



Transcript provided as a free resource by



## **Data Mesh Radio Episode 21: Data Mesh in Practice: Insights from Zalando's 2 Year Journey with Data Mesh**

Interview with Max Schultze

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](https://www.starburst.io/info/data-mesh-resource-center?utm_source=DataMeshRadio)

### **Scott Hirleman**

I want to thank my guest today, Max Schultze, data engineering manager at Zalando. In this interview Max shares a lot of great info about how to get started implementing a Data Mesh, how to maintain momentum and an implementation, where some major pitfalls, especially early in your journey, so much more. Lots of great points to consider regarding your own journey but also lots of sensible perspective about your mileage may vary or since he's in Europe is it your Kilometerage may vary I don't know. I think the two big takeaways were one have and build empathy, and two try not to get ahead of yourself. Thoughtworks like to say, "Think big, start small, move fast," and I think that applies in many many aspects of Data Mesh.

Make sure to check out Max and Dr. Arif Wider's book, Data Mesh in Practice, which is free via Starburst sponsor download, or also free if you've got an O'Reilly subscription. And, of course, check them out at Datanova: The Data Mesh Summit on February 9th. They'll be presenting "From Theory to Practice" on their O'Reilly market research book. That's at 1:30 PM EST, 6:30 PM GMT. Max and Arif are also doing a Data Mesh related training via O'Reilly in February. Links are in the show notes as per usual so please do check that stuff out.

Back to the interview! Zalando was in a bit of an analysis paralysis in the early days when Max first came across Zhamak's initial blog post. And it really spoke to the pain points that they were having. Max realized that, even as a technologist, the technology side is not the most important aspect to fixing what was ailing in them regarding their Data Lake setup. Pulling an inverse conway maneuver for the entire company wasn't going to work, but small inverse conway maneuvers would help move things forward. Their first steps were to take a step back and make very conscious decisions about data, something that I've referred to a lot as intentionality. Some people have given me a little grief for how much I say intentionality, but it's very important relative to Data Mesh. So they started building more scalable and scaled data products with teams that were already building kind of data assets, data artifacts in some sense. You know, you might call them data products if you're not



Transcript provided as a free resource by



looking at it from a Data Mesh terminology perspective.

Max used a three-pronged approach to move things forward. First, start with knowledge sharing about why this all matters, e.g. to an application team who is using your data, especially how crucial it is to that team and how the changes have impacted your team. A blameless approach is key. We discussed the need for empathy, but that empathy isn't built via haranguing and demanding. It's built via sharing a vision and context. The second prong was a clever, I think Max called it slightly devious, use of technology. The starting realization here was that 1 to 1 knowledge sharing could not scale, if you're at a large company, I think you understand this, you can't have 1 to 1 meetings with everyone, so on the techbase solution. For every pipeline of data, there was an automatically generated schema. But, every schema was strictly controlled by the data producing team. So if a consumer of the data wanted a change, say to add some new column that existed but wasn't part of this schema, the only people who could make the change were the producing team, the data engineering folks couldn't make that change. So, the consumers were forced to go and speak with the producers and thus build direct one to one relationships, and that created much more communication up and downstream about what needs were, what changes were coming, things like that. The third prong was more a result, rather than something that they planned out ahead of time, but as data producer teams saw more and more usage of their data and wanted to serve their newly realized internal customers well, they realized they would need a data engineer or two within their domain to really create the highest quality data products that met the demands of those consumers. Not every team needs a data engineer but there are some where the domain just didn't have the deep level of expertise, especially there are really complex data product considerations in some domains.

Max closed on a few key points, and I really like these, so this is something to keep in mind, maybe write down in your Data Mesh Journey journal. First, Data Mesh is a journey. Even though Zalando is much further down the path than most they are still far from an end state and there probably is never really an end state. There's not a stopping point. This isn't something that you have an end specifically in sight. You're constantly evolving as new things with your business change. Second point would be that empathy is crucial. Having it to start with everyone that you're talking to—everybody is trying their best in general. You have to assume that positive intent, and having that empathy from the start, and then working on fostering it between other people as well. So much of the challenges with data are producing teams not even knowing about how people are using their data. Focus on a blameless culture when educated, not you broke this, it's you made this change which was important for your application, and as a result the downstream data consumption was broken. So let's talk about how we prevent that in the future and how important your data is to



Transcript provided as a free resource by



others. Third point technology is important but far from the most important aspect to getting Data Mesh right. Starting with technology only is likely to lead you in a wayward direction. Fourth, start from the knowledge share, but look for ways to scale that knowledge sharing. Again, Max used that inverse conway maneuver. But, in general you want to look at how you create that data literacy and data sharing culture ,but that it's not all one to one conversations. And the last point is, try not to get ahead of yourself. Max mentioned this on the governance side specifically, but I think it's important for all of the pillars. If you just give the domains data ownership but not the tools and resources to handle sharing their data, it won't go well. If you try to make a platform capable of dealing with a thousand data products before your first one is in production, that's headed for disaster. Have patience,move forward together again. Think big. Start small. Move fast. With that, I think you'll really enjoy this episode.

