Transcript provided as a free resource by 

Data Mesh Radio Episode #112: Driving Buy-In and Finding Early Success - Kiwi.com's Data Mesh Journey

Interview with Martina Ivaničová

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?utm_source=DataMeshRadio

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 Adrian Estala- Starburst

Welcome to Data Mesh Radio, with your host, Scott Hirleman, sponsored by Starburst. This is Adrian Estala, VP of Data Mesh Consulting Services at Starburst and host of Data Mesh TV. Starburst is the leading sponsor for Trino, the open source project, and Zhamak's Data Mesh book, Delivering Data Driven Value At Scale. To claim your free book, head over to starburst.io.

0:00:39 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.

0:00:57 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 re-emphasize 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. The contact information can be found in the notes below. Thank you.

0:01:49 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.

0:02:39.2 Scott Hirleman
Bottom line up front, what are you going to hear about and learn about in this episode? I interviewed Martina Ivaničová, who is a Data Intelligence Engineering Manager at the Travel Services Company, Kiwi.com. Here are some key takeaways or thoughts from Martina's point of view. Number one, the most important and possibly one of the most difficult aspects of Data Mesh implementation is "triggering organizational change". It's gonna be hard to really get the ball rolling and really get the organization in line with moving forward on this. Number two, driving buy-in for something like Data Mesh is obviously not easy. As you are getting started, look to leverage one-on-one conversations to really share what you are trying to do and why and how this can impact them and the organization. These one-on-one conversations are crucial to developing your early momentum around a Data Mesh implementation. Number three, on driving buy-in for Data Mesh, really think about how to limit incremental cognitive load as much as possible on the developers or software engineers. If you can keep cognitive load low, you are much more likely to succeed; succeed in driving buy-in and succeed in delivering value as well.

Number four, when sharing internally about Data Mesh, it's important to focus on what it means to the other person. Using the actual phrase of Data Mesh can lead to a lot of confusion for people not on the data team. If people aren't really familiar with data and they start to really dig into or read up on Data Mesh, it's pretty easy to get lost in that if you don't have the background and experience around it. Even if you do, for a lot of folks out there. So make it clear what you were trying to accomplish, the what, the why and the how. Using data as a product, the leading concept resonated and worked well inside Kiwi.com for Martina and her team. Number five, Kiwi.com started driving buy-in by working with engineering upper management, then found a few valuable and achievable first use cases to move forward. And they have kept the incremental cognitive load low on the engineering teams while those engineering teams learn how to deliver data as a product. I think it's important to really think about where you are trying to effect the change. You wanna go to upper management whenever possible.

That's what's gonna be able to drive the change from top down. Number six, if possible, the easiest way to drive buy-in is by finding a use case that is beneficial to the producing domain. If not, then look to spend one-on-one time to really share why this use case matters, why doing things in the data as a product in the Data

Mesh type way, why that matters. If you can find something that's beneficial to the producing domain, they're gonna be much more likely to be in line on doing this. Number seven, Kiwi.com is getting software engineers in domains to commit to simply sharing their data, not even really structuring it into data products. So the software engineers in most cases, are really only focused on maintaining the high-quality data sharing mechanisms. You can read that as pipelines, but it's not relevant.

You can talk all different things about exactly how they're actually doing that, but it doesn't really matter in this case. So that is a relatively low initial cognitive load, low workload ask. They aren't at the data as a product stage, but they're still seeing initial value from what they're doing. Number eight, analytics engineers are creating the data products from the source data to satisfy consumer needs. Martina and team want to move more to software engineers handling more of the data product creation management over time, but it's a process. They plan for the analytics engineers to upscale the software engineers by pairing with them closely, but right now, it is the analytics engineers taking that data that I talked about earlier, that was getting shared, and creating the data products out of it themselves.

Number nine, it might initially be more important to find a way to evaluate and iterate on what data is shared and how than it is to get to the most complex or valuable data product. You want to build the muscle around sharing data first before trying to go too big too soon. Number 10, it's important to know what you are trying to prove out in your initial Data Mesh related deployment. It's okay to prove out you can produce data products before proving out you can build out the full mesh. Number 11, a key success metric for a Data Mesh journey could be how many direct conversations and then actions come from data producers and consumers speaking without data engineering involvement. This is something Max Schultze talked about as well. At Kiwi.com, these conversations are still usually driven by analytics engineers, but that might also change in the future. It might be that you don't need the data engineering team or the analytics engineers for a lot of the conversations. And that's a good place to be in. And finally, number 12, data governance centralization didn't happen overnight. When you look to decentralize and federate your governance, you should look to be patient instead of trying for an overnight revolution in trying to go back or trying to go in a way that is decentralized and federated. With that bottom line up front done, let's jump into the interview.

0:09:00.1 Scott Hirleman
Very excited for today's episode. I've got Martina Ivaničová here, who's the Data Intelligence Engineering Manager at Kiwi.com. We're gonna cover a lot of different things Kiwi is doing. They're heading down the Data Mesh path, and they're kind of early in their journey, and we're gonna talk about how they're transitioning in. They're

working on driving some of the buy-in and sharing what they're planning to do internally, including the whole thing of my unicorn farts principle that they're using as well of not really calling it Data Mesh internally versus showing people, what are they trying to do? Not lead them down the confusion path. So I think we're gonna cover a whole lot of ground, and it'll be a very interesting conversation. Martina, if you don't mind, before we jump in, if you could give people a bit of background on yourself and then we can jump into the conversation at hand.

