I do?

0:24:25 Kate Carruthers

They just want the sausages.

0:24:27 Scott Hirleman

They wanna know what's in them, a little bit about the like what flavor am I gonna get and things like that, the quality level and stuff. But yes, they're there to eat, they're not there to learn how to make sausages. So it's funny that you're doing this kind of in the background. I'd love to hear a little bit about how you did get that buy-in. Like who did you have to talk to because there are a lot of people that are struggling to do this. I know this wasn't exactly what we were planning on talking about, but this is fascinating. And so like, 'cause people are going out there and are



Transcript provided as a free resource by



having the conversations that aren't being nearly as fruitful as what you're doing. So, was this just kind of, hey, we're doing this or did you have to go and get a lot of approval or did you just kind of have that?

0:25:16 Kate Carruthers

We are not part of the IT department, but we comply with their practices. So, we comply with all of their architecture review practices and stuff. So we had to go to the architecture review board and present what we were gonna do and explain what we were gonna do and get all the architects to buy in. But one of the things that we always do in my team is we anchor it in a business need. So I keep telling the team, if you're doing anything that you can't anchor to a business need, then why are you doing it? So if nobody needs it, why are you doing it?

So for us it was we've got a big AI project in the student experience domain that's running. And for that to work effectively, that means that we needed to have our pipelines running so that we could support both BI and AI and from a single pipeline because we've got a really small team. So that was the imperative for us. How do we do it with existing resources and how do we do it? And the other thing that we've got coming up next year is we wanted to start to democratize access to our data. So we want people to be able to self-help data in a secure way and straightening out our pipelines and making the data warehouse go away was a big part of that. So we can now expose the data to people through our security model so that they'll be able to self-help for their data needs, which is long. It's been a long a dream of mine to do that, but it's taken years, you know, it's taken since 2018 to get all of our ducks lined up so that we can actually do this.

0:27:02 Scott Hirleman

Yeah, I think that Benny Benford, who is formerly the CDO at Jaguar Land Rover talked about their transformation journey and it was like it was a six year thing, right? I think a lot of people are looking for this to be like historical IT projects versus transformation needs momentum and trying to do a huge, huge shove in that transformation. You know, it might work, but a lot of times it's high risk. So I wanted to, I mean, I wanna give you space to react to that as well, but I want to dig in a little bit on what you talked about of anchoring to business needs. 'Cause I'm more of a business person than I am a data person. And so, yes, of course I fully, fully agree, but how are you working with your team to help them understand how to find those business needs or anchor it to the business strategy? We had a panel on tying the data work all the way up to the business strategy via the data strategy and kind of up and down as to make sure your data work is actually tied to the business strategy. So I'd love to hear how you're working with your teams to know that.

0:28:12 Kate Carruthers



Transcript provided as a free resource by



Formal stuff, the formal joining of strategy up and down is one thing, but it's actually the human relationships that really matter. And so one of the things that I say to my team is, I'll never be cross with you if I see you out having coffee with our customers. So they've all built really good relationships with their business colleagues and they go and have coffee with them, they have lunch with them, and they've built solid relationships, which mean that they can have meaningful dialogue with them about what they need. So we are hooked into what our business partners actually need to help them deliver on their strategy and their jobs. So I think that's the important thing 'cause the formal bit's the formal bit, but at the end of the day, it's human beings.

0:29:04 Scott Hirlleman

I mean, are you recruiting for people that already think like that? Or are you training that just because I know this has been, this is something that kind of came back from Big Data London of data people building data things for data people instead of doing data work for the overall organization and building tooling for the overall organization for people to use. So like, are you just finding that it's a pretty simple conversation or do you have to kind of break people from some historical bad habits?

