Transcript provided as a free resource by 

Data Mesh Radio Episode #76: A Skeptic's View of Data Mesh and Learning Your Data Product ABCs

Interview with Tim Gasper

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

Scott Hirleman

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

Adrian Estala- Starburst

Welcome to Data Mesh Radio, with your host, Scott Hirleman, sponsored by Starburst. This is Adrian Estala, VP 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.

Scott Hirleman

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

Bottom line up front, what are you going to hear about and learn about in this episode? I interviewed Tim Gaspar, VP of Product at data.world, and the co-host of the Catalog & Cocktails podcast. We covered two main topics, a skeptic's view of Data Mesh and Tim and the data.world teams', ABCs of data products framework.

Let's start with the skeptic side, so skeptics have a few main push back on Data Mesh in Tim's view. First one being Data Mesh isn't for everyone based on size or complexity of what data you're dealing with. The second one being tooling doesn't exist. Is that yet or won't exist? To make it easy for domains to easily take over data ownership and just in general, to do Data Mesh. Number three, there shouldn't be anyone who "owns the data." Number four, there aren't enough case studies out there yet of people doing Data Mesh, especially people who have been successful. Number five, not clear guidance on how to handle the combination of data across domains, how does that ownership work? And number six, Data Mesh will create

data silos.

So let's go down the list one by one and talk about some of the points Tim had made as well as some things that in general are kinda coming up in the Data Mesh community as kind of push backs against these push backs. So the first one, Data Mesh isn't for every organization, depending on size, number of domains, data/problem space complexity. Tim said this, I've said this probably 30 plus times on the podcast so far, Zhamak has said this. Literally almost everybody who is pro Data Mesh, who is as a Data Mesh advocate has said this, other than maybe some vendors trying to sell to people doing Data Mesh. This is one of the myths of Data Mesh, that it's designed for everyone. Data Mesh isn't for everyone. If you don't have these problems, that would cause you to actually need to go to Data Mesh, don't go to Data Mesh, don't do it for the sake of doing it. Much like distributed systems, people will tell you don't distribute your systems unless necessary, don't go to a decentralized data set up if you don't need to. Tim made the very good point though that we need more conversations and better guidance on how to actually measure and how to think if centralization of your data team and processes is what's actually causing your challenges. Is that becoming the bottleneck or is it something else, like your overall culture, your data culture, your architectural set up, your level of data literacy etc.

So the second push back was the tooling doesn't exist yet to make doing Data Mesh easy. We talked about how a big conceptual issue of Data Mesh is that it has to solve every data problem, even the most difficult, right out of the gate. It's just not true. Tim mentioned that your implementation needs to really think about selfservice being empowerment, not necessarily a big red easy button. And your implementation will evolve, it must evolve for this to be successful. Yes, the tooling to make data modeling easy for application developers isn't there yet, nor is really the governance tooling, the data catalog space is really just starting to emerge here, data discovery tooling in general, it's all kind of still early days, it wouldn't be bleeding edge if there weren't some bleeding, if there weren't some pain. You're gonna cut yourself because this is so forwardleaning that if you're doing it now, you kinda have to accept that. If you aren't willing or honestly capable at this point to build a lot of the Data Mesh plumbing yourself, Okay, you can jump in later. It's not like Data Mesh is leaving the station and you either get on now or you can't ever do Data Mesh. It's fine to say "It's not for us right now, and it might be in the future."

Number three, there shouldn't be a "owner of data." Tim made a really good point here on accountability to sharing your data versus the kind of fiefdom thinking. On that fiefdom thinking someone has complete control over how the data is used. Yes, someone shouldn't be able to prevent other domains from using their data in general, there might be some governance cases and things like that, but in general,

having that model isn't Data Mesh anyway, it's not what Data Mesh really ascribes to. So if you're doing that, you're doing it wrong. Why would you make data reusable and discoverable if people can't use it or discover it, that's putting out a sign that says free cake and then not actually letting people have the cake. Why would you do that? I want the cake.

Number four was not enough case studies on Data Mesh, people doing Data Mesh well, and what's then the results. Tim mentioned this briefly in the episode, but I fully agree with this one. But if we're waiting for people to be done with their journeys, it'll probably be another five years. Max Schultze at Zalando, it's pretty far down the path. They're over two years in, and he said, “We're about 25% of the way through our journey,” right? If you want to wait until there are those case studies, go ahead. But when you think about Data Mesh, it's are you an explorer or are you a settler? It's fine to be the settler, but the explorers get the stories told about them. Gold Rush is probably a bad analogy because of how many people were negatively impacted by it, but if there's a rush, do you wanna be there first or do you wanna get there when everything is settled and there's no longer that kind of big big value creation there it's something that you're more easily able to implement, but are you gonna see the same kind of returns? Are you gonna be that much further ahead of your competitors? It's something to think about, I'm not saying do one or the other.

Number five, lacking guidance on exactly how to cross domain data, how to do those cross domain data combinations. Tim mentioned that there is the question of how do those combinations get managed right now in a data warehouse or a data lake world? There are clear owners, the data team. But in Data Mesh, is that a new aggregate domain? Is that a consumer aligned domain? Do you have that still on a centralized team, like Adavinta is doing. I think this is one of the vague points in Data Mesh that is actually intentional. You have to figure this out for yourself. It's situational. There isn't a cookie cutter approach to this.

Number six, Data Mesh will create data silos. Sure, if you have the data mark model concept of old where data is only created only for the domains themselves to use. 80% or so of the guests on the podcast have recommended doing two plus data products across different domains as part of your proof of concept. So you understand how to do that interoperability, how to do that. You can build muscle and figure out how to do this, just like the response to that kinda push back, number five. You can learn how to do this, it's not that you have to have everything laid out ahead of time for you. Tim talked about how important iteration and collaboration is to prevent data silos. So much is about the intent to not let it become a problem, and if you're really focused on that, it probably won't be. Overall, we agreed that a lot of the push backs are probably coming from a place of where somebody is in an organization where Data Mesh would create a lot of friction in their existing culture,

and as Tim said, changing culture is very hard and "fixing culture is even harder." And as Tim's colleague, Kayda, mentioned on his episode, Data Mesh is often an exercise in finding where you should fall on the centralization decentralization spectrum for a whole lot of different decisions. There isn't that cookie cutter approach, and people have to be comfortable with that if they're gonna try and implement.