0:10:00.6 Martina Ivaničová
Yes, sure, Scott. Thanks for inviting me. So my name is Martina Ivaničová. I'm leading data intelligence at Kiwi.com. Kiwi.com as a company in the travel business, which helps travelers around the globe to find itineraries. So we combine flight and other means of transport to find the best and cheapest option. And my journey, my professional career has always been somehow connected with data. Back in time, maybe even more than a decade ago, I was a database developer on Oracle. Then I was part of teams which were building data warehouse solutions in a classical legacy sense for telecom companies and government agencies. And then I worked as a tech lead for a full stack team in a smart buildings segment where we collected data from senders to visualize them. So really, my whole life, professional life has been somehow connected with data.

0:11:13.4 Scott Hirleman
Yeah, that's funny, the number of people that I talked to that are driving Data Mesh things forward for their companies, it's kind of 50-50 or maybe 60-40 of people who've been in data for most of their careers or their whole career versus a lot of people coming from other backgrounds. And I think those perspectives are really helpful. 'Cause you felt the pain of what's that classical data warehouse approach of as it was becoming less and less fun to be doing that. So I think, let's start from the beginning about Kiwi.com's journey around heading towards Data Mesh. I know a lot of people think, "Oh, we're not that far down our journey." And it's kind of like, well, nobody is, right? So let's talk about the reasons that you started to really think about Data Mesh. What were you considering and what led you to look for a new solution and maybe choose Data Mesh itself?

0:12:21.4 Martina Ivaničová
So I will maybe connect it with what you said that the old classical warehouse solution might be painful. So it was painful in some sense, but back then many, many years ago, 10 and more, the landscape was much simpler I would say, at least in the warehouses I was working with. If I compare it now to Kiwi, the data landscape is much more complex because we have micro-service architecture. We have multiple domains which are organized around business capabilities, so literally we have several tens of databases. Apart from that, we are buying data from external parties.

About the market competitiveness, we are getting data about from social networks and other sources, and we combine all this data, or try to combine all this data together. So the landscape as such is much more complex and this brings new pains. So that's why this is something which was manifested in Kiwi when I came, this landscape of one central theme, which is trying to take care of all the data pipelines. It is extracting and loading data from these multiple data sources, as I just mentioned, to one single place. There were several hundreds of these data pipelines and quite a big frustration on the shoulders of the central team. So this was like the main motivation why we started to think about a different approach and why my colleagues and I got inspired by Zhamak's blockbuster by her book.

0:14:20.8 Scott Hirleman
So I think it's kinda interesting for a lot of people, was it more you saw the Data Mesh approach and said, "Oh wow, this could help us," or a lot of people were like, "I'm trying to figure out what I could do. How can we change this versus, "Oh wow, this could solve what we're doing."

0:14:47.4 Martina Ivaničová
I think it was like the click of ideas because people somehow felt, people in those data engineering teams, they somehow felt that producers should be responsible for the data. Because they know, they have the knowledge, they should share intentionally. And when we saw it in Zhamak's articles, really similar thoughts, it just clicked, and we wanted to go this or to at least to try to go this way.

0:15:25.5 Scott Hirleman
It's funny because so many people say, "Oh, the reason why Data Mesh was such a big thing for us, with Zhamak's articles, was because it was put down on paper. It combined all of the challenges we were seeing and where we wanted to head because it wasn't point solutions, it wasn't, okay, the producers just now own the data. Well, what is that? They don't know how to own the data, so how do you enable them?" People had just been like, "Oh, we're just gonna put producers owning the data and that's it," versus, "Okay, when you do that, you have to also think about managing it as a product, and you have to enable them with self-service, and you have to stop making governance this centralized function that becomes the bottleneck," you know, it's the interplay of the different pillars. So it's funny that it's kind of that same thing of everybody in data was hoping somebody would come up with a better solution and that would work, and that it wasn't just a pipe dream.

0:16:32.4 Martina Ivaničová
Yeah, so when I heard of all the data governance, I'm always a bit scared because exactly as you said. I cannot imagine this being a central function in an organization which wants to scale. Because how can you actually ensure without understanding

the data and how can you understand the data if there are so many different domains, so many different business contexts. So I think this is really a crucial point, to decentralize the data and context and centralize the platform, tooling, policies and so on. So this is the most amazing concept I see.

0:17:26.1 Scott Hirleman
Yeah, I think this is the thing where people misinterpret Data Mesh. I don't think you are, I think a lot of people misinterpret Data Mesh of, "Okay, it means that it's doing decentralized data, therefore, everything is decentralized," and exactly what you said of like, "No, you centralize the capabilities, you don't centralize the things where you need actual context to make the decision." And you do centralize the capabilities to support those people, right? So if somebody does have a governance question, that is centralized, that you can go to a team that can help them with it to enable them. You put the decision making in the hands of the people who actually understand enough to make the decisions. And that's where I think it's difficult to say what exactly should be centralized and decentralized in every organization. But if you keep that picture in mind, I like what you're saying of so many people were like, "Oh, but this is how we've done data governance." And it's like, "Yeah, but it doesn't make any sense." I don't care if that's how you've done. It doesn't make sense.

