MR: Mike Rabinovici here and this is Dimodelo's "Conversations with Data Warehouse Experts" podcast series. Our mission in these podcasts is to speak to the best minds in the data warehouse and BI space, and to get their take on the state of the business and find out what they think the future holds. We are also committed to do it in a way that adds value to the tech profession, while at the same time is clear and compelling for the business managers and executives they collaborate with every day.

In today's podcast, our guest is preeminent business intelligence analyst, speaker and author, Dr. Barry Devlin, founder and principle of 9sight Consulting. Many consider him to be the godfather of the data warehouse, defining its first architecture in 1985. He's joining us from Cape Town, South Africa. Welcome Barry.

BD: Oh, it's great to be here—a pleasure to join you.

MR: We're glad you joined us. I'd like to start by asking you to tell us a little bit about your background, for listeners who have not had the chance to hear you speak or read any of your books.

BD: Absolutely. Let me say first off that you can probably hear that I'm not South African, or at least the South Africans among the audience would know that instantly. I'm actually Irish—most of my life was based in Dublin, Ireland. Back in the early '80s I started working for IBM as a young ingénue, and one of the first things they asked me to do was a piece of architecture for decision support systems within IBM at the time. And you may remember, back at that time that was a mainframe company, essentially, but also had a few mid-range computers that eventually became the AS400.

So they were trying to figure out how to manage their sales and do decision support, and out of that came data warehousing back in 1985, before Inmon wrote his first book; the first architecture paper I published was in the IBM Systems Journal back in 1998, which described the data warehouse architecture. Since then I worked for IBM for many more years, as a manager, a consultant, and eventually ended up being something called a distinguished engineer, which is a fairly grandiose title in IBM, but I've never met an engineer who's distinguished, so I've always wondered about it.

Anyway, by 2008 I'd tired of being distinguished and an engineer, so I left and set up 9sight Consulting. I've always been involved with information and data and its use within businesses and beyond businesses, and so that's where I focus still as a consultant and an analyst. Mostly these days as an analyst, which means working with software vendors, but also wanting to keep my hand in with real companies doing real work.

MR: Where have you seen the most significant changes in the data warehouse space over, let's say, the last five years?

BD: Yeah, I think the thing that has really struck me most has been the emergence of externally sourced data as a huge input to data warehousing. Now most people would call this big data, although "big data" is probably not a good name for it, although most of it does have characteristics of being big. But I think

the more interesting thing about this data is that it's coming in very different structures, with very different quality characteristics, with very different rates of feed, sizes, etc., and it is a very different type of data than we fed into data warehouses in the '80s and '90s and even into the early 2000s.

The availability of this data has enabled a move from business intelligence to analytics. Some might hesitate as to what that actually means, but for me, business intelligence was knowing what questions to ask, where analytics is about drawing statistical inferences from the different parts of the data you've got. And I think what that's done is, it's really changed people's way of thinking about the use of data within the organization, the use of information in the business, and how you can begin to predict the future and indeed begin to proactively influence future behaviours using it.

That shift is a big thing that Gartner would talk about in terms of digital business and so on, but I think there's a real downside to it. We've moved away from the old days of asking questions and doing analysis in order to figure out market issues, or why our sales dropped in the Midwest, and we've started doing things that are very impactful in terms of privacy, in terms of public perception, in terms of the economy. We've really moved from being a small internal type of internal organization/function within businesses to really being on the front row, and I don't think we're thinking enough about the ethics of what we're doing.

MR: Looking at the current slew of data warehouse products in the marketplace, do you feel based on your historical experience that they're addressing the increasing complexity of data and the exponential increase in data flows and sources?

BD: I think it depends on how you define current data warehouse products. Many products talked about a few years ago as big data products are now talked about as doing some form of data warehousing. You see that in the emergence of relational function and analytic function with the Hadoop sphere and the NoSQL sphere—all of those things are to me addressing the increasing complexity of the data and the exponential increase in the data flows, but they're probably not the products you would have thought of as being data warehouse products in the past.

But that doesn't mean that many of the longstanding players in the data warehouse field aren't also upping their game. Take, for example, the recent announcement by Teradata about their Teradata Analytics Platform and IntelliSphere. Teradata was the original data warehouse vendor—now they're really stepping up to analytics, to different types of data, to different platforms, the ability to bring data in from other platforms. It's interesting to see companies like Teradata—Informatica is doing the same—really adopting a lot of the changes, the new data types, the new approaches to data. Informatica has introduced this thing called CLAIRE, which is an AI program embedded within their products to help in terms of modelling, ETL—I'm sure you know about that from the Dimodelo side as well. These, I think, are examples of both the new vendors and the old vendors really stepping up to this data complexity.