0:29:33 Kate Carruthers

No, no. So, I think we've got a really strong normative culture now. So, our team culture is so very strongly set now that when you come in, you get absorbed into it or you go away very quickly. So you are there, you come in and you go, oh, I don't like this. And turn around and go straight away again, if you don't wanna work like that. Otherwise, you tend to just adopt our ways of working and our ways of working are written down but they're all so strongly enculturated. So you're in no... There's no way you can not understand the imperatives in our team. And the teams are just hilarious. They run themselves like, we run it kind of democratically. So, they get to set the agenda, they get to set the timing for the work that they do, but it's all gotta be anchored to a business need. And so, we take the whole Agile approach with user stories, with a human being who's articulated that need behind it really seriously.

0:30:39 Scott Hirlleman

Okay. Yeah, I think yeah, there's a lot in there that I really liked. But I think just the managing themselves, that's that whole are you a leader or a manager? And it sounds like you're a leader more than you're the manager. You know, you are obviously from the title perspective, but it's like, hey, I'm gonna help you go in the right direction, but I'm not gonna micromanage. I'm not gonna kind of make sure I understand absolutely everything. I trust you as long as you understand these roles.

0:31:07 Kate Carruthers

It's just the thing, you can't know everything. It's, you've gotta trust your team



Transcript provided as a free resource by



because the only way you can achieve large things is by teamwork. You can't do it alone. If you wanna build stuff alone, then you go and have your one person's, you know, small business. But if you wanna build big stuff in big organizations, you've got to do it through people. And you've got to build trusting relationships so that people can feel that they can achieve, and they need to know what's not achievable and also be not afraid to say it.

0:31:38 Scott Hirlleman

Yeah. Benny Benford also said, "If you wanna go fast, go by yourself. If you wanna go far, go together." And so, you were talking a little bit about it earlier, but I'd love to hear about this, trying to expose data to people without having to do the work as the data people, but like, how do you limit the risk, not just from risk of PII and data leaks and things like that, but also kind of the self, the ability to harm yourself with data by not understanding what you're dealing with. So I'd love to hear how you're heading down that road, because it's, one, a lot of people are finding that overly indexing on self-serve has hurt them because they're not over indexing on making sure the data is super, super well documented and struct so that people aren't really sure what they're getting, but they can get access to it.

0:32:34 Kate Carruthers

Yeah. Well, we're attacking that on a number of levels. The first thing that we do is, if people wanna use any personal information, they actually need to complete a data sharing agreement, and they need to get the data controller to agree to their use of that data. And our data controllers are very diligent human beings who take every request very seriously. And they say things like, you can have that field, but not that one. I don't agree that you should have that one. So like, they will not routinely give people student personal emails, even though they may get other personally identifiable information to students, they will not give that, because we've known that once people get that, they tend to use it.

So, we're very careful about what data we give out to people, and then they actually sign a contract, there's a data sharing agreement. They get told, these are the terms, you've got to secure the data in this manner and form. You're not allowed to pass it on. You're not allowed to use it for secondary purposes and stuff. So there's that side of it. So that's a bit of command and control sort of thing. The other thing is, we've also taken the idea that people need to know certain things to do their job. Like some of our teachers, they can walk into a classroom and know who's in their classroom. So why shouldn't they know who's in their class? So what we've tried to do is actually take the idea that people need to know what they need to know to do their job, and we should just give them that. They shouldn't have to ask for that. So we're trying to turn this whole, on one side there's the, I'm asking for data and I need to do a data sharing agreement, but the other side of it is I'm doing a job and there's



Transcript provided as a free resource by



a lot of other people like me doing a similar job, and why am I even asking for this? Because I already need to know it to do my job.

0:34:26 Scott Hirleman

So, I like a lot of what you're saying, especially the data sharing agreement thing. I've been trying to say any user needs to register their use case. Even if you have non-data sharing agreement, you have data products that people can access. I as a producer, I need to know what you're actually using it for. One, for that making sure you're not using it inappropriately, but two, oh, you're doing this? Hey, you know, we've got this data that would actually be better around that or I can add this, or you're untransforming this column, why don't I just stick in the raw as well as the transformed column? And so you don't have to do work and so I think that's, I like a lot of what you're saying and people think that something like a data sharing agreement is too much overhead and it's oh, blah, blah, blah, and I'm just not seeing it. I'm not seeing that when you have those little checkpoints that it's really slowing people up. It's creating those relationships so then that data owner actually does know what it's being used for, and they are being cautious to not just give access to everybody. Like has that been, has that caused any kind of friction or are people pretty, I mean, in the US I think that definitely would, I don't know if Australia is a little bit better about protecting people's data.