0:18:39.1 Martina Ivaničová
Yeah. So it all sounds so cool but it's far from happening. It just doesn't happen from day to day. So you just cannot wake up one day and say, "Come on, you are now decentralized and we are now managing all of the platform." So I see this as like a big road, which we step on and we are trying to still go this path and it brings a couple of challenges, many, many challenges. But I think what is important is to get the buy-in as a very first thing. And so it's hard to get buy-in from people who never felt these pains, who are not like data engineers and so on. I saw that it was really important to start explaining, to start to talk to engineering managers, other engineering teams. To start to talk to product managers and explain what is happening, explain that let's try some use cases. Let's try to go this new way.

0:19:52.5 Scott Hirleman
Yeah. I think that's a great place to head into how you are driving buy-in. We talked about if anybody listens to a number of episodes. I keep bringing up the concept of unicorn farts; which is anytime you're gonna say Data Mesh , or put Data Mesh into your documentation, that's seen outside of the data team or the people who are working very closely with the data team there. I've had a couple of people that are like, "Oh, I'm the data enablement tribe lead, but not in the data team." That they're embedded in the business, but they're part of the data team. Everybody else, I say, call it unicorn farts, because then you won't call it anything. You'll remove that 'cause you won't actually say that, 'cause it's really stupid to say that. But I would love to

hear about how you're driving that buy-in because you are going to the engineering teams and saying, "We want you to take on extra responsibilities, extra capabilities. Is that we're gonna give you more resources." I would love to maybe walk through a couple of the different ways that you've had that conversation and gotten to an at least okay outcome, if not a perfect outcome.

0:21:12.0 Martina Ivaničová
Yeah. So there were a couple of things which we did in parallel. So one thing was we tried to get a buy-in from upper management, from engineering VPs and so on. Then we tried to come up with one or two real use cases. For example, in our context if you go to Kiwi.com and you search for some flights, you'll get a couple of itineraries and these are represented by so-called deep links. And we build a model around these deep links. We promised that let's put these data together and make a product based on these deep links where you can monitor your conversation rate and pay out to partners. And we just defined one or two data products like this, which we want to build on top of domain data.

And then when we got the buy-in from leads, then we tried to approach, or I tried to approach product managers and also team leads, explaining how cool it would be if we had this. Slowly trying to get to their backlogs some capacity to work on this together with us. What shows to be super crucial, I think people might think that this is implicit. But what is super crucial is to make the cognitive load really low, as simple as possible. So to build such a tool, to share data is super simple even for people who are not data engineers, who maybe know the same codes, the same stack as you are using, but that they don't need some deep knowledge when they want to share data. So we somehow set our dream job or dream timeline that you should be able to teach someone to share your first data, domain data in two or three days.

0:23:46.6 Scott Hirleman
And I wanna dig into a couple of aspects of that, but I think what you're saying there is a trend that I'm seeing, especially in some of the more start-upy and kind of not massive companies that are doing Data Mesh. There are a lot of companies that are in the 50,000, 100,000 people type of companies, I'm not sure, I can't remember how large Kiwi.com is. But the ones that are in the 500 to 10,000 range a lot are kind of starting with, "Hey, we're gonna work on finding these small, smaller use cases or these easier, lower friction use cases," because we're trying to get towards sharing data and we need to get people used to sharing data, and we need to learn what people don't know about sharing data, like what explicitly they don't get. And so we're gonna start with lower complexity initial use cases because we don't have to go in and do this super complicated data modeling with them. Where we have to train them for multiple weeks, and where we have to build very complicated tooling.

We wanna get to a place where they can share data, and then we can work with them to evolve what they're sharing to add additional value. But that we can get them to a place where they can get small wins quickly and you build that momentum because then people get the positive feedback and you're driving that. So is that kind of how you're looking at it?

0:25:32.1 Martina Ivaničová
No. That was our thinking that we have to start somewhere and demonstrate the value. Otherwise, maybe if you're a bad company, then that's it. It might be easier to get the buy-in here, let's go do it, do the new thing, you have two years to finish it. And for these two years, nothing else will happen. But if you're really a company of several hundred engineers, no, it's not expected that you just stop your business. So it has to go continually and bring value continually, this is what I believe. And also, as we go, we can validate the approach. So do the full cycle and reflect on what we did wrong, what we should improve. So really, this seems to me really important to start with some practical use cases.

0:26:39.8 Scott Hirleman
And so when people are kind of heading down their Data Mesh journey, there are a couple of things that people prove out earlier, or look to prove out early. Some people look to prove out that this single dataset itself is valuable, which I think is a step that sets you in the wrong direction because you're proving out your minimum viable. Whatever you're proving out, your proof of value is that the dataset itself is valuable, versus we're proving out that we can create data products that are valuable, versus we're proving out that we can create a Data Mesh. And a lot of people don't have the luxury of proving out that they can create the Data Mesh versus like, "Hey, we can get people to a place where they can actually create something that is viable from an ongoing perspective." It's not a data set that is valuable today, but starts to decompose very quickly. What were you thinking there? I don't wanna put words in your mouth. But it sounded like you were more trying to prove out the data products while you're building out the capability to scale out the mesh, but the first thing you needed to do was prove out, "Hey, we can make these products and that they can have value." And then we'll start to further expand our mesh capabilities. Is that where it kind of fell in your approach?

