



Transcript provided as a free resource by



Data Mesh Radio Episode #254: Easing Into a Data Mesh Journey - Ocean Spray's Pre-Data Mesh Preparations

Interview with Paul Cavacas

Listen ([link](#))

Transcript provided as a free community resource by Starburst.

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

To get their Data Mesh for Dummies book (info gated), please see here: <https://starburst.io/info/data-mesh-for-dummies/>

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

0:00:00 Zhamak Dehghani

Hi everyone, this is Zhamak I am the creator of Data Mesh, the founder of Nextdata, we are reimagining what data sharing could look like, we are growing our team rapidly and we need you. If you are a distributed systems engineer, if you're a product manager or designer of a large scale PaaS SaaS infrastructure, please check out our page at Nextdata.com and look at our open roles, we'd love to hear from you.

0:00:31 Scott Hirleman

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

0:00:38 Adrian Estala

Welcome to Data Mesh Radio with your host, Scott Hirleman, sponsored by Starburst. This is Adrian Estala, VP and Field CDO at Starburst and host of Data Mesh TV. Starburst is the leading contributor to Trino, the open source project and the [Data Mesh For Dummies](#) book that I cowrote with Colleen Tartow and Andy Mott. To claim your free book, head over to starburst.io.

0:01:08 Scott Hirleman

Data Mesh Radio is provided as a free community resource by Data Mesh Understanding. It is produced and hosted by me, Scott Hirleman. I started this podcast as a place for practitioners to get useful information about data mesh, and we're at over 200 episodes. I have now left DataStax, thanks for all their help in founding things, but I've left to start Data Mesh Understanding, which is also helping practitioners to get to the information needed to do data mesh well. We have free implementer introduction and roundtable programs in addition to the more advanced yet affordable offerings, so please do get in touch if you're looking for more information on how to do, how to approach data mesh. Just check



Transcript provided as a free resource by



datameshunderstanding.com for more info. There's also a helpful organization of past data mesh radio episodes there, if you want to dig into specific topics rather than digging through 200 different episodes. So with that, let's hit the funky intro music and listen to what you'll hear about in this interview episode.

Episode 254: Easing Into a Data Mesh Journey - Ocean Spray's Pre-Data Mesh Preparations. Bottom line up front, what are you gonna hear about and learn about in this episode? I interviewed Paul Cavacas, Senior Manager of Data and Analytics at Ocean Spray. Quick note, before jumping in, Ocean Spray is just at the beginning of their journey, kind of in their pre implementation phase. There hasn't been a lot of resistance yet internally that might make a few people out there jealous. Right? But there's a lot of interesting things Paul is doing to ensure that they are ready to decentralize what makes sense to decentralize at the right time. There's a lot to be gained from not rushing in and Paul also told me he's listened to a whole heck of a lot of episodes and he's picking and choosing things from a lot of different episodes. So people might wanna even reach out to him and be like, "Hey, what have you learned from listening to so many of the episodes?"

Also, apologies that my audio in this is a bit weird. I've yet to build my kind of makeshift sound studio in the Netherlands. So there's a little bit of kind of strange echo things that got somewhat taken out in post processing, but not all that well. So some key takeaways or thoughts from Paul's point of view. Number one, as many have stated, asking the data team, especially one person to become an expert on many different areas of the business, just to complete some data work for a project just won't scale. At best, it creates incredibly concentrated tribal knowledge, use this point to drive buy in for decentralizing data ownership, letting people know, like nobody can really learn all of this stuff. And if they do, then they're a point of failure if they end up leaving or moving on.

Number two, having someone who really knows your internal IT application landscape well can really help in choosing which initial teams to start with for a data mesh implementation. That person already has good relationships and a deep understanding of your operational plane, so you can pick good problem areas and partners. Paul is that he worked for something like 15 years on their IT and application side. So he knows so much about that, that when he moved over to the data team, he's able to help kind of navigate where they should start and kind of even who are the people that are gonna be the most friendly. Number three, similarly, build your early buy in momentum with people that are more likely to be excited to participate in a data mesh implementation. This comes up in a lot of episodes, you don't need to convince the most difficult teams to participate at the start.



Transcript provided as a free resource by



Number four, central ownership isn't necessarily bad until things stop scaling. Having that central ownership means less flexibility and agility to react quickly to market changes or opportunities, but it also means less cognitive loads on teams, it's a trade off. Number five, many of your domains really won't understand data ownership, find ways to slowly transition them into understanding what ownership entails, such as starting with documentation and SLOs. What data are they sharing and why does it matter? This isn't going to happen overnight. Number six, if you aren't building overly complex data products, look to find people within the domain that are somewhat technically savvy, especially if they want to advance their careers and start to prepare them for data ownership. Those might be your data product developers or data product managers. Scott note here, Brian McMillan talked about a plan to do exactly that. He wrote a book on doing exactly that. So he talked about that in episode 26.

Number seven, potentially controversial. Even if you aren't looking to move fast with your data mesh implementation, look to put at least a proto platform in place so people can at least test out ideas if they want to move more quickly. This can be a little bit dangerous if you get people that are putting things into production that shouldn't be into production yet, but getting that sandbox and letting people get towards being more comfortable with data is always a good thing. Number eight, setting up your testing and data contract framework will probably be a big challenge. Talk to people about what they need and iterate towards it, but it will be difficult to figure out at the start and you'll start with something suboptimal, be prepared and go forward anyway with that suboptimal approach to your data contracts, I can't emphasize this enough. Number nine, focus on finding the right partners. Those willing to really partner and learn and try rather than "the right domain" that has the most valuable data or use case. Again, this comes up so often.

