[music]

**Paul Thies:** As our world continues to accelerate all things digital, we are seeing a greater emphasis from organizations on the productization of data. While data scientists undoubtedly have a role to play, the commercialization of data will require thoughtful consideration on the various skill sets and strategies necessary for success.

Hello, I'm your host, Paul Tiffs. On this episode of *IF Win*, our topic was Data Products with an emphasis on an emerging new role. The Data Product manager. My guest for this episode is Dr. Tom Davenport. Distinguished Professor of Information Technology and Management at Babson College. Visiting Professor at Oxford Business School. Fellow of the MIT Digital Economy Initiative, and senior advisor to Deloitte's AI practice.

Tom's many accolades include authoring 23 books and over 300 articles for Harvard Business Review and many other publications. Being named one of the world's top 25 consultants. One of the top three business technology analysts. One of the 100 most influential people in the IT industry, and one of the top 50 business school professors in the world.

Tom, thank you so much for joining me today. I'm looking forward to talking with you about data products and data products strategy as a component of some of the emerging technology that we're seeing in data analytics and artificial intelligence and whatnot. Now, I've been looking at some of your recent papers and research you've been doing work for AWS and publishing stuff, of course, as you always do on the MIT and Harvard Business Review.

You've been talking about, I think recently, within the last month or two pieces I've been around data product strategy. Can you start us off, can you describe a little bit, what a data product strategy is?

**Tom Davenport:** Sure. I guess I should first say what a data product is and there's a little bit of maybe not controversy, but disagreement about that topic. I always thought of data products as the combination of data and the tools that you need to accomplish something specific with it, which usually turns out to be analytics and AI. The ways we make sense of data, but there are some people and organizations that argue a data product is just the data itself.

They might refer to an analytics product or an AI product, but to me, data by itself isn't terribly useful. I like the idea of a data product being both the data that you're using and the tools that you need to make sense of it or get some value out of it. Then a data product strategy is things like which data products are you going to develop? What is the business purpose of them? Who's going to do it? What are the different capabilities that you have to have as an organization?

Including data product managers, which I think will talk about at some point. There are a number of other groups that are necessary to make data products ultimately successful so that's a big issue. I think those are the basic things. There aren't that many companies yet who have data product strategies because a lot of people are

saying, "We're going to decide on what analytics or AI use cases we're going to develop and we're just going to call them data products."

I do think that just as companies who produce things for their customers have product strategies, it makes sense to have data product strategies.

**Paul:** I hope I'm not off base here, but it sounds like it's kind of productizing context and being able to look across and creating a more holistic solution. You have the data, you have all the raw data and then you're making sense of it, but then applying that sense that's being made and then productizing that and selling that to the market. Is that fair?

**Tom:** Yes. Except that it may not be sold externally. It might just be for an internal purpose. Those are often called data products too. Maybe I want to do a better job of assessing my risk under certain circumstances. You could have a risk-oriented data product or you could have a marketing data product to produce recommendations. Or some companies I've worked with have created these next-best action data products.

I think when data products were initially discussed, it was in digital native companies, mostly in Silicon Valley. For them, it was typically, something that you take to customers and sell or at least give away with the intention of getting eyeballs and selling something else. Within legacy companies, it could be just an internal application or something for customers.

**Paul:** Now I think, in concert with this, we're seeing the rise of the chief data officer. It's a fairly recent phenomenon among the C-suite, I think within the last 20 years, I think is what I had seen. I know largely, or I suspect largely really was emergent in the financial sector because there's a lot of any money laundering and data privacy and a lot of things where data just naturally lent itself to the risk of the institution.

They were looking at it, the board was demanding greater governance over data. Anyway, that said, you recently offered a report for AWS on chief data officers and prioritizing business value creation. Now, how do you see data products folding into the efforts of the CDOs to create value for the organizations?

**Tom:** I think a key element of creating value, in general, it's hard being a chief data officer if you don't have responsibility for analytics and AI and even business intelligence. Is a tough job because data's an abstraction to many people. They don't fully understand what's right or wrong about it, how to make it better, et cetera. Many chief data officers have become chief data and analytics officers.