### **Scott Hirleman**

Hi Everyone! I'm really really excited about this episode. I have the most prolific content creator in the data mesh space. This is Max Schulze from Zalando. Max, if you could give people a bit of an intro to yourself.

### **Max Schulze**

Hey Scott, it's a pleasure for being here with you today. I'm really really happy to drop on the podcast together with you and it's been amazing that I got the invite. And yeah, absolutely I'd love to give a short intro to you folks out there on where I'm actually coming from and what's my background and what makes me talk about Data Mesh at this point. Generally speaking I am a data engineering manager at Zalando, as Scott already mentioned, for those of you who don't know Zalando, it's the biggest European online platform for fashion and it's also a huge tech company. They have had a very strong focus on data for quite some time. And that also brought me into the picture around six years ago if I'm not mistaken. Basically at that time I started working on the first version of the company's Data Lake and that was just after the company started moving to the cloud so it was a very early journey, like a really greenfield project that we started with a very junior team playing around with some first like data processing technologies and cloud computing and what have you. But then, basically I've stuck around with the topic ever since.

So I've stayed with Zalando for the whole time, and really started rolling out some bigger data computation engines like Spark, like Trino. I started rolling out bigger storage layers for the Data Lake, most of our stuff is on AWS or S3 where by now we have like several petabytes of data and a massive inflow of data massive usage of data each and every day.



Transcript provided as a free resource by



### **Scott Hirleman**

Yeah that's awesome and, as I said, Max is the most prolific of the content creators. I think he's done something like seven or eight different meetup and/or conference presentations and so there's a lot of great information out there. But Max, today, the reason we wanted to chat was to share the Zalando story and also Max and his sometimes co-pres presenter Dr. Arif Wider. Hopefully I pronounced that well.

### **Max Schultze**

Yes, that's fine.

### **Scott Hirleman**

They have also put out a really great kind of book. That's not a 300 page type of book but a compendium around kind of what is Data Mesh and that you can really dig into that but not have the full 300 page experience. So would love to talk about that process as well. But why don't we start with what you were seeing at Zalando and what led to the challenges around why you started looking at Data Mesh. Is that a good place for you to kind of jump in?

### **Max Schultze**

Yeah I think it's actually a perfect starting point, so when, as I already mentioned right, I've been working in a Data Lake setup for several years before we started even considering some things like Data Mesh even before the term was born for that matter and it was quite interesting because of course like if you're working in a centralized Data Lake setup you run into quite a lot of challenges as well over time. You fix a lot of things, you build a lot of amazing products for the data partitioners within the company. But of course the organizational setup also gives you some restrictions on what you're actually able to do from the technical side and this is really something that I also have to say, yeah I'm a huge tech geek, I'm really coming from the engineering side of things. So, I've really always been looking at how we can turn this infrastructure into a platform, into just like engineering topics for that matter. But also I had to realize that eventually there are a bunch of things that come more from an organizational angle. That's something that I also realized at Zalando, and this is something that we've been discussing a lot with several colleagues within the company and eventually this Data Mesh term popped up, right?

It was around the time when Zhamak published her first article on the topic that people started talking a bit about this new Data Mesh and the need that was coming out there, and they were discussing a couple different approaches on how to address certain challenges. The interesting thing for me then was to look into where this is actually coming from to understand why Data Mesh? What is the problem I am actually trying to address? And this was when I realized that a lot of these challenges that were mentioned there were actually very similar to the challenges that we



Transcript provided as a free resource by



already ran into over the time and even more so. Some of the solution ideas that were presented there as well were also partially overlapping with the things that we came up with already and that's what I found very very interesting because for the first time it made me realize that it's not just us who are having some very specific Zalando challenges that only we have in our setup, but that it's actually something that a lot of people and a lot of companies are actually facing the same challenges at the same time and this is where the whole discussion around Data Mesh really started.

### **Scott Hirleman**

Her article, I know a lot of people have said they felt seen after it because they were like: oh we're not alone here and then yeah she was combining a lot of different aspects from software engineering as to how they might work together for you know data management. But then, if we do this it is going to introduce this challenge. So it's kind of shaved off a number of years for people's iterations. Okay, you just push the domains to own the data. Okay, but they don't own the actual infrastructure and they don't own the actual pipelines so there's still that cross boundaries. Yeah, all that exact stuff. That's really helpful and you know do you have any thoughts on where you started you know, looking back, I know it's been a few years, but like if someone were to think about where they were to get the momentum behind this where to jump into their POC, is there, I know you've given a couple of examples of stories of some of the exact challenges that you saw I think that might be useful around that as well. But, you know if you can interweave some things to help people get over the hump to actually get going because I know a lot of people are in analysis paralysis around that.