Number 10, if you are looking for the right partners, there are good signals to watch for such as who presents their internal results in a quarterly basis or monthly basis. And they use advanced data graphics because they're likely a domain that really cares about the data and want to dig in more. If they're out there already analyzing things and showing that they think that the data is important and not just throwing up some graphs or figures, they're the more likely to wanna lean in on this. Number 11, consider how much need there is to decentralize certain aspects of your data practice. And if there is enough capacity and/or work for a domain to actually decentralize into that, those might sound different, but if there isn't a ton of work, that's a lot of effort for a domain to learn to own data for a small amount of data work. And then Scott note here, Yushin Son at JPMorgan Chase mentioned they are creating a central team for a handful of small domains where they can't justify a full data product team in that domain in the panel episode on number 233. I think that's a really interesting insight that you can overly decentralized. This doesn't mean you have permission to not decentralize at all, but there is a balance there. Data mesh,



Transcript provided as a free resource by



doing data mesh, right? Isn't about trying to decentralize absolutely everything.

Number 12, when creating a satisfaction scoring system for data products, consider what is the goal of the satisfaction score and how long that feedback is useful. So when a data product is in beta, harsh, but useful scores are valuable, but a two on data quality before version 1.0 might not be relevant once that data product is actually released as an actual product, it's okay to wipe past scores clean when there's a new iteration or something like that, you don't have to have those Amazon reviews from 10 plus years ago. Number 13, potentially controversial. When looking at your first data products, you don't want to lift and shift existing data assets to your platform, but you can go after existing use cases as your early products if you are providing the same thing in a better, easier to manage format. This is controversial because a lot of people have said don't try and target lift and shift. So I think it's interesting if you can find people that are like, "Oh, this would make my life easier, so I will move it over."

Number 14, don't let perfect get in the way of good or progress, it's okay to put in solutions that won't work in the long run, as long as you acknowledge they will need to be replaced and give yourself the room to replace them, it's taking on technical debt through a conscious choice and not locking yourself in. Finally, number 15, potentially controversial, when you are getting started, make sure to work with a few different areas to prevent yourself from delivering a solution even at the mesh level that is overly targeted to one domain's needs. Okay, enough of just me. Let's hear from our awesome guest in this interview episode.

Okay very, very excited for today's episode. And I've got Paul Cavacas is here who is the Senior Manager of Data and Analytics at Ocean Spray and if you're not familiar, they make delicious drinks and for a while I was relatively addicted to, I can't remember, it was like cranberry and blueberry or something. So we're gonna be talking about a lot of different things. So Ocean Spray is starting their journey and but kind of going at a pace that makes sense for them, right? It's a very, very interesting company, the structure and stuff is very interesting. We might get into that a little bit, but like trying to not just rush into doing data mesh for the sake of doing data mesh, how can you make that journey into it? How do you plan ahead, but also how do you know when you're kind of done planning and all of that? So I'm really excited to kind of have, more of these conversations about people going at a reasonable pace instead of trying to absolutely blitz through the data mesh journey as fast as possible, but before we get into that, Paul, if you don't mind, if you could give people a bit of an introduction to yourself and then we can jump into the conversation at hand.

0:12:10 Paul Cavacas



Transcript provided as a free resource by



Sure, thanks for inviting me. So it'll be an interesting episode, I hope. My name Paul Cavacas. I've been with Ocean Spray for about over 20 years now, and I've been in software my entire life since I was like 10 years old, I started writing software and started writing code and playing writing games and all that. So during most of my career, I've been a software developer, up to about probably about five years ago, I kind of switched more into data analytics and kind of doing APIs and kind of managing our data right now, Power BI stack and all of the platform that it sits on top of and in the process of kind of managing that and looking through that sounds of look at data mesh and looking at it really just, it kinda like resonated with me because it's all about software development for data. So, since I've been in software developer most of my life, it really made sense to me. So that's really what we started a few months ago. We started looking into that. That's kinda where we're at now.

0:13:12 Scott Hirleman

Awesome. And I would love to hear about that kind of... Let's talk a little bit about what were the specifics around what made you realize couldn't keep doing what you were doing? What was the specifics that was the trigger to say, "Okay, it's time to start to look at this." Was there like an aha moment or was it kind of a building moment or how did you start to say, "Let's start to look for something different?"

0:13:38 Paul Cavacas

Yeah, so there were a number ones, but a number of different things that really kind of hit the mark there but the biggest one was really we working on this project, it's a financial project, so getting all of our financials together for the manufacturing section and in trying to do that, basically I had to become the expert about manufacturing and receiving in transportation and logistics and understanding all that data. And I'm like, it's great man. Now, I'm to the point where they're like, "Ah, yeah, when you're done, you want to come work for our department." But that obviously is not scalable and I'm like, "I can't, I go rather learn, go all of marketing and then go learn all of," whatever, so on and so forth. So that's probably the biggest one is really just the fact that like understanding and distributing out kind of who owns data and what it really means.

But beyond that, even though I've been in software development like 15 years ago, I wrote an analytic, I wrote this, application that we use, still use today that does our analytics, it's essentially kinda like a pivot table, but it's a custom application that we built and I supported that, I was 24/7 on calls support and everything else, I gotten all the calls at night at two o'clock in the morning, three o'clock in the morning and got bad data, you get stuff out of we use SAP and we get to, date fields out of SAP that are 000-00-00. That's not a date that ain't gonna work very well, so I've been on both sides of supporting data, supporting analytics and transformations and all that type of data platform, data flow all the way to now, like I said, working with the business



Transcript provided as a free resource by



teams to understand not only what they're trying to do, but actually understand their data. So those are probably the two biggest pieces, yeah.

0:15:31 Scott Hirleman

And was this that it's like you are the one that's leading it, or was there kind of like a group thing of everybody realizing, "Hey, wouldn't it be great if we could do more?" Or like, has it been that there's building demand for this or is it more that you're kind of driving it because you're like, "I see the straining points and I need to address the straining points," or 'cause I'm wondering for that conversation about buy in, right? If everybody's coming to you and saying we need something different, or if they're happy, but you're so overloaded that you're like, "Hey, we've gotta change this."

