

Data Mesh Radio Episode #129: Iterating Data Governance for Data Mesh: Lessons Learned from 'The Data Governance Coach'

Interview with Nicola Askham Listen (<u>link</u>)

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 Mesh for Dummies book (info gated), please see here:

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

Scott Hirleman

The following is a message from George Trujillo, a data strategist at DataStax. As a reminder, DataStax is the only financial sponsor of Data Mesh Radio, in the Data Mesh Learning Community at this time. I work with George and I would highly recommend speaking with him, it's always a fun conversation.

George Trujillo

One of the key value propositions of Data Mesh is empowering lines of business to innovate with data. So it's been really exciting for me personally, to see Data Mesh in practice and how it's maturing. This is a significant organizational transformation, so it must be well understood. Empowering developers, analysts, and data scientists with downstream data has been part of my personal data journey that reemphasized the importance of reducing complexity in real-time data ecosystems, and the criticality of picking the right real time data technology stack. I'm always open and welcome the opportunity to share experiences and ideas around executing a Data Mesh strategy. Feel free to email or connect with me on LinkedIn if you'd like to talk about real time data ecosystems, data management strategies, or Data Mesh. My contact information can be found in the notes below. Thank you.

LinkedIn: https://www.linkedin.com/in/georgetrujillo/

Email: george.trujillo@datastax.com

Scott Hirleman

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

Adrian Estala, Starburst

Welcome to Data Mesh Radio, with your host, Scott Hirleman, sponsored by Starburst. This is Adrian Estala, VP and Field CDO at Starburst and host of Data Mesh TV. Starburst is the leading sponsor for Trino, the open source project, and the Data Mesh for Dummies book that I co-wrote with Colleen Tartow and Andy Mott. To claim your free book, head over to Starburst.io.



Scott Hirleman

Data Mesh Radio, a part of the Data as a Product Podcast Network, is a free community resource provided by DataStax. Data Mesh Radio is produced and hosted by Scott Hirleman, a co-founder of the Data Mesh Learning Community. This podcast is designed to help you get up to speed on a number of Data Mesh related topics, hopefully you find it useful.

Bottom line up front, what are you going to hear about and learn about in today's episode? Interviewed Nicola Askham, data governance consultant known simply as The Data Governance Coach, and she's also the host of the Data Governance podcast. So some key takeaways or thoughts from Nicola's point of view. Number one, the key point of data governance, ensure the data we use is the right data for the right people to better address our business challenges. Everything you do with data governance should circle back to that key point.

0:03:10 Scott Hirleman

Number two, in doing data governance right, you need to set yourself up to take in feedback and iterate. You absolutely won't get everything right up front. That's okay. It's crucial to set expectations for yourself and others that your data governance approach will evolve as you learn more. Number three, If you see Data Mesh as being about making better data more accessible to your current data consumers only, that's a very big opportunity wasted. Aim to significantly expand your pool of data citizens, not everyone should be a data scientist, but data should play a role in many more people's jobs. Number four, to get going in Data Mesh, you need to get your data governance to "good enough" and start moving forward. This might sound familiar, a lot of people have said this, it's very important to hammer on this point so that we don't have to get it perfect. Think about what you need, is it the very complicated standard for interoperability across every domain ever that's gonna last the next decade or is it about getting people to understand and trust the data they can now access, and so you want some data quality standards that people can measure their data quality against or something like that.

It's probably the second, right? Number five, to drive buy-in for data governance, you should tailor your message to the audience. It's very hard to have universal appeal around a specific selling point of data governance, but data governance can and should drive value for everyone. Does that sound like Data Mesh? It sounds like that to me, you need to tailor your message on Data Mesh too, right? Number six, every data governance approach should be tailored to the organization, but it should start from a few building blocks, policy mandates is the first, processes and standards is the second, and roles and responsibilities, there's much more about exactly what that means in the episode. Number seven, central governance teams are crucial,





they make it easy for federated teams to do what's necessary to comply with regulations and things, and internal standards so there isn't just the kind of silos. We don't wanna have silos around data governance, but their job is to have as little friction as possible in preventing those silos, the central governance team should be a value-add not a gatekeeper. They're not making the specific day-to-day decisions on data, they're federation that out, they're giving the domains and things, the ability, the capabilities to do that with little friction.

Number eight, make sure teams understand data governance can add significant value to them. Participation is not just some mandate, it has a benefit, then make sure you are actually providing that value, right? You can't just talk the talk. Number nine, in Data Mesh, you will likely need new roles to handle these new data governance needs, where previously there were some vague ownership kind of requirements that nobody was really looking at. You should really look to get really explicit around who owns what and why. And really put it down on paper.

Now, tying into that point number 10, you will likely have different sets of requirements under roles across your domain, that's okay. Look to create a standard model for roles and responsibilities and adjust it where it needs to be adjusted. It's okay to have non-uniform rules, but there needs to be a starting point for domains to go from. And finally, number 11, when starting out in Data Mesh, look for a relatively simple first use case, but don't only stick to simple use cases early in your journey. It will make it much harder to tackle the difficult use cases later. You don't want a mesh that can't tackle the hard but high value use cases, right? I think that's a unique point that really hasn't been made in a lot of places, so I think that's one to really think about as you're heading into listening to this episode. With that bottom line up front done, let's jump into the interview.

Very, very excited for today's episode. I've got Nicola Askham here, who is a Data Governance Expert and Consultant, and she's also the host of the Data Governance Podcast. And I reached out to Nicola specifically, we've been talking a little bit, but then I got a little bit more insistent, when I saw that Andy Kenna at RenRe had posted on LinkedIn about working with Nicola about the governance aspects of what they were doing with their Data Mesh implementation. So we're gonna be talking about a lot of things, data governance at the high level, but also how do you actually think about doing this in practice? How do you take these highfalutin concepts around data governance and really think about how can we actually look at how this evolves in the day-to-day, and that there are so many folks out there that I talk to where data governance has been a blocker because they think they have to get everything absolutely perfect upfront. And personally, I don't think that's the case. Nicola may have arguments with me or push back on that as well. But I'm really excited to kinda dig into a story of how we actually think about applying this,



especially in a highly regulated space, like insurance, right? That's a financial services, tough space. So with that, Nicola, if you don't mind, if you could give people a bit of an introduction to yourself and then we can jump into the conversation at hand.