They've taken responsibility for both the supply of data, but also the demand in terms of analytics and AI and business intelligence to make sense of it and use it in the business. The problem has been that we started out, I wrote an article about a decade ago with a person who ended up being the first chief data scientist of the United States DJ Patil in Harvard Business Review and nobody remembers the title of that article, but they remember the subtitle. It was called Data Scientist: The Sexiest Job of the 21st Century.

**Paul:** Yes.

**Tom:** I know maybe we created a little bit of a monster, but the problem is companies thought, "I'll put data scientists in charge of the whole process of producing valuable things with analytics and AI." It turns out many of those data scientists aren't really interested in anything but creating a great model. They love manipulating the numbers and the models in order to get high area under the curve, percentages, and so on.

Many of those models just never got deployed in organizations. There were various surveys about the percentage of models that got deployed, 13% in one study. Some of the companies I've talked to said we never deploy any of them. We create them, but they never get deployed. Data products obviously don't work if you're not deploying something. It's like you came up with an idea for a new product, a new form of toothpaste, but you never introduced it to any customers, so you never got any value and that was the problem in the data science space.

Pretty critical I think for chief data officers, if they're going to get value to take an approach that says, "We're going to fully deploy this model into production and we're going to get economic value and we're going to even monitor it over time in case the model drifts or doesn't predict as well anymore." All of those things. Not every chief data officer has embraced this idea or chief data and analytics officer, but I think in our survey, 39% say they were working with data products and were moving toward using data product managers.

**Paul:** That's interesting. I can't help but think that there it's somewhat analogous in some ways with, a lot of times you see innovation efforts within corporations. There's a lot of energy to stand up innovation efforts, but then they don't necessarily get productized or they don't actually, they don't have demonstrable value.

It's a culture thing, but it's like, you've got to have somebody who can help get it through the pipeline and get it out the door. I know we'll touch on that here in a minute when we talk about data product managers, but--

**Tom:** Yes, exactly. Same thing. We think of products as being developed by R&D people or something like that, but you have to have a very cross-functional approach. If you're going to bring a product to market, it's got to involve marketing and sales and supply chain and so on. What a lot of product focus companies have done have created this overlay of the product manager.

Those are increasingly popular in software companies too, who job it is to look at the entire process and make sure that it makes it to market and that it's what the customers want, and that it's successful.

**Paul:** Absolutely. Now, to that point, how should an organization productize its data? What are maybe just a couple of examples of that that come to mind?

**Tom:** Typically you're taking some collection of data that you think is useful for a particular purpose and combining it as I say, with some data, I mean with some analytics or AI or maybe business intelligence. It might be something if you're talking about a company that is in the business of data products, I once asked the chief economist, Hal Varian of Google, how many data products do you have?

He said "I want you to go to the website for Google and the Wikipedia page for Google and it lists all of our products and I want you to count them all. That will be the number of data products that we have. [laughs] Because everything we do is a data product basically." When I wrote about this a number of years ago and I talked about things like LinkedIn, people you may know Facebook also has the people you may know. It's some combination of a social graph with the social graph data that you have within the organization and then something that does matching.

You can tell people, you probably know this person since this other person and that person knows this person well. Chances are pretty good we'll make a triangle out it and you might want to connect with them because that's going to make the site more valuable. It turns out people you may know was a very successful offering for LinkedIn. I think we're all bored with it now, but for a while, it was really quite exciting.

Or it might be something as simple as patient dashboard in a healthcare institution where you can quickly see what the state of their body is along various dimensions, it doesn't have to be anything exotic, it's just a set of data and some way of making it meaningful.

**Paul:** Now I think to that end, you have put forward or you have been writing about recently, I should say the concept of reusable data sets. How does that fit into this discussion around data products? How-- what are reusable data sets and how do they-- what part do they play?