Tim talked about how we too often think about data implementations, whether macro and something like even the Data Mesh implementation level or something more micro, like at a data product level as a singular event, something that's once gonna happen and that doesn't evolve. Data implementations aren't a house, they're much more like a garden. Seasons change, you might have to weed a little bit or a lot, some of the places I've been, you might even change what the focus of your garden is. Are you sick of zucchini? Is this data product or this report dashboard no longer relevant? Think about that as something that is not just a one off and that this can't ever evolve, set yourself up for that evolution as you learn so that it can improve.

We transitioned into the data product ABC's framework that Tim and the team at data.world put together. There's a link in the show notes. An important aspect of this framework is that much of Data Mesh, it isn't about providing specific answers, but more the questions you must answer for your situation to get to a good outcome. A key point Tim made at the end was just how many data challenges come from implicit expectations and knowledge, versus getting very explicit to make sure everybody is on the same page. Tim basically summed it up in kind of a way of get in the room, negotiate, come to a conclusion and shake hands and then document it so that it's not implicit for anyone, that there is an explicit documentation of what you're trying to accomplish.

So the ABC's framework for data products, A is for accountability. Who owns the data product? And what does ownership specifically mean? B is for boundaries. What is a data product? What interfaces does it use? And crucially, what isn't a data product? And also what isn't part of a specific data product? What are the very specific boundaries around data products in general? And then what are the specific boundaries around each one? C is for contracts. What are the explicit expectations? Who can use it? What are the SLAs? We've talked about data contracts a lot on this podcast. Abe Gong mentioned in his episode #65, how often these contracts at least start as implicit. Let's get to communicating and negotiating and really get this too explicit. D is for downstream. Who actually uses the data product? Who might wanna use it? And why are they using it? And why might they want to use it? What is the road map?

E is for explicit knowledge, partially because A, B, C, D, K doesn't sound as good. Exactly what we've been discussing in this entire episode, don't believe your data

products are self describing, document things, explain what are the relationships to other data products or concepts outside the data product. Really get explicit, put that down so people can be on the same page. So many of our challenges are, again, because of implicit expectations or knowledge. Get it down so people can understand it and can collaborate more effectively. And even if they don't get it immediately from the documentation or whatever, they can get to the right person who does, and they can ask much better questions. With that bottom line up front done, let's jump into the interview.

Okay. Very, very excited for this episode. Here today, I've got Tim Gasper, who's the VP of Product at data.world. And also the cohost of the Catalog & Cocktails podcast. And what we're gonna be talking about today is Tim and his cohost Juan had done an episode about Data Mesh on Catalog & Cocktails, and this is why I had originally reached out to Tim to talk about, he took on the role of the Data Mesh skeptic. And we've talked about this and he's positive on the Data Mesh, I'm positive on the Data Mesh, but we're both skeptics as well. There are specific things that we haven't figured out and that need to get figured out, and there are caveats, there are drawbacks, there are different things where Data Mesh isn't necessarily the right choice. It's not fully baked yet. We've all kind of talked about that. That's a big thing. And then we're gonna also move into data.worlds, had put out a framework for thinking about Data Mesh in general, their ABCDE framework, which I guess the F is the framework. But with that, Tim, if you don't mind giving people a little bit of an intro to your background and then we can jump into the topics at hand.

Tim Gasper
Yeah, absolutely. Well, thanks Scott so much for having me on the show. Big fan of the Data Mesh Radio podcast, and Juan and I over at the Catalog & Cocktails podcast, really love this community of data oriented podcasts that we have going on here. This is a lot of fun. I think this information is so helpful to the community. Yeah, thanks for the introduction, Scott. VP of Product over data.world, a cohost over at Catalog & Cocktails. I've been in the data and analytics space for probably 13 or 14 years now as a product manager working in data, working in customer identity management, and now over a data.world, and cataloging and governance. And yeah, this is a super interesting topic, this idea that not only are we trying to figure out what Data Mesh is, how it can be valuable, but also where its pitfalls, where is it not applicable? There are some naysayers out there who are very much like, what's this Data Mesh? This is the new sensational hype thing. Here's all the reasons why it's wrong, and you should just throw that out the old way is the right way, or this other way is the right way. And this is an important topic for us not just to dispel myths, but also to embrace the skeptical points of view that actually are good and accurate skeptical points of view.

Scott Hirleman
And I think that's the fun differentiating point. I think there are a lot of skeptics who think that their point of view is immediately correct and it's like, No, there are questions, have those questions. And I'm seeing, especially from the data warehousing folks, I think I'm gonna write an article on this of why I think the data warehousing folks, or the enterprise data warehouse folks. Data warehousing as a practice is still alive and well, look at Snowflake, but the enterprise data warehouse just got these kind of fatal flaws when you're looking at flexibility and scalability. And if you're not able to change what you're sharing very easily and very quickly, 'cause everything has to conform to a single model or maybe two models, it means that you lose all of your context. So I could go on and on, but let's talk about what are the viewpoints that you think of when you think of valid skepticism that comes relative to Data Mesh?

Tim Gasper
Yeah, great question. And I think there's a few areas that we can look at around some of the skepticism here, some which I think are easier to dispel and some which I think are more valid. And so one of them that I think is more valid, and it's an interesting talking point, and I know Scott, we've talked about this. Is sort of like the size of an organization. That there are some organizations that maybe are smaller, maybe you're a startup, maybe your data ecosystem is much simpler. And one skeptical perspective is just like, why do I need a Data Mesh? Why would I need to have these different domains and push ownership to these different domains or have a microservices architecture here where my data is pretty simple? I just got a couple of data sources, or maybe it's, I'm a product company and I've got this event data and I just need to stick it in my warehouse, and I need to pop some dashboards on it. It's like, do I really need a Data Mesh?