MR: Speaking about the essence of data warehousing, I've read in recent years some articles here and there by—fairly smart folks, I guess—heralding the death of the data warehouse. What's your take on that?

BD: Well, let me tell you Mike, anyone threatening to murder my baby will have to deal with me first. I'm not the illegitimate grandfather of data warehousing for nothing. I don't think that the data warehouse is dead, but I think it's morphing in many different ways. I think many people forget that the data warehouse is actually an architecture, and being an architecture, how it gets implemented is actually less important that following the principles of it.

So what are the principles of the data warehouse architecture? Well to me, it's always about consistent and timely data that can be used in decision making support, and I don't think those needs are ever going to go away—they certainly haven't yet. More to the point, the technical environment and the business environment today is so much more complex than it ever was. So whatever we call it, this set of architectural concepts that emerge with the data warehouse will continue to exist and will have an even bigger role. You know, some people also talk about the data lake as the successor to the data warehouse. I've heard that a lot—I'm sure you have too—but people who dig data lakes still have to address issues of timeliness and consistency, so renaming the concept from a warehouse to a lake may sound cute, but it really doesn't help when addressing the issues.

MR: I'm glad you brought that up, because one of the things we're doing in our podcasts is not simply talking to tech professionals, which of course, we are, but helping the C-suite understand things a little bit better when they speak to their technical teams, as to what they're trying to convey and the value proposition. I've seen the "data lake" being brought up by executives at meetings, but it's not always clear that they understand the difference between the data lake and what the data warehouse does. If you could give the nub of that, that would be great.

BD: I'm not sure you'll like the answer that much; to me the data lake is just a conglomeration of technology products from an open source stack standpoint that have been thrown together in order to deal with the topic of big data. I think it's a very different concept from the data warehouse, which started as an architecture, as a set of rules, a set of things that you need to address—principles you need around how do I make data consistent, how do I manage its quality, how do I assure its availability, how do I balance timeliness and consistency. All of those issues are architectural design issues that you actually have to think about when you choose your products.

I think the problem for me with the data lake . . . although I must say this just to honour the people who are doing it: it has evolved over the years from being just a gloop of product stuff to some sort of architecture, so the Cloudera's and the Hortonworks and those kinds of guys have thought about architecture, but it was only after a few years of selling product that they started doing it. So, to put it in a nutshell for a business person, if you want to have decision making that's based on the best available data, you need to do some architecting of your data and your data feeds in advance. That means doing a data warehouse no matter what you call it. It could be done on a set of technology as it was in the past,

such as relational databases, or as the Hadoop or NoSQL people would say today, do it on the open source stacks. That doesn't really matter at the end of the day, but from the principle of getting quality and consistent data, it's really important that decide how you're going to do that and then decide what the physical implementation is going to be.

MR: That's a great summary, Barry, because what I was driving at was that I would go to these events and I would hear from VPs or different people that are considering these initiatives, that the data lake essentially replaces the data warehouse, so I'm glad you clarified that. Now I'd like to turn and talk a little bit about the cloud. More and more companies are adopting it in one form or another—what do you think the impact will be on their data warehousing projects?

BD: To me the impact is going to be a real increase in complexity. Now I know from a business point of view there's a lot of other things going on—you say, we get more flexibility, we don't have to invest capital, we can do opex instead, etc., etc., but I think to me when you say impact, the impact will be the increase in complexity.

Essentially what you're doing here is you're moving from an environment where the data that you needed and that you used for your data warehouse was all internally sourced, internally managed, internally used. You're now moving to an environment where the data you want to use is both externally and internally sourced. You're talking about storing a lot of it in the cloud, which is obviously external to your enterprise boundaries, but you're not putting all of it out there, so some of it is going to be internal. Then you've got this issue of data moving back and forth across the boundary between on-premises and the cloud, which is really an expensive and troublesome procedure. You've got all sorts of issues now emerging around geography, such as where is your data actually going to sit. If it's in the EU, you're going to have to follow these rules, and if it's in the US, you're going to have to follow, well, probably no rules at all.

But all these issues will start coming up, and I believe that the complexity of data management will have a major impact on the implementation. And then of course, the physical architecture, the technology and so forth. I've focused a little bit on the negatives, and you said before we started that you're a guy who sees the glass three-quarters full . . .

MR: I did.

BD: I think I'm a guy who sees it empty. [laughter]. So I think these are the sort of impacts that you have to deal with. There's a lot of technical selling that goes on out there, saying the cloud's going to solve everything, if only we had the cloud, it's going to really make the difference. There's certainly good things about it, but I think we have to keep in mind that what we're really doing is building a very complex distributed data environment, and that's going to be more of a challenge than building a single-environment data warehouse.