0:28:16.2 Martina Ivaničová
Yeah, it goes in this direction. And really what you said Scott is really important. And I am also like, I'm trying to wrap my head around this. How do you prove if it is a product or not? Is it already a product, does it already have a value, is it not just like native data where the boundary actually lies. So this is a really important thing. And I don't have a proper answer on this. At this stage, I'm happy at least with the fact that we brought producers and consumers together, so that we approached the

engineering teams. And it's sad that this data is needed, because if we combine them with the other domain data, we can build this and this and that we triggered this discussion.

And the engineering teams are deliberately exposing some data and they are thinking about what they expose, and they are bringing semantic meaning to this. Even if this brings super crucial value, I think that you can rely that someone deliberately exposed some of their data and described it in a broader business context to it like I would put it like this. The actual product I would envision as a dream state is that we will also bring together analytics engineers and embed them more closely to the domain. So we will be able to share that the aggregated data models will be part of that domain data product. This is like how I would like to go in the near future.

0:30:21.7 Scott Hirleman
And I mean, there's so many different places to go in there. But I think what you just said is you're changing the way that the producers are thinking about what they've got. That you're getting them bought into that art of the possible of, well, what this is something that it's... You're further along than a lot of people even that I talked to that are multiple years in because they're still trying to get a lot of the engineering teams to really think about, "Well, what else could I add to this?" Like not just what is the specific ask, but what more could I add to this? And then you're getting them to exchange their context with the consumers and say, "Well, if we did add this, would this add more value?"

Or what would this do and that they're really starting to say, and if they're taking on the ownership of, "Hey, we've seen this change, right?" We're not changing the data product right now. But I wanna share with you that the information embedded has changed slightly. And so how do you start to share that kind of semantic drift, even though the schema and the model hasn't changed, and in all of that, and that it's really difficult? But yeah, I love what you're talking about. You're probably further along than you probably think, I think that's the case for everybody. I think everybody's very early days and very quite immature with their Data Mesh implementation, but it's great that you're seeing initial value.

0:32:00.3 Martina Ivaničová
Yeah, I like what you meant about semantic drift, because this is something which I really think about a lot in the past month, and I also liked one of the episodes of the Data Mesh Podcast, when I think it was with a guest who was co-author of The Great Expectations.

This is important: how we actually define the contract between consumers and

producers, when schema changes are not so difficult. We also started here NKV by building a tool, which notifies you about schema changes, which is integrated in each good report and sends notifications and downstream consumers can subscribe to intended changes. But semantic drift is something which really, we don't have a good solution to this yet. And we saw it happening a couple of times, and we saw the burden, or a mess which happened when we didn't detect it in a timely manner. And one way to prevent it is like this. I would say the organizational aspect is that someone deliberates who shares the data and they are aware that someone else is doing something with that data, which was not there before. Because if you had just a central team, which is pulling data from many, many databases and the actual database owners are not aware of that, this just doesn't happen, that communication. So one way to prevent it, I see at this organizational level, but the other way, like more technical means, how to really set the contract is something I would like to figure out.

0:33:58.4 Scott Hirleman
Yeah, when you said, at first when you were saying, "We don't have a good solution to this" I didn't know if you were talking about Kiwi.com or the industry, and it sounded like you were talking about Kiwi.com, but what I would tell you is, the answer is the industry doesn't. It's not that you're behind, it's not that you're missing anything. The more that I dug into this data contracts concept, the more it's just like, "Yeah, nobody really knows how to do this semantic drift." I'm hopefully gonna have somebody on relatively soon from the ML world who's gonna talk about drift in general as a concept within ML and how we can start to think about drift and in its model drift and things like that, but how can we start to get people to think about something has changed?

How can we get people to understand who is consuming what, and that you do have that regular check in of if you see something say something, like, "Hey, there's something that's kind of interesting." An episode that was recently released was with Andrew Padilla. And we were talking about the experiences of the organization. The way that we shared information about what is happening with the organization itself. What are the experiences, how is it interacting with the real world? So we're always gonna have to put that into data in a certain way and so you always have to abstract it. But we can get better about detecting when things have changed and we can get better about, "Okay, was this a blip or was this not?" People think about...

A big one, I used to work in the tech space and we tracked monthly revenues for Taiwanese companies. And depending on when the Chinese New Year fell, it would either be in late January and some of February or all of February. So depending on when that fell, if it had been in only February for two or three straight years, and then it fell in half January, half February; that January would look terrible. And so it would

be the year over year comparisons that look like this is gonna be a really weak year. It's gonna look really terrible versus like, "Hey folks, the numbers say this, but the reality of the world says this." And I think that kind of thing of, "Hey, I know that this number is coming through. Our sales yesterday looked like they were insane, but it's because there was a time zone change so we had an extra hour for this specific set of information." If you're not working on UTC, I do think everyone should work on just UTC, but that's my own personal thing.

0:37:01.0 Martina Ivaničová
Agreed.