**Tom:** In general, I think we've not done terribly well on the data management side. We've focused on earlier stages of the data supply chain. How do we collect this data, how do we store it, and so on, and not enough on consumption? If you care about consumption, you need to package up data sets, collections of data, maybe involving a set of different variables or data elements, and say, "All of these are useful for accomplishing this particular purpose."

Was talking with someone from AT&T recently and they were saying, "If you want to study churn, we've got a set of different churn data sets for different markets. Then we've got the churn model that you can plug into with an API." If you're interested in preventing a particular type of customer from leaving us is we're going to make it very easy for you to do that. The reusable data, I think goes quite well with the reusable features or if you're talking about machine learning or variables, if you're talking about traditional analytics, same thing actually.

Everything you need in order to do this analysis is available and we're going to make it possible for almost anybody to do this analysis. The reasonable data sets might involve customers, might involve products, it might involve particular regions. There are all combinations and I think, maybe we're going to talk about this later on, but I think it could also involve external data and probably it should involve external data.

**Paul:** Yes. Because I think that given that data sets train the AIs and there's so much emphasis now on training sets and growing that, it just seems to me that there's an opportunity there to cross-pollinate with other organizations. You can get around the data privacy concerns and that sort of thing and then maybe there's got to be some interesting ways that organizations can collaborate and help their AIs muscle up.

**Tom:** Yes. Just for example, in banking, everybody's interested in not only how much money of this particular customer I manage, but what's my share of wallet. To do that you really have to have some collaboration with other banks to say, "Let's share data to some degree anyway, maybe through a third party. It often is the case through a third party so that you can see you have 20% of their assets but these four other institutions have percentages as well and might be an opportunity for you to gain some share or there's some data that's just much more likely to be generated external, like weather data for example."

Weather data is quite useful in predicting what things you might want to buy, which is one company where they did those lenses that turned dark in the sunlight, transitions lenses, and obviously weather plays a big role in that. If you have a really rainy summer, people aren't going to buy as many sunglasses. Weather data's usually supplied by, it was supplied by the National Weather Service, but now IBM bought weather company, a number of companies are getting into supplying whether data AccuWeather does it too. External data I think is generally a little overlooked by many companies and I think it's a good idea to apply it when you can.

**Paul:** Yes, no, that's interesting. It's not to get off track, but it's like that, it makes me think of supply chain implications, for going to your transition lens if I'm a transition-- if that is a part of my product portfolio and I know that it's going to be an exceedingly rainy season and I don't need to buy X amount of components, I need to buy Y and I need to adjust my supply chain and then there's downstream implications to all of that.

**Tom:** Yes, a lot of companies now are interested in not only what things might be going on with their suppliers, but they're trying to look further up the chain and say, "What about my suppliers' suppliers? Are there floods where they're producing as their political unrest?" Maybe something is happening. My supplier hasn't had a problem yet, but they will. You can really anticipate a lot more of what's happening. Obviously, we should have done better about that during COVID-

**Paul:** Oh yes.

**Tom:** -than we did.

**Paul:** There you go. You live and learn. Now you mentioned earlier, when we're talking about data product strategy and you mentioned, the digital natives, Silicon Valley, they're really the leading on this, but that maybe, I don't know, more traditional is the right nomenclature you hear about. More traditional companies, large companies are trying to make this shift towards data products.

Can you talk about maybe some of the challenges that you've found among large companies that are trying to implement a data product strategy?

**Tom:** The ones that have really embraced it seem to be doing pretty well with it, but probably the biggest complaint is, it's hard to find the right people because data scientists, for the reasons I was discussing earlier, generally are not the best at it because, they tend to be focused on building great models and algorithms, but some data scientists might be okay at it if they willing to take this big picture and not have to use their highly analytical skills every minute of the day, but you have to know

something about analytics and AI, you have to understand something about data structures and different types of data.

