AI LEADERS PODCAST: DATA MESH FROM THEORY TO PRACTICE TERESA TUNG, ABETH GO, LIAM DONOHOE

AUDIO TRANSCRIPT

ABETH GO: A growing need for us to able to share data with three different parts of BP and essentially the old model was not going to work.

TERESA TUNG: Hi, welcome to another AI Leaders Podcast. I'm Teresa Tung. I'm Accenture's Cloud First Chief Technologist, really looking at the future of cloud. Super excited to be talking about the hottest topic in data, Data Mesh. And really excited to be joined with my friends in the data space, who's working on this Data Mesh journey. So welcoming our partners from BP.

ABETH GO: Thanks Teresa. So I'm Abeth Go. I'm the VP for Data and Analytics Platform at BP and our team is responsible for building and operating the technologies that connect the world of BP with data. Thanks for having us, Teresa.

TERESA TUNG: Thanks, Abeth. And we're also joined by Liam.

LIAM DONOHOE: Hi, Liam Donohoe. I'm a principle architect and working along with Abeth on our data and analytics platforms. So really trying to figure out how we're putting together and the technology architecture.

TERESA TUNG: Thank you for joining.

ABETH GO: Liam's very modest, Teresa. So Liam is actually the one responsible for getting us into this Data Mesh trouble. So, yeah, he's the master of this whole thing. So it's thanks to him that we have been proactive with Data Mesh.

TERESA TUNG: And that's why I'm so happy to have both of you. So both from a business perspective, really owning the tools, as well as an implementation perspective. You're working with the tools in the organization. You're one of the few companies that have actually started on this Data Mesh journey. So hoping you could just start with why are you interested in Data Mesh? And maybe is this a new concept?

LIAM DONOHOE: Yeah, so I think the trigger for us started with some work we were going around the data fabric and BP has been in a journey that's really evolved our platforms through two key dimensions. One's been around technology and architecture, but seen this transition from data lakes, big data that everyone's familiar with to data fabrics where we're really trying to tackle the unification and standardization of varied and distributed data ecosystem and part just trying to realize and support data marketplaces.

The other aspect and dimension is really around the data organization transformation that's going along with that. And we'd seen the establishment of an organization we refer to as BP Data Works, that was setup to really facilitate the development of data products within various data, the main team is underpinning our digital agenda.

As we got to the point of trying to understand where do we go next, the Data Fabric 2.0, so to speak. So we considered as part of our data platform diagnosis, what were the key shifts that we needed to make to really realize the full benefit potential that we hadn't quite made it in previous implementations. And think recognition there and realization was that we really needed to bring together both the technology and the organizational aspect at one time, rather than dealing with them separately.

And very much the dimensions or principles that went along with that resonated with what we were seeing from the emerging Data Mesh thinking that had come out from domain and so, for us essentially, and were embodied in a leading focus data platform capability that was integrated across the whole data value and the theme around (inaudible) oriented, decentralized data ownership, as much as anything to be able to achieve a scalable architecture that was very focused on data products.

And in that final theme around self-service data infrastructures platform, so how do we enable these data works teams to operate individually, but do so in a way that's been governed through a central federated base policy. So that resignation with those Data Mesh teams led us to drop the Data Fabric 2.0 term and from then on, it's all been about Data Mesh and really embracing that.

TERESA TUNG: So, Liam, you pretty much just made the case version looks attributes, right? You went from a Data Fabric, which it sounds like it was much more technology driven to the four attributes that she mentioned. So that decentralized domain driven ownership having a federated model data as products. I heard you mention that. A self-service infrastructure to really democratize and scale data and then, I don't think I heard you say, global governance, but I certainly heard federated and I assume as part of that is that governance model?

LIAM DONOHOE: Absolutely. And I am that governance, although we talk a bit federated, it's really trying to deliver essentially defined global governance and policies and being able to roll that out. So it's effectively applied across the various pockets of where the platforms been realized.

TERESA TUNG: So, Abeth, I'm going to give you the opportunity to answer the same question. So why are you interested in Data Mesh given your role?

ABETH GO: Actually, it's very much what Liam said. That the key thing there is the fact a lot of the initiatives before has been very much technology led. We needed something that would work with our organization. And if you think about BP, we're such a federated organization. It's engrained in terms of the way we operate, in the terms of how the way we work. And so, we need to think about global requirements, including regulators and compliance measures. And at the same time, cater to the unique nature of knowing each of our businesses and how our geography works.