0:35:53 Kate Carruthers

I think it was really funny when we first implemented them and it was a number of years ago that we did. The data controllers, the people who are data owners, it was the first time they'd ever been listened to because people used to just go to IT and grab their data and they never had a say in its use. And this really empowered them. So they really liked it. And the other thing is that it's helping us to uncover a whole lot of uses of data that we weren't aware of. So third-party software as a service that people are downloading CSVs in an uncontrolled manner and they're uploading it to. So we're actually managing to corral those and get them a cyber review and a third party security review. So, a lot of, it's actually adding a lot of value to the rigor that we're able to bring to those relationships because the people that are doing these things, they're trying to do their job, but they don't know what they don't know. So when we surface it and go, you need a data sharing agreement, and then we go, what are you doing with it? Oh my God, you're doing this. Have you had a cyber review?

So it's opening up all sorts of interesting conversations around the organization and allowing us to bring a whole lot of other people to the table. So one of the important things we do in the data space is we always bring not just ourselves, but privacy, our records and archives, our cybersecurity, our information security people, our risk people together. Because for us, it's kind of a team sport because if we're not



Transcript provided as a free resource by



covering off all of those bases, then the data is at risk.

0:37:41 Scott Hirleman

With those data sharing agreements with unauthorized copying and all sorts of those things, there's some people that are building platforms that absolutely prevent anybody from doing anything. Disney had a really interesting platform where they give people a lot more access than they might in another organization because literally their data mesh is completely locked. And so you can't exfiltrate any data no matter what. So they kind of have those reviews that come through. But is it, hey, we're gonna trust you. And if you do the wrong thing, it's on your head, so be it, right? Or are you building a lot?

0:38:23 Kate Carruthers

No. So we're doing a data loss prevention program next year. So we're doing all the controls, the proper cyber controls that any organization would do in addition. But one of the things we wanna do is give people access to the data that they need to know to do their job. And for some people, that's quite high levels of sensitive data because they need to know that. On the other hand, there's things that they don't need to have, which is within our power to redact or not show or scramble. So one of the things that we're gonna do next year is go through methodically and do that across the data that we've got. So there's no good reason for anybody, our tax file number, our TFN is the equivalent to your SSN. And nobody needs to see that. So there's no good reason for anyone to have that except the finance person who needs to put it in to report to the government.

0:39:28 Scott Hirleman

Yeah. I heard this one a long time ago about... And it was more relevant in the US when you had kind of landlocked phone numbers, but somebody was like, "We need to get access to people's phone numbers." And the other side was like, "Absolutely not. There's no way that we're gonna allow that." And they went back and forth, back and forth. And it turned out that all they wanted was the area code so they could get an approximate location of these people. And so that becomes slightly PII, but not really PII. It's pretty anonymized when you think about it especially at a very large sample size of tens of millions of people. It became something that wasn't nearly as big of a deal. So how much are you getting involved in these conversations versus it's the two other people, the data owner and the proposed data user, and they come to you only when they have questions? Or are you like banks or...

0:40:29 Kate Carruthers

I've got a data governance manager who does all of the coordination for that. So she manages those conversations and I'm usually an escalation point. But the thing is that our data owners are really so diligent and they genuinely wanna protect their



Transcript provided as a free resource by



data. They want to look after it. So they do a really great job of saying, "Look, we're willing to give you this and this, but we're not willing to give you that for these reasons." And there's those kinds of conversations where you need to keep asking why, why do you need that? Like that example you just used of, they just wanted the area codes.