### **Max Schultze**

Yeah, yeah, yeah, no absolutely because that is exactly where we were coming from at this starting point as well, right? and I think we can also consider ourselves lucky that we had been at that point already when the whole discussion in the community essentially started. Because that is essentially also where I picked it up when I realized, there's a lot of discussions about the theoretical foundations on the topic. But what people were really lacking at that time was some practical examples. So all I did, and that was like really my first involvement with the whole topic, was I went to a couple conferences and said this is how we are doing stuff and this was not necessarily even like largely entwined with the strict theoretical foundations and following everything by the book but it was more like we are picking up the general idea. And we are picking up some of the things even from the technical side again because that was also my background to really put these things into practice and this is something that I think we have been one of the first to actually share our practical story for that matter. And that was something that was incredibly appreciated by the community to really look at some practical examples and look at



Transcript provided as a free resource by



some solutions not necessarily just to identify if this is the exact challenge that we now also have and to see the exact path out of that, but to really trigger a thought process right.

### **Scott Hirleman**

What are the questions, like that's the thing is if we're trying to provide all the answers for everybody. It's not going to work. But if we can work towards providing the questions: here are the questions you have to answer for yourself. I just interviewed Juan Sequeda yesterday and he was talking about centralization versus decentralization, and there's 50 different things that you have to specifically know, some very high level, very important, some not nearly a high level or is important but you have to figure out that sliding scale for yourself because it is dependent on your organization. But exactly what you said you've been talking about the practical and in practice and what does this actually mean and and it has been very helpful to a large number of people. So I think everybody, including myself, is very appreciative of that. But say somebody is feeling these pains, did you start from the people process side? Did you start, I mean you're a technology person you probably wanted to start from the technology side. Did you do that or would you recommend that you know, kind of 2 different questions there.

### **Max Schultze**

Yeah, yeah, one of the most painful realizations that I actually had very early throughout this process was that even though I really wanted to make this about technology, it absolutely was not. Like it was really being driven by the people process, by the organizational setups, and they were in the end much more important than what I wanted to make this about. Of course being in the position that I am in the Zalando perspective, somehow I needed to take technology as a lever to start pushing organizational boundaries and this is really something where on the one side. We try to get creative and we try to understand what are certain things that we can do to force people to change certain behaviors or start moving into that direction but there was like more a trick I would say to leverage technology to push for organizational changes for that matter.

### **Scott Hirleman**

So it's a moderate scale reverse conway maneuver rather than trying to do the full, not trying to do the full lever but it was like the little stuck parts that you're going around with the crow bar instead of the huge like kind of Jack to lift up the ten thousand pound thing. Okay, that's interesting! Can you take me through some of the things that you were doing there or where what your learnings were about, the kind of initial process of how you were dealing with people and and some of that? Those friction points and what you found useful and maybe if you've got anti-patterns as well like things that didn't work that could be very helpful.



Transcript provided as a free resource by



### **Max Schultze**

Yeah, yeah, so it was quite interesting because of course like Data Mesh is coming from a couple different angles but like one of the core things that I always keep repeating is to bring yourself into the position that you make conscious decisions about data. That really starts with coming from, let's say a data swamp setup where you have lots and lots of data with completely unclear responsibility, where you have no clear ownership for who's actually taking on certain challenges for several thousands of datasets that you still have by now. But you don't have clear products. You don't have clear ownership or anything like that and this is of course one of the hardest things to change because this is something where you really need to get into the head of the individual. And this is really something that we took as a first angle.

Let's say to start doing some POC, some pilot projects where you really started to address some of these things and to really start building out your first data products so that you can have examples that you can share with the rest of the company and that you can showcase as well. This is an approach that is actually working, that is actually leading to valuable input for many other teams that are using this data. This is what we did with a couple of our teams that were already really close to us. On the one side, I have to say we were comparatively lucky with the organizational setup that we were coming from because from the time that we still had business intelligence teams that were mostly taking care of the data analytical side of things in the company, and I think there was like even before I joined to be honest, the company made a decision to put decentralized BI teams in place that are responsible for the analytics within the respective business units.

At the same time, the company is very well structured in the sense of having business units that are responsible for certain parts of the business, certain processes, certain areas. Which already gives you a very nice organizational setup where you have a certain business area and a certain domain and you already have an embedded data team that is responsible for the data of the domain. That helped us a lot to reach out to some of these teams and start building the first data products. Some of them already had data products even in place they were like BI teams with product managers that were already talking mostly about tables in a data warehouse and there needed to be a huge upscaling in terms of technology and bringing them like to the newest state but that was something where we could already rely on a very nice foundation when it came to the organizational setup.

### **Scott Hirleman**

Yeah I think that's a common challenge as well, is people trying to go to the domains and get them to share their data but that they're not necessarily bought in as to



Transcript provided as a free resource by



what that gets them and so it's nice that you already had somewhat around that but do you have any insights as to when you were talking to them about when they were upscaling right saying okay or you know, upgrading their what they were sharing. What were the domains going to get additionally out of doing this? Was there anything that you could share that people might be able to take to go and talk to those domains because this friction of saying hey domains you need to share your data but not going back to them and saying and here's why here's what you get here's what we get from it. But here's what you get from it has been something that I've just seen people complaining that it doesn't work and it's like well of course because you're just asking them to do more work and you're not giving them anything in return, so.