0:37:01.7 Scott Hirleman
But you say, "What were our sales yesterday?" And on those days when there's an extra hour or there's an hour fewer, and all of a sudden it's like. "Our sales were down 5%." It's like, "Yeah, because we only had 23 hours instead of 24." All of those little things, if you see something, say something. That communication, we haven't had that with data. Sorry, I just get very excited about this.

0:37:30.6 Martina Ivaničová
Yeah. So just to add this, we are patching it, maybe by trying to set some DQ checks. But even this is something I think the industry requires a better solution as such. Yeah, I agree with you.

0:37:46.5 Scott Hirleman
Yeah. Chris Riccomini talked about with the schema changes that they literally had a CICD system that even before somebody could propose to commit a change, it would go through and say, "No, this breaks the downstream." But can we have that as well? If we push the ownership to the producers, hopefully they'll know when those changes are happening so they won't make those changes without letting people downstream. But sometimes you don't really know that people are consuming, even if you've got this product aspect. I think we do really, really need to get better at that. So I did wanna circle back to something because you have been successful in driving buy in. Most people aren't, so I wanted to dig into a little bit, or a lot of people, that's where they're getting blocked, is to get domains on board. You talked about how you found very good use cases.

Like, how did you find those use cases; is it that you're not going to the ones where there's necessarily gonna be the most value. But you found the ones where there's value and they're willing, or how did you talk about driving the buy-in at the highest level. Like, you've learned a lot. You might not think that what you've learned is extremely broadly applicable, but I'm sure it is. If you can tell people about what you've learned and what might be some advice to try some things, I think that could

be really, really helpful to people.

0:39:33.0 Martina Ivaničová
So first of all, we are not in a green field, maybe this would be much easier, but we are in a brown field and just maybe a bit of context. Luckily, there are data engineers building the data models, which was set up when we started with this. So, like thought about what is the next requirement, and which we knew it's needed, or it was, combining data. We took from the backlog something which was already there, and said, Let's do it in a new way, and we will enable this and this, and we will prevent these pain points which are from the past. And really the success lies in talking to people, talking to people one on one, explaining what we are doing. Important is to really create one one-pager, where you try to explain your thoughts on one A4, and crucial points, not really saying, this is the Data Mesh concept and so on. We just mentioned some of the most important thoughts. So people don't need to read blogs and books, just really highlight what and why and how we are trying to achieve. And I used this one-pager a lot, like sharing it whenever some question pops up on the flag. And apart from that, as I said, crucial one-on-one talks with individual stakeholders, whoever it is, either engineering leads or product managers or analytics engineers, and so on. Really one on one.

0:41:45.4 Scott Hirleman
I think Max Schultze talked about this in his episode, too. If you're trying to scale and you're trying to really take this very, very wide, one-on-one conversations don't scale. Right? They just don't. But to start your initial momentum, Newton's second law of motion, those things that are at rest tend to stay at rest, and those things that are in motion tend to stay in motion. To move from at rest to in motion you have to put in a little bit more effort, you have to do those one-on-one conversations to build the momentum, and then you can build out more on that one-pager. Right? I'm assuming, just unless you're the person with the most brilliant foresight in communication and Data Mesh and everything, that that one-pager has evolved, right? That you've changed what's been in there, you've changed when people are saying, Oh, what about this thing, or, I don't get this, or whatever, that you've evolved that considerably. Even if it isn't in English, especially if that one-pager is in English, if you'd be willing to share that with folks, so other people can look at how that's being shared, I think it could be really helpful. Because so many people are confused about how to have these conversations, but I think a lot of what you're saying is to extract their context.

Instead of selling, you're like, "What are we actually trying to do here? Why does this matter to you? What is expected of you?" So, we're not saying we expect you to deliver the world, but we're gonna work with you, we're gonna give you the support. Is there anything that's kind of stuck out in those conversations too?

0:43:33.3 Martina Ivaničová
When you say that the one-pager evolves, so did the title evolved. The text evolved as well, but the title really evolved because really we realized that when you start the conversation and you mention Data Mesh. It's like just a buzzword, and as you said “unicorn farts”. But what people understand, what I observed that people really understand, is seeing data as a product. Because the product is something almost everybody could relate to. Okay, it has a responsible, let's say, product owner, someone cares whether customers are satisfied, someone cares how much effort it costs to produce a product. So, really, if we talk about data as products, this is what people can easily get.

0:44:36.5 Scott Hirleman
Yeah, and I think there was somebody who posted on LinkedIn recently about, we should just stop using the phrase Data Mesh and just start calling it data as a product. And I fully disagree when it comes to talking to the data people, but when you're talking to the non-data people I think that's exactly the point, like when you are talking to them, what does it matter to them? There are people who are saying, "Oh, it's very disrespectful to somebody to not tell them that you're doing Data Mesh." And it's like most of the people don't care, and if they really do want to dig into it, they'll figure out that you're doing Data Mesh and you're not hiding it. But you say, "Hey, we're taking this new approach, and what we're gonna do is data as a product." But most of the people don't care that you're doing Data Mesh, and it is very easy if you say, "We're doing Data Mesh." If that's what you're selling, you're going to fail, because most people are gonna get very confused. I use DuckDuckGo as my search engine, and when I type in "Data Mesh", the first I think of Zhamak's article, I think it's not even her first one, I think it's her second one on Martin Fowler, is like the seventh result. And the first six are all vendor crap.