0:09:28 Nicola Askham

As you've already said, my name is Nicola Askham. I'm known as the Data Governance Coach, and I help organizations understand and manage their data better. I've been doing this for nearly 20 years now, and I've helped loads of corporates and more recently, some public sector as well, reduce costs and inefficiencies and remain relevant, is what I like to say, and I think that fits nicely to what you're saying about I view data governance not as a blocker, but very much as an enabler and a facilitator.

0:09:56 Scott Hirleman

Yeah, and I think a great place to kinda start this conversation is a little bit around that. When we were talking in the pre-call, we were talking about a kind of a "normal governance" approach, right? I think with Data Mesh, so much of this is about evolving the way that we work, evolving the concepts that already exist, instead of trying to come up whole cloth with everything new. Because I've talked with this a lot from people trying to just completely come up with everything from scratch, it's a big ask to, as an understatement. So can we start with what your concept is of what has become the normal way of doing governance? Is it the centralized, by committee, and slow approach, or what do you think is decent practices and then how do we start to look at how we apply that to Data Mesh?

0:11:00 Nicola Askham

Yeah, I think that's a really good question. And yeah, I've read loads of things online. There's so much if you google data governance these days, and they'd tell you about a centralized model or distributive model or federated model. And I think what I have found over the years is actually a combination of both, because I think when people say a centralized model, they're talking about one small team of people making decisions on all data across an organization. And I don't think that's useful, I don't think, particularly when you're talking about a larger organization, that any small number of people could actually have all that knowledge to make the right decisions all the time. So I do like to have a more federated approach, where the right people in the business are signing up to own the data and make decisions about it. But I do think that they need some central support.

0:11:50 Nicola Askham

So I suppose this is where perhaps I go for a bit of a hybrid that you do need a central data governance team, and you probably do need to get your data owners together regularly on a data governance council or committee, whatever you want to call it.



And I think that kind of approach seems to work well for everybody. I've not changed from that for many years now. So whilst the individual framework I might help my clients design will be subtly different, because it has to work for that organization and their particular data and whatever challenges they're facing. But I think that concept of having some central support, but actually data ownership being out in the business in various places, wherever it's appropriate, is the one thing that remains the same.

0:12:36 Scott Hirleman

I think that's a really important question of the central teams. You said the central committee, but you also said central support. Is the central team making decisions on any one piece of data or are they there to only provide support or are they there to make the overarching, "Here are the ways that we're going to do this, so that we have standards and practices," versus, "We're actually gonna be making the decisions on anything day-to-day," or, "We're going to collaborate with you." I think that's what's coming through of the support and the centralized practices and making it so that it's easy to do governance at the federated level, but I'd love to hear what you've seen that works and what doesn't in that.

0:13:31 Nicola Askham

Yeah. So I think generally, your central team, which is gonna be headed up by your data governance lead, and depending on your size of organization, they may have some support from data governance analysts or data quality analysts. So I think about what they're not probably gonna be making any decisions on anything, and that's the one thing that I always find is important, if I was your data governance lead, how could I make a decision on what your data quality standards should be? The only people who know that are the people that are using the data in the business.

So what I think the central team is doing is almost like, they create the framework in the first place, and then they make sure it happens. And in the early days, there is gonna be a lot more hand-holding of the data owners and the data stewards. Because if you're just starting, and you're really early on, and you go to a data owner and go, "Oh, Scott. You know you're the data owner for customer data. Could you give us some data quality standards, please?" You're probably gonna look blankly at us and go, "Well, how do I know what to do? What should I do?" Kind of thing. So whilst I don't think the central team can say to you, "Oh, here you go, Scott. Here's all the data quality standards you need." What they can do is say, "Well, there's different ways we can measure customer data," and they can take you through it all and help you work out what would be good enough for you and help you set standards.

0:14:48 Scott Hirleman





Yeah, I think that's the thing that comes through in a lot of the conversations of people who are comfortable with what they're doing with their governance in Data Mesh is that it's about helping and it's about facilitating, and it's about like, "We're gonna teach you how to measure this. And we're gonna provide you some tooling to help you so you can measure your data quality." But if your customer, whoever the use case is, doesn't require really high accuracy, but requires it to be very, very complete, because if there is not that completeness, then that completeness is the thing that breaks everything. But the accuracy, they only need it to be 90% because they need it within this time frame. And so helping people to understand that when people think about gold star data and they think about MDM, and they think about all this stuff, that it's like, it must all be 100% perfect, 100% accurate, 100% right, 100% complete, everything. It's like, "That's not feasible or reasonable. So let's talk about what you are trying to accomplish?" And I think that's where governance really is headed, or like, "What are we actually trying to do with this? Let's make it so that we can do business appropriately with data, but with the proper controls in place so we're all communicating and understanding."

0:16:16 Nicola Askham

I think that's exactly it, Scott. I think a lot of people think that there should be some kind of central standard. I had an email even just last week saying, "Could you tell me what data quality standards we should apply in my company?" And you're going, "Well, no, 'cause I don't know what your company does. And I don't know what you're doing with the data." And the only people that know what you're doing with that data, and as you say, whether accuracy is important or completeness is more important, are the people who are using it. And that's why there are always gonna be some external standards you can use, like perhaps the ISO country codes or currency codes. But there's an awful lot of other data quality standards that are going to be reliant on what your organization is doing with that data, and therefore what makes it good enough for them to use.

0:17:00 Scott Hirleman

Yeah, I think we'll see more and more standards around measuring it, right? And I think that that's the only thing that we really should be looking to have standards of, like, "How do we do interoperability? How do we standardize communication, so that we can actually have high context exchanges, and that we can trust and feel like the data lineage is that, "Sometimes it is like, "I wanna understand how this was computed." But a lot of times it's just, "Can I trust this?" And if we remove that feeling of, "Do I need to trust this? Can I trust this?" What would that unlock for your business?

0:17:41 Nicola Askham

Well, exactly. And I can't tell you how many people I've spoken to over the years who



waste so much time and effort trying to fix data or find data before they do whatever their job is with the data. And if we can eradicate that, think about how much extra time people will be doing what we pay them to do, and to use the expertise that we recruited them to have, I kind of often call it data wrangling. You know? 'Cause we have to go and wrangle this data before we can do our day job. And sometimes it's horrendous. I was doing some training for a client the other week, and I was saying that, "Over the years, I think nearly every analytics or business intelligence team that I've worked with tells me that they think they waste around 50%, some go a little bit higher, of their team's time data wrangling, not doing the reports and giving the insights to people that they need." And there was a data scientist on this workshop that I was doing for this organization, and he just suddenly said, "I'm sorry, I've just gotta come off mute and say, actually I think we spend up to 90% of our time data wrangling." And you go, "My God, Think of how much money that organization is paying their team of data scientists, and they're doing most of the time data wrangling, and not doing these amazing insights for the organization."