And perhaps a very valid perspective here is like, well, there's certainly valuable concepts around thinking about data as a product and thinking about how do you establish ownership around your data and really accountability around your data, but maybe you don't need a Data Mesh if you're a small company. Right?

Scott Hirleman
Yeah, I think that's the big one that I push back on the most is I even think that it's not like a 200 person company. I think if you're at a 1000 person company, there are a couple of interesting niches that could require you to have a Data Mesh. I've talked to a couple of people who they're a rollup company, so not fruit rollups, but they're going out and buying a whole bunch of different companies. And so that ability to have these different companies not try and integrate everybody into the exact same format and that everybody can share their data in an appropriate way, but then you're not really necessarily even talking about the interoperability versus again, it's

that ownership and that there is ownership of the data matters. We have to think about how we're sharing data and how we're consuming it, but do you need to build the self service platform? Do you need to move away from the decentralized data team versus centralization has benefits until it doesn't, right? It's a bottleneck once it's a bottleneck, but to start, if you're pouring out a small amount of water, the bottleneck isn't a problem. It's kind of a feature, right? It means you've got checks on and balances and all sorts of things in place. You've got expertise around data so that every domain doesn't have to develop that expertise.

Tim Gasper
Right. No, I think that's totally right. And I think that there are certain pains that trigger you being like, oh wow, the centralized approach is not working right. That's where if you have a couple of data engineers, you have a couple of data analysts, and you actually find that that centralized team kind of building out your warehouse, your modern data stack, however it is that you think about your smaller company or your startup kind of architecture here, or maybe it's just within a single department, let's say, maybe that centralized team is working out well for you and actually it's efficient and it's effective, and there's no need necessarily to break something that's working. It's when you start to find that the data engineers are becoming a bottleneck or that the data analysts don't understand the data, and they're constantly having to paint the different domains and the application owners to understand what's going on, or they're constantly having to do tons of cleaning and cleansing and reworking of the data in order to get into a form that people can actually be able to leverage. That's when maybe you start to say, "Do I need a Data Mesh or not?" I don't know, right? Think about that really truly, but there's certainly something that you need to think about now in terms of how do we get more of that responsibility pushed out up to those applications, how do we actually break past these bottlenecks that we have in place?

Scott Hirleman
And shared ownership models always come with their own issues and things, but one of the biggest issues that I see when I'm talking to people is the domains or the product side or whatever you wanna call the people that traditionally have just had data be exhaust and haven't had to have that ownership. A lot of them just don't even know how their data is being used. Chris Riccominis episode when he was talking about what happened at WePay, people would drop a column and their CI/CD servers would kind of flag it and go, "Hey, people are using this column." And that person would have no idea, right? So we need to get to a place... I hate trying to solve everything with tooling. I think we all kinda do. I think the people who really understand this problem, tools are helpful, right? You're at a tool company, I'm at a tool company.

Tools are helpful, but they aren't the solution, but we need better tooling and practices around making it so that they can understand what impacts their changes will happen. We keep treating that the product is the only thing that matters in so many companies, or the application level is the only thing that matters. And so we have to get out of that mindset. You've got a team of three software developers and you've got a monolith. Do you really need to have them fully own and learn how to do all of the data modeling and all of the... No, not really. You can get them kind of understanding what people are using their data for and things like that. But yeah, I think I am very skeptical of that size of domain issue or size of company and a number of domains and all that. If you're not having the problems of complexity, why do this? Just put in a better people process side, not the full difficult implementation.

Tim Gasper
There's things that you can do to really help yourself that don't necessarily require you to jump head first or fully into Data Mesh. And I think your point, Scott, around tools is an interesting one, 'cause I think this is another moving on to a second, sort of skeptical point is around, well, my domains. Like, okay, cool. The idea of empowering domains and things like that, that sounds interesting, but I've heard a skeptical point. They're like, “Oh, well, the domains don't have the data. It's sort of tooling capability or they don't have the data developers that could actually develop this sort of products within the domains or ownership in the domains. That's a centralized skill. Me trying to push that through the domains would be silly. That would be difficult.” I hear that is one thing that people are challenged with. I'm curious about your thoughts on that. I certainly have a couple of opinions on where that's valid and maybe where it's not.

Scott Hirleman
I'd love to hear you start first.

Tim Gasper
Yeah, sure.

Scott Hirleman
I don't wanna be leading on all of these. I think you've got a very valid point and you probably actually talk to almost as many people as I do about Data Mesh, right? Being a VP of product.

Tim Gasper
No, absolutely. So on this particular thing around tooling, I think what people don't always appreciate and one of the things that we talk to some of our customers at data.world is you want to have this interplay between the self service tooling and the empowerment. That's one of the things that I really like about the Data Mesh

framework is that it's not just about one thing or another. It's about the interplay of these different pillars together. And so, for example, data lake and data warehouse technology, right? How do you make that really accessible and easy for different groups to use? Tools like Snowflake, for example, have done a great job of really democratizing and lowering the barrier to getting a warehouse going and being able to have multiple warehouses going that they can work in conjunction.

Catalogs have been important to really lower the barrier to being able to define data to sort of have a place to be able to put data products and communicate about data products. And so I think when you think about empowering domains, when you combine it with the self service tooling, again, it's not all about the technology. It's people and process, of course. That's very key to this, but when you combine people processing technology together here, you can actually empower domains, different business groups to take on more responsibility around the data and to leverage the self service tooling in the timing that makes sense for the groups that make sense. It's not all or nothing. Some groups can take on more ownership and some things can stay more centralized, but the interplay of technology and the people and process is really important here.