And so, data is very much the same. And we need to scale our data operators at the same time, be sure that the data is governed securely and properly. And there's this growing need for us to able to share data with three different parts of BP and essentially with the old model was not going to work. It was not possible to scale with the old model that we have. And so, this Mesh operating model, really truly speaks to organizations like BP and our ambition. And in order for us to make this work and actually, it will help us make it work because when we started going into implementing the Mesh, what is becoming really clear is really the concept of data management, data governance and also starting to elevate the conversation at the data product level. Because if you're talking about things at the technology level, we're always at the raw data set. When, in fact, what we want is for people across organization to recognize the data product as the asset, be able to share it and we just couldn't go (inaudible) and not in the scale that we want to be able do it without the Mesh operating model.

TERESA TUNG: So maybe you could back up a little bit and talk about the BP data landscape? So, clearly, you're talking about needing a data fabric. Can you maybe just level set, what are the platforms that you have today? How advanced are you prior to this journey?

LIAN DONOHOE: Yeah, so I think it is important to start with the realization that as a global company, BP and as a technology direction that's founded on a dual cloud strategy. So we have our data platform by necessity is to distributed across both those cloud environments and not only within that, we have multiple incidences of data technology and platforms within each of those clouds as well. So it does make for an interesting challenge that we have this very disrupted foundational platform based around some key cloud trained or technology sets. But then, we've additionally will complement that with a number of niche vendors that bring competitive advantage and capability that we don't yet find within those foundational cloud platforms.

So it's a fairly distributed environment, a fairly rich variety of IT of technology that we have to be able to through necessity make that interoperate in a way that is really the heart of what Data Mesh is trying to achieve and it's fairly important for us that given that diversity, that we're able to really recognize and identify which each of these different components of the overall platform can realize and bringing and delivering data products and their relative strengths and how we mix and match that as part of the platform as a whole.

TERESA TUNG: So you are needing to do a Mesh because it wasn't your platform or mine. It sounds like you were already dual platform, multi-technology, lots of data, lots of parts of the business. So it sounds like Mesh was perfect for you.

LIAN DONOHOE: Absolutely. This is very much a necessity, not a luxury. It's not a theoretical exercise. It's something that we really have to make work to be able to bring that diversity and fundamentally achieve the interoperability aim around these various technologies and varied distributed data products and to bring them to life together as a mesh.

TERESA TUNG: So, Abeth, the next question is for you. We know that we need the Mesh at BP, but from a business perspective, what's the vision? Like how did you make the business case?

ABETH GO: Well, maybe I'll start with the last question. We did not make a business case for the Data Mesh. So the business case is really to do with building a comprehensive data platform service to address the company's data needs. We need to increase velocity in terms of how we deliver data products. We need to reduce the fragmentation of our estate. We need to accelerate BP's ambition to become an integrated energy company. So that's the vision and the mission. And it was not our Mesh. It was more to do with the Mesh is going to help us realize that because we need to move fast.

I mean apart from the limitations that Liam mentioned about the data fabric, essentially, we needed an architecture that can support and sustain the growth that we're anticipating in BP. And because of the new businesses that BP is getting into, modularity is going to be key in our digital approach. We need to be able to respond to new ventures and the Mesh operating model also promises exactly that. And Mesh or no Mesh, we need to be able to abstract the underlying complexities, so that our users can focus on what they do best.

And also, we need to be able to – a very key part of this is streamlining how we deploy secure data platform services. So that's the key. So I think the Mesh is more of a way to achieve all of that, but we didn't set out saying, oh, we need a Mesh. It just so happened that the underlying concept and how it is actually aligned to the way we want to be composable in BP and be able to bring in new business, new data and that sort of thing and that federated approach is actually very much aligned to how we want to operate going forward.

TERESA TUNG: I love that answer. You just gave a very Data Mesh answer to why Data Mesh. So you started with data as a product. You started with the vision that really made data product stand up and having the self-service infrastructure to differentiate the product and drive the business outcome for BP. So I love it. It wasn't a technology led story at all. It was very much around making BP that data driven business.

I guess, Liam, maybe I'll also from a technical perspective. So you've talked about this vision, where are you on your journey? Can you give some examples of your data products? And specifically, how did you find your first? Maybe you could just talk through your story there?

LIAM DONOHOE: Yeah, so I think building our best point there as far as our strategies are integrated energy company and that modularity theme underpinning digital and a lot of the innovation was core to realizing that. And at the heart of that work is an element around sustainability and linking and driving our ambition as a net zero company. And we've been very fortunate as we've been tackling and starting to deploy and realize our Data Mesh, that there's been a couple of really great examples where we've been able to tackle that and see the benefits of that and directly targeting that understanding how we're getting on with that ambition.