0:19:04 Scott Hirleman

And I think sometimes we get a little wrapped up in the data wrangling, how much is it actually getting the understanding of it. But if we can create the standards so that the people who are actually sharing the data, if we can create easy or maybe not even standards, but that we create low friction ways for them to make it so that people can trust it and understand it, how much does that save? And yet, we're always in such a rush to get to the insight and put the insight into production instead of the safe sensible. The long-term approach, it's like, we're so focused on what's the next week, not even what's the next quarter. You've been working with one specific client, but let's not get into anything specific about that client, but how do you think about taking what you've learned, you've got 20 years of doing data governance stuff.

You've seen all the ways that don't work, so how do you think about starting to take these approaches and how would you adapt them to Data Mesh? And I know every organization is gonna be slightly different, but like let's think about that federated approach and how do we do that gently instead of it's a hard break. I think most things in Data Mesh shouldn't be about big things, so I would love to hear how you thought about approaching it, what you'd think about, if you were to talk to anybody generically rather than airing any dirty laundry or anything like that.

0:20:45 Nicola Askham

I think it's really interesting, 'cause the one thing that I've learned over years, and it's actually kind of gone into the methodology that I teach when I'm doing my training courses is that there is no standard approach for data governance, but there are some key building blocks that I've learned, and I'll be honest, the hard way, by doing





it wrong, that you have to address, but how you address them is gonna be slightly different depending on the organization. So I think I always start every new client engagement with the, well, these are my building blocks, and let's see how I'm gonna adapt them for this organization. I think what was really interesting doing data governance over Data Mesh for the first time was that Data Mesh is still relatively new, it's an emerging technology, but it's not just the technology. I think that's what really got my interest in it, it's much more about socialization and getting the people's side of it. And that's what I love about data governance, is the people's side of it. I often say that data governance is more about the people than it is the data in the early early phases. So I was really interested to kind of get my head around Data Mesh because it is new and as always new technologies come and go, but because this one takes the technology and the people's side of it together, I was really kind of interested in it.

And excited to learn more about it. So obviously the first thing I was doing was researching as much as I could to try and understand Data Mesh. The first few things I read, I thought, "Actually, this is far too techy for me, I don't get this." And then I realized that I had to not concentrate on the techy side, somebody else is gonna worry about that. But I need to worry about how we make sure that the data that goes on to the Data Mesh as a data product is, as you say, trusted, is the right data and is trusted. And how do we govern going forward, perhaps that's, based on what we've said already in our conversation, governed isn't always the right word, because governed makes it sound like we're gonna put obstacles in the way. But actually what we want is who's the right person to make decisions on this data, so that it's there and available, so that everybody knows what this data is. It is really well-defined. They know that if it's there on the Data Mesh, they can use it, and if they've got any questions about it it's also clear who to go to. 'cause I think that a lot of the problems of people is that they use the wrong data or poor quality data, because they just don't know who to go and ask if it's the right data, or, "How do I make it better?"

And I think that's what data governance gives you, is that understanding of the data and who to talk to about it. So I think whilst I'm always thinking about how to adapt these things, normally some of the amendments I would make are subtle. But I think with the Data Mesh, it took some really, actually, good conversations with my client. We were kind of bending it backwards and forwards. And I'd say, "We should do it like this." And they'll go, "No, you can't be causing it, no." And then they'd say something and I'm going, "Oh, that won't work because" And it would be based on perhaps one of my previous mistakes I had made in years gone by and I'm going, "It never works, so I can't see that's not gonna solve the problem in Data Mesh either." So it was a really collaborative, great way of working because they'd obviously done a lot of research into Data Mesh as well. And I think that the big thing that we worked out is





that you do need this more federated approach, and if anybody has read the book by Zhamak Dehghani. Is that how I pronounce her name? I never know how to say it properly. But she has this lovely section on it that says you must move from having a centralized data governance approach to a federated one. Kind of when I read that, I went, "But I've always had a federated approach," so good news is I don't have to move from.

I don't think a centralized data governance approach works very well, anyway, as I already said. So I think what we're trying to do is think, "Oh, well, that's great," 'cause my way of doing it in the past isn't too dissimilar from it. But I think what we did find is that there are some perhaps new roles that we hadn't thought of before, and I always recommend that you keep your data governance framework as simple as possible. I find the more complicated you make it, the harder it is for people to get their head around it, and then it doesn't get embedded and then it doesn't deliver any benefits. So I always love this, keep it simple, and I'll always challenge my clients if they try to put too many extra roles in, but when it came to Data Mesh, what we actually decided is we did need some extra roles. And clearly you'd got roles like data engineer, which is mainly the technical side of it, but we wanted to make it clear that they can't just go and choose whatever data they like, so we needed something else in there. And they were gonna have a data product owner as a role, just to kind of meet their needs from doing the Data Mesh. And I realized that this should also be a data governance role. And we've had some really interesting conversations about moving some of the data owner and data steward responsibilities that you'd expect in a more, I was gonna say, traditional?

Usual data governance approach over to the data product owner. But there were some that I didn't feel comfortable moving over to, so I don't think you can lose the data owner and the data steward, you still need them. But what I think has been really interesting is that even when I do data governance in a more traditional set up, I can almost guarantee you won't always get it exactly right when you start. Even if it's a case if you didn't get the right data owner, or they suddenly say, "Well, I can't decide this because " And you take on board what they say and you tweak things. So I always say that whatever framework you design on day one won't be the same even a month or two later. And I think that's what happened, and I would kind of say to people, "Don't expect that you're gonna get this right, particularly because there aren't a lot of people talking about data governance for Data Mesh." So nobody can say, "These are the best practices, you're gonna get it perfect straight away."