Scott Hirleman
So there's a lot of things that maybe weren't exactly explicit in what you said, but I think are incredibly important because these are the ones that I find are really hidden in these conversations of people who have talked to people that are going down this route, right? One is you don't have to get to the most complicated complex data products from day one, right?

I had an episode about whether you should embed data engineer, analytics engineer in your domain or not, and I think it gets you to a place where you can have much more complex data products much quicker, but do they become that single point of failure if they're the single data product developer, right? Versus are they teaching everybody, but if you're putting somebody into a domain to teach everybody to do the job that you're doing currently, that's kind of a rough spot to put somebody in. That's not really a high empathy role of, I'm replacing myself. So what do I do after I teach everybody to do this? But talking to NAV and talking to Talkdesk and a lot of these companies that are in that kind of smaller or less necessarily complex needs, they're doing a data product as a table or a view, right? It's not a full data set. It's not that. So you can evolve your complexity relative to your needs.

Tim Gasper
Right, totally, totally agree with that. And depending on what makes sense for your organization, you may approach it differently. One of our customers is Prologis, really large logistics organization, real estate warehousing, and for them, data products is

what they call building blocks, and they're like tables or semantic concepts or a particular transformation, and it's quite granular, and that's sort of the way that they're approaching those data products. It's what works for them, right? And we work with really, really large consultancy, global consultancy and for them, data products are quite literally like there is a data set of multiple tables that either has been purchased or has been derived that is used for client needs, and that's what data products means to them, and things that are upstream of that are the supply chain that gets you to the data products. So what makes sense is what makes sense.

Scott Hirleman
Yeah, and I think when you talked about giving domains the capabilities and that empowerment, Jessica Kerr or Jessitron, her episode of Data Mesh Radio, she was talking about we need to get to a place where this is easy for them, and the tooling isn't really there in a lot of cases right now, but it's again, you can evolve towards that. You can see what are the things that really matter for folks, right? Do we need this to be that we drop in the tooling and it solves versus like, "Hey, we're having difficulty with X, Y, Z. Okay, then let's do it manually?" Or, "Okay, we're gonna have the data engineering team handle this for now as we build out the tooling or... " A lot of this stuff, it is that if you're looking to just solve it by purchasing, you're going to struggle, but you're gonna have that no matter what. Just throwing tools at things doesn't have to work and having your evolutionary path, I think, matters. But yeah, I constantly run across the lack of tools as being a big, big detractor to Data Mesh, and it's like, "Yeah, if that is your viewpoint right now, it's fine to wait." Right?

It's not that you jump on this bandwagon now or it's leaving the station. This train is leaving the station. No, presumably this is gonna be around for decades, right? This until we can get to a place where we actually can have that kumbaya of operational and analytical in the same plane. We're just not there from a complexity standpoint right now that treats all of the operational use cases and analytical use cases from the same data products and things. It's just too complex right now, but I think that's kind of beyond the Data Mesh, and I think that's 15 years down the road once we figure out how to do this well.

Tim Gasper
Right. Yeah, once we figure it all out, I have a feeling you will be able to sort of buy the key components that allow you to build out your Data Mesh in sort of a more plug and play fashion, right? But I think that right now, I really believe strongly and I know you push this and really everyone who's sort of leading the Data Mesh community right now pushes this. Like, you're not buying a Data Mesh. There's no one tool that you're gonna buy and you're gonna plug in and it's gonna empower the Data Mesh and I think that also is something that feeds the skeptics a bit because they look at something like the logical data warehouse and they're like, "Oh, okay,

cool. I plug in my warehouse. I plug in my virtualization layer, and this is my warehouse here, and then I apply my modeling concepts." Things like that, right? And I think some folks are wanting to figure out like, "Oh Mesh, so that's a catalog." Or, "Oh Mesh, that's a virtualization tool. So I gotta buy myself a virtualization capability and then that's my Mesh." And it's like, "Oh, hold on now." I think you're oversimplifying this and you're taking too much of a technology perspective on it.

Scott Hirleman
Yeah, if you're trying to solve it via technology, that's what we've been doing for 50 years, and in certain circumstances, it's worked relatively well for a few companies, but 90%/95% people have been struggling with their data for five decades, right? So why do we keep the definition of insanity of throwing tooling at the problem instead of saying, "Let's think about how we could actually do this in a scalable way." And tools matter. They really, really do matter and we don't want people to have to build everything themselves, but yeah, exactly what you're talking about. So we've covered a couple. Do you have any that you think are the ones that are really, really valid? I think the tooling one is valid for certain companies if they're not at a maturity level where they can build that glue in between, right? If you can't build the glue between the tooling and understand and extend your tooling, you don't have the team capabilities or you don't have the head count or anything like that to do that, Data Mesh isn't for you right now, and that's fine. Start working on the ownership, the people process side, but you don't have to go full Data Mesh. You're gonna be in a better spot if you can establish data ownership and data practices, but you don't have to go full Data Mesh right away.

Tim Gasper
Totally, yeah. I think, Scott, there's a couple of points that I think are very valid, but I think skeptics think that they're very incompatible with Data Mesh, but I feel like they actually can work together if you're thinking about it in a broad enough way. So one of those things is this concept of ownership of data and the idea that there isn't ownership of data or there shouldn't be ownership of data. And I think about comments that folks like Doug Laney, for example, will get into where he'll say things like, "I get the concept of letting data owners control access to data, but within an organization, data should be an enterprise asset. It is not owned by any individual or department." And I think that that's trying to kind of push the idea that Data Mesh is saying that you have to have explicit owners of the data and that domains must be defined as domains on the data.