You have to be a good communicator because you're really coordinating across this broad spectrum of cross-functional activities. You have to be knowledgeable about what customers in your industry want because producing products that appeal to customers is a big part of your success. It's hard to find people with all of those attributes who are available. There are no-- I think I'm working with one university that's starting to create data product managers, but in general, there aren't any even--

**Paul:** Programs around that? The academic programs that help. Because you've articulated this is a new role, that it's not the data scientist but it's somebody who's maybe like you just said, they have that domain expertise around AI and analytics, but then also they're able to wear the hat of a product manager and they understand marketing. They understand the big picture of what it takes to put something together and all the constituent parts and bring it to market.

**Tom:** Exactly. There was one related role that people had been talking about for a while, the analytics translator role. I'd written some stuff about this but never really, I call them purple people, half red half blue, half technical half business-oriented. A McKinsey article in *Harvard Business Review*, I think used a new term, the analytics translator, and a number of companies have experimented with that, but my guess is that data product managers also are capable of translation and they do a lot more [chuckles] and their value is a lot easier to establish than just translation.

I think it'll ultimately be a more popular role than the translator role was.

**Paul:** Let me ask you, going back a little bit on the idea of data privacy, I know that was obviously a big concern, particularly in 2018 with the release of the GDPR and its lingering effects, but do you see data privacy concerns creating any significant barriers for the deployment of a data product strategy? If so, how are those overcome so that value is not lost?

**Tom:** I think that could be yet another thing that you make data product managers responsible for, that there are privacy safeguards in place and that you have to have somebody to point to when something goes wrong. Data product manager would certainly suffice for that. If you're in healthcare, you can have an overall person who is responsible for HIPAA compliance, for example, but you really need somebody who's close to the individual applications or use cases, and that's going to be the data product managers.

I think it can help more than hurt. Obviously, data privacy can be a difficult thing to address, and I'm not saying it makes the job easier for data product managers, but it does make dealing with data privacy somewhat easier if you have a structure in place for data product management.

**Paul:** Circling back to what we were discussing about the concept of collaborative organizations, can Productized data be combined with other organizations' data to expand its value? Have you seen any particularly interesting examples of that?

**Tom:** A lot of companies now are trying to look at, it may violate some people's idea of data privacy, but they're trying to track their customers across a variety of different platforms. They're internal platforms, but also social media and loyalty programs, and so on. You have a class of companies emerging that do that. We always had third-party data brokers, but now they are working across channels in many cases.

If you want to get a sense of, is this person who appeared on your website somebody who has bought a lot from other companies in the industry in the past or whatever, you're only going to get that through a third-party provider. That'd be one example. I think the number of potential external data suppliers is really endless, location data, health data, and I think driving data.

A lot of the automobile companies now have connected vehicle data, there's going to be even more of that in the past as cars become more electric and autonomous and so on. They're going to be fantastic opportunities as long as you don't annoy your customers too much to take advantage of all that external data that's available.

**Paul:** That's interesting. I had spoken years before with, he was the chief digital officer for a European car manufacturer. I think one of the things they were trying to do is get to this personalized digital memory. Basically, you create this profile as a driver, and then whenever you rented a car-- you're traveling, and wherever you rented a car, you had a key fob or something.

It would, regardless of where you were in the world, if you rented one of their electric and autonomous vehicles, it would remember your Netflix preferences and logins, all your social media, and everything. It extended beyond just memory of, "This is how you like your seat," and the ergonomics and all that kind of thing to where it was your whole digital life followed you in your whatever vehicle.

It wasn't just your vehicle, but any vehicle that you rented from them across the world, which it was interesting, but it was like, "Wow. That's, again, data privacy." That might be a head-scratcher for some people.

**Tom:** I wrote an article recently, it hasn't been published yet, I don't think it's in one of these management journals with long lead time, but it was on AI and personalization. It's how I reviewed some of the literature on what consumers think about personalization. It turns out they're very schizophrenic about it. They want it if it can save them a lot of money, or save them a lot of time, or bring them free goods or discounted goods or whatever.

On the other hand, they don't want to give up their data in order to make it possible. It's a little hard to know where the line is. I do think that having your customer's permission is increasingly important and providing some value exchange. "I'm going to use your data, but I'm going to give you something valuable in exchange for it." I use Gmail as my email client. I know they look at my data on occasion.