So I think you need a considered approach, I don't say, "You're gonna change it, so do any old thing," 'cause then that'll go wrong. You'll put people's backs up and they won't like this data governance, but have a considered approach, but be open to the fact you might change it. Because I remember at one point I said to my client, "Do





you not think the data product owners are going to be data stewards? So could we just conflate the roles and make it one role?" And they said, "No, absolutely not," and gave me what was a really good reason, so they convinced me and I went, "Okay, no, that's fine. Let's go for two separate roles." And then when I caught up with them recently, we were discussing some other things, they said, "Oh, funny thing is we had to move some of the responsibilities around because it wasn't working quite right and we weren't getting the results we wanted, and it wasn't frictionless as you were saying it "And it shouldn't be about adding friction, it should be about making this as easy as possible to get trusted data available to people.

And so they hadn't got rid of any of the responsibilities, they just moved them around the roles a little bit. And it was quite funny, as they were talking about it, I suddenly said, "Oh, that's interesting. So given what you've just told me, are the data product owners sometimes the same people as data stewards?" And they said, "It's funny you should say that. We were thinking that only the other day that they are." And you're thinking so we've almost gone like a full circle, 'cause when I suggested that to begin with, it didn't make sense with what we knew and understood about it, but now they've got data products live in the Data Mesh and they're using it, they've realized that perhaps the initial people they had as the data product owners were not the right people, and they've found new people who coincidentally were also the data stewards for that data.

0:28:32 Scott Hirleman

Yeah and Sarita Bakst at JPMorgan Chase who's leading up their governance for their Data Mesh, when she was on, she was talking about that within governance, especially around Data Mesh, you have to be even more flexible because you can't have a cookie-cutter approach by domain. Everything is like, we can be flexible around this. So the way you just said is that in certain ones, it could be that the data product owner is the data steward, but in other ones, it should be separate. Which frustrates people because everybody wants a cookie cutter, they want to be able to stamp a whole lot of things and just do this in a lot of ways.

So you said a whole lot of things that I could react to, but I think a good one would be talking about that evolution and talking about when you started to get started, what do you think people need to have in place to get their first data products out the door? And yes, it's gonna be different for each one, but people are constantly telling me they have to figure out all aspects of governance, and then you also talked about that tweaking. How much do you think you want to do up front versus prepare for iteration. Is it just starting with nothing and getting that out there, or spending at least a little bit?

0:30:06 Nicola Askham



It's a difficult one. It's a bit like that chicken and egg. Though I do, I've always been a fan of an iterative approach, but I think you've got to have something in place, you see. I've worked with clients over the years that have almost become paralyzed by the analysis, and that's not just on Data Mesh, just data governance, generally, trying to design the perfect data governance framework, and I just said. "Just get it good enough and start going kind of thing." It's almost like the minimum viable product, isn't it? And I think that's a good way of thinking is that we need to have an idea of what the roles and responsibilities are gonna be, and we need to have some idea about the data standards, and if we're going to have an idea about that, then we need to have engaged with the people in the business. But I think it is really useful to actually test them out on your first data products because it's all theory and particularly 'cause this isn't something that a company has done before, so until you actually start trying to do it, it is all theory. So you can come up with this beautiful framework that would be really lovely.

And then the very first time you use it, somebody is gonna go, "Oh, I can't make that decision," or, "I don't know that. I can't tell you that." So I think it's really good to have your idea of it, but not have it set in stone, it's a kind of, this is what I do. And I think that's what I'd always recommend is, I absolutely think you need to have a policy, you need to have all this stuff mandated, but don't think that's the first thing to do, get an early design done and try it, do it on your first phase. Do it as a proof of concept, and that's when you realize that I asked Scott to make this decision and he just looked really bewildered and went, "How do I do that? I don't know that," kind of thing. And we got, "Oh, perhaps Scott shouldn't be in that role. Or is it that Scott is right because there's five responsibilities that we've given him and he can do four of them, but this one is just, we've given it to the wrong person, kind of thing. And then we can move the role, and I think that's how people have to think. The trouble is, if you're involved in it, whether from the governance angle or the Data Mesh angle, we're all data people and we all like things to be perfect and just right and neat and tidy, and I think we can tie ourselves up in knots in that. And sometimes we've got to say it's good enough, let's get started and then let's go back and refine as we learn from our experiences of trying this.

0:32:27 Scott Hirleman

And I think the cost of change in data has been huge. You think about the cost of a change in the data warehouse, and that if a lot of times what you'd have to do is re-implement the data warehouse, 'cause it's like, okay, everything is so rigid, and we need to get to a place where we're able to evolve. And I think getting into that mindset and then also setting yourself up in such a way. I tell people, get to CYA, or for your initial thing, like what's gonna get you not in trouble?

Shawn Kyzer and Gustavo Drachenberg

from Thoughtworks, when they were on, I asked them this question and they said,





"Huh, yeah, every single of our first data products in all of our implementations, we haven't had PII in any of first things, because it's very difficult to really build that early on of the masking and things like that, so we found use cases where it wasn't valuable. And then once we figured out, we got rid of other friction, we could start to tackle that challenge."

So I would love to give you time and space to react to that, but I also wanna kinda circle back on one thing you said about your building blocks. I don't know if that's proprietary TM TM TM type of thing, but would love to hear about those building blocks as well, because I think that does factor into getting too good enough. What are the things that we need to kind of have a firm idea around, but be malleable, right, that it's firm but malleable so that we can change it if it makes sense to, but that people can get going and actually not feel like it's all chaos?

0:34:27 Nicola Askham

Yeah, and of course. Well, yeah, it's not. It's obviously something that I teach in my training course, and so it's not something I can tell you all the details quickly anyway, but I always think that particularly when you're starting off, is to think about three things. And it is the policy, but bear in mind what I've already said, maybe not first. But your policy mandates it, because otherwise, it's just a nice to-do, and I think particularly with things like Data Mesh, and we've got data science, AI, ML, everybody gets really excited about these in a way that they don't get excited about data governance. And we get this unfair image of being blockers, which I think, sadly, some people have designed their frameworks that are blockers, and that's not what we should be doing. So we're never gonna get very far if we just ask people to do it nicely. So we do need a policy that mandates it, that says, "At our organization, we take data very seriously, therefore we have a data governance policy, and where you would do the following things." Now, I say you can't write the policy first, but I do think that's a really important building block, and any of my clients that I work with now, if they don't have one, I will call it out, because I learned that the hard way.