0:16:08 Paul Cavacas

So right now it's mostly me leading it, but really what actually happened is we have a consultant partner, a company called Nagarro, that they were like, "Have you looked at data mesh? What have you done? What's your plans to do anything with data mesh about six months ago?" So that's when I started looking at it and looking at that and looking at our current workload, everything that's going on. I was like, these two could really fit together so my director and the CDTO, the Chief Data Technical Office, I forget what exactly what it is, but they were on board with like supporting this and we started talking about how we could push this out to the rest of the company.

The good thing is that at the same time, we're also in the process at Ocean Spray of really going through... Actually let me take a step back. We gotten to the point where we basically, the pendulum had swung very, very far to the left where everything was centrally owned, tech owned everything, tech knew law, the business processes. Then even the business did in some cases, not all cases, but in some cases. Tech fully ran entire process, business processes, knew the data, knew that know what was going on. Obviously people left and things have moved and now nobody knows anything 'cause it's not well documented. And now we're trying to kind of keep everything going and understanding what's going on.

So we're in the process of going through the entire... Basically the entire stack of all the data that we use of all the software and applications, the ERP systems and all the stuff that we use, and figuring out what our futures will look like. So in doing that, we're... Without even really knowing what they're doing. They're basically defining domains in defining business processes, and defining the data that owns like it's holding out each of those business processes. So I'm working into that process, so that we can take it, all the way into the data mesh, so that they are not only just defining the business process now, but that they audited the future.



Transcript provided as a free resource by



0:18:17 Scott Hirleman

Yeah. So it sounds like essentially, they're already finding the data products for you, so to support the business processes and you just kind of work backwards into that.

0:18:25 Paul Cavacas

Yes.

0:18:29 Scott Hirleman

So as you said, this, this came up kind of six months ago of hey, have you started to look at this, you started to say, hey, you know what we are having these problems so let's look into this a bit deeper. How did you start to think about that decentralization journey, right? Like, we were talking in the pre call, we were talking about, like, how do you plan ahead? A lot of people are stuck in this planning phase. But like, let's talk first of not just starting, right, where you go, okay, we want to do data mesh, let's start day one, versus like, how are we going to roll this out? How are we going to think about that? Like, how did you start the planning journey? Was it around collecting outside information? Was that about you going and doing a lot of research and homework? Like, what did you start to do that that made it so that you're kind of preparing yourself to do this story?

0:19:19 Paul Cavacas

Yeah. So we did... There's really two main things that we kind of have done, I've done a lot of research, a lot of reading, a lot of watching your podcasts, a lot of.. Across the board, just to kind of understand what all the different capabilities are. And I started to write a couple of blog posts about what I've done and what I'm planning on doing. So I've kind of put, from a technical standpoint, I've kind of like laid out a roadmap of kind of what I think a full data mesh implementation might look like, obviously, it's gonna change, six months from now, it's going to be probably a lot different than even what I think it is today. But it has a lot of the key different points, the testability, all of the kind of all the different things that you'd expect to see, documentation, testing, and all of that pieces. So I kind of laid out the roadmap from that standpoint as far as what we're doing, what we're going to do and we've started to build some of that.

From a company standpoint, and data products, and yeah, from that standpoint, it's very much we have a lot of stuff today, obviously, Ocean Spray has been around forever. So we've got a couple of different things today, which are kind of, I think a little bit as maybe data products, they're not data products, but you think that they could be data products, probably a better way to put it. So what I've done is I've started with a lot of these key pieces that are out there. So things around you about like revenue reporting, and financial reporting, and manufacturing, and some of the different pieces that might be up that we already have out there today, that I know



Transcript provided as a free resource by



something about. So those ones there, what we've done is started to look into those domains really, invoked those down into different pieces of what could potentially be data products, inside of each of those. I know from what we already have today, I know where they come from, I know who's responsible for them from a business standpoint.

So we've started to work with each of the different teams and kind of transfer ownership to them, and help them out so that they can help document it, help define use cases for our, define what the SLOs and the things might be. Some of this, obviously, this is all kind of work in progress, and so different pieces are at different stages. But that's really what we're doing is we're taking kind of the things that were pre assets today, data assets, just things that are there that we're maintaining from the central team, and breaking those down into the individual pieces and say, "Okay, where does this come from? Who can we work with to get that and make this into more of a data product instead of just like an asset that's out there?"

0:21:56 Scott Hirleman

And when you're talking to these domains, I'm sure they call themselves lines of business or something like that... Yeah. So, whenever I'm talking to, I have to switch between those two phrases, but when you're talking to them about, "Hey, you're gonna have to own your data." Is this something where they're pushing back or are they kind of bought in? I'm imagining it's kind of across the board that some are like, "yay," some are like, "nay," some are excited, some are pushing back. But like, how in general are you finding that willingness to start to take ownership where you don't just go, "I'm throwing it on your plate?" It's like, we're going to help you get to a place where you can actually do this.

0:22:43 Paul Cavacas

Okay, so what we did with this, is we really kind of concentrated on the people that we know would be more receptive to this. Obviously, I know there's going to be some people that are gonna be less receptive, we worked with them, they said, "You've been doing this for 15 years, why should I start now?" And so we're not really concentrating on that area, we're concentrating on the people that we know are going to be more receptive. What's kind of happened just naturally, in each of the different business and teams, there's been one or two people that have joined more recently that tend to be on the younger side, they tend to be more technical savvy and what not. So they've kind of taken over the what will be, kind of the curator of the data for that team. So we're working mostly with those people, 'cause they, a lot of them might come from another job, where they've used different things and they've had different access to stuff. So we're using kind of a lot of that experience that they have from other places and the fact that they're just more just, more savvy and more kind of want to learn and advance their careers, and we're working with



Transcript provided as a free resource by



those people to kind of really drive home who we're working with first. Obviously, in some of the big areas, manufacturing, that's a big area for us.

