INTERVIEW WITH PATRICK CONNOLLY

Eric Green: Hello and welcome to a new podcast series from IBM software that explores the challenges IT managers and business professionals are facing today. I'm Eric Green and I'll be talking with a range of experts to discover new perspectives, approaches and examples that can help meet these challenges and introduce you to the capabilities of smarter software from IBM. So let's get started.

Welcome back to our next episode focused on information integration and here to share his thoughts and knowledge in this area is Patrick Connolly, Program Director with IBM's Infosphere.

Patrick Connolly: Thanks Eric. It's great to be here. I'm looking forward to sharing some thoughts and concepts about integrating information.

Eric Green: Excellent. So on that then, maybe you could help our audience understand a broad definition of what information integration is.

Patrick Connolly: Sure. One of the ways to answer that question is to look at the evolution of information integration over the last few years. You know, it all started with products designed to load data warehouses. We called them ETL tools, and vendors like IBM realized, in fact IBM sort of led the industry in realizing that these tools had a sport utility flavor to them, that we could use them for a host of different use cases. And those include not only loading data warehouses, but also for loading master data management repositories, for doing large scale data migrations and application consolidations, and even delivering information as a service or real-time information. And when you combine that with data quality built into the same solution, it makes for a pretty powerful, adaptable tool that does a lot of different tasks.

Eric Green: Thank you for that. And so feeding off from that and what organizations are doing, what are the challenges that organizations are facing that demand this type of capability?

Patrick Connolly: That's a great question, Eric. Put simply, I mean, I think the number one challenge that the organizations are facing is, you know, they're being asked to do more with less, particularly within IT. You know, they've got less resources, less staff, and unfortunately we've also got an increase in the data volume, variety and velocity. And business certainly isn't going to slow down just to wait for information. So they need to be able to do more with less, but also make better decisions with the information

that they've got. And part and parcel with that is this movement and growth around information governance, and the fact that we need to start looking at information more as an asset like you would a piece of capital equipment or real estate. It has a life cycle, it has, you know, a certain value that's associated with it. And we treat it as such and make sure that we're able to leverage that for better decisions. There's a lot of challenges also in terms of the change of pace in the industry, grappling with new sources of information. I talked about the variety, volume and velocity – there's a huge growth and interest in areas such as cloud computing or accessing applications that run outside of the four walls of the enterprise. And as well, IBM's been leading the way around the big data initiative. So taking what the _____ find, or accessing and making sense out of very large, very distributed disparate internet based content and data repositories for better actionable information.

Eric Green: So that's all very interesting. So it seems to me that we're in a unique space these days, because we've been increasing the volume of information exponentially in general, but now we're also increasing the direction from which that information is coming from. Right? People are coming with new information from the web, they're coming from their mobile devices, they're hitting from social media. I mean is that sort of deluge of information coming from all of these different locations really adding to just, you know, the stress of having to, you know, with the challenge that we were just talking about?

Patrick Connolly: Definitely. And I think there's two ways to look at that problem space. One, you know, is new tools for new these jobs, or new tasks. The other is looking at established, adaptable platforms, like IBM provides with Information Server and looking at how we can leverage and include those capabilities to the existing platform. The benefit to clients is that it complements their existing infrastructure. So we don't want to, you know, create a totally separate island of information by any means. But yet we want to be able to combine and make sense of that information in the context of that information that it's very important to how we compete, to the critical elements of our business. And so what it does is it gives us access to a whole new world of information that doesn't simply fit into the rows and columns of a typical relational database. So we're seeing a great increase in the number of sources of information and the nature of those sources that can lead to trusted information and better insight to help organizations compete and win.

Eric Green: And I was just going to ask that question. And that's on trust itself. I mean by the very nature of information integration, you're taking disparate information and bringing it together and how do you determine and are challenged with dealing with, you know, governance and trust of the integrity of that information itself.