And they're all not correct, not appropriate things. And so, I think that it's so easy to head people down the wrong direction by trying to be like, "This is what we're doing." Versus, "This is what matters." And it's good that you're having that same thing. I know a lot of data people want to share all the really interesting aspects of Data Mesh but I don't think people really care about, "And this is the way we're federating governance." Versus, "Hey, we're giving you control. We're gonna make it so that you can make the decisions and that you don't have these bottlenecks and that we trust you."

0:46:33.5 Martina Ivaničová
Yeah, and even so within data people, I'm still actually hesitant in my context to call it Data Mesh. Are we really doing Data Mesh? Because we are covering all the ideas which Zhamak brought up. You have incentives, you have immutability, the data

should be bitemporal, so there're so many concepts that we should wrap our head around and the tooling, we don't have it, at least we don't have it. So still sometimes I just, "Okay, can we call it Data Mesh or can we call it just data products? Are we actually allowed to call it a Data Mesh?"

0:47:29.4 Scott Hirleman
Well, I think that self-reflection means that you are on the Data Mesh journey. And that it's like, when is a Data Mesh a Data Mesh? When is a person an adult? Do you say that it's when they hit a specific age or when they've matured to a certain level? And so it's like, when is a Data Mesh a Data Mesh? Well, if you're on the journey and you're really focused on that, I think it's okay to say, "We're working towards Data Mesh." And that it doesn't really matter at the end of the day. I don't really care. One thing that a lot of people expect of me is that I'm really, really bought in that Data Mesh is the way. I'm like, I'm bought in that we should really, really evaluate if it is the way. It deserves us to really put it through its paces to really consider this, but it might emerge that there's some big holes in it, or that there's other ways to approach this that works better. And that the more that we can get people sharing information, the more that we can iterate towards what works best.

And that it doesn't really matter at the end of the day if you're doing Data Mesh or if you're not, if you're delivering business value. If you're doing the right things for the business, that's what matters versus this ultra pure of, "Is this Data Mesh or not?" What I don't like is when people jump in and say, "Oh, I'm gonna completely change the Data Mesh concept for my own purposes and still just call it Data Mesh." But I think people on the journey like yourself, it doesn't matter. It's not a big deal if you're like, "Well, we haven't put bitemporality into every data product." Especially 'cause you said the tooling, you don't have the tooling, the tooling isn't there. Zhamak said this multiple times, "The tooling really needs to evolve." And vendors are still just trying to sell you what they've got and say, "Oh, but it's Data Mesh." Versus like, "No, we've gotta rethink the way that everything should work together so we can head down this direction." Sorry. It's just, this is a big passion thing for me.

0:49:45.9 Martina Ivaničová
Yeah, so these are my doubts. On the other hand, when I can highlight when I really felt happy was when I saw that things are happening without data engineering intervention, I would say. So it really happened that the product manager came up with an idea, "Oh, I'm interested in the data about how many times the ancillary price changes." And he just put it together, like he approached an analytics engineer and software engineer, and they just did it together. So the software engineer from their respective domain was able to expose additional data, the analytics engineer was able to build a wizard output on top of it, so really when things start happening within the domain is when I have a really good feeling.

0:50:47.7 Scott Hirleman
Yeah, I think that's an important success point of, part of what we're trying to do is extract the data engineer from playing telephone and being involved in everything. And I bring it up quite a bit, but like the Office Space thing. I don't know if you've ever seen Office Space, but there's a guy who's describing his job and he's like, "I take the requirements from the customers and bring them to the product or whatever, to the engineering." And it's like, "But do you actually take the actual physical paper?" "Well, sometimes when my assistant isn't there." But that's kind of what it can feel like for the data engineering people, or like, "I'm not quite as... " Not in the useless sense, but that you're constantly in the middle of a conversation when, if you can extract yourself from that and get the people who. Instead of trying to translate from what the consumer wants to then into what that would mean from data and then translate that to what the producer wants. If you can get the producers to a place where they can share that and where the consumers can talk the producers language as well.

And that it's easy to put something out there and test it, and that you have that change in culture from, "This is data, therefore it is right, and therefore, it will never change." Versus like, "Hey, let's talk about what your needs are, let's talk about meeting your needs." Instead of, "I'm going to put this thing out and you can't change it and evolve it." I think that you're sparking again the positive, happy feelings of somebody who's being successful, even though it's early, right? Like we've talked about that, you're still early, so I wanted to dig into exactly the actual machinations, like, how the data products are created. Because you mentioned analytics engineers, some people think that analytics engineers are data engineers and things like that, so like, what are you seeing? Who is producing the data products? Who's developing them? Is it that, it's the data engineering people originally and then you hand it over to the analytics engineer. The analytics engineers are building the downstream data products from source aligned data products. Like, just would love to understand, kind of, how people are playing in this world, in this space.

0:53:17.7 Martina Ivaničová
Yeah, one of the toughest things in Kiwi is to explain to everybody that data engineers are not doing anything with data. Because they are data platform engineers, so another role which we have, are analytics engineers, who really take source alignment data and build data models on top of them, which is like, some transformational calculation. This role is something which is quite crucial, it was very crucial before we started with this world Data Mesh idea, because the output is something which can be used in a self-service way. So, really an end user can come and use the data product prepared by analytics engineers.