I spent about two years doing data governance in the bank where I was working when I first started doing data governance, and I thought I was making really good progress, 'cause I just got sheer energy and enthusiasm, and everybody came along with it. And then people started to say, "Well, we're really busy at the moment. Do we have to do it?" And I go, "But you said you'd do it. Why are we not doing it?" And it was, "Oh, no, well, we've got this big transformation program going on. We haven't got time for your data governance thing." And I was going, "But why are you not doing it?" And they went, "Well, let's put it like this. I need to get this done, it's in my objectives, I need to get to deliver this. If an audit isn't gonna come and beat me up for not doing what you've asked me to do, why would I do it?" And I was going, "Oh, that's brutal." And that's when I realized that we have to have a policy. I would always





rather sell the benefits of data governance to people, but we do need something that makes it mandatory for an organization. So I think within the first, I don't know, six months to a year of doing data governance, you should try and get your policy formalized and signed off.

But apart from that, I think you need some processes, because we need to be consistent in what we're doing around our data. And I think those processes will differ on what you're trying to do and achieve. And clearly the processes that we've designed at RenRe that have been around getting the data ingested onto the Data Mesh. But for other people, it might be their master data management, or their data warehouse solution. So you gotta think, "What are the processes? What is it I'm trying to achieve from doing this, and what processes do we need?" Because again, something I didn't necessarily worry about too much, I just thought in the early days, I could inspire people to do it and they'd all go and do the right thing. But, if I don't give people very clear instructions on how to do that, they're gonna do things in an inconsistent manner. And you might, for all the right reasons, decide to fix and improve some data, but if you don't tell somebody who uses that data downstream, you might cause them more problems. So it's all about having a process so we can be consistent, but I actually think that the roles and responsibilities are the key of all of the building blocks.

And the reason I say this is, I'm sure, Scott, you have been to the same kind of meetings I have over the years, where everybody is in violent agreement, we all agree what is going to be done, nobody has any dissent at all in the room, and we all go away really excited. We've agreed on something, it's gonna happen. And then we come back, and everybody says, "So, did you do that?" And everybody's looking at me and I go, "Oh, I wasn't gonna do it, I thought Scott was gonna do it." And Scott says, "Oh no." And you look at somebody else. And that's the whole problem with data governance, is not having people sign up and know that it's their role and their responsibility to do certain things. Yeah, over all the years that I've been doing this, I've had countless conversations with people about why they should do data governance at their organization. And I can promise you, I have never had anybody tell me that they shouldn't do it. Everybody agrees it's a very sensible idea and that they should do it. But I can't tell you how many people then tell me that they can't do it because, and they come up with all sorts of reasons why we can't do it, or they always assume that somebody else will be doing it for them, maybe even the central team, going back to our very early conversation in this.

So what we need to do is to make it very clear what the roles are and what the responsibilities of those roles are. And then we need to identify who fills these roles, and I think that is kind of the key to it all, because if we don't have people doing that...



0:39:25 Scott Hirleman

Apologies, we had an audio issue here.

0:39:28 Nicola Askham

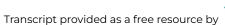
I think it's fair to say that over the years, I have spoken to countless people in organizations, trying to convince them of the need to do data governance at their company, and I don't think anybody's ever said to me that they shouldn't do data governance. If you explain it properly, everybody agrees that they should do it, but they always have this impression that it's difficult or it can't be done. So I'm often told that it can't be done for whatever reasons, or that it should be done, but never by the person I'm talking to, always by somebody else. So going back to perhaps a conversation we had very early on in this session about, perhaps they have this belief that the central team is gonna do it all for them. And I think this is why the roles and responsibilities are so crucial to a successful data governance implementation. Because we need to make it very clear what each role is responsible for, and then we need to tell the people which role they're playing in that and get them engaged and signed up to do that. And that is the only way to make it work. I've seen too many people say things like, "The business is responsible for." And I go, "But who in the business?" It's not gonna happen if we don't know who in the business is gonna do it.

0:40:47 Scott Hirleman

Yeah, and so one would be, I wanna ask you a little bit about how do we actually make it so that it's easy? How do we remove the friction of actually doing data governance so people can make it very, very simple to actually comply with that? But I'd also like to dig into a bit about, again, where if people were getting started, things are going to change. That hasn't been something that I've seen that's very often communicated about data governance, is that we are going to change, we understand that this isn't right at the start. So, how do you take in that feedback, how do you get people bought in that that's a thing, that it's not just that we are all knowing, we are the central brain and everything must comply, versus like, how do you get them so they're bought in, that you're taking in that feedback appropriately?

So I think people underestimate significantly how much effort you have to put into communications when you are doing data governance, and I think that is one of the key things that you have to do. You have to manage people's expectations, and I think particularly so, because it is really common that an organization, for whatever reason, will decide it needs to do data governance. And they'll often look to people already in their organization to take that on and start doing it. But these people haven't had data governance before, until all of a sudden, they're, "Oh, Scott, you know about data, you can do data governance for us, can't you?" And so quite often, you know, these people are scrambling on the internet, listening to podcasts,







whatever, trying to get their head around what data governance is, and then work out how to do it for their organization as well. So I think that if you pretend that you're all knowing, "I know how to do the data governance perfectly," it's gonna go wrong. And I think what I've seen is that people who do really good communications explain that data governance, there is no standard framework that works for everybody, there's no one-size-fits all.

So we're working through everything, the materials available to us, and we're trying to design something that's right for us. And I think it also fits in nicely to this iterative approach that we talked about, you can't get it right in one go. So you say to people, "We're doing this iteratively, we're going to work with something." And it's picking up what you were saying about that other podcast you guys just had from Thoughtworks, saying that they don't do PII first 'cause it's too hard. So I would always say to clients, do something fairly easy first, that's great, then perhaps do something hard in your second attempt. Because if you do something only easy for the first few iterations, it's gonna come as a shock six months, a year down the line when you try to do data governance over something a bit more complicated and it doesn't work, and you've got to rip up your whole framework and start again, which is not what we want. So I quite often like to get all of my potential data owners identified fairly early on, even if I'm not gonna come to you and talk to you about data governance very actively for months, because I want you to be aware of what we're doing.