And it's one of the reasons why in one of the first reports that I got built during COVID was a demographic overlay for the student data because people wanted to see where students were when they applied, where they were during term, because in the old days, pre-COVID, we used to just know that people would be in a classroom, the start of class. When the term started, people would just be sitting in a classroom. With COVID, there were people all over the world. We had no idea where anybody was. So we had to start caring about where people really were during class. And hilariously now, because we've got rules for international students, we actually care if they're in the country or not. When they're studying, because there's rules for how many units they're allowed to study online and stuff. So there's government regulations. So understanding that and giving people that insight without giving access to the underlying data is actually really important to them.

0:42:20 Scott Hirlleman

Yeah. Again, you were talking about there is a job to be done. So you figured out a safe way to do that rather than just being like, hey, we're gonna drop all people's IP addresses for when they log into the Zoom. That's probably not the best thing to be doing. Have you found that you're... One of the things that I'm finding when I'm talking to a lot of organizations is the business people aren't as data literate, aren't as data savvy. And so a lot of times they'll come in with a proposed way of doing something and it's absolutely not feasible or it's not the right way from all sorts of ethics and legal issues. Are you finding that people are now comfortable coming to you and saying, "Here's what we're trying to achieve. Let's work backwards towards a thing." Because I'm finding that a lot of people are struggling having those conversations 'cause people are like, "I wanna do it the way I wanna do it," instead of, "I wanna achieve what I wanna achieve." Are you struggling with those conversations at all?

0:43:22 Kate Carruthers

No. No, because the best investment I ever made was in my BAs. So I've got BAs, people requests, make requests through an online form. And then my BAs triage it and speak to everybody and have an actual conversation with them that tries to tease out all of those things. Well, what are you trying to do? Why are you trying to do it that way? Did you know we could do it this way? So the best investment I ever made was in a bunch of business analysts.



Transcript provided as a free resource by



0:43:52 Scott Hirleman

Yeah, it's funny. A lot of people doing data mesh, they're saying whether their data analysts or their business analysts, are leading the upskilling or they're handling a lot of that ad hoc and training somebody to actually do that themselves or kind of almost in a data sherpa type of role of like, "Hey, let's go and explore what's out there and kind of help them." It's something where a lot of people are really trying to cut at the BAs, but it just seems a little foolhardy in a lot of... That means sometimes you just have to cut budget, but it seems like those are the ones that are gonna be your best points of leverage for actual business value.

0:44:36 Kate Carruthers

And you need to keep them around long enough so they get to understand the underlying data because my lead BA, he's just amazing. He's a great human being. He's really nice to talk to. You find yourself telling him everything, but he also understands the data and can whip together a demo Power BI to show you what things might look like and stuff like that. So he's worth his weight in gold.

0:45:02 Scott Hirleman

We'll make sure that he doesn't hear that. So then he doesn't...

0:45:04 Kate Carruthers

No, no. I tell him that all the time. He knows he's good.

0:45:09 Scott Hirleman

So one thing that we were looking at talking about as well that I think has been woven throughout this conversation is the idea of, what do we have to decentralize versus what don't we have to decentralize? Because this is something I fight a lot at data mesh, but a lot of what you're doing is somewhat centralized, somewhat decentralized, and you're kind of... It seems like you're finding the balance and it might be different even in each domain and things like that. So how do you think about having that conversation with people where data people are typically used to be having everything centralized versus you go, "Okay, it's just free for all self-service." How do you figure out that balance? And how do you figure out when you've become out of balance? Maybe you found a good initial equilibrium and now it's no longer there. I know it's a very difficult question.

0:46:02 Kate Carruthers

Look, I think the piece of work that we've just done jointly with IT, which is to develop a target architecture, data architecture for the university for enterprise data. It's the first time we've done this and it's been a really important piece of work. And we've come up with... So the enterprise, we would need all of the data centrally so that we can surface it to people as and when they need it. Because what we've discovered is



Transcript provided as a free resource by



that there are people around the place who are getting a copy of a bit of data, they're meshing it up with their local data. And the thing is that if it was all in one place, we would just be able to expose that to them and they'd be able to develop insights from that.