### **Max Schultze**

Yeah, yeah, yeah, and on this point I clearly have to say that we have essentially two major setups that largely differentiate from each other. On the one side, we have exactly what I just mentioned before- already embedded BI teams that for some of them already had product managers as well and they had like a great great product-driven mindset for how to work with that data and for them of course it was more or less just a change of technology just to to like upscale them to bring them to newer standards to also be able to work with unstructured data larger amounts of data and things that didn't fit into their good old data warehouse anymore. That was the comparatively easy part because the basic mindset was already there and here was really about technology. And if something is actually about technology that's amazing because that is something that I'm very good at influencing.

The very hard part was of course all the others out there. Yeah so I mentioned right? This was like mostly the stuff that we had in the data warehouse. But for instance, we have a company wide event bus, and it has something like between 4 and 6000 datasets that are flowing through there on a regular basis which are largely created for service-to-service communication. So this is like really the transactional world speaking with itself to make the business run. But for some reason we put a pipeline in place that actually archives all this data into our central data lake and of course now you have this huge archive of like petabytes of data in the end thousands of datasets where nobody feels responsible for right. This is the part where it's much much harder to get into. This is like the part where you really need to dive into specific use cases, to convince specific teams and specific business domains one by one to actually make them adapt like new habits, and take on new responsibilities.

### **Scott Hirleman**

How did you start that conversation or how did that conversation go? Were you the one that was leading it or how did the organization get them kind of bought in? Because exactly what you're talking about, a lot of domain teams are just not familiar



Transcript provided as a free resource by



enough with the data. So you're going to them and saying hey I need you to do these things and not giving them the additional resources both on the technology and the people knowledge side. So you're just giving them responsibilities just adding more work to them, so, of course they're going to kind of push back. How did that go or how did you? How did you get over that hump?

### **Max Schultze**

Yeah, so there were several layers to that I would say, starting with some like initial knowledge sharing awareness to really drive people towards understanding how data is being used. There was a second part where I was basically drawing one of these technical tricks to actually force people to talk to each other that I will get into in a second. And then of course lastly there is really the part where you speak about adapting something on an organizational level. Bringing people to the point that once they understand certain responsibilities. They can also get into the position where they can for instance, get more resources and get additional engineers to actually take care of these responsibilities right?

So these are more or less the 3 steps that I want to touch on. The first one is of course something that I really started running around in the company and like started giving training, started giving talks and started to have one-on-one conversations with people where I knew that this is actually very impactful. Let's say a specific data set that has actually many many users. To really get the buy-in of certain people, explain to them what are the challenges that we're actually facing in the current situation and not even necessarily like blaming them for something that they are not doing but just presenting them with a situation of saying, like hey, there is this other team that is actually using your data, but they had a problem last time you actually changed the schema because it was completely breaking their process for instance. Many many times the people actually came back to—Oh oh oh oh somebody is using that? Oh yeah I didn't know that was actually not meant to be used that way, right? and this was like one of the most common reactions where people were simply lacking awareness but like completely and this was like really really tough on the one side to get this reaction so many times. But also to see actually what kind of motivation and drive it triggered just by getting people to be aware of the situation and this was already really impressive to see that all of a sudden once people realized that there was a dependency to them, they actually wanted to help.

### **Scott Hirleman**

I've been talking about this a little bit of the empathy for downstream users is they don't have any insight into what people are using and why, and so they can't have any empathy for the downstream because they may not be aware that it is or if they are aware that it's used. They have no idea how it's used or that so they're either stuck in. I can't evolve my application. I can't evolve my schema because I might



Transcript provided as a free resource by



break people downstream or just okay so then I can't do my job. So, no I've got to do my job, so we've got to provide that kind of feedback loop. Exactly what you're talking about. That's great to hear that I'm not crazy about that concept, so.

### **Max Schultze**

Yeah, yeah, and the interesting part here was really like that this was an amazing starting point, right? Like that you could actually have impactful conversations, to push single use cases forward. To really make sure you bring people in touch with each other you have them to have a continuous conversation. You can eventually even withdraw yourself from the conversation because they continue talking to each other because now they actually want to help each other. This is really something that there was that was quite impactful to push that.

### **Scott Hirleman**

Yeah, and that blameless aspect too of you said you went in and you said this is not anybody's fault. It's not you know I'm coming to you and saying oh you're doing these things wrong. It's hey we need to enable you to have empathy if you don't have empathy that's a different challenge but we have to enable that so that's fantastic. So you said the first point is knowledge sharing. What was the second point that you kind of were jumping around?