So, in particular, the first one is really something that we refer to as our BPMs tracker that's looking at how do we bring together and measure emissions towards net zero and actually capture, not surprisingly, data that allows us to measure that from a whole range of different businesses and energy sectors within BP and really create trust for the emissions related data products that we've got for a rigorous data lineage about – and are trustworthy and that we can then reuse and discover and share in a reusable way. So Data Mesh clearly underpinning that is one of our key objectives in tracking the progress that we're making along that ambition.

The second aspect to that, as well as tracking how we're doing emissions-wise, is how do we optimize and reduce emissions in the first place? We've seen some great success with digital twins, a new data capability as part of our digital platform where it's more been about how do we commoditize that capability, embed it within our Data Mesh platform and make it available for other parts of our business to apply to the other areas, so that we can optimize methane and other emissions in the first place and really start driving the agenda.

So what we've seen is the timing has been right, aligns to our ambition and this change in our approach with some very important and focused elements of a Data Mesh right at the heart at being able to measure and influence that agenda.

TERESA TUNG: I love it. So it starts with that data product to support the sustainability strategy. You have unique trusted data and then, you're really thinking about making that product better. Who's the consumer? How does it connect as part of a digital twin? All that makes that product easier to use for that end user.

What about from a technology perspective? Were there new tools that – there's a lot of tools when people talk about Data Mesh and especially then there's wanting to be a part of that. Did you see any tools that were really needed to be added to make this happen?

LIAM DONOHOE: Yeah, so I think the one area that as an enabler from a technology perspective more than any that's been challenging is really around data access management and protection. Really, in order to be able to provide a consistent user experience in accessing and consuming data products are widely distributed has really remained a pivot from some of our traditional approaches of very role-based access models embedded very localized and within compute platforms to really centralizing how we do policy-based access and how we deploy those entitlements. And there's probably three key aspects to that.

One is actually the definition of those data policies in a central way and capturing the data attributes. So that's required both strong data governance and successful application of data cataloging technology. We have the user attributes and identity management and, again, requiring new and innovative ways that we up our game around identity management and how we describe users that can feed into that. And then, finally, there's a technology dimensions on how do we manage those policy definition in a very centralized way, but then are able to federate those out and enforce them at the various compute and storage capabilities within our platforms.

So it's easy. I think we found and underestimated is probably to be fair, the complexity of the interdependencies of each of those aspects. It's hard enough to really try to bring them all together in a greenfield in an estate, but recognizing that actually we're trying to apply the new world of Data Mesh to some of our legacy or existing data platforms and really trying to get those into interoperate. It's been challenging and definitely an area where I think there's been a level of pragmatism and recognition that we just have to be patient and just take our time and consider carefully how we evolve from that old world to a new world and probably just temper back just a piece at which we can deliver and realize some of our ambition and aspirations around that security space is probably the most challenging aspect.

ABETH GO: Yeah, and I can echo that. I mean the hardship is real because it's not just the complexity and the sophistication in terms of how we need to execute it. It's also the fact that the solution is not there. There's no template that you can just follow. What we have are concepts and ideas around and like accurately based access control. This is going to help us with how we want to secure secret data, how we do this and that, but it's just really, really hard and to couple with that, I think the challenge that we saw and we probably knew this all along, so apart from the technology not always being there and we're having to think about the solution, it's also that the actual business process, we're not used to, oh my God, we can do this. So now how do we do this?

So in terms of the defining the actual business process and the policies around how we actually govern data and just understanding the – you know, when we talk about data governance and making sure that, oh, you are the data owner now, you're going to be responsible for making sure that we approve or you know who are the people who are accessing your data. So all of a sudden, it's a mindset shift. So while the technology is hard, it's almost like,

oh God, we're back to basics. The process has to be there and we need someone to take ownership of the data and the actual data management process because we can't define everything and we need the data owners to come forward basically.

TERESA TUNG: And so, you've given great answers, both about of how your role has changed, both from a technology perspective, Liam, talking about a lot of the security patterns. Many that we're lacking that needed to be redefined for the federated model. And then, Abeth, you touched on something really important about the talent, the people, the process and really enabling the team. So really that your role as a central organization it sounds like that's part of the shift that has occurred.

ABETH GO: Yes, definitely. And I would also add that the shift, the change is happening and everyone is feeling it. So from our central teams, like we have this huge responsibility to get this federated model to work. But at the same time, having to scale up ourselves to work in that mode and also, the different business domains because they are needing to adopt and actually absorb some of the data engineering capabilities, at least from the node operations perspective.