So the question that we've grappled with is how do we... For all of that enterprise-like data, we think it all just needs to be in one place. And realistically, it's a very simple process to add another data source into a pipeline. But one of the things we want to stop is people taking some of our core source data and making copies of it, what I call data puddles. So we really wanna stop all this proliferation of data puddles because they're not very well secured. They're not always up-to-date. So they're not always timely in their updating. So we wanna give people that central data and then they can mash it up with their local data. So that's gonna be the project for us over the next couple of years of sorting all of that out, I suspect.

0:47:52 Scott Hirlleman

Yeah. Well, I like that you said literally over the next couple of years. It's not over the next quarter. It's not... It's, "Hey, this takes the time," right?

0:48:00 Kate Carruthers

Well, one of the things that we worked out with it when we did that piece of work was we needed to make some fundamental decisions about things like where does integration sit? Does it all sit in the one platform or not? And so we've worked out all of that and now we've just got to work out, how do we do this? And it's very much a joint exercise between my team and the IT team to work that out. So it'll be interesting to see what happens. I don't know what will happen, but I think it's essential for it to happen for the future of the organization.

0:48:35 Scott Hirlleman

I'd love to hear your thoughts. I'm gonna get into the GenAI as we wrap up and stuff, but I'd love to hear your thoughts around what you've learned and your approach as to how you would share that to others, right? Like what you're talking about right now, a lot of it is kind of this state and steady approach rather than trying to rush through. But sometimes people have very specific goals, budgets and they're more pushed on a quarterly cycle. But how do you think about maybe that communication if someone were to come and ask you for some advice? I know it's hard to distill it all into some advice, but is there anything that you kind of think about when you're thinking about this approach of doing this in the right way over the right time period?

0:49:19 Kate Carruthers

Two things. First of all, I think people overestimate the power of making change



Transcript provided as a free resource by



really fast and underestimate the power of sustained incremental change. So you can make a change really, really fast and it makes a big hullabaloo and everybody gets out of joint and everything. But if you just plug away at that every day for a year, you can make that change over a year and nobody is been upset. Everybody is... Because it's sort of gradual but it's not endless, you can actually achieve an awful lot if you... So we have around about 200, 220 business days a year. You make change every single one of those days, so that's 200-ish changes a year, times that by the number of people you've got. That's an awful lot of change in a year. So you can make big change over reasonably short timeframes using incremental methods that don't require the same level of change management, of reconstituting people's mental maps.

And that leads into my other thing, is to make change, you need people to have a mental map for the change. And that actually takes time. And to implement all of our technology changes that we've made over the years, we've had to sit down and spend time with people to just even get them to understand what we're even talking about. With that whole serverless data lake idea that I had all those years ago, nobody even understood why that might be a good idea. But now everyone just accepts it as real. But we had to buy that mental real estate and build up that mental model in their mind so that they understood what we were trying to achieve.

0:51:13 Scott Hirleman

Yeah. I've been using the alliteration phrase of maintaining mesh momentum. And that momentum is what drives change. It's not this giant shove to try and do that because then you kind of have that Sisyphus analogy of it's just gonna roll back on you but versus building up the momentum to actually move forward. So I wanted to wrap up around how you're seeing GenAI. It's obviously I just got back from Big Data London and everybody was talking GenAI. I mean, it was...

0:51:47 Kate Carruthers

Everyone used it everywhere.

0:51:50 Scott Hirleman

Yeah. And I have my skepticism. I also think there's some really interesting aspects to it. So I'd love to hear how you're seeing GenAI being used in universities and how you're especially looking at this from the AI Institute. So a broad question, but it's kind of wherever you wanna go with that.