0:24:03 Scott Hirleman

Well, and that makes sense. I think, especially with, it doesn't sound like you're in a huge, huge rush to change everything. It's like, this is an opportunity because everybody is starting to say, "Okay, we are seeing that we're under strain and we need to change some things, but that we're not like rushing into this." And so I think that's a really, really helpful point for other people too, as well, look for is, finding those people within the domains where you don't necessarily have to have it at the highest level. You don't wanna come in and say, "Hey, I'm undermining you and I'm going to take your people's time and allocate it to what I want." But at the same point, there might be some people that get you, if not data mesh ready, if not that whatever, that they're at a level where they're going to be helping you advance what you're trying to do, they're going to prepare the domain to move forward. And it sounds like is that the right way that to think about kind of where people are headed right now?

0:25:16 Paul Cavacas

Yeah, that's very much kind of where we're at is we've got, in a lot of the key areas, they've hired when I'm gonna hopefully turn into some sort of analytical engineer type person. But today they're more like they, we use Power BI that's, that frontend visualization. So they understand Power BI so they understand the data, and they understand from a technical standpoint, how to build that and things like that, and kind of a little more data literate, little more, getting their hands dirty with understanding what the data means and how to work with it. So we're using those people to kind of help drive forward what we're trying to do.

0:25:54 Scott Hirleman

Yeah. Seeding the domain so that it's ready for moving forward, but not being like, "Okay, we've now handed all of this off, especially to somebody that's maybe a little bit more junior," and you go, "Hey, we've handed off all data ownership in this domain to your new person that's the junior on your side."

0:26:13 Paul Cavacas

Yeah. Yeah. It's very much about getting everybody ready, getting things ready, using our own internal team and resources since we've been doing this and supporting it for who knows how long that, we know a lot of the data, we know a lot of the pieces of the things that we're, like I said, those are the first ones we're doing, it's the ones that we already manage and maintain today.

0:26:34 Scott Hirleman

And when we were talking in the precall, we were talking about the concept of



Transcript provided as a free resource by



moving at the right pace and like, it sounds like you're just being very pragmatic and that it's maybe your nature to be like this of, "Hey, I'm not gonna get ahead of myself." But like, if somebody else is out there and they're like, "This is a problem, I must fix it and I must fix it now, I must fix it as quickly as possible." How do you think about having that conversation with them? Or what are you learning or what could you share about kind of figuring out what is that right pace for the organization? Because, organizational change is tough, especially if your organization isn't used to change. So like how are you thinking about that right pace and not getting ahead of yourself?

0:27:22 Paul Cavacas

Yeah, so I think we have kind of, there's two things that are really driving this forward, besides all the kind of, people and process stuff that we've talked about, like I said, I'm doing the technical piece as well. So by having kind of a technical base that we'd be able to fall on, if somebody wants to try and move a little bit faster, they say, "Oh, we really want it fix this. We really wanna get these going." Or, maybe they've read some stuff about data mesh, and they think it's really cool and they wanna try figure out how to do it. Then by having some of the technical underpinnings kind of there, we should be able to move faster with those people, 'cause we have that that we can fall back on. But right now we're doing more of it kind of on the side, more of it on our own time. A lot of this is really work that's being done in addition to all of my normal day work day job that I do. So I'm doing this and it's kind of an extra piece, but by having kind of that technical piece, that's really the... I know it's people process is the biggest, most difficult piece, but having the technical piece that you can fall back on and you know what's there, you know what capabilities are there, allow you to push people down into whatever speed that they want. They wanna go slow and will wanna slowly do one data product at a time. Great, we'll do that. I'm okay with that. We'll test everything out as we're doing it. If somebody wants to move faster, having a underpinning of what's there, will let us move faster as needed.

0:28:54 Scott Hirleman

And are you finding that if somebody wants to move faster than your... Is there anybody moving faster than you're able to kind of put support in place for them? Now I'm thinking of like the old Looney Tunes cartoons or whatever of the railroad going down the track and there are, Bugs or whatever is laying the track at the same point? Are you finding that people are trying to get ahead of themselves? Or are you finding that people are willing to go at a pace that makes sense for the organization?

0:29:26 Paul Cavacas

So far, I haven't had any problems. So we'll be, everyone's going, a lot of what we're trying to do, even beyond just, the normal, understanding the business process, understanding the data, trying to put a lot of like testability, a lot of observability and



Transcript provided as a free resource by



stuff in place, most of the people that I've talked with and that we've kind of worked with, haven't really thought, they don't think about this on any, any sort of regular basis. So just kind of saying, "Okay, yeah, these are the things that we're talking about, the things that we're thinking of being able to do." Some of the times, sometimes I ask questions like, "All right, what do you think would be the right thing to test for your, on your data products? What tests should we do?" You're like, "I don't know. Should be all there and it should be good and it should be right." "Okay. How do I know that it's all there and it's good and it's right?" So it's a lot of new concepts that we're introducing, even to people that are, like some of those questions that I was just talking about, I've done to kind of local tech teams as well, kind of own like customer or some of the kind of more core governed type of data that we have.

0:30:33 Scott Hirleman

Yeah, it's a difficult thing to kind of say, "Hey, we're doing this entirely new approach." And something that you haven't thought about, especially that I had a recent recording with Sofia Tania from Thoughtworks about like setting up a testing framework for data. And especially when you're thinking about, "Hey, we're making a change to this column. How's that gonna flow up and downstream, and all of that." And it's like, I don't think anybody's doing consumer driven testing, but everybody kind of agrees consumer driven testing should be a thing. But so like, I think, are you having conversations with people and they're afraid of all of the new things? Or are they just kind of like, "Hey, we'll work with you, we're gonna pair." 'Cause the reason why I'm digging into this is a lot of people out there that are listening are probably pretty jealous of everything you've said thus far, which is that you've got the buy in of management, you've got people are leaning into having these conversations. They're leaning into potentially owning their data.