So all of a sudden, it's the, oh, we need to upskill and then we need to also staff correctly. And so, just understanding that mix of between the data domains, the nodes of infrastructure, all of a sudden you have this matrix of an organization and just trying to simplify that, so that we don't trip over each other before we can execute this properly.

TERESA TUNG: And, Abeth, you used a great quote. You said there's a key to managing the Data Mesh, so it does not turn into a data mess. And I think some of it's what we just talked through right now, but I know you've done a fundamental shift and just into even better data management. Can you give some examples of some of the things that you're doing?

ABETH GO: So some of the things that we are trying to as an example, is really trying to or the operating word is actually trying, trying to get a proper data office set up because in terms of just defining some of the standards and also, again, we're not standards practice centralized organization. It's some level of consistency for operation and it's really important for us that the business domain owners come forward.

I would say if we were watching a movie or a TV series, one would think that when we talk about the Data Mesh and data platforms, that the primary actors and the stars are the data scientists and the data analysts. Well, actually, in this situation, you could see that the data managers, the data owners and the data stewards are going to have a spinoff because this is where they come forward. And so, that's part of that big challenge that we have is actually - well, first, identifying who those people are and working closely with them. Because when we started this journey, it was not that obvious.

I mean don't get me wrong, right, BP has a lot of data managers and data stewards. We have data coming out of our ears and from a traditional – we have geologists and seismic people, so watching and managing data and having good data libraries is part of our DNA, but the scale at which we're trying to get to and how we want to automate all of this is actually pretty hard. So, yeah, that's some of the challenges that we're starting to face right now is almost of making it more official and actually bringing

some formality into this setup because without them, we might fall back to just doing the tech. So it's good data management that we need to make sure is present, so that we can be successful.

TERESA TUNG: And I love this point because I think some companies are scared of Data Mesh because they think that it's the Wild West and you're allowing all the domain owners to do whatever they want. And, in fact, describe that to make a better data product and a better data outcome for the business, actually one way, again, that you're differentiating is making sure that you're thinking about the users and bringing together groups to have integrated data products and some of that does bring in some of the traditional, almost a bad word like master data management. Somehow, they think when you do a Data Mesh, suddenly that's not important or data labeling and everything we do is not important. That's not the case. All of that makes a better data product. So really glad you brought that up.

I wanted to ask, it's clear that you're well on your journey and as one of the organizations on this journey, I want to ask for some practical advice. So let's start with some of the thorns. So what are some pitfalls that people should be aware of?

LIAM DONOHOE: The first thing, you're quite right, that it can seem very new and intimidating at first and I think the key pieces to realize that in many ways, Data Mesh is just really bringing together a number of best practice and thinking that's been around individually for some time across different areas. It's just how it's brought together and applied it to the data problem and found that balance and focus that really seems to work.

So I think it's easier again to assume this is about technology and get focused on that again. And for us, we prefer to view the Data Mesh really as a mesh of data products. It's not a mesh of some new technology. And we probably started off where the initial reaction just saw this as a new technology and different architecture and why are we doing this differently?

But ultimately, whilst that technology plays a key role in the ultimate realization of that future agenda and where we see mesh will get to, the harsh reality is the technology's just not there quite yet in a number of areas and I think as long as you realize that and recognize that it's going to be a journey and then, it makes it easier to accept that you can just start somewhere. And that in many ways, is the beauty of the Data Mesh that because of the way it can naturally evolve and grow over time, you can start with one data domain, one data team and a very simple individual or limited number of data products and just applies Mesh thinking to that. So pick an area and that you know that they're starting with, whether it's an operational data plane and domain or it's an analytical plane and start with it. Don't let the technology and a belief that we need to have this all singing and dancing, find a link, stay on the underlying technology platform, pick the data product, step into it, learn and iterate and I think it can just grow from there.

TERESA TUNG: I think that was quite a positive story, even though I asked about a pitfall, you're saying that don't be scared. I think there's a lot of ways to ease into it with the projects that you already have and by applying the data mesh lens, it gives you a way to evolve into when the time's right for you and the business need is right for you to modernize and adopt some of the newer technologies and the newer tools.

LIAM DONOHOE: Yeah, absolutely and I think that our learning data realization that partly came for a reflection of what is it that stops some of our realization or the benefits of past architectures and approaches is actually really trying to step back as a senior leadership team and organization and understand what does this mean to us and really try to get that shared understanding before we step into it. So we spend quite a lot of time particularly as we've gone through our new company and establishing a reinvented BP the first half of this year really taking that time to step back and explore and understand what it means to us before we really got stuck into it too much. And I think that top down understanding and buy into the approach and the potential benefits and really has given us the runway to be able to tackle it and see some of the successes we've had so far.