0:52:07 Kate Carruthers

Oh, well, so for the AI Institute, they're kind of inventing the future. So they're inventing the next big thing. So we'll see that when it comes out. But across the organization, one of the things that we're seeing is integrating ChatGPT with other



Transcript provided as a free resource by



things like cognitive services, like text mining, like all the other AI stack tools. Because a large language model, an LLM, only knows what it knows. And they're getting better every day. They're getting better intraday sometimes. Sometimes I log in in the morning and they've done one thing. And by the evening, they've done something new. But it's all evolving so quickly.

But what I can foresee is a future whereby we have... We've got all these apps on our phone. I can foresee a future where we don't have any apps on our phone, except we have a ChatGPT-like interface where we ask it to do stuff. And it knows contextually what we're asking. So you might ask it to play your favorite song, and it will know what service your favorite song is, and it'll go and get it. And it might know that it's on Spotify or maybe even Amazon Music. Who knows? There's got to be some people using Amazon Music. But it'll know where your particular favorite song is and be able to play it. And it won't need an app to do that. So I see the future for us with Generative AI is moving us towards an API-driven future, which we've talked about for years. But that's actually gonna become real because I can see that this... And it'll be a series of chatbots that you'll be able to talk to. And all the APIs will be where all the power lives.

And I see that we're gonna have LLMs on our devices. So we will have a custom LLM on our mobile phone. We will have one on our laptop, on our PC, as well as large corporate models. And so we are going to see a proliferation of LLMs across our devices, and they will change the way we think about how we interface with computers. And I think that's the really interesting thing.

0:54:27 Scott Hirlleman

How do you think about GenAI and disrupting of learning? Because we've seen this already with... Some people throw the Zoomers, Gen Z under the bus and everything like that of, that they can't do all of these things, that they don't know how to do any research. They don't know how to go and find information and assess if it's good information or bad information. I don't know that I fully agree with all that stuff. I think there's just some examples of some not so bright people, and that happens in every generation. But how do you think about that risk to academia of just the way that people learn isn't actually learning, it's just asking and getting an answer instead of actually going through that process of learning? Are you seeing that as a risk or...

0:55:22 Kate Carruthers

You're sounding kind of Homeric. I can just imagine Homer, the guy who wrote the Odyssey and the Iliad. 'Cause that was a spoken word world. And you memorized your verses, and they were in verse to help you memorize them. And suddenly people started writing stuff down. And I'm sure they were saying exactly the same thing. Kids these days, they're writing stuff down on these stone tablets. They're not



Transcript provided as a free resource by



memorizing anything. Oh my God, the world is, woe is us. What are we gonna do? I think that this has always happened. And then when books came, we said the same thing. Suddenly we don't have to remember anything. Our ability to write it down has changed the nature of learning. It's the same thing. It's the same thing happening again.

So it's just another shift in the technology from writing on stone tablets, writing on parchment to the Gutenberg Printing Press. It's just the next evolution in it. So I don't see the problem. One of the funny things is, when I was studying law many years ago, we had closed book exams, and we had to handwrite closed book exams. And it was just to test your memory. It was ridiculous. And my brother did a law degree recently, and his exams were all handwritten still, but open book exams. Because they realized that if you don't know where the answer is in the book, you're not gonna find it during the exam.

So there's been evolution in that one thing, in just the assessment of tasks. We're gonna have to find new ways of doing it. We're not... Like as an institution, most of the Australian institutions have said, "We know you're gonna use ChatGPT. Here's how we want you to cite it. You can't write your whole assignment, and you can use it for segments." So we're trying to teach them. But one of the good things is, it's put the whole contract cheating industry out of business. So there was a whole industry that you'd go into the toilets, into the bathrooms, and they'd have the little tear-off tags for people who would write your assignments for you for money. And it was a really bad thing because students would do that, and then they'd get blackmailed for the rest of their lives.

And that entire business has gone out of business now because of ChatGPT. So it's disrupting many people. And there are many jobs that will go, but there will be new ones that emerge. So the example I always use is the jobs that I do now did not exist when I left high school. There was no way to train me for them, but I can do them anyway.