And I get that criticism, I get that criticism that like, yeah, if one group is the domain for the data, then it shouldn't be that that one domain gets to dictate to the rest of the organization sort of how their data products are gonna be used, how their data is going to be leveraged. That really what you want is the free flow of this information

and while there might be accountability around the data, you don't necessarily want this sort of the one product manager and the one domain to rain down on the parade of other parts of the organization that could really benefit from and use that data. So I think that's a very valid concern. Now, I think that the way that you mitigate that is you don't let there be the sort of fiefdoms that form, right? I think if you've implemented a Data Mesh in a way where you've reinforced silos and you've made governance harder, not easier, and you haven't embraced that principle of interoperability and standards and actually trying to create value from your data, then you're doing it wrong, and that's not the way that you should be implementing your Data Mesh. So I think it's a very valid concern, but I think that we in the Data Mesh community gotta make sure that that's not the result. That's not the outcome that we're driving for.

Scott Hirleman
And I think there's some really obnoxious nuance in that. So if you listen to Sarita Bakst episodes from JP Morgan Chase, she's leading their governance charge for their Data Mesh. So pretty big, complicated role at a big bank that's heavily regulated. So data owners should be able to dictate how data is used in a governance fashion, but what she talked about as well was that when you federate that governance... I think the big thing that I really think is underlying a lot of what he's talking about here and I'm not a huge fan of Doug based on some, you could call it trolling, I wouldn't even give it that benefit of the doubt, but that there's a lot of people that are very afraid that if we open this up that it's chaos and that we don't have any governance left and that everybody then just gets to do whatever they want with the data, and you're opening yourself up to a lot of risk. And I think that we need to have that accountability but we need to enable that governance via the platform whenever possible, right?

The self serve platform isn't for consumers, it isn't for producers. It's for both, right? When I ask, "Is this also a platform for producers or consumers?" The answer should be yes, because it's for both. And so if you do have somebody who's saying, "This is the only interpretation," that's bad versus, "We're sharing with you how we view the world, we're sharing our context, do with that what you will." If you're completely misinterpreting it and we see that, we're gonna come and tell you that You're misinterpreting it, and that's fine because we're sharing our context. But I'm not coming and saying, "Well, you use this in this way that I never thought of, and that's wrong." If you don't have the ability for that serendipity... My executive sponsor at DataStax calls me a serendipity engine, 'cause I just go out there and create random... Even just this Data Mesh community, it was supposed to be a two week thing for me, and we're going on 14 months not 14 days. So I think you bring up a really good point, but it's one of those where I think people are skeptical for the sake of being skeptical, 'cause they haven't seen it work. It's like, "Well, but it can if we just have the people in

the right way," so.

Tim Gasper
Well, and to see it work is actually one of maybe two more skeptical points that I'll mention here, so I think there's a valid statement that there aren't enough case studies of Data Mesh, there aren't enough examples of it really sort of in the wild working. I think we're starting to see more of that. And you've had a few of those on your show. But I think folks wanna see that working in action, so I think that is a valid piece of skepticism here. And the last item of skepticism here that I think is very valid that, honestly, I'm still thinking about it. I'm still piecing through what it means in the context of Data Mesh, is that there is a ton of interesting analysis and data value that comes from... I don't wanna call it the data lake or data warehouse approach, but the bringing of the data together and the combining and the analysis of the data when it's in combination across the domains. I think kind of what Data Mesh thinks of is sort of these either aggregated domains or in some cases, consumption domains.

And you think about the history being important there, time series data, time slices of data, being able to combine your customers with your support data with this data, with that data. How should a company go about managing these sorts of in between domains? Does that stay centralized? Do you try to empower that to two domains? I think things get kinda complicated there, and that's an area of valid skepticism where I think today in more of a centralized model, it becomes clearer how you approach that, in a Data Mesh model, it becomes a lot more complicated.

Scott Hirleman
Yeah. Well, on the case studies, yes, anybody who's out there doing it, you need to be sharing what you're doing, even if you don't think you're very far along the path, everybody, you need to be talking about this, right?

Tim Gasper
Yeah. War stories to get 'em out there.

Scott Hirleman
If we're just doing this in the dark, it's gonna take an extra three years, so get out there and share. You're gonna have some scars. You're gonna have some worts. It's fine, everybody understands that. But yes, on that, I think that's a really interesting point, because the way that people have consumed data historically has been as just a report consumer. And so in the new world... I'm not a fan of the citizen data scientist, but in the new world, we're gonna have people that still want that, "Just tell me what it means. Tell me what I should be consuming." And so in a Data Mesh world, can we get to that? I think we can. I think Zhamak has really relaxed a lot of her constraints from her early thinking around, can you put these kind of combined

queries as just not necessarily even a full data product versus it is kind of a data product, it's a data product light or whatever of like, "Hey, we see this query constantly, so let's do this." It's kind of Wannes Rosiers talked about it kind of being a data application of there's some use of data that is outside of the mesh and we put that on the mesh. But I agree that it's not prebaked and everybody has to figure it out for themselves. But kind of, so? That's kind of your job, you have to figure out relative to your organization.

Tim Gasper
Wait, you're telling me Data Mesh isn't a magic bullet? Yeah, no, I think that's very fair and very valid, and I think it can work out. I think we can make this all make sense, and I like the idea that while all your data may be part of a domain, not all of it is a data product, and that there are different levels of data product. There are the things that you are really trying to build strong constraints and SLAs and expectations around, and there are things that really need to be thought of like a product. There needs to be reusability, the visibility around it, there needs to be the right contacts, whether directly or indirectly, sort of owning those products that you can reach out to, you understand what they are, you should be collecting that meaning and documenting it and sharing it. Not all data products have to be perfect, and not all of them have to be broadly applicable to the widest set of use cases. It can be a query that somebody wrote that maybe other people can benefit from, put that out there. And that's the kind of stuff you can build upon.