MR: I want to turn now to your book, *Business unIntelligence: Insight and Innovation beyond Analytics and Big Data*, which came out in 2013. Those are always huge projects, so what compelled you to write it?

BD: I think it was interest in the principles of data warehousing. When I started thinking about this book, around 2010, the data warehouse architecture at that stage was about 25 years old, and that is a hugely long time in the technology industry. When I started doing data warehousing, I had a lot of hair—now I have none. The changes in business needs and the technology possibilities have been huge since then. What drove me was I really wanted to address the issues raised by these data types and sources, like, as I said, externally sourced data, data volumes—I wanted to really see what that would mean in terms of issues in data warehousing. I started out, even a few years earlier, writing down what I thought were the postulates of data warehousing and then coming up with new ones for the modern era, which are much more relevant to what people are trying to do with data today in the business sense.

Back in 2010, 2012, I think that people were mostly doing one of two things: they were either trying to push all this new stuff—this new information, these new needs—into the old architectural model. Inmon's DW2.0, for example, is one way of doing that. Or they were just thinking about technology—"What's the problem? Oh, Hadoop has the answer." I wanted to really step back from that and come up with some modern principles that would get back to the first principles of architecture.

I sort of came to this phrase, and it's reflected in the book and in the architecture, that there are three thinking spaces that you need to deal with: **information**, as the foundation; **process**, as the middle piece that you use to do stuff to information; and then there's the top piece, which is **people.** So information, process and people are the three principles we have to think about in terms of this architecture. And it turns out that when you then reverse that around, you get a nice sentence that says, "People process information." And that's essentially what we're trying to do—create an environment to allow people to process information to run their businesses.

MR: I particularly like the concept of "business unIntelligence" in the book's title. How would you explain that concept to businessmen or people who are conversant in tech but are far from being experts?

BD: That name "business unIntelligence" actually came about quite late in the writing of the book, and to be honest, I'm not entirely sure how successful it was. I have a feeling people look at it and think, this is a glass three-quarters empty. The reason it came up for me was that as the book evolved and I started to deal with the people aspect I just mentioned, I got very interested in the question of how people make decisions. I began to discover that the belief system that management consultants and schools of management have is that businesses make decisions based on rational choice theory. We probably don't have time to talk a lot about rational choice theory, but it is a way to balance out all the aspects and probabilities of things happening. The more I thought about it, the more I said, do I really make decisions that way? Is that really the way others make decisions? And it became obvious to me that no, emotions are involved, physical energy is involved, bias is involved, societal interactions come into

play—all these things play a part in how people make decisions, so I just wanted to show that there was more to decision making than rational choice. Of course, that meant that it's not just about intelligence, it's about something beyond intelligence, and that's where the idea of unIntelligence came from. It actually also came from the idea of the "undead," which is a diversion we should probably go into another day.

MR: [laughter] Ok, we can do a round two on that for sure.

BD: Anyway, there's a lot of thinking around how do people get from the information that's in their warehouses and BI systems to the contextual information—the knowledge, if you want to call it that—and then on to meaning. But that's something that's very difficult to incorporate into BI technology, and we've never really thought too much about it. And this is really something I'm passionate about at the moment. It's really important now—more important than it ever was—because we're now at this stage where there is a story going around that we're about to use AI to . . . eliminate bias, let's say, from decision making, but we've no idea what biases we're trying to eliminate. We're going down a path, as we usually do with technology, of "it's going to solve all the problems of the world," without considering what the problems really are. We're going to augment, we're going to automate, we're going to solve all the decision making problems with AI, but we haven't even figured out how it is we would represent business decision making in a sensible way.

MR: I want to turn and deal with a topic that's closer to our hearts, because we deal with it every day—how do you think data warehouse automation is changing the delivery of information to the business?

BD: I've been really impressed with data warehouse automation as a way of thinking, as a set of concepts, since the early days when I came across it in probably the early 2000s.I think what really struck me was that, in contrast to previous thinking, it was what I might call a whole-process view. It considered the whole process starting from the idea that, we've got a whole set of structures and prior art, if you like, in our operational systems, and we're not just interested in how to get it out and how to cleanse it, but we're actually interested in how do we get it all the way from the operational world into the warehouse or whatever it happens to be that you're trying to populate.

And you have to take all of the steps into account. You can't just focus on the technology—"how do my engineers translate data from this form to that form?"—but you have to actually figure it out right from the business point of view all the way down to the technology. Data warehouse automation is also, to me, great in the way it insists, at least in most of the tools, that business involvement is a vital part of what you're trying to do. So this, first of all, holistic view of the process, and secondly, this idea that it is a collaboration between business and IT to make it work—these are the two things that have most changed the delivery of information.