### **Max Schultze**

Yeah, and the second point was then really to think about how this works really nicely, but this doesn't scale. I cannot have one-on-one conversations with every single team in the company to convince them that there is actually somebody that needs their help. And that they should get involved with them to actually push them forward this simply didn't work from that perspective. And that is, again, where my platform thinking came in where, like me as an infrastructure engineer, I tried to figure out how I can start a technological setup that forces people to talk to each other. Let's just say and that's where we where we pulled a really neat trick when it came to the very same event data sets that came from our company wide event bus that were archived and in our Data Lake were on the one hand side, a lot of that stuff or actually all of that was json data very very poor performance for analytics, so very tough to actually work with that on an analytical scale. So eventually, we put a pipeline in place to convert this into a binary format so that we could have much faster analytics and much better performance for the users that actually wanted to make use of the data but to convert something into a binary format, you need a schema.

And you need a schema that you can rely on that needs to be provided from the input side and that cannot break all the time. Because people that would actually break the conversion and again break all the downstream use cases and what we did



Transcript provided as a free resource by



is we put schema evolution system in place that relies on the schema that is actually registered in the event bus, that is under full control of the owner of the dataset and that is where the fun part comes in because whenever something was changing in the schema in the event bus, it would automatically be adapted through our pipeline and also make these changes available in the binary version of the dataset. Which is a very cool feature and the cool part specifically is that the input is 100% under the control of the data producer so now and that happened many many times in the past a consumer comes to us as the infrastructure team and asks us– Hey, I'm using this data set, but this new field is missing. I want to use this new field as well and I'm like well is this new field registered in the schema that is under the responsibility of the producer they're like, well no, we are like well then ask them to do it. Because our pipeline has this automated feature where once they register it, it automatically gets put in place and then you have your data and that was pretty neat trick to now say, okay, we are actually giving up the control about understanding how the schema actually looks because everything is automated but the ones that can control it are the producers.

So now if a consumer comes in and wants a change. They only have the option to actually reach out to the producer and ask them to change it. So just by like setting up the pipeline and the way we did, the only option that the consumers have is to talk to the producers and this again essentially forced a situation where more and more and more consumer teams over time started reaching out to the teams that they were actually relying on and this triggered a lot more of these conversations where of course whenever there was the possibility to like jump in and help somebody and explain something in more detail we are still doing that we are still like consulting the folks.

### **Scott Hirleman**

Still a big support team within the company as well when it comes to that but triggering these conversations and bringing the people together without needing to convince them one by one was actually a large driver to raise the adoption. Yeah I think that's great. That's an interesting thing. I've literally been saying that the one to one convos don't scale. So this is exactly what you're talking about and I'm kind of also trying to figure out a good way to maybe even record one to one conversations where there is an interview about a data product or things like that. So that way people when things are debuted or when there's a major new feature change or whatever, there's just like a little video people can watch but they go– Oh okay I can do that. So it's great and then you start with the knowledge share. You kind of used reverse conway and in a very clever way it is kind of the second and then the third point that you were kind of talking about going down for this as to affect the organizational change that needed to happen.



Transcript provided as a free resource by



### **Max Schultze**

Yeah, that was then something where, like after going through these first two steps, I slowly saw the things starting to take off, so I slowly started to see more people get involved in the conversation. Not just from the perspective of wanting to learn but also from the perspective of wanting to share. We actually had the first data producers that wanted to share how to build a data product like what were also the technical things. What were the platform capabilities that we were offering? They were using it to build their specific data product, so that others could ease the setup to actually do the same for their respective business domain and that was quite interesting. The one insider made me incredibly happy to see that there's like more and more people that are seeing and feeling that it's actually becoming a success and that are interested in sharing the success story where I also honestly have to say I am very very happy for the general culture that we have at Zalando, in the tech space. Which is usually about openness which is about transparency which is also about sharing failures to really understand and learn from the mistakes of others, but to then also advocate the behavior of having people share like what are new approaches that they tested. What are new approaches that they actually were able to pilot successfully and that they now want to share with others and this is like something where the first people started to speak up to raise awareness and started to become like multipliers within the company to really push the topic forward.

And it led to one very interesting thing which by now I think already like 2 or 3 teams reached out to me personally, discussing for instance, the budget round of the next year where they wanted to hire a couple new people and then they realized that there's more and more data consumers, that they are always asking them to do like certain things on the data side like fixing a schema like helping them with some content descriptions or what have you where they realized that their team of backend engineers was not capable of actually performing all these actions in an efficient way. And then they came back to me and asked me if it's a good idea for their team to actually hire a data engineer and this was like really where I was like throwing a party inside of myself because I was so so happy that they wanted to adapt the idea of cross-functional teams and they actually came up with it by themselves. They really wanted to take on their responsibility but they realized that they are not efficiently set up to do that and they wanted to change the setup, and they wanted to use some of the budget that they had been granted. To grow the team in the upcoming year to actually invest into the data space because there was enough people that actually told them that there's the need that there's enough value to actually dive deep into the space.