That's a very atypical, right? You're in a very atypical spot. And so, where are you having the things that you're concerned about or where are you having things where you're like, "Hey, I think people are just smiling and nodding instead of actually getting it and saying yes." Is there that kind of thing that's going on? Or is it just like everybody's understanding that things have to change, but that it's gonna change not overnight and that you're gonna kind of all collaborate, work together to go forward?

0:32:16 Paul Cavacas

So if I painted a super rosy picture, I didn't mean to, because I very much have targeted kind of the who we're working with at first, who are the first couple of people are that we're working with. And I've done that intentionally so that I knew it would be people that I would be able to work with that were able, willing, and ready to take on additional responsibilities of ownership and kind of being responsible for data all the way through to the front end reporting pieces that they have. So that



Transcript provided as a free resource by



part, that's probably the biggest reason why things are rosy. I know as I move out, and move out to different domains that we're gonna run into lots of different kind of issues and push back, as far as, especially ownership, anytime you try and push more work on to anybody, then you've gotta be, you're gonna get pushback. But these first couple of people that we're working with, and like I kind of talking about the precall, we're very new to this, I don't have a ton of stuff in production, it's more of things that we're working on or we're building and it was starting to kind of get ready to move up into production. But we're not truly anything really in production at this point.

So that is the goal, that is where we're trying to get to. But I don't think we'll have a ton of the initial people... Like I said, I don't think we'll have a ton of problems with the initial people, as we move out with that, that's where we'll run into other issues with people as far as trying to get them to understand the goal and the vision of what we're doing and why we're doing it.

0:33:56 Scott Hirleman

I still would say it's a rosy picture, but that's just the reality. So it's nice to hear that. And this is something that I think it's nice to have a conversation with somebody who is so early in their journey. Most people are very unwilling to have a conversation until they're like, a number of times people have said, give me six months in production before I'll even consider it. But I think there's some good points about the momentum you're building, right? You are building up the momentum before really getting going. You're starting to say, "Okay, what are the tools? Hey, do people know how to use these? Let's get people kind of leveraging some tools, but not necessarily that they're experts in anything." And when you think about the platform you should build and the capabilities for people, you wanna make it so they don't have to be experts in the tools. They're just experts in the data.

But I think that where I'm seeing a lot of people struggle, is that they focus too much on going for, "the right domains," instead of the right partners, right? And so finding those right partners, finding those willing partners, the willing and able and capable and that that is something that a lot of people are struggling with. Was it that you already knew these because you've been there for 15 years, right? So you know the people very well, a lot of the domains, you were able to kind of do that, which at a larger organization that probably 100,000 person company or I just talked with somebody from Volkswagen who has 700,000 people. So it's maybe not that exact thing, but I'm trying to find how others can leverage what you've started to do of finding those capable people. Was it you just kind of knew them ahead of time, had the relationships, or was there some, like, some sign that others could look for or?

0:36:12 Paul Cavacas

Yeah, and from what we've done, it's very much, not only did I know the people,



Transcript provided as a free resource by



Ocean Spray, you might think it's a big giant company, but it's really not. It's smallish, relatively small company. So I kind of already knew a lot of the different people that are in the company, again, been here 20 years whole thing, all that. But one of the things I would say probably to look for is. Look for the teams that, so we have, I'm sure most companies have something, so we have MBR Manfred Business Reviews, or Monthly Business Reviews. So in those MBR, try and find somebody that is not just pointing up a bunch of numbers. So like we have certain groups of people where the review is just, "Here's the number, yep, blah, blah, blah." Just road column number. Find somebody who has a MBR that looks a little bit more involved, a little bit more, whether it's graphically evolved or whether it's, 'cause if they are taking the data and manipulating it and making something nice and not just like rows and columns of numbers, they typically understand something to do with data. Now maybe it might just be, there might be graphics designers and they'd be able to be, make things pretty. But if they are doing something visually, so look for something different, that's really probably the biggest key is look for something different whether it's in like a monthly business review or whether it's in some sort of presentations that they do. Find somebody that, the final output might be the best... Not the best, but different than everybody else.

0:37:49 Scott Hirleman

Yeah. They're leaning in already on. So they're at least... And I guess one thing that's come up in some conversations is a lot of people that are struggling, are going to domains and saying, you have to give your data to everybody else versus we're gonna make you more capable, we're gonna make you better with your data. So that's going to mean that you are better and as part of that, then it will also mean that the rest of the organization can access the data and we can do this. Is that a selling point that you're going with? Or is it not one that you even need to go with?

0:38:30 Paul Cavacas

So that's not one, you've had that... Sorry, retract. You've said that a couple of times in different podcasts and that's what I've never really fully understood. Now there's always data that people have that's off to themselves that they are not necessarily sharing, but from 20 years ago, our core data has been available to the whole company. Now maybe just rows and columns or what not. So like, you want sales data, we could open that application that I wrote, 20 years ago and see all the sales data and export it and put it into Excel and take it, do whatever you wanted with it. Trade data has been available. So there's a lot of different data that's been available.

One of the biggest things that have come up more recently is that more and more people are trying to get access to the data. Now, it's not that people are holding it to themselves, it's that we are having a lot. Ocean Spray uses a lot of off the shelf applications. We don't write a lot of custom out anymore. We did that, for the first 15



Transcript provided as a free resource by



years I was here. But now we're very much used to kind of off the shelf third party kind of tools, and that's what's generating all the data. So people have to like, get access to that, export it out of there, do something to get to the data. So sharing data has not... Well, convincing people that sharing data is the right thing, I've never really seen as a huge problem. It's that we've done that in some fashion for some core data, as long as I've been here.

0:40:01 Scott Hirleman

Yeah, that, I mean, that is somewhat unique and where I'm finding a lot of people are getting bought in is when you go to somebody and go, we're gonna make you better with your own data. And then as part of that, we're also gonna share it with the rest of the org. But it's great that you've got that kind of sharing culture, which I think is one of those signs that you're also going to be more ready, right? If you already have a sharing culture, if you already have these, like, I mean the monthly business reviews that is not something, I mean, it's a smaller organization, so you can have that 100,000 person organization. Not everybody can sit in on every domain's monthly business review.