So, the stage we are currently at is that, we have learnt from some of the software engineering teams, they are sharing source aligned data. Analytic engineers come into the game and they put additional layers. Either they are like aggregated data models or data needed for some particular use case, but they put an additional layer on top of it. This is my mission for the upcoming month is like, inject more. They actually embedded some of them in the domains already. But to inject more of these analytics engineers in those domains, so this becomes one thing, like really like software engineers and analytics engineers working together. Not just like, I exposed my source aligned data and you came and took and did the next step. No really, this has to be like one tandem, working on the same thing.

0:55:42.0 Scott Hirleman
Yeah, 'cause it sounds like what you're saying is that the producers are committing to sharing data, but not actually their data as the product, right? They're sharing that information, then the engineer actually shapes it into the product so it can be consumed. And that you want to move more and more, so that the software engineers can be partnered with them and paired with them. I think this is a common issue of... I did a community meet-up literally around who should be your data product developers, because if you just say, "Oh, the software engineers should." Do they really understand how to model data? How to share it? Like, what do people want? All of that. So you start with very un-complex, uncomplicated data products, so they get used to sharing, but is there really that concept of, "I'm going to build something that I know people want."? Like, can they really understand what data consumers want?

Which is what the analytics engineers are translating for them, but how do you train the software engineers more and more? Because the analytics engineers, even if they're embedded, they're obviously very intelligent, but so much of the context around what's going on, is in the software engineer's heads, so how do you translate it then to a product? It's difficult.

0:57:08.6 Martina Ivaničová
Yeah, it's difficult. Like, for the very simple things, we already saw that the software engineer was able to go the World path. Really the analytics engine was not needed and the outcome was really like a usable product, accessible by anybody in the organization and usable. But for more complex things, I don't have this figured out, to be really honest. Yeah, this is something which is on my radar.

0:57:43.4 Scott Hirleman
And there's an episode that hasn't come out yet, but it's coming out relatively soon with some folks from Leboncoin in France, and they were talking about this same challenge of like, "How do we train the software engineers to do this?" There isn't

anything out there. I haven't found any content around how you make it so software engineers can really start to understand data. It's kind of this thing I've talked about where software engineers can't share their data as a product in a lot of instances, because they don't understand what data consumers want. They can't even empathize with, "Oh, this change is gonna break things downstream." Because they don't know about it, so now we can start to share the information about, "Hey, here's what's being consumed downstream." But like, the why and how and how could I make what they're consuming better and easier, I haven't seen anybody who's concretely talked about this.

I'm sure people have figured it out somewhat behind closed doors, but nobody's sharing it. And maybe they think that it's their crown jewels. But in a lot of cases, people are just like, "Well, we don't have it fully figured out, so we can't share yet." It's like, "No, no, no, you absolutely can share, you don't have it all figured out if "Like, a lot of what you're saying, even just the stuff around what you're doing to drive buy-in, that's gonna be super helpful, even though you're like, "Well, we haven't figured out for every domain and we haven't figured out exactly "

Where you are, you've progressed a lot further probably than you would've thought coming into this conversation compared to where everybody else is. You're a lot more mature than I would say than a lot of people on this journey, even where you are, even though it feels that you're early in it. So I think a lot of what you are saying is you're reflecting back what other people are feeling. You're not alone.

0:59:48.6 Martina Ivaničová
Yeah, I think we should be happy at least from the baby steps, because at this stage, the success is already there. If the software engineer thinks about the data and puts the really meaningful description to the shared data and even maybe things about reshaping it within a small transformation like something which is normalized. To normalize there are really small baby steps as something I am happy personally to see. That we should aim for this bigger alignment, and yeah, let's see how it goes in a year or two.

1:00:33.9 Scott Hirleman
But that continuous incremental value. Karolina Henzel in her episode was the first person that I at least noticed saying this. And I've been saying it more and more of big bang approaches and data don't work. They don't work for many, many reasons, especially 'cause the sources are constantly changing as you're trying to set up this big bang. So everything just keeps shifting and all of that. So you're delivering good value already. And so that just builds your momentum. And it builds you being able to go back. From an incentivization standpoint, when you do have domains that are successfully sharing this data, is it that they're typically the consumers of their own

data or is it that other people are consumers? 'Cause that incentive is tough. Getting people bought in, not just this benefits the organization, how does it benefit them? 'Cause people are kind of inherently selfish.

1:01:35.4 Martina Ivaničová
So absolutely easiest is when the domain is directly consuming the data. So because then you have really strong motivation. But this is often not the case. Usually, data from a couple of domains and then the end consumer is sometimes someone who is even out of R and D organization. So I don't know. Maybe in the finance department, revenue management and so on. So this is really, really hard to tell. One group of people have to do it and someone else will benefit. That the buy-in is not so strong as if they can directly benefit from it. But the huge question mark I had and I think I got it wrongly from the beginning was like what to do with the data which are actually from outside and how to find the owner for them? And I had one specific case where I thought that there is a specific engineering team which is closest to the nature of the data and they should own it. And I failed to persuade them that this is their thing, their data product. And this is something where I see, for example, that I was successful in persuading consumers to take ownership of those data. But it took me some time to actually figure out who should own the data.