MR: When looking at data warehouse projects, which are notoriously over budget and take too long, if a business C-suite asked you, "what is the one thing we have to know or must do before launching into a data warehouse project?", what would that be?

BD: I would say make a plan—which is a very South African phrase, to "make a plan." I think it's really important that you start from the beginning and make a plan. When I start consulting on a data warehouse project, there are very few greenfield projects these days; they are generally now what we call brownfield projects, in other words, this is the seventh time we've tried to build this thing. And the thing I always overlook is that the clients are always in a hurry to get to the product choice, to populate the database—basically to just do stuff. I really have to stop and think and say, no what you really have to do is to think about how you're going to get all these things working together. You have to make sure that all of the business people, who are coming from many different departments, are all aligned behind that same direction. They have to realize that somebody's going to get first delivery and somebody's going to get last delivery. How do you make sure they're all going to agree on this—and continue to agree on this?

As you said, these projects and processes take a long time. So making a plan, figuring out how to build a team that will manage that plan—they're often call a BI Centre of Excellence, or Centre of Competence. And who are the people that need to be on that team, and what are the personalities that you need on that team to make sure that it works? Then you can start thinking about whether you're going to use data warehouse automation, will you use ETL or go with Teradata or Hadoop. All of those things are much further down the path. If you haven't got a plan that tells you how to get from the beginning to the end of the process, I sort of think you're wasting your time.

MR: So it's an imperative of yours to plan your work and work your plan?

BD: [laughs] Absolutely.

MR: I want to switch gears a little bit to a different segment of the podcast where we get to know our guest a little bit better. And it has nothing to do with tech. My first question, if you're ready Barry, is: what is the one non-tech book that you would recommend most to people?

BD: I have to say to you, and this is maybe going to come across a little strange at this point, but about 10 or 15 years ago I came across—not one book, but a series of books—called *Conversations with God* by an author called Neale Donald Walsch. Given my background having a Ph.D. in physical chemistry, you would probably not think I would recommend a book called *Conversations with God*, but it really opened my eyes to looking at the world in a different way, to thinking about what we do in the world and how we are in the world. It's a completely off-the-wall topic, I would say. So have a look at Volume 1 of *Conversations with God* and see what you think.

MR: I've heard of it, but I haven't read it. That's a great recommendation. Who was the author again, Barry?

BD: Neale Donald Walsch.

MR: Neale Donald Walsch. Excellent. Another question: what is the one TV show that you're watching, or binging on, as the case may be, right now?

BD: I'm not going to tell you what I'm binging on right now because it's a bit too gauche. The thing is, I'm really a fan of European and English TV, so certainly there are some great things there. I love Nordic Noir, which as a genre has some really great stuff in it. One series called *The Bridge* is a crime series set in the areas surrounding the bridge between Copenhagen and Malmo, which is in Sweden. It's a very good series, very dark. That one sticks in my mind.

Another one is based on books by a guy called Robert Galbraith, which turns out to be a nom de plume for J.K. Rowling. The series is called *Strike* and two of the books, *Silkworm* and *The Cuckoo's Calling*, have been made into TV series. And these are also very good series in terms of the development of character—you know, I like development of character, I like to see the underlying psychology of what's going on as opposed to people chasing around at 130 miles per hour in cars, hitting the wall, and walking away.

MR: It's interesting that you bring up Nordic Noir—there's another show that's on Netflix call *Occupied*—

BD: I shall write it down.

MR: Yeah, it's a one-season thing, but it's pretty interesting and it's the same genre. These are great recommendations, thank you. I want to conclude our podcast by letting our audience know where they can find out more about you and the work that you do.

BD: Well, I'm afraid we're back to now the boring stuff. First of all, go buy my book. Could I avoid saying that? Absolutely not. *Business UnIntelligence*—it's out there, it's available, it's well worth a read. If you have time to sit down and think about architecture, principles, what we're doing with information in the world, I think you'll find it very interesting. At another level, I say go to my website 9sight.com. Everything I publish goes there. I don't actually ask for your email address to get to it. It's all freely available. I'm happy if you write or email me about it to ask questions—and of course, follow me on Twitter because, hey, I'm more interesting than #therealdonaldtrump.

MR: What's your handle?

BD: It's barrydevlin. Very straightforward.

MR: Thank you so much for your time today, Barry. It was great and I think there's a lot more questions I would love to ask, so maybe we can do a round two down the road.

BD: Mike, it's been a pleasure for me too.

MR: Thanks for listening everybody, and when we upload the podcast to the website, we will have a list of the URLs and resources Barry mentioned during the show, including his favourite books and TV shows. Thanks everybody for listening—'til next time.