But that is I think a sign of can you even get access to information to understand if people who would be leaning in and like going out there and finding that so that you're prepared and kind of doing the back channel research instead of just going to the head of that domain and going, you now own your data. You have to share all of your data with everybody. Like here's the way. Go. I think you're building up a lot of this momentum and not making it scary, right? It doesn't sound like when you're talking about ownership, that it's honestly true, true ownership because it's not that you're saying you absolutely have to own everything from the start. You're still gonna be supporting them, you're gonna be working with them, and then they're going to transition into that kind of true ownership over time. So it sounds like the reason why it's so rosy is because you've got a good culture, but you're also setting itself, you're setting up the people to be like, we're gonna make you successful. We're gonna make this work for you.

0:41:53 Paul Cavacas

Yeah, and a lot of the kind of work, especially probably engineering work, I probably have my hand tied, truly getting that fully distributed. We'll probably have a more central engineering team because, the team's a small team now, I can't imagine that every, the way it's going they're gonna have their own small team. There's just not enough time and work. So there'll still be a lot of centralist, even though we're gonna try and distribute out like the ownership, distribute out the kind of responsibility for it. But a lot of the more technical stuff would probably remain more in the central team as opposed to being fully distributed down, resource wise.

0:42:36 Scott Hirleman



Transcript provided as a free resource by



And, I had somebody from Iterable, Riya Singh and she was talking about, they're kind of embracing some aspects of data mesh, but kind of not because they're at a scale where they don't need to fully distribute that right? JPMorgan Chase even talked about on a recent panel, I think it'll be out, before this episode, I always get kind of confused as to when things are coming out, but that there were some domains that were small enough that it just didn't make any sense for them to own their data. So that data ownership fell back onto a central team. So you might have a couple of domains that truly do own their data or own it 80%, 90%. And these other ones, you've got a domain of 10, 12 people or something like that. I imagine logistics and shipping is much bigger for you, but like for a software company or something like, that type of things or facilities management when people have like two offices and they don't have a big manufacturing facility. You don't need like them to fully own if it's going to be way too much overhead. And so I think having that balance is really important. Now, I don't want to give people permission to just not decentralize any of that ownership. Right? You have to test whether it makes sense or not. But I like this approach of you're not doing data mesh for the sake of data mesh and you're not saying I have to do all of it, because clearly all the signs are saying you don't fully need to go all the way here, but 100,000 person company probably doesn't want to go and do the same amount of decentralization here.

0:44:25 Paul Cavacas

Yeah, I would definitely agree with that. And even to the point where like we'll have domains, even in my situation, we'll have domains where they're 80-90% of the way decentralized, and some that are maybe 30, 40, 50% decentralized, but even in those ones that are less, we're still going to try and shift the ownership of the data itself. So like defining what it means kind of from a business standpoint, shifting that... At least that responsibility onto the domains 'cause they have the context. I don't think it... Context is everything like you say a bunch of times. And I don't have the context for all of those, especially the little domains 'cause I've never worked with them before. So I don't know what a lot of things might mean.

0:45:11 Scott Hirleman

And there maybe some things that get lost in translation between that, but there is somebody that's at least overlooking it versus there's this clear sharp handoff of, okay, the data comes out of my application. It's now on you data team. Instead of like, hey, we're collaborating to make sure that these things are owned by someone and that there's a maintenance kind of concept to this, that this doesn't just it's not as if this is done, this is a product like that it's ongoing and that as things change, Abhi Sivasailam from, I think, episode nine was talking about when he was at Flexport, the concept of an order had changed and had grown in complexity 10x.

And so like having people that are kind of aware of this, it sounds like you're really



Transcript provided as a free resource by



setting yourself up for success, but not setting yourself up to try to be the perfect data mesh use case instead of we're taking what does work. So many people say they're just taking what works from data mesh for them and it's like you're just taking the easy parts. It doesn't sound like you're doing that. It sounds like you're taking the parts that are going to work for you. And if you grow and all this, it doesn't sound like Ocean Spray is really in that kind of we must hyper grow type of mode, which is kind of a nice organizational approach but that you're not if things do change then you're going to continue to decentralize, if that makes sense or you'll kind of shift the pendulum back and forth as you're testing. And have you communicated that with people or they just kind of inherently know that or?

0:47:00 Paul Cavacas

So I haven't really communicated that. I don't think we're really far enough along to the point where things are going to shift back and forth. They will. I definitely expect them to shift back and forth across different teams and different times. One of the things that I want to try and do is a lot of people have score like a you have a data product and you give this rating 1-5 or whatever like that. I'm going to do that. People can vote on it and give it a score, but I'm going to also actually make that a composite score. So I'm going to do things like if there's test defined, there'll be a test score. If there's usability, there'll be a usability score. So kind of you'll see, not only will the data product change over time, but you'll see if people stop using it, this thing isn't going to slowly go down. If it's changing all the time because people don't really know what they're talking about. They're just making version 2,3,4,5,6,7,8 the score would go down again and so you could kind of see kind of reliability of what those data parts might look like over time and you got to see what it changes and all the things like that. That's not built yet. That's a theory.

0:48:10 Scott Hirleman

Well, and it's a little bit beyond kind of a net promoter score of like, hey, were you satisfied with this? And I think having one thing that I really liked from the folks at NAV, which was Audun and Gørn, they did something they called cake driven development. And that was that anytime somebody brings out a new data product, their whole team gets like a very nice gourmet cake that gets brought in and then and that gets posted into a Slack channel and then people can ask questions about their data product and that type of thing. It feels like that's kind of the same type of attitude that you're going for of we are seeing that you're putting in the effort. So we're going to reward that and we're going to give you good feedback. But like I'm sure the initial data products, people aren't going to know what's good or bad. But it's not like your data product is bad, therefore you're bad. It's like, hey, the data product needs work like let's talk about that. It sounds like you're just kind of setting yourself up for a collaborative process instead of kind of a fly by the rules type process.