### **Scott Hirleman**

That's great. Yeah, there's been a lot of conversations recently about an analytics engineer where it's somebody that's, you know, tied to a data product, should each



Transcript provided as a free resource by



data product be developed by a data engineer or should there be a data engineer in a larger domain that kind of floats between multiple data products or you know I think it's an interesting conversation that every company has to kind of figure out. But yeah, that's great that they were saying hey we're not serving our constituents well enough. How did that kind of reflect back on how the domains are measured? Did you have any thoughts on domains now measured on how well they're serving not just the application, the external facing applications and things but the data consumers? Is that something that Zalando overall has kind of reflected into whether that's literal compensation or just you know, the internal feedback?

### **Max Schultze**

Yeah, so, it's quite interesting because it's still a work in progress to large extents also on that side, and just to throw that in from the side Data Mesh is a journey right? And we are nowhere clear and nowhere close to actually be finished on that side. We are absolutely on the journey ourselves. We might be one or two steps ahead of some others but we are still very much in the middle of it and what we are actively discussing at the moment is data product KPIs. Like what are really like metrics that you can measure that showcase how valuable is the data actually that you're providing to others like how many users do you actually have that are using this for their respective downstream processes. How many other data products are potentially built up on top of the things that you are, but also like just the operational metrics like in the sense it does it make sense to invest additional money into making this particular data product available earlier every day because there are certain downstream consumers that would largely benefit from that. And like once you get into this direction of actually measuring the impact that you have and measuring the value that your data is bringing to the company. You're getting into the situation that a lot of business decisions become much much easier because then understand this one dataset is the most valuable dataset for the whole company and everybody is working on that. Yeah, then it becomes pretty easy to say okay, we should probably grow the team by like two more people to put the additional effort into actually increasing the quality or like lowering the latency or whatever. Because, you have a clear understanding of how much impact that would actually have and this is something again like absolutely work in progress. We are still like working on certain standardizations and of the setup working with some peer teams to pilot specific setups for their current, specific domain, but yeah, like it's absolutely necessary to measure what you're doing there and to have the metrics about how well you're doing with your data products. How the adoption is going and what the impact is on the respective downstream users.

### **Scott Hirleman**

Yeah, when I spoke with Juan Sequeda he made an interesting point about, especially downstream data products, if there's a really popular feature of a



Transcript provided as a free resource by



downstream data product should that be pushed further and further upstream right? Let's say you've got data product A and data product B is built off of that and data product C is built off of that and data product C has a really really popular, you know, aspect or whatever could that just be moved into data product A and maybe data product C goes away even though it's like this very valuable thing. Can you push that up further and so that everybody who is accessing this you know, kind of main data product now also has that? And how do you think about new features of a data product like how do you think about that marketing recommendation engine around data products? You liked this data product, you might also like it's an interesting concept. So I wanted to make sure that we also jumped into the book and things like that but was there anything else that you wanted to touch on before we kind of jump into kind of the reason for creating the book and what what you've done there.

### **Max Schultze**

Yeah I think we can draw the bridge here as well because those two things actually for me personally tie together very very well because like I already mentioned at the beginning, when the whole Data Mesh topic actually became popular I wanted to just bring up the idea and share with the community. What were some of the things that we were already doing in that space? And that was perceived as incredibly positive. And they got so many people that wanted to talk about the topic afterwards and dive deep into the experiences, share their own experiences as well and give input and just have open discussions on the topic. And eventually after one of the conference talks that we were giving on the topic as well, we were also approached by O'Reilly if we would be interested in doing a training on the topic and that was a very interesting idea at that moment because like I had honestly never thought about doing something like this like going into the area of like teaching or like more than just giving a presentation or like talking to someone but actually like preparing some educational material on such a large and upcoming topic. We were quite surprised to actually get the request but we decided to actually follow through with that and what we specifically wanted to focus on was exactly the part that we discussed at the beginning which was how to get started from the angle of actually at first understanding the perspectives of each other and developing empathy for the different people that are working with data and this is like really where our first first training was focusing on, to present the different perspectives of the data consumer, the data producer, the data infrastructure team that is sitting in the middle and really showcase, what are the pain points of each of them so that you at the same time guiding the people through like some practical exercises to really feel the pain for that matter.

And to really create the awareness that there are others out there like you or the other way around that there are others out there that you are depending on that



Transcript provided as a free resource by



have no idea that you are actually depending on them to really kick off the ball by picking up these conversations and push the people towards just starting to talk about it. This was like the very first thing that we did there and by the way we are still running that training as well. I think the next one is in February at some point and there will be another one later down the year so there's still the opportunity to jump into those as well. But then further down the line. We also had the idea to basically expand on that. We wanted to not just stop where we had the starting point, where we were talking about, getting people started from just from the perspective of empathy and awareness but to really follow through on a topic and give a bit more extended angle. Really drawing out the Data Mesh journey and what are the different steps that different companies are usually going through which is very close to what we have actually discussed earlier in the podcast from the starting from the angle of like starting small like having first pilots that you're really going into, starting to work with like single teams on specific use cases and then just slowly but surely convincing others and starting other use cases over time so that you can understand how to actually scale the whole thing. And how to again, bring in the platform anchor. How to support the people and how to actually provide the features to make the data practitioner journey earlier similar to the point that you touched upon earlier does there always need to be a data engineer in every team ideally not, if the platform tooling is so easy to use. That any business analyst can actually just sit there and click together their data product that would be amazing of course like very few companies are there yet, but like moving into that direction to just improve the platform features that you're offering and lowering the entrance barrier for the people to actually work with data is like an amazing amazing next step when it comes to that part.