And I want you to get the chance to say, "Well, when you come to finance or HR or wherever, that won't work for us" So that we've already got that feedback in. So it won't be perfect, but I'm trying to get everybody engaged in understanding that we're not designing something that's rigid and that will just be shoehorned in whether you like it or not, we're taking on board everybody's feedback. And we're communicating all the way along, explaining why we're doing what we're doing, explaining that and encouraging people to come to us if it's not working. I've made several of my clients panic over the years, of going into difficult, colder meetings with them to support them. And then people say, "But it's gonna be more work and I haven't got time and everything."

And I've got a habit of saying, "Well, yeah, that's a very valid concern. And I promise you, if you start doing this and it takes you more effort to do data governance and you get backing benefits, you can come to your data governance lead and tell them to change the data governance framework." At which point the data governance man just sitting next to me, totally unprepared for that, normally goes white, then red, and then white again. But it's the kind of thing you have to say, and I think that's what I really did wrong those years ago, as I was so passionately bought into my framework that I designed, and was so proud of it that I wasn't willing to concede



that I might have got it wrong. And that caused the whole thing to fail, whereas if I had said, "Oh, okay, that's not working. What do I need to do differently?" I could have saved it.

0:45:54 Scott Hirleman

And I think that value-add governance is something that we don't talk about enough. I think in Data Mesh, there is the computational aspect of it as well, that we need to make this so that it is easy for people to comply with the processes and that the processes add value. And some of that is risk mitigation, some of that is, Okay, we have PII here, so we do have actual manual access control because of what the people need or we have something that could be misused. If this information could be misused, you have to register what your actual use case is with the data owner so they can make sure this isn't gonna get themselves in trouble. This needs to be something... But outside of that, really, this can be something that really adds a lot of value because you are making it so that they don't have to do the data wrangling. You are making it so that and it's not just the reverse or the prisoner's dilemma of, "Well, if I don't participate but everybody else does, then I get great quality data from everybody else internally, and I don't have to do the extra work." How do we make that so that it's not that, that we can actually find ways where a data producer gets value from their own data that is produced? It's not necessarily easy, but it's something we have to figure out.

0:47:27 Nicola Askham

I think so, and I've come across this saying many times, you shouldn't be doing data governance 'cause somebody told you to. There's got to be some value in it. And as you say, the producers, sometimes that's the hardest thing. You made me think of another insurance company I worked with many years ago now, where for some reason, the underwriting teams didn't put the tax codes in when they set up new clients. And it was causing the finance team, the tax team, a huge amount of pain and effort every quarter when they had to do the regulatory reporting to the various tax authorities around the world. And so they asked me, they said, "Well, it's a data quality issue, 'cause the data is missing, so can you beat up the underwriting teams for us?" And I went and said to them, "Look, you need to be putting this data in there." "We don't have to. It's in our procedure." So they even waved the manual at me that says, "You don't have to do this 'cause finance does." And I said, "Well, finance does because you don't." And they went, "But we're not interested in it, we don't care about the tax code, we've got other things to do." And it's that same kind of attitude that you see time and time again from the data producers who are responsible for putting in data that they don't use. So they don't care about it.

But luckily, I was able to find out that it did have an impact on them, and I said, "So when finance put the codes in for you, did they cause you any stress?" And they





went, "Oh God, yes, because they always ring at the quarter end when we're really, really busy, and they demand that we have to give them the tax codes with no notice whatsoever. And we have to go rummaging in the paper files to find the tax codes." I told you it was a few years ago. "And then we have to email it to them, so they do it." And I went, "So would it not be far quicker when you have that form in front of you, to not put it into the system and then finance won't ring you at a really bad time for you?" And they went, "Oh, yeah." So I think it is always that tying it back. So yes, they didn't use that data, they thought they had no interest in it, but the lack of it being there was causing them pain, and I managed to get them onboard, and they started doing it and realized that there was something in it for them in actually putting the data in correctly.

0:49:34 Scott Hirleman

I'm just loving that, "no notice," 'cause it's like it happens every quarter, what do you mean? Have you not recognized this pattern?

0:49:45 Nicola Askham

No, and another one is people always telling me, "Oh no, we don't have a problem, we have a manual workaround," or even say like, a monthly manual workaround. You're going, "Yeah, okay, did you not listen to what you just said?"

0:49:56 Scott Hirleman

Yeah. I mean, sometimes, I do think in some cases, what is the return on investment? I used to work in covering stock stuff and things like that, and there were certain reports that we could automate, but that it would take me 20, 30 hours to automate, and it was something that took me 10 minutes every week to do, so it just never would pay back. And it was like the actual automation wasn't a big thing, but we had one report where we were tracking like 70 different companies, and the data entry took four hours and it was often filled with lots and lots of manual issues. But I had come up with a way to scrape the data and we expanded it to 300 companies, and it took one second, because it was just a copy paste that just made it. So then you just did this, this, this, and boom. So we were able to 4x and give us way more information, but automate it 'cause it paid for itself.

So I wanted to maybe do a little bit of a hard shift here, 'cause there is one question that we were talking about in the pre-call that I think is something that has come up a whole lot. So I wanted to give you some space to answer about how you looked at it and how you're thinking about it. More people are gonna come to you and say, "Okay, well, you fix this for me too," but I think it's just useful to put in front of people, which is In Data Mesh, finding the owner of the source data is relatively easy unless there are five different domains that have the same "source data," and then it's like, okay, how do we deal with that? But when you have downstream data products that are



from multiple upstream data products, like who manages that, who owns that, how do you do that, how do you figure it out in the early days versus once you form that process, you might have a little bit of a better framework around it, but the first time you run into it, it's like, okay, but now what do we do?

0:52:16 Nicola Askham

Well, I think it's not too dissimilar to thinking about just normal data governance over perhaps on a data warehouse. Because I think I've even made videos on it, because the number of people that say to me, "Well, who owns the data on the reports?" And again, we're the original data owners, unless you've changed it. So if you have transformed that data, perhaps you've aggregated it or perhaps you've done some kind of derivation to create this data, then perhaps it's not the original data owner. But if you've got a report that has gross written premium and gross written premiums owned by Fred, then Fred still owns it if it's in your report or if it's in your data product on the mesh. But when it's different, when you're saying, when we suddenly get them, perhaps we pull them all together and we create some new thing that didn't exist before, then we need a new data owner for it.