ABETH GO: And I just want to add, Teresa, that what Liam said around that is almost like evangelizing in helping people understand this ambition. We cannot underestimate that because we actually spent a lot of time, almost distracting sometimes, in terms of focusing on the build and solving problems, convincing people, explaining to everyone why we have to take this step. And I think it's obviously we have milestones to deliver, but at the same time, we have people not being on the journey, it's really difficult.

So, yeah, conversations like this help us because it helps bring the community together and also, hopefully, other people speak up and we listen more to their perspective as well in terms of how they did. Because when we started this journey too, we said it's not – I don't think we knew of anyone implementing this when we started. And so, not having someone else to ask about, okay, what did you do, what worked well for you?

And so, a lot of the learnings that we have apart from just helping educate and inform our stakeholders and our customers is also helping – so it's also making sure that when we talk about what we want to implement, that we're speaking of the capability rather than discussing the brand or the technology set because I think people have biases with regards to which tool they want to use, which vendor they're affiliated.

TERESA TUNG: So maybe that leads me to asking about the benefits you're seeing today, as somebody who's started and I think one of the great things about the Data Mesh is the concept and the vision. But can you share some of the benefits that you're seeing today and was any of this unexpected?

ABETH GO: I don't think we realize its benefit yet, but maybe I'm jumping ahead too much. Once I really understood what the Mesh is all about, I realized that when this whole thing was conceptualized, I think – and correct me if I'm wrong – I think the initial emphasis is really providing data for the use for analytics. But what we realize is that we break this case, you know, we solve this, this actually opens up an entire new world because all of a sudden, this is how we could be provisioning data within BP and outside of BP.

So imagine the way that we are able to now curate data and do data sharing across different domains and such, we all of a sudden, this could be the only place where this curated data set exists in our company. So these are not data created through transaction, this is a data product that came from different parts of the organization and all of a sudden, this could be valuable and not always for analytics.

And so, it's almost like there's something else that would come out of this. We do this right, the price is actually much, much bigger than just bringing data together for analytics.

TERESA TUNG: Yeah, I definitely think the data product story is much bigger than analytics. And I'm really glad that you're seeing initial promise there.

I think, Liam, same question to you. Do you see any sort of benefits that you really want to highlight?

LIAM DONOHOE: I think, I guess, looking at how we interoperate and deal with that legacy platform estate and so, we've seen a lot of what we've been talking about is the new and there's some fairly easy and obvious cases where we've actually been a point of end of service life and moving some of those old systems and whether it's analytical and other data platforms onto a new infrastructure, it's made a relatively easy for us to apply data product and thinking to that and Data Mesh thinking and understand, okay, which part of that new platform do we overlay those onto?

I think what's surprised us is when we start to consider the gray area, so where we get newer platforms that apply some of the same principles and thinking about how do we put together that starts to look at that Mesh like, but aren't quite aligned as far interoperability from a data governance and access management perspective, for instance. Do you actually tackle those by trying to do a full retrofit and trying to get alignment with new policy management, workflows and ways in working or do you put them into forced migration and say, no, everything is now in new Mesh, you must move across to that?

And we realized that there's probably a pragmatic in between where we just have just enough alignment to some of our new policy management and governance to allow them to interoperate, but then properly recognize per mesh thinking, what are the data product domains and boundaries and the value that they're adding as far as delivering data product and how can they then co-exist with some of our new platforms?

So we're being able to complement all data products on these old platforms that we've been able to interoperate with our new estate and really get into complement and add to it from anything we're delivering new. So it becomes a bit less of a technology question and do we have this legacy and new estate and more a case of actually how do we get the data products that are being realized by these two platforms to build and complement each other?

And that's been a surprising outcome where we thought we're going to have to migrate because it's new technology and actually we've been able to federate how we crack that interoperability and really let the old co-exist with the new.

TERESA TUNG: Well, thank you, Liam. Thank you, Abeth. And thank you for sharing the outcomes and the lessons learned at BP, whose really been a pioneer in this Data Mesh space. And I love that in pretty much every answer you really stress on data as products and how that change in mindset seems to be the big change, whether you're starting from a legacy system, looking at a new enterprise sustainability goal, but regardless, that's really the driving function between how you make the case for adopting this way of working to how you're rolling out your first data products and upgrading your legacy or not. I think all of that is a key lesson learned and I really want to thank you for sharing that journey with our audience.

ABETH GO: Thank you very much for having us.

LIAM DONOHOE: Thank you.