### **Scott Hirleman**

It sounds like you're actually just reading off some of the points that I've been making but you're not so it's great that we're this well aligned because when I talk about pushing this responsibility onto the domains you have to give them additional resources is that the knowledge share is that that they're that you take certain aspects of what they have to do off of their plate is it giving them additional people. Is it those abstractions at the technology level like whatever? But if you're just giving them additional responsibilities. It's a very not nice move. So, how do you balance that it is interesting and it's a conversation for every organization. If somebody tries to tell you there's a blueprint for this is the way to do it, I'm highly skeptical, but exactly what you're talking about and end the incremental approach right? Where there's been a few people that I've talked to who've tried to say– Okay, we've done our POC, we think our platform is ready, we're going to go to all the domains and tell them they have to start publishing all their data. It's like whoa that's getting way far out of yourself. That's basically heading for a data swamp but that in it Data Mesh setup like it's the data mess right? It's the you're just creating too much by trying to



Transcript provided as a free resource by



move too fast and so I love a lot of the points that you're talking about. So, the book specifically you're saying it kind of digs into a lot more of the specific aspects of how people could go about these things. Is there any one highlight that you would point to for the book that was like your favorite aspect that you got to where you're like okay you know, Chapter 2.5 or whatever is like your absolute favorite? Is there anything that you think, yeah, we obviously want people to go out and do their reading but at the same point is there anything that you think came out of reading the book where you really reflect or writing the book that you reflected on and went oh that's really like a really interesting useful point?

### **Max Schultze**

So, there's two different angles that I want to touch upon the first one is the general structure of the book so that potentially some people that are interested in and picking up also have an idea what it's actually about on the one inside of course like we are starting off just with a general. Ah, first part where we introduce the whole Data Mesh concept and absolutely if you are already like an expert on understanding what are the different pillars feel free to also skip that part like, it's really, I still love that part because it's like a concise part for bringing these things together in not bad, extended chapters. Where then the part of the word is really the one where we all outline the journey. What I love the most are some of the practical examples that we've actually like thrown in there. Which to large extents are things that I have actually experienced myself or latest have people experience that I was talking through talking to in the process when we actually develop the whole book and again. Saying kind of calling it a book is always a bit of an overstatement like the whole thing just had like fifty, sixty pages so don't be overwhelmed by by actually diving into that but like the the second big aspect that I wanted to to point out is, actually the support that we got for pushing this thing through. I've been talking to so many people to do our research work and to really like get the the impressions of like what are other companies doing what are like common patterns that we can maybe see what are some of the some of the pitfalls that are repeating in each and every company. What are what are some of the best practices that we can maybe draw from the different setups that we've seen there? And this was really like where we got so much support from the Data Mesh community to really bring this thing together and even to an extent of course that we got amazing reviewers that really helped us also to drag ourselves out of all our thought bubbles to make sure we are. We are really getting the big picture here. So it has been like an amazing journey to actually go through that lots of effort went into that but I'm super super happy where we arrived with that. And I absolutely think it's worth a read if you're interested in the topic.

### **Scott Hirleman**

Yeah, and that was some of that was kind of a concept of why I really wanted to



Transcript provided as a free resource by



create the community and push for it is that so you're you're the bleeding edge of the bleeding edge right? It's kind of like when I look at it's it's pretty much like you and DPG Media are the two first ones that really moved quickly and a lot more people started moving in late 2020, early 2021, and you know we're seeing more and more people pick it up. That we need a place and ways to share this knowledge and to say what are the pitfalls. So hey I fell into this pit. Literally that's what a pitfall is is there is a pit here I'm going to tell you. What what happened when we fell into it why we thought that it didn't exist and then fell in and and move that so that's amazing that that there's so many people out there that are are willing to share and and, you had mentioned that the the book is on O'Reilly so anyone with an O'Reilly subscription can get to it. But you also have a mechanism and I'll drop it in the show notes both link to O'Reilly but also Starburst has sponsored it as well. So you can get it through that access as well. Correct.

### **Max Schultze**

Yes, exactly. It's also publicly available through that angle where you would need to provide your personal details but you can then get the full Pdf copy for your own access which is something that is I think a really nice opportunity. Even if you don't have an O'Reilly account that you can access the material through.