Scott Hirleman
Yeah, exactly. And I think there's Emily Gorcenski's episode, which is coming up in a month or so. She talked a lot about that. We talked about fit for purpose data products. There are certain data products, Justin Cunningham also talked about this concept of just getting data out there and then people can see, okay, what is this, and then they can go back to the owner and go, "Okay, I want this in an actual quality way." So much of what's happened with the enterprise data warehouse is you have to really get it right up front, and we don't have iterative processes in data. So you tell people that it's like, "We'll figure it out along the way." We've seen that just not work. And so it is in a way where if you're sharing data in an effective way as an owner, great, and you're gonna add value and then you can figure out how it interoperates, and you can figure out how to do that. You don't have to answer everything up front, but that's always led to disaster historically. We're building the modern data stack, people are building stuff on quicksand, and so that quicksand keeps falling underneath it and their data stack applications just keep breaking. And so like, yes, we need some intentionality, we need to think through this, but we also need to be like, "Hey, let's actually have the conversations instead of I put it out and you either consume it or you don't," it's like, "Oh, we need to find the consumers, we need to have those conversations."

Tim Gasper
Yeah, there's actually a collaboration that needs to happen here, I think that that's sort of like, if you wanna take it to an extreme, that's like data as product thinking gone wrong. Where it's like, "Build it and they will come," and "Build it once and forget it." That's not the way it's supposed to be.

Scott Hirleman
I think I posted on a Bill Schmarzo post or something, and I mentioned, that's the data field of dreams, right? Exactly what you just said. The data field of dreams is that I have and I will build it and they will come and it will be amazing instead of like, "No, just talk to people, have that conversation."

Tim Gasper
Right. What problems are people trying to solve? Right. What are the business questions?

Scott Hirleman
So I think that's a good summation. I think that the data silos one is the other one that people always bring up, and it's like, "Yeah, you're gonna start out with data silos because you don't wanna create all your standards up front," but you're in a better spot with data silos that are high quality data to start with, and with an idea to iterate, that you're going to get towards interoperability when you find those good points of interoperability. And that you set yourself up and you look for interoperability, you look for reuse, everything should be reused, like reuse of data, but reuse templates, reuse of processes, reuse of everything, everything about Data Mesh, should be looking for those. So you can have the micro value of the data product versus the macro value at the Data Mesh, but we need that macro value to start... Like the incremental value of each data product is more about not the data product value itself, but it's greater part in the Data Mesh, but that's once you're to 50, 75, 100 products and you can evolve, these things don't have to lock in., yeah, a lot of exactly what you've said. So is there anything else on the skeptic side, or did you wanna jump into your fun framework of ABCDE?

Tim Gasper
The data product ABCs. No, I think, this has been a great conversation about some of these skeptical points, and I think one last comment I'll make before we kind of transition just to zoom out, is that when you talk about things like silos, like you just mentioned, and we talked about whether you're too small or your perspective on ownership of the data, like a lot of these things may have to do with culture, culture in our organization. And our people in your organization really complaining about and bringing up skeptical points of view that are valid and need to be worked

through or is this just challenging because it's a different perspective, or you've already had a siloed culture, and so when you look at something new like this, you say, "Uh oh, this is gonna just perpetuate these silos or deepen these silos," and of course, culture is hard to change, it's hard to fix.

But I think that's an important aspect of this as well, and one of the reasons why I'm a big fan ultimately of Data Mesh is because I think it provides us a lens and a framework by which to look at our data culture, bring some new concepts in that break our frame a little bit and say, "What if we addressed product or data a little bit more like a product? What if we thought of this as a spectrum between centralization and decentralization?Where are we on that spectrum? What makes sense for us? Do we feel like we have standards, do we feel like we have interoperability? Or how important are those? Are those the most important things for our organization right now? Or maybe they're not." These questions are important, and evolving our culture is probably ultimately one of the deepest and most important things we have to do to not just address the skeptical points of view, but really take the right path forward.

Scott Hirleman
Yeah, and I think people want answers instead of questions to answer, and that's the other frustration, is that in data people have just been sold, "We will solve it," and so when people go, "But here's your homework to go do." It does frustrate people. And you look at some of the push back. A lot of the pushback is also from people that are kind of... I don't remember the economically captured or whatever, their bread is buttered by vendors that are against this. I think Bill Inmon just posted something about... And he was just saying, "Oh yeah, Data Mesh, you just push the ownership up and it doesn't matter about the silos and blah blah... " He's being somewhat facetious, but it's like some of it is this actually how you think, is it just that it's changed or is it that your bread is buttered somewhere else? Part of the reason why people view me as complete chaos is that DataStax doesn't really care about me trying to push anything direct to DataStax from this. This is all learning information, so I get to be chaos and just push what I think is actually good for people, and I don't have that buttered bread kind of challenge.

Tim Gasper
Everything's gotta fit into this box kinda thing, right?

Scott Hirleman
Right. And so you kinda have to look at, are those skeptics really skeptical or are they just like, "This threatens my way of living," and that's a valid pushback, it's exactly what you talk about culture too of, "Does this threaten my job?" Right? If I've just been a report consumer and now I have to be an active data consumer and figure

out what I wanna consume and why my world has changed. That's frightening to me. So how do we get people along that path, is difficult.

Tim Gasper
And I'm actually glad and excited about all the skepticism that we have, not just about Data Mesh, but really about anything. You can bring up Data Fabric, you can bring out any of these concepts of people like, "Ah, Data Fabric... " And obviously, there's proponents for these, but there's a lot of skepticism as well, and I think that's a beautiful thing because we're actually having a more critical conversation in the data space right now. I think we've all been a little bit... I connected back to the big data movement, I think we all got burned a little bit by being like, "Oh, Hadoop's the savior." And like, "We just put everything in the big data hub, it's gonna solve all problems, and we'll do predictive analytics, and we'll solve world peace." And the truth. Now, looking back, it's like, "Well, duh." But the truth was, that didn't happen, right? Big data is a piece of a much, much larger and more complicated pie.

And that's good that we're having this critical conversation. We're gonna come out better for it, and I think we're gonna really start to embrace what you mentioned. This idea that we do have to think about this more upfront. We have to design the future, we wanna live in, it needs to be flexible, it needs to be adaptable, it needs to be resilient. It can't just be, "Let me slap these four modern data tools together and voila, I'm done." It's a journey, not a destination.