Transcript provided as a free resource by



0:49:20 Paul Cavacas

Yeah, that's what I'm trying to get to is very much like I'm talking about that score a second ago. When we are first developing something, if it's changing a lot and tests are failing all over the place. We'll just drop all that data when we can release the data product so it doesn't look bad 'cause again, the whole point is to succeed. It's not to follow exactly strict rules and be like, you you're not doing this, you're doing this wrong. You're doing this. It's very much of like you have whatever, you have sales data, you have logistics data. Other people can make use of that. Let's get that out there so that everybody can make use of it however it makes sense for them.

0:50:02 Scott Hirleman

Well, yeah, and then you talked about kind of... I'd like to talk a little bit about what you're planning on the user interface 'cause it's not like, it doesn't sound like you're taking your people and putting them through these, extensive bootcamps to get them to high data fluency and things like that. Are you just exposing them to the Power BI and that you're gonna try and work behind the scenes to make sure everything just is there? Or how are you thinking about the data consumers and also making sure that they actually really do understand what they're consuming, if they're gonna make key decisions off of it?

0:50:41 Paul Cavacas

Yeah, so there's two kind of two phases to that and you touched on both of them really. The first one is Power BI is what we use. So one of the output channels of the consumer facing data products that we do will be essentially a Power BI model. So people will be able to say, "Okay, I'm gonna go connect to Model X and that will have sales data." Helps definitions and descriptions. They can understand what it is that all the different tables and fields and relationships might mean, what the different measures and calculations might mean and how they will use use cases. So that would be one of the output boards I am trying to build. That's actually the first use case that we're really pushing all the way through. There's two of them that we're working on. But one of the first ones is all around sales data. This just general data that people have access to that is used a thousand different ways. So that's kinda the first one. So that'll be the bulk main population of the company would probably just use Power BI. Go look at reports, go look at data coming out Power BI which is an output channel that is driven by all this data mesh stuff that that some people probably wouldn't care. Most of the people probably wouldn't even care about.

The next level below that is what we wanna build, which we haven't built yet. We've talked about it and did a little bit of design work on what we want is we use Microsoft Teams. So you can embed a Teams app right inside of that. So I'm gonna create a central data hub application that people could go into and they can look at and they



Transcript provided as a free resource by



can see all the data products that are available. They could do searches and filters and say, show me this domain, this domain type data and see a description of it. See use cases, see different pieces, see all the tables of the fields that might be in that data. And then from there, be able to connect to different output channels, whether it's download a file that it might download a file to use and play with it in Excel or connect to it to either through Power BI or through an API or through any other different method that we end up building.

0:52:50 Scott Hirleman

Yeah, and it sparked a reminder of me of a little bit of what Adevinta with Xavier Gumara Rigol was talking about of, "Hey, we're gonna create some actual show and tell type things". They've got sample notebooks that come with each data product to be like, "Hey, here are the queries that we think are interesting from this." Not these are the only queries that you can do. But that type of of thing I think is really interesting and useful for a lot of folks to just spark their creativity and not say this is the only way to use this but...

0:53:32 Paul Cavacas

In all honesty, that's where I got that idea from. That wasn't in my original design, it was from that episode.

0:53:42 Scott Hirleman

It's good that the podcast is helpful for folks. Yeah, I really like that concept a lot and I talk about it still, even though that was literally 200 plus episodes ago. I think that was like episode 38 or something. Yeah, so I wanted to wrap up on this idea of... And we've had it as a through line, but of getting going versus getting it perfect and that it's okay to not be perfect at the start and where you're going with these initial data products. That it's okay for it to be simple views and things like that. How are you thinking about communicating that to people internally? Because people typically think of data as right or wrong and so this one or a zero, like how do you make it something that actually resonates with people that they feel like they can rely on this data, but that you're also not trying to go overly complex.

0:54:49 Paul Cavacas

So if we take that first use case that I was talking about basically the sales data, they can add general sales data that are available. The first data product that we're building, which will probably be more internal than really external, although I know a few people that will probably use them, really like you said, simple views. They're just views that are sitting on the top of the data. That data is the same data that's driving our models today. So our Power BI models things today. So there'll be some differences we gonna probably create structurally know different things. We use Snowflake, so different structure in Snowflake, but it'll be the same data and the



Transcript provided as a free resource by



same values. So I know that, from a data perspective, it's as right as it is today, whether that's how it was sent right or not, that's different story, but that's for another episode.

So that's the throughput there. So we're using data that we already know. We know the data's right. We know the data's good and the model that the Power BI model, that final data product that sits on top of all the rest of those will be pulling in these space data products. So they'll produce data products that are really coming directly out of SAP or various other systems will be merged together into a more cohesive model data product that's a consumer aligned data product that way people can use. That's what people will really more likely use is that one 'cause that one's got relationships between different tables and whatnot. But there are people that I already know that will be able to take the raw data and do stuff with the raw data that they can't do today. People are exporting stuff out of Power BI and stuff today to get at the raw data. We'll work with those people to switch them to say, "Hey you can do that directly through this new method here. And we will work with that to get it."

0:56:44 Scott Hirlleman

Well, and it's funny, the word raw data, can have a couple of different meanings because I talked to Michelle Conway who's a lead data scientist at at Lloyds Banking. And we were just talking offline about this and I said with Data Mesh, it provides you a clean source of data so you know this data that you want. It's actually cleaned for you and prepped for what you need it for. And she was like, "I don't I want the raw data. I want I don't want it to be transformed." And I'm like then you but you still want it cleaned right? Where you want it that it's a reliable source but that it's just non transformed. And that just clicked in her. You could just see that the physical click of like yes having a reliable source of clean data but that it's not already transformed and I have to untransform it would be perfect because that would just solve so many of my challenges.