It's never really bothered me all that much, although now I notice there are some things that I get that I can't declare spam, they just keep coming day after day, so maybe at some point, I'll tire of it, but for a number of years, it's been a good value exchange for me. I still don't get very well-personalized offers from [chuckles] them because of It.

**Paul:** I think there's a desire to see personal data get monetized so that people can make-- It's like, "If you're going to use my data, then what's in it for me?" You're saying it.

**Tom:** Monetized, being the owner of that data is the individual from whence it came, and they get some money. Unfortunately, a lot of companies when they're talking about monetization, it's, "How do we make money off of our customers' data?" That, I think, is not really ultimately going to succeed.

In fact, I even tell companies, "Don't use that term monetization if your customers hear it, unless it's providing money for them," which is probably another case, they're going to be annoyed.

**Paul:** I do think there are data advocates out there. I've talked to some commentators that think that that might be coming where it's not-- The monetization goes back to the person. It's like, I recoup, I get monetized for you, ACME incorporated using my data.

**Tom:** Tim Berners-Lee, the inventor of the Web, is working on some products that would allow that. 'Sandy' Pentland at MIT, a friend of mine has argued for that. There's a new venture by Frank McCourt, who used to own the LA Dodgers called Project Liberty, where they're trying to create that. It's a hard road to hoe. It's a lot of change necessary in

how the internet and the web work to make these things happen and I hope it does. I think we need some changes, but I don't see it happening quickly.

**Paul:** There's going to probably be countervailing forces that I try to get around that, but we shall see. Tom, I do have a couple of questions that have been submitted from our audience. Pretty fascinating, just a little bit off-topic from data products, but these are questions around AI and data sets and data science. If you're game, I'd like to just ask you a couple and get your take on this.

The first one comes from Dr. Gin Bloom of Germany and she writes, are there any current IT trends that you view as fads such as trends that you think are getting a lot of hype but won't necessarily enact meaningful change?

**Tom:** I guess I'd put the Metaverse in that category and I was skeptical about it all along, not just because Meta and Facebook have fallen upon hard times recently. I just didn't really think that the technology was yet up to the task of making a highly engaging visual experience for consumers. I guess I was also a little bit skeptical about Web3 and cryptocurrency.

I do think that there's it will and the blockchain will play some role in the future, but it's one of the things that's amazing to me is this is supposed to be such an appealing technology because your assets are secured on the blockchain, yet everybody gets hacked. [laughs] I just don't quite get how that works. Clearly, a bit of a contradiction there and I think the reason over the last few days vanishing of this very large and once successful crypto exchange FTX is going to put another short-term nail in the coffin of Web3. I don't think we're going to talk about it nearly as much as we have over the last couple of years.

**Paul:** It seems like there's been a lot of talk about blockchain and cryptocurrencies for the last several years, but it just hasn't replaced currency as we know it. Tales of the death of the dollar are greatly exaggerated yet, so we'll see.

**Tom:** If it has replaced your currency, then you have a lot less currency than you had

[laughter]

a few months ago.

**Paul:** Oh, wow. Our next question comes from Noah Pakinis of London, and he writes, what is the most eye-opening realization about AI today?

**Tom:** This turns out to be a very good question, Noah because today, I and a co-author from Deloitte published an article in Harvard Business Review about what I think is the most interesting thing happening in AI today. It's called How generative AI is changing creative work. It's about these, I don't know, some people call them large language models, although now they're far much more than language.

Some people call them foundation models. I think generative AI is, to me the best term but they can be trained on a wide variety of content types and they can generate in a very impressive way, another wide variety of content types. You can train it on emails, you can train it on websites, you can train it on images, et cetera and it will produce with only a short prompt, quite impressive paragraph or produce an image or produce a video, or produce a blog post. For us content creators, it's a little scary to see how this might evolve over time.

**Paul:** My next question comes from Hamid Adib of Connecticut. Hamid, writes what is the optimum solution to keep bias out of AI and machine learning?