Scott Hirleman
Yeah, exactly. Sorry, I'm just writing down the... I only got down reliable, resilient, flexible, but I think you threw one other superlative in there, but I fully agree. We've had too much locking into, "This is the way that we're gonna do it, and we've made our decision," and that versus like, "We need to set ourselves up that we measure and change." We've kinda had the, "Okay, we're gonna measure twice, cut once, and it's done," and it's like, "But this is changing and evolving." It's not that the other thing that you've created even six months ago is necessarily the right thing. "Oh, this report hasn't changed in five years." Major red flag, right? Is it still even measuring anything that's important? Right?

Tim Gasper
Right. Yeah, data is not a house that we're building, it's more like a garden. You're planting things, but seasons change, problems occur, the dirt is good one year, it's not the next year, and you can't be expecting that you're gonna measure twice and cut once. It's a process. It's a process, it's a journey.

Scott Hirleman
And you need to keep working on it. You need to keep pruning. You need to keep

weeding. You need to keep doing that. I like that analogy a lot. I think you should write something on that, or you should do a little post on data as a garden instead of as a house.

Tim Gasper
Yeah, absolutely, no, I love that analogy, and actually came up with our guest earlier this week on Catalog & Cocktail, so hey, I gotta credit that in terms of where it came from. But I love that analogy too. Yeah.

Scott Hirleman
I wanna be cognizant of time, 'cause I know you gotta hop here in a bit. I would love to give more time to this framework because I think it's a really good one, but let's go through your framework at the high level, and I might poke at a couple of questions here.

Tim Gasper
Yeah, absolutely, no problem. So it goes really back to a comment that you made earlier about having all the answers versus having the right questions. And we've been asking ourselves like, okay, well, Data Mesh, obviously, there's a lot of excitement and interest in that, there's a lot of skepticism as well. What is a good perspective on this that isn't trying to feed you all the answers, it's trying to help you ask the right questions? And really get people into a frame of mind of thinking about their data, like a product or what do they need to solve in order to treat their data more like a product. And so this framework is called the Data Product ABC's. And it's a little quirky, and actually Scott and I were joking in one of our prep sessions about maybe there should be like a kids book, like where it's like A, B, C, D, and each one of them has a funny, either sarcastic or fun term about either Data Mesh or just the data space in general. So I think we should get on that. We should loop in a designer and build this kid's book, I think that would be a lot of fun. Pass them out at conferences.

But the Data Product ABCs, basically has A, B, C, D, E. And A is first of all, accountability. So that's like, who's the owner of the data? Who defines the requirements of this data product? Who fixes the data product when it breaks, so thinking about... Who's the technical owner, who is the business owner, who's thinking about the road map of that given product? B is for boundaries. So every product, and this really comes more from my background as a product manager, products aren't just about what it is, it's about what it isn't. What is this product? What is it and what is it not trying to be? And what are its interfaces? What is the box that you draw around this particular product? What are the things that go into it? What are the things that come out of it? What are not the things that come in and out of it? For example, if you have an API product, you're not expecting to work with

that API product in the same ways that you're working with a product that lives in your data warehouse, or your data lake. And that's okay. Embrace that. Be specific about it.

So B is all about boundaries, C is contracts, so different data products have different contracts and expectations around them, is there an SLA around this, is there a performance characteristic that you're guaranteeing around this product, is there a certain security stance that you have around this particular product around sort of like how it's gonna be stored. Who's gonna use it? These are all contracts, and sometimes data products are gonna have looser contracts, sometimes they're gonna have stricter contracts, and that's up to you as an organization in terms of how you manage your governance, how you own and manage your data products. D stands for downstream, downstream consumers. So who's gonna use this? Who uses it today? Who might use this? What are the use cases that have been considered for this? What's the roadmap for this, like are you planning to handle more use cases in the future? And then finally... E, it actually used to be K, But A, B, C, D, K, felt kind of weird. So we were like, Okay, well, it's E, and E stands for explicit knowledge, and explicit knowledge is really like what is the meaning of this data product and the things within it, what is the schema of this data product? How does it relate, what are the relationships between this data product and other data or other data products and provide the documentation around it.

I think one thing that we see is so valuable, if you can establish it within your culture is more of a culture, not of knowledge stuck in people's heads, but knowledge put down on paper, put into systems. Institutionalized, right?

Scott Hirleman
I'm gonna guess that that was Juan that was jumping up and down and saying, we must include this one, we must include this, 'cause he's got the whole knowledge first thing.

Tim Gasper
That was a good conclusion there. Yes, definitely the knowledge piece came from Juan, and that fits into this whole thing of like, okay, so you've got the ownership, you've got the contracts, the boundaries, the downstream consumers of this, and ultimately you want it to be steeped in in knowledge, take a knowledge first approach.

Scott Hirleman
Well, and I think that explicit word actually hits almost every single thing that you actually mentioned. I know you did it so that it wouldn't be A, B, C, D, K. But explicit accountability, explicit boundaries, explicit contracts, thinking about, I guess, maybe

not explicit downstream, 'cause the downstream is a little bit more open ended, but so much of what I'm finding is that people need to actually have these conversations. So much is implied when it comes to data. I had Abe Gong on talking about Great Expectations, and they talked about defensive data contract creation, using Great Expectations to say, I expect that this data that I'm consuming from is going to be in this format in this way, and it's gonna meet these kind of different SLAs and things like that inform me when it doesn't happen, and then once that contract is broken, then those consumers go to the producers and start to negotiate an actual explicit contract, and it's like stop doing the implicit jump to the explicit. Have the conversation.

The number of people that I have who just say, “Well, just set up some time with me to do a call,” and nobody does, I'm worried about their implementation. I'm worried about your implementation if you're not willing to have a call, I don't have anything to sell you. So what's the drawback of having a call with me, if you're unwilling to get into, again, that culture of like, let's actually get in the room and exchange our context. Exchange what we're thinking, and so that there's no more of this implicit BS, right?