And having that conversation of those people using the raw data do they need it as untransformed? And so you go that's gonna be in the data product or do you actually want it transformed 'cause you're having to do these same transformations, Khanh Chau when he was at Northern Trust had an episode and he was talking about they had data services and the quality just wasn't there. And then they'd talk to the four different users of of a data service instead of a data product. And three of the four would've been doing the same transformation on it. So that transformation got done three different times and all three of them had to do quality checks and all this stuff instead of pushing that back up into the product level. And so I think like you're you're again I think yes I think you'll have challenges when you start to talk to domains that aren't as bought in but I do think that there's a lot that can be learned from yours but there's also a lot that's nonapplicable to people out there because so



Transcript provided as a free resource by



much of your stuff is in this sweet spot of you're not under this massive time crunch, you're not having this big big pushback of resistance.

You've got enough institutional knowledge where it's not like you're dropped in on day one and you're trying to meet everybody and hand over responsibilities to them. But, I think there's a lot of, like I said of good learnings as to how can you set yourself up to be successful? Because I mean I can't guarantee that you're gonna be successful but from everything that you're saying I think you've got a very very good shot because of kind of going through these and doing the prep work but also not being like I must have everything perfect versus hey we're gonna learn and iterate together. We're gonna go forward we're gonna... Carlos Saona talked about this a couple of times but he said like my data product is not stable until it has hit three months. I have to be able to make breaking changes. I have to be able to do this stuff as I really figure out what people are are doing. So don't become overly reliant on exactly what what I bring out as an initial data product because I need this flexibility. And it sounds like you're setting everybody up to be like we're in this together and we're we're creating this flexibility which I think is something that a lot of people can take a lot from but, I do think a lot of people are just gonna be jealous of you as well after this episode.

1:00:25 Paul Cavacas

Yeah, yeah, that's kind of like I was saying. We're going to do all the testing, we'll do everything else. And if the results are bad for the first month, two months - and then it stabilizes, I'm okay, we'll throw those results out and start fresh so that it's not penalizing all the future use. Because it's not about what is the score, it's about what is the usage, what can people do?

1:00:51 Scott Hirleman

Yeah and I think about this with physical products it's really hard to kind of iterate on physical products right? Like but well and you know this with manufacturing but you could even say oh okay we tested this con... Or this mix between the three different fruits or whatever juices and we're finding that this mix is way better received than this other mix and so that's what we're gonna do. But when it comes to software and data and things like that well when it comes to software we're used to being able to iterate when it comes to data we're not and so setting yourselves up to be like we're gonna have these changes because we want it to serve what people need instead of we want to lock in exactly what we serve from day one.

1:01:42 Paul Cavacas

Yeah. And that's the historical problem is locking in this is my schema this is... I can't change. That's the problem that we're trying to resolve. Yeah.



Transcript provided as a free resource by



1:01:51 Scott Hirleman

Yeah. I'm seeing that with data contracts where people wanna lock on schemas and I'm like, we've had that with APIs on the software side and if you don't have APIs that can be somewhat flexible as things are changing and iterating you're just it's you're gonna have a bad time.

1:02:10 Paul Cavacas

Yeah. I mean contracts are useful and we're gonna need them but yeah we gotta figure out the right way to kind of implement them. I've done a basic implementation of contracts but it's very much is schema driven right now. It's like do all the columns match what's in there? And if they don't then it fails the schema. I know that that's not sustainable in the long term but again it's an iteration. That's the first step. Does it work?

1:02:36 Scott Hirleman

Well, yeah. And that's the testing thing of like hey we're gonna get our arms around that this stuff matters. And then but like, I don't know that I've seen anybody that has actually done it. Not schema locked like I don't know of anybody who's doing that but I think we all know that that shouldn't be what we do but it is the way that everybody's doing it. So well Paul we've talked about a whole bunch of different things. I'm really excited for you and your journey. I don't get to talk to many people this early in their journey so, I really enjoy talking about this. Is there anything that we didn't cover that you wanted to or any way you'd kind of wanna wrap up the episode?

1:03:15 Paul Cavacas

No I think we covered pretty well everything that's there. I just think if people are early in their journey just word I know this is all gonna sound kind of productive to most people but work on a couple of different areas at the same time. 'Cause you'll learn from one and the other. You like you can piggyback on top of them. If you just focus on one then your solution's gonna be targeted towards that one person and that one person's need. But if you have at least two different people from different domains and different areas then it's yeah you can bounce ideas and things back and forth between them to see what makes sense in total.

1:03:51 Scott Hirleman

Yeah. That building up the bigger picture to what is the actual enterprise need instead of what is the domain or that use case need.

1:04:00 Paul Cavacas

Yeah.



Transcript provided as a free resource by



1:04:01 Scott Hirleman

Yeah. So Paul I'm sure there's gonna be a lot of people that would love to follow up with you. Where's kind of the best place to do that? Anything specific you'd like people following up about?

1:04:09 Paul Cavacas

Yeah so best place is LinkedIn. So you can come to meet just to talk about this, in LinkedIn and they could follow up really pretty much anything. I'm open, I've been like I said I've been in computers and development for a lot longer than most of my life let's put it that way. So I know a little bit about a lot of different areas so I could talk to people about it.

1:04:33 Scott Hirleman

Yeah. Yeah. Well and again Paul thank you so much for your time today and as well thank you everyone out there for listening.

1:04:41 Paul Cavacas

Yes. Thanks everybody.

1:04:42 Scott Hirleman

I'd again like to thank my guest today Paul Cavacas, Senior Manager of Data and Analytics at Ocean Spray. You can find a link to his LinkedIn in the show notes as per usual. Thank you.

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

We also have some free programs around introductions and round tables that



Transcript provided as a free resource by



Starburst

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