0:58:03 Scott Hirlleman

Yeah. Well, and my joke was gonna be until ChatGPT becomes sentient and then starts blackmailing people and going, "Hey, you used my thing." I don't think that's actually gonna happen. Well, Kate, this has been a fantastic conversation. Is there anything we didn't cover that you wanted to or any specific way you wanna wrap up the content of the episode?

0:58:26 Kate Carruthers

I think I wanna bring it back to the whole idea of data mesh, which I think is a really important and really significant one. And one of the things that I wanted to just say



Transcript provided as a free resource by



is, I love this idea of the technology, the people, the process all integrated with governance in a really organic way. And I think that that's something that we as data practitioners really need to try and build towards, because if we don't start to consciously build towards it, we'll never get there. It'll always be a thing that people talk about that we never really do. And a lot of the plumbing stuff that I've been doing is kind of building towards it, but I haven't told anyone we're building towards it. They probably won't listen to this podcast.

0:59:14 Scott Hirleman

Yeah. I have a lot of people that I've talked to that are kind of doing things and adding more capabilities to their developer platforms around data and things like that where you're just creating, again, that capability for people to do what you want them to do instead of, "We are now doing data mesh and it's a proclamation with a bunch of horns and da, da, da, da." Like no. It's that momentum. It's that incremental. It's that building towards better instead of trying to make huge shifts. But in certain organizations, I understand politically that's not as easy. But hopefully people can find...

0:59:52 Kate Carruthers

'Cause maybe we can coin a new term, perhaps it's stealth data mesh.

0:59:57 Scott Hirleman

Yes, I like that. There's even one that's in your neck of the woods, ANZ Plus, that was talking about kind of just adding more and more things to the platform so that people are like, "Oh, I can do this awesome thing with data that I couldn't do before. Great." And they're learning more and more about data. Well, Kate, I'm sure there's gonna be a lot of people that would love to follow up with you. Where's the best place to do that? Anything specific you'd like them following up about?

1:00:23 Kate Carruthers

Well, people are welcome to follow up with any questions that they have. The best place to find me at the moment is my DataRevolution Podcast, DataRevolution, one word, dot T-E-C-H.

1:00:34 Scott Hirleman

Okay. And we'll drop a link to that in the show notes as per usual. So Kate, thank you so much for your time today. And as well, thank you everyone out there for listening. I'd again like to thank my guest today, Kate Carruthers, who's the Head of Business Intelligence at the UNSW AI Institute and Chief Data and Insights Officer at UNSW or University of New South Wales. You can find a link to her LinkedIn as well as her podcast in the show notes as per usual. Thank you.



Transcript provided as a free resource by



Hopefully that interview episode was really useful for you. Please do consider getting in touch with guests from the show from these episodes. Most have said they'd really love people to reach out to them. And please, as well, if you've got a minute, rate and review the podcast somewhere. It really is honestly super helpful for other people looking into kind of data podcast to kind of get this in front of them. Data Mesh Radio is again, provided as a free community resource by Data Mesh Understanding. It's produced and hosted by me, Scott Hirleman. April of 2023, I left DataStax who are wonderful in getting the data mesh community stuff started, so give them a shout for streaming and real time AI needs. But I left to start my own industry analyst kind of information as a service firm. Our offerings are affordable, and you can do them on a one-off or a month-to-month basis. Read, kind of throw it on the credit card. Don't worry about going through purchasing and things like that. The services include lots of practitioner roundtables, one-on-one data mesh kind of planning or feedback sessions, and tailored introductions to other data mesh practitioners that are focused around your topics of interest, what are you actually running into challenges with? We also have some free programs around introductions and roundtables that people can kind of check out as well.

Check the show notes or just go to datameshunderstanding.com for more info or helpful resources. As always, if you have suggestions for guests or topics, please do get in touch as well. Have a wonderful rest of your day. Now, let's hear that funky outro music.