### **Scott Hirleman**

Yeah, if you do have an O'reilly account. Please do actually go through that mechanism at least to click on it just so that way they get positive feedback that it's a valuable thing. We want to make sure that O'Reilly folks know how much the community appreciates doing this or you know everything that you've put together all the presentations. All the work that you've put into this book, so you know we've covered a whole lot of amazing things. Is there any one point that you would really focus on when talking to somebody who's early in their journey? And maybe you've got somebody who's maybe done their POC you know two different ones but somebody who's super super early and trying to get going and, then somebody who's kind of gone through their POC and trying to figure out what next. Is there any nuggets of wisdom that you could provide as somebody who's gone through this and really thought through a lot of what's going on out there?

### **Max Schultze**

I think the biggest thing that I can recommend and I think we've even already touched upon it partially is, try to not get ahead of yourself like I've seen many many people that started to to build the roof behind before they even started building the house. And especially from an angle that I was talking to so many different companies that were always asking me questions about the federated governance and they were always like diving deep into like yeah, but how can we adapt the federated governance because that's the savior for all our problems quote unquote.



Transcript provided as a free resource by



And then just like diving deeper into the conversations, I realized well they don't have a single product person in their whole organization. They have like really not started with the absolute basics and of course you can think about fixing your governance situation. In the setup that you currently have in your organization. But if you don't draw the foundation straight, it's a completely different conversation right? And like what I can absolutely recommend is really start from scratch like start with the basics and go one step at a time. It's always good to keep the next step in mind. So that you're not running off totally in a different direction at first but it is really important to do your homework and to really get the basics straight before you jump to the advanced level and I can absolutely say talking about federated governance is a luxury. If you really want to dive into this topic you should already be in a quite advanced topic I can say from my own experience. We are still only scratching the surface of federated governance even at Zalando where we've been under where we've been on that journey already for 2 years because we are still struggling with the basics in many many different areas. Company and like this is really something that you should keep in mind, don't jump ahead of yourself, but really do your homework, build the foundations and then take it from there.

### **Scott Hirleman**

Yeah I think this is amazing advice because I see it a lot with domain-driven design where someone says well we have to map out all of our domains before we start. It's like no, you have to map out one or two if you want to build your POC with interoperability across domains. You need to do two if you only want to build your POC with interoperability within a domain across a couple of data products. You only need one domain and your domains should have firm boundaries but that are malleable and flexible right? That they're set so, this data is owned by x y z. But even when you're starting, it can be pretty fuzzy for okay I've mapped out one side of the domain and so that's where I'm going to make sure that I stay inside that but I don't even have to map out the full domain and or the platform side where people are like I'm gonna build out this full platform that can do these 85 different things and it's like okay, but like your first data product only needs you know, Kafka or Pulsar or whatever and and some Snowflake or Databricks or whatever, you need technologies. And you're trying to build it out to handle absolutely everything. So I love that advice and it's phenomenal to talk to somebody who's kind of giving that practical of, hey it's okay, it's a journey you don't have to. Get a little frustrated by some of the industry bluster around if you're not data driven in the next year you're going to be out of business. It's like no, that's not the case. Yeah, you might have missed some opportunities for growth, but so this is phenomenal. I obviously want to give you some chance to kind of sum up your thoughts here if there's anything else and also let people know where they can get in touch with you and what you're up for for people reaching out to you about as well. So if you've got any kind of final parting words. Great if not let's make sure that we can get people to see all the



Transcript provided as a free resource by



amazing content. You've done and get in touch with you if you want that.

### **Max Schultze**

Yeah, yeah, First of all, thank you so much for inviting me for the podcast. It's really been a pleasure talking here as well and going through some of these things in detail and of course, for you folks out there, if you're interested to follow up on some of the things that that I and specifically also my partner Arif I've been doing over the last two years feel free to check out some of our talks that have been recorded and have been put on Youtube, some of them are very much specific to the Zalando related side of things. Some of them are also going into the theoretical angle. There's a bunch of upcoming conferences as well where we will be speaking so check some of those out as well and of course if you're interested in the training that is still running on. O'Reilly you can still sign up for those so I mentioned the next one is around mid of February but there will be another one later this year at least as well. The book we already shared like where you can find it so that it has really been a great journey to actually make that happen and to work that out. So I really hope that you folks are enjoying that as well leave some feedback. Yeah, if you're actually diving into that. So far I've received very little feedback from only a couple folks that were telling me that they enjoyed it but I would be super interested as well on what are the parts that we can do better moving forward because again like it is a journey like even doing something like this like this booklet is something for me where I want to improve on as well and get better and make this really useful for the community. And last but not least feel free to reach out to me on social media either like on Twitter or on LinkedIn. Maybe we can drop the links below as well. So that we can also have a chat. I'm always open to having a chat for like sometimes even having some follow-up conversations because that is really what the community is all about exchanging the opinions and the situations that we are in and really learning from each other.

### **Scott Hirleman**

Yeah, sharing that context with each other is so crucial as well. Thank you again. Max has been so phenomenal and I think it's going to be so useful to a lot of people and thank you to everyone for listening along.