And I think that's the way to try and think about it, but I think the thing that I found perhaps the hardest to get my head around with the Data Mesh is that it will make data so much more accessible, and this is really, really great. But you have a data product that you've rightly said, you might have multiple data products that have some of the same source data in. And each of those data products may have a different product owner, a different data product owner, so I don't want them making decisions about the source data. They might make a decision about how we bring it together, how we calculate it, how we derived something, but I don't want three different people making potentially inconsistent decisions about that data. And I think that was one of the hardest things that I found, because up until that point, I was thinking, "Oh, well, we'll just scratch the data owner role and it's being replaced by the data product owner role."

And then I realized that I can't, because the more I was talking to my client and getting my head around it, is that it won't always be necessarily the same data product owner that owns every product that has that same source data in. And I want somebody, just one person to have that holistic overview of the data in an organization. In an ideal world, outside of the Data Mesh as well, they own that data wherever it is, whether it's on the original source system, yourData Mesh, your data warehouse, in a report somewhere. I want them to be the one person who really gets what that data is.

0:54:53 Scott Hirleman



Well, and I think this is where data literacy comes in and can be a challenge, is, if we're trying to make data more accessible to more people, they might not know how to track the SLA of SLA of SLA. This piece of data is updated in the original source system every day, and then it's pulled into this on a weekly basis, so how fresh. And then somebody else pulls it off of that and says, "Oh well" So you've got the original tier data product, somebody has created something that's a data product and that's only updated every week. But then somebody tries to pull off that and say that their data product is refreshed every day, but it's got a source that's only every week, and so it's not. And that you get through these things of trying to go directly to the source, so that you're the one that is not tapping into downstream of downstream of downstream, but then are you re-doing the same combinations of data where you are crossing these three things of data and you're doing a bunch of the same computational work.

It becomes a challenge and it becomes a thing, and I think every single organization, it's gonna be slightly different. It's something that everybody needs to look at. But if we're just trying to do that in a computational way, somebody who's not as familiar with data may not know, may not be able to really think through, what are the actual ramifications of doing it in this one way versus the other? And what does that mean for the data that you're ingesting? It's a difficult thing to get your arms around, even if you really know data, and when we're trying to bring it into the hands of more people. Marisa Fish on her episode made a really good point of when we're trying to share information, not just the ones and zeros of data, but when we're trying to share information, what are we trying to do?

Are we trying to give people the ability to augment their own insights, to create their own insights, augment their own understanding? Or are we trying to share with them our or somebody specific pleads their packaged insights? And so I think this is something that comes up a lot, especially around data governance. It's like what are you actually trying to achieve with this, and it's fine to lower the bar to people getting access to insights and if they wanna swim upstream to get to the data and form their own things and stuff, but we can package these things into better dashboards and things like that. The whole death of the dashboard thing is ridiculous to me 'cause it's like, if you've got packaged insights and they haven't changed, somebody needs to own. But it's just all kind of jumbled together in my mind, but data governance to me is about exchanging information and doing that in such a way that is compliant and useful and valuable, and that people can actually understand what it means and all of that.

I feel like a lot of these conversations go into putting out data for data people instead of putting out insights for people that aren't data people. And so I'd love to talk about you a whole lot there, but is there anything that you're seeing that's been helpful or



because you've been on the ground, right? I haven't been on the ground doing this.

0:58:38 Nicola Askham

So I think you're right, and I think it's a mixture you say. And I think that if you view Data Mesh as just a way of getting the data to the same people who were using it before, then it's almost like a wasted opportunity, isn't it? This is getting it available and accessible to people as you say, to get insights who perhaps wouldn't otherwise. But I think our problem with that is exactly what you said, is that these people aren't data people. I think that's why data literacy plays a huge part in it. Obviously, I've been doing data literacy training for many, many years. I've probably only called it data literacy more recently. Because for me that was part of doing a data governance implementation, is that we had to start changing the mindset and the culture in an organization to get them to understand that data was an asset, and that we needed to manage it the way we do our other assets. And so I think I've been doing that as part of the comms and training around data governance for many, many years, and it's only in these last couple of years, everywhere you look on LinkedIn, there is something about data literacy.

And like a neanderthal, "I have to look into it and see what it was." And I'm thinking, "Oh, I've been doing that, I just didn't call it data literacy." And so it's that kind of I'm quite pleased that there is this kind of focus on it at the moment, because I think we're getting to this. Everybody should be data people, but there's gonna be different degrees of it. Some people just need to know that actually, it's just important that they put this data in correctly because somebody else in the organization is using this data and it's really important to them. But somebody else needs to understand what that data is, where it came from, what can I use it for? And they are the people that will be then looking at our data glossary that we've built as part of having data governance. So I think having these different levels of literacies are really important, but I think we need a base layer level for everybody, maybe not the cleaners or something like that, but generally everybody needs to understand that their job does involve data, even if they don't think it does.

1:00:46 Scott Hirleman

And teaching them how and why it's of value and why it's a value to their career, why it helps level them up, all of those things. I really hate the citizen data scientist concept, and I think that was something that was a flash in the pan. But I think the data citizen concept is like,"Well, sometimes in the US, it doesn't feel like it, but we live in a society, being part of a society, it's important to think about how you impact others and what role you play in that society and that" Exactly what you talked about. Are you somebody who's, it's important that we get this right? And you may not be consuming a lot of the data, but it's important because it's important for your domain, it's important for your role. And making sure people understand that, but





also making it so that more people can form their own insights and that we can have this almost like when you talk to data people that are really against Data Mesh, a lot of times, it is somewhat, we don't want those are kind of the cave people, and we don't want them to do it 'cause they're not gonna know what's going on.

And it's like, well, they're the cave people because you all have kinda kept them the cave people. Somewhat, is that we haven't had the ability to do it. Some of it's not the tooling that's been there to make it. But a lot of it's just because you're like, you're just keeping them down and so. But you can't just give them access to data, East Oldfield's episode talked about this of, if you just give them access to data and not the understanding, the literacy of how to actually use it, then they're going to create chaos because they're going to get the wrong insights 'cause they don't understand what they're actually looking at. So it is a balance, and I think a lot of what you talked about is that. Rome wasn't built in a day, but it sure was burned down in one. Of, we need to build towards this, we need to iterate towards this, and you don't have to get it all right and you're going to add value, constantly look for that value, constantly look for, why are we doing this? Don't make this a three-year thing before you see value, but how do we actually create those conversations? And that we're open about this, 'cause this isn't the way business has really been done. Governance has been viewed as this gatekeeper, even if they haven't been.