**Tom:** That's a very good question. It is a very challenging issue because bias results from data. The data we use to train these models, data only comes from the past. The past typically involve human decision-making about who to hire or who to give a loan to or whatever and humans are very biased so it's a big issue. Sometimes we think that that bias is okay.

For example, if I were to apply today for a long-term care insurance policy, mine would be less expensive than my wife's, and what could argue that's a case of gender bias because we men tend to live shorter lives than women do. Is that bias or is it just the way the world works? It's a complex issue. I think really there are companies that have produced software that will identify potential types of bias, which groups may suffer as a result of applying or deploying this particular model.

It's sometimes hard to know beforehand before you have any outputs or predictions, just because the models that are increasingly used to make these predictions aren't very transparent. They may have hundreds or thousands of different variables or features and not be very interpretable even by a data scientist. I think we're making slow progress and you can at least look at, okay, which features tend to be the most influential in these models, but it's still in part, at least a human activity to say, "Who suffers from the use of this particular model and can we live with that? Can our

customers live with that?" Is this something that is an acceptable form of bias that we've been using for decades like the insurance one I mentioned?

**Paul:** This last question comes to us from Ian Sharp of Sydney, Australia, and I think it ties to the question Hamid just asked is, what safeguards should organizations put in place as part of machine intelligence design to help ensure appropriate and legally defensible decision outputs by the AI?

**Tom:** I'm very interested in that and I have a new book coming out in January called *All Lane on AI*, where I talk about companies that are really aggressive in their use of AI. Most of them, I would say, they say we care a lot about ethical AI, AI fairness, et cetera, responsible AI. They've created policies, but the rubber hasn't really hit the road yet in terms of day-to-day governance.

I do talk about Unilever in this book, and they now, for every use case that somebody wants to develop, if you want any corporate resources at all, or really if you want to even introduce this into the marketplace, you have to send them the description of what the use case is going to be about and they don't evaluate it internally. They have an external partner called holistic AI which looks at it both with human eyes and also with various automated assessment tools to say, "Is this likely to be biased? Is it transparent? Is it likely to be effective?"

Some of the proposals that come in are not likely to work very well. Some people think that AI can do more than it really can so they might come back and say, "It would be fine. Ethically, it's just not going to work very well." I think just as we do that for financial performance and public companies anyway, we'll probably end up having external auditors. My friends in Deloitte and the other big four companies say, "We can't really audit it yet because there are no government standards."

We're moving in that direction and Europe has an AI Act. It's not implemented fully yet, but I think there will be progress in that regard, and maybe we'll have official criteria by which companies will be able to audit the performance of their AI use cases.

**Paul:** That's interesting. We don't have time to unpack it today, but I think that opens up all kinds of freedom of speech and all kinds of different and interesting discussions, like whose bias, are we talking about who gets to decide, who are the arbitrators of what is true and ethical?

**Tom:** Oh no, we all agree on what's truth in this society, Paul

[laughter]

just kidding. Those generative AI systems are also going to be very controversial in terms of one, they'll be able to generate deep fakes, I think quite easily. Two, who owns the intellectual property? There's already a lawsuit against one of the tools that Microsoft has introduced that generates code because it scrapes existing code in order to come up with automated code generation.

This is the GitHub business unit. They have a tool called Copilot, and they've already been sued even though it was just introduced to the marketplace a month or two ago.

**Paul:** That's interesting. I don't know why, but that reminds me of record sampling and how that--

**Tom:** Yes, exactly.

**Paul:** The late '80s, early '90s with groups like the Beastie Boys and suddenly it was like, "Whose music is it?" but anyway.

**Tom:** Yes, same idea.

**Paul:** Tom, thank you so much. This has been a fascinating discussion and congratulations on the new books and also obviously all the material that you're putting out this fall. Data product strategy, fascinating data, product manager. Kids get your resumes together and if any universities are out there listening, it sounds like we have a new potential new job market to train up for. I really appreciate you taking the time to talk with us today.

**Tom:** My pleasure. Thanks for having me on.

**Paul:** Absolutely.

[music]

**[00:41:42] [END OF AUDIO]**