1:03:17.3 Scott Hirleman
Data hasn't really been owned historically. It's been like maybe that the data team has owned the actual technical ownership, but nobody's owned the concepts. And then trying to shift that off of the data team, they're like, "Well, shouldn't the data team just own this?" And it's like, "No, 'cause you actually own the concept. You understand what this is." The data team can't understand the concept of exactly what's all changing. So, yeah. And I wish we could all have the financial services model where literally the teams just literally pay each other. Instead of like, "Okay, we're gonna work with you to make sure you have the resources." No, they've literally just transferred money to and from each other within the company. It's kind of crazy, but it works.

1:04:06.5 Martina Ivaničová
Yeah. That I can imagine that, actually.

1:04:10.4 Scott Hirleman
It just makes it clean. It makes it a little bit more mercenary style of just all about the money, but. So we've covered a whole lot of different things. There was one thing that you touched on a little bit earlier that I think would be a good kind of wrapping up point, which would be around the tooling. You were the data platform lead before this current role. So right now we don't have the tooling. What tooling do you think is missing? How have you evolved your stack if people think of the tangible side of Data

Mesh of, okay, I'm just gonna focus on the architecture and the tooling. So what advice would you give those people, and what's been helpful for you this far?

1:05:00.5 Martina Ivaničová
So we just used what we had. If I can really bring it to the particle stack, we are running on GCP. And we just said that to domain teams, like, "Use pops up to stream your data." Then consumers can subscribe or we can then publish them to BigQuery from pops up and then they will be available in, yeah, a scale interface for like other use cases. For those teams which didn't want to stream their data, they had some technical issues or they didn't really fill their application. We told them to use Airflow to expose data as a BigQuery table, basically. So to extract them from the source system, do some small transformation, exclude data which are not needed, write a DAG which publishes the data to either BigQuery or a GCS bucket. And we are trying to use the tooling which we have, such as Google Data Catalog, where you can find the described data sets. You can see how much time the particular data table is accessed, and so on. So, we are trying to persuade everybody to stream data, because then, it allows for multiple protocols or types of ports, which you can access. But if this just doesn't go, we are also fine with the other approach.

1:07:09.3 Scott Hirleman
Yeah, and what tooling are you thinking that you're missing, right? You talked about not having bi-temporality right now. You can do that somewhat in the source, but it's also difficult, like doing time series data in general, they're not good. There are a few databases that try to do it, and it's still very difficult, and things like that. So is there anything where, right now, you think that is the biggest gap out there that you haven't seen a good solution for. Even though you're using what you've used, so you haven't gone out and really tried to find a bunch of new things, but is there stuff where you're like, "If we had this, we would buy it, but there's not anything out there"?

1:07:53.0 Martina Ivaničová
Yeah, I mean, like a lot of tooling. Really, as we just discussed, there's something which will do this feedback loop and point us to not use data products. Which are not incentivized, which will be then decommissioned or not put so much effort in for the development. Also, there's a unified layer where you can approach. Like, you can get access to the same data product, but using different ways of approach, different ways of querying or consuming those data. So, yeah. I think we are missing a lot of stuff, and this is the reason why I'm still so hesitant saying what we are actually doing.

1:08:49.3 Scott Hirleman
Well, and I think that if there was somebody to do a unified data marketplace, as well

as additional complex interaction points for the data scientists and things like that. Where you can give people feedback about a product, but the owners can see a lot of very clear specifications or very clear metrics around what's being used. And that there's one thing that also, the people at ABN AMRO are talking about, that they've got this integration layer. Or it's the data integration and access layer. And it makes it so that you can actually combine data from multiple data products within the marketplace in an easy way. And so, I think that's something that a lot of people really need much more of, 'cause it would just make things so much easier, so that you don't have to do custom integration and things like that. Well, again, Martina, this has been so phenomenal. This has been a great conversation. We covered a whole lot. Is there anything we didn't cover that you think we should have, or is there any way that you'd kinda want to wrap up the episode and kind of everything that we talked about?

1:10:12.5 Martina Ivaničová
The buy-in is the most important thing, or the triggering of the organization change in it, and trying to get the people understand the whole concept, or at least give it a try to understand it. And yeah, this is my takeaway so far.

1:10:36.3 Scott Hirleman
Yeah, I think, yeah, if you've got the momentum of people who are bought in, you can figure out ways to do it, right? That's super crucial, so I'm sure there are gonna be lots of people who would like to follow up with you. Where's the best place to do that? What would you like people following up with you about?

1:10:55.9 Martina Ivaničová
So, the best place is probably LinkedIn, so please contact me there. And I would love to talk to any of you if you figure out. For example, the semantic drift question, which we discussed before, or even any other points which we touched. Like how do you ensure immutability, how do you build this integration layer, how do you work with these incentives, do you have incentives on data products? So, I would be happy to have a chat with you.

1:11:29.5 Scott Hirleman
Awesome. Well, again, thank you so much for your time today, and as well, thank you, everyone, for listening.

1:11:36.5 Martina Ivaničová
Thank you, Scott. Thanks for having me.

1:11:39.5 Scott Hirleman
I'd again like to thank my guest today, Martina Ivaničová, the Data Intelligence

Engineering Manager at Kiwi.com. You can find a link to her 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.