1:03:24 Nicola Askham

I think you're spot on, Scott. And I think the challenge you have is that it's not the same message to everybody, and everybody thinks that we can do one thing, so in a few weeks time, I'm gonna do a data literacy lunch and learn for a client. And that is a really hard thing to prepare for, because I'm having people from every level across every part of their organization could potentially join this session. In an ideal world, I want to know that I'm talking to Scott, Scott's job is this, so therefore I need to talk to him about this kind of data. Whereas if you are from the finance team and I'm talking all about the product data and the sales data you're gonna go, "Oh, she's not talking about anything interesting to me." So it's really hard, 'cause I think people think, "Oh, this is really easy, we'll just do one communication. Tell everybody data is important and they'll all get it and we'll all happily ever after," but it doesn't work that way. We've got to kind of tailor the messages. So obviously they look to learn. What I'm doing for the client is gonna be like a starting point. And I know that their plan is then to go to individual team meetings, to then say, "Well, taking what you learn on that, what data do you use? Why is it important, how would understanding and using that data better or thinking about it differently help you?"

And I think that is the challenge because that's an awful lot of work, and an awful lot of communication to make sure we get the right message to everybody.



1:04:52 Scott Hirleman

Yeah, I've talked about this in a lot of episodes and I keep using the phrase unicorn farts, and that to me is any time you go to write Data Mesh in something that's going to be in front of anybody that is not on the data team, is not really deeply involved, call it unicorn farts, like copy, find, replace. Because you're gonna remove that. People are like, "Oh, you say you want me to put out documents that say we're doing unicorn farts?" No, I don't want you to put in Data Mesh because it's not relevant to them. You need to tailor it to what matters to them, why is this useful. All right, Gabby wants to make an appearance on the podcast. But so much of exactly what you just said, if you try and have the same message to everybody, it's kind of like when you're thinking about who are your data consumers, right? Should the same information be in front of everybody or in the marketplace, it's for that lowest common denominator of people to get a basic understanding. But you might have advanced documentation that the data scientists are the only ones who are able to see.

And it's not that you're being exclusionary, it's just they're the only ones that are gonna be relevant to. So do you try to make it so that everybody tries to explain Data Mesh. When I'm trying to explain it to my parents, they have no idea of what data governance means, any of that stuff. So I think that's so important. I think it's really crucial to think about, it's not excluding somebody, and it's not talking down to them, but it's like, "Do they really have to care about this? Does that matter to them?" If not, then why cognitively overload them. Why put that in their head if it's not gonna be anything that's good to them? So exactly what you talked about trying to tailor that message is so difficult, because if you're talking to 50 different kinds of role types, how do you make it so that other than data is important to your role.

1:07:02 Nicola Askham

Yes, and then that's kind of really bland, it doesn't mean anything, so nobody takes it on board.

1:07:08 Scott Hirleman

Yeah, exactly. So this has been a phenomenal conversation, I really enjoyed it, and I think it's gonna help advance a lot of the way people are thinking, and also just feeling comfortable about that it's okay to not have all the answers to be moving forward. It's like, cover your butt, make sure you're not gonna get yourself into regulatory compliance, but look for use cases where that's not the case. And then I like your thing of, don't just go for the easy things to do for the first x number of months, 'cause then it is gonna hit you like a ton of bricks, it's like doing exercises and you're doing really, really light weight stuff, and then you're like, "Okay, I'm gonna do a heavy day," and then you're like, "Okay, I can't move the day after that." But is there anything that we didn't cover that you think we should in kinda wrapping up, or is



there any way you'd like to kind of in general wrap up the episode, put a button on it.

1:08:06 Nicola Askham

I think it probably came across during the discussion, but I just love to tell people, just to make it really, really blunt, is that if you're doing Data Mesh without data governance, I believe it will go wrong and you won't get the results that you need, so really, it's really really vital that you do data governance, if you're doing Data Mesh.

1:08:27 Scott Hirleman

Yeah, I fully agree. It's one of those where it's one of the pillars for a reason. It wasn't specifically called out in Zhamak's first paper, but in her first blog post, better, second one on Martin Fowler's site. It's like, no, this is one of the four pillars. If you try to skip one of the pillars, you don't have to build your columns to the sky, when you think about pillars, you don't have to build all of them up entirely, but if you just try and start it without this. Nothing works together, you're not reducing the friction, you're following the same path that has caused us to break. So I fully agree with you. I think in general, if you're of any size and you're not doing data governance, please rethink. And I don't wanna say shame on you, but please rethink.

1:09:13 Nicola Askham

Yeah, absolutely.

1:09:15 Scott Hirleman

I'm sure there are gonna be a lot of people that want to follow up with you about this and other things, as well as being a consultant and obviously, if anybody would like to do that, you've got training courses and all that stuff, where is the best place, what's the best way? And I'll drop links to stuff in the show notes.

1:09:34 Nicola Askham

Thank you, everybody's welcome to follow me or connect with me on LinkedIn, but I'd probably direct you to my website, which is just nicolaaskham.com, because I've gone out of my way over the years to try and create loads of free resources. They are all the kind of things I wish I had in the early years. So there's loads of good things on there as well as links to my podcast, my YouTube channel. So there's loads of hopefully helpful resources, so please, feel free to go and avail yourself of them.

1:10:03 Scott Hirleman

Yeah, they are very useful. I've listened to a few of the podcast episodes and they help to cement some things and for somebody who's new to data governance, it's really helped me from that aspect. But again, Nicola, this has been so great. Thank you so much for spending the time today and thank you as well everyone out there for listening.



1:10:23 Nicola Askham

Thank you for having me.

1:10:25 Scott Hirleman

I'd again like to thank my guest today, Nicola Askham, the data governance coach. You can find links to her website, podcast and LinkedIn in the show notes as per usual. Thank you. Thanks everyone for listening to another great guest on the Data Mesh Learning podcast. Thanks again to our sponsors, especially DataStax, who actually pays for me full-time to help out the Data Mesh community. If you're looking for a scalable, extremely cost-efficient, multi-data center, multi-cloud database offering and or an easy to scale data streaming offering, check DataStax out, there's a link in the show notes. If you wanna get in touch with me, there's links in the show notes to go ahead and reach out. I would love to hear more about what you're doing with Data Mesh and how I can be helpful. So please do reach out and let me know as well as if you'd like to be a guest, check out the show notes for more information. Thanks so much.