Tim Gasper
I think that's so important. That's so important what you're saying here. 'Cause I think this might even be like if you could define the central struggle of data organizations and what they're dealing with for decades, but certainly also today, right? It's that so much is implied in the work that they do and the things that they produce, and that's why you've got data engineers waking up at 3:00 AM on Sunday morning because there was an implication that something was gonna work a certain way and it no longer works that way, and somebody's pissed off downstream. And so how do we avoid that? You have to get explicit. And maybe that's something that we can really draw from the world of software, there's a lot of these analogies that we're pulling from the software world and trying to apply it to the data world, this is an important one, which is like in the software world, you have an API, that API is gonna provide that certain payload and it's gonna do it in a certain amount of time, and it's gonna be across a certain protocol that's very explicit, right? And we need to get explicit with our data too.

Scott Hirleman
Yeah, yeah, Emily Gorcenski's episode on SLAs and SLOs and kinda talking about resilience engineering, nothing in data is resilient, nothing is set up to be resilient, everything is one off, everything is about this specific use case, so there's not even reuse and then everything is like, okay, well, we just thought it was this way, and it was like, well, okay. When something doesn't meet your expectations, you throw a ticket to the data engineering team and then that data engineering team goes to

the data production team and you're playing telephone back and forth. Get in the room, have the conversation, do it with high empathy. This is not combative. This is cooperative, right? This is co-op, you're in two player mode in your video game, you're in co-op mode. The goal is to get to that data, informed decision making and the right decisions and all that.

Tim Gasper
Right, you should be shaking hands at the end, you should document it and know that you're gonna come back to it at some point because meaning changes.

Scott Hirleman
Oh yeah, and when that meaning has changed, make the explicit call out, that explicit reset, reach out. If you see something, say something, something has changed, who do I know that's consuming this that would wanna know that? It might be a new feature, but it might be, hey, this concept doesn't really make sense anymore. Let's reevaluate this, let's have that conversation. And consumers have to be flexible and not lock into what they're consuming as well. But it's again, that culture that's a difference in the way that people have done data. It hasn't been working for a long time, and it's especially not working now, it's accelerating. It's not workingness, is that right?

Tim Gasper
Well, if we're approaching new architectures with the same sort of roles and perspectives and biases that we had before, we're just gonna repeat history in new and creative ways but it'll result in the same ultimate problems, and so we gotta break through that and we gotta be willing to build more of a resilient approach to data that we know can adapt and be flexible as we go forward here.

Scott Hirleman
We're throwing band aids over things that that's not gonna fix, I don't wanna get into gross analogies around the wounds or anything like that. So well Tim has been awesome, really, really appreciate the time, and I think the perspective is very helpful for a lot of people, including people that are running across skeptics internally. I think addressing skeptics as, what are you really afraid of, is important and it's valid and it's okay to have those conversations, it's not that Data Mesh solves everything for everybody.

Tim Gasper
Absolutely, I don't think you need to come away thinking, oh, we need to adopt a Data Mesh. I don't think that that's the answer here necessarily it might be. But the answer can just be like, Let's look at Data Mesh, let's understand it, let's see what's valuable here, and let's build the things into our data roadmap that makes sense for

us.

Scott Hirleman
But it is evolution, it's not revolution. It's not you decide you're gonna go down this Data Mesh journey and everything has changed the next day. Like you're figuring out how to do it, you're figuring, you're finding your reuse, you're making it better, you're evolving and adapting.

Tim Gasper
That's a feature not a bug.

Scott Hirleman
So if people wanna follow up with you, where is the best place and what do you want people following up with you about?

Tim Gasper
Yeah, absolutely. Well, first of all, check out Juan Sequeda and I's podcast Catalog & Cocktails, you can find us on all the main podcast platforms as well as social platforms, we broadcast every Wednesday, so check that out, and if you're interested to learn more about the data product ABC's, we have a lot of cool materials over on the data.world blog, so that's another great place to learn about the data ABC's, and you can find me on Twitter @timgaspar, shoot me a note any time and would love to continue the conversation around Data Mesh.

Scott Hirleman
And as always, I'll drop links to that in the show notes just so that people can jump to things very easily, but Tim, again, this has been so helpful, I think, to a lot of people, and I think it's just been a fun conversation on my end as well, so hopefully it's same for you.

Tim Gasper
Absolutely, it's fun to look at these different perspectives and Scott, your show's awesome, really love it. Really excited that I could be on the show and looking forward to hear the communication and the community chatter that comes out of this.

Scott Hirleman
Yeah, same here and yeah, I've got my two favorite data related podcasts right now are Catalog & Cocktails and the Data Engineering Podcast. So I do recommend people checking it out, and it's a little bit more of not quite craptalking, but it is a little bit more free flowing and Juan likes to kinda go for a little bit of controversy as well, I think you gotta be a good straight man.

Tim Gasper
You gotta be okay with a couple of rants, definitely a lot of cocktail action and overall, lots of good, honest no BS, non salesy talk about enterprise data.

Scott Hirleman
Well, I think wasn't Zhamak's drink mushroom coffee or something like that?

Tim Gasper
Yeah. Juan and I were like, What? Oh, interesting. Well, you know, it's troubling or difficult is that for a lot of folks like on the West Coast and stuff like that, we always record at 4:00 PM on Wednesdays, and so people don't always wanna have a cocktail at 2 o'clock in the afternoon. They're gonna bring their coffee, they're gonna bring their mushroom mixes, that sort of thing.

Scott Hirleman
So well again, Tim, this has been awesome. And so, thanks so much for the time. And thank you everybody for listening.

Tim Gasper
Absolutely, thanks Scott for having me on.

Scott Hirleman
I'd again like to thank my guest today, Tim Gaspar, the VP of Product at data.world, as well as the co-host of the Catalog & Cocktails podcast. You can find links to Tim's Twitter as well as some links to the content that we had mentioned in the show, including the podcast and the ABC's of data products framework 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 fulltime 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.