Patrick Connolly: And that's where I think this notion of governing information is very appropriately linked to organizations that understand their information assets. And the cornerstone of that information is really about metadata. I kind of get excited about metadata because it defines, you know, where the data starts, everything that happens to it en route to the final application or report, analytic or whatever I'm looking at. And those clients that I work with that recognize metadata as a key differentiator in terms of how they approach information are those that outperform their peers. Put simply, it's the Rosetta Stone or the secret decoder ring to their information assets that also allow direct linkage and shared understanding between business and IT. You know, if I'm looking at a report as a business user, wouldn't it be great to right click on a term, say, or something that I don't understand in a report, identify who the data steward is, in case I need to call them if I think something's wrong or up with the data. But also to see a data quality rule and the max, the mins, the limits, etc. to tell me whether or not I can actually trust that information that I'm looking at. Metadata makes all that possible.

Eric Green: So continuing on, I think when it comes to our listeners, it's really good to give some concrete examples. Can you talk a little bit about some scenarios or case studies, or how are enterprises doing this?

Patrick Connolly: Sure. Most of the organizations that I work with have some approach to metadata but, you know, typically it's leveraged solely in the context of IT. You know, to be able to identify or inventory their information or their data sources. We need to be able to extend that to the business community. So we utilize a powerful solution that's part of Information Server called Business Glossary. And that Business Glossary can also include pre-loaded terms and definitions from the IBM industry models. So you don't have to start from scratch. And I think one of the challenges in working with and leveraging the power of a metadata repository like this is that it takes some investment, it takes some diligence. So these clients that I work with, across industries, including financial services, including retail, insurance, healthcare, etc., are those

organizations that have taken a more governance-centric approach in linking IT and business so that, you know, IT typically doesn't make a move without business and likewise on the business side. That they're working together to facilitate to build that shared understanding that lives in that business glossary. So, you know, the clients that I work with that see the greatest benefits are those that have invested not only up front, but created policies and standards and processes for making sure that, you know, they're able to continue to leverage and enrich that metadata repository and those assets and their business glossary to achieve substantial, definable benefits.

Eric Green: Excellent. And how about if you could go into a little bit of detail? I mean we talked a little bit about what you guys are doing in this space. But how is IBM really innovating in this space?

Patrick Connolly: Well IBM has been on an amazing pattern of innovation around information integration, and it all started back in 2006, when IBM introduced Infosphere Information Server, a platform that brings together leading capabilities around understanding your data, applying quality principles. And including that native to the fabric of the transformation and delivery of information are our solutions that include Quality Stage and Data Stage which are really for clients that use both of them for one solution.

Going forward, what we've done is we've managed to raise the bar in terms of our capabilities for real time or right time delivery of information, including our change data capture abilities, as well in terms of working with a flexible, adaptable, services-oriented architecture approach, our Infosphere Information Services Director, which is part of Information Server, allows clients to quickly and easily build and deliver information services around their data quality and data integration processes and subprocesses, to deliver trusted information as a service to the organization, without having to have, you know, dozens of JavaScript programmers to build these things out. So it makes it a lot easier.

And I think that's a recurring theme that we've seen in our latest release, including version 8.5 which came out last October. And that includes capabilities for automating the discovery of data and relationships among those heterogeneous sources, the discovery of metadata, and also for automating the processes of creating ETL logic, and creating quality logic and identifying data quality rules. so we're making the solution easier to use, we're automating things so that we can get clients up and running faster, with less

risk, with lower cost of ownership over time. And of course because this is a platform that's built with reusability in its DNA, the fact is that we can start to define standards and policies that live as quality and integration subprocesses. But then we can distribute across the enterprise, and get that reuse benefit about it, centrally maintain. That is what information governance is all about.

Eric Green: Excellent. So we're running out of time here, but I think it would be really good if we could touch on, I guess what your top best practices for organizations are in this space.

Patrick Connolly: Sure. I think I'd want to focus on a couple of different areas, and these are very closely related to the things we talked about with information governance. The first thing that my clients can do to gain advantage out of more effectively integrating information is to help them get their arms around their data. So, you can't govern that which you don't understand. So often with clients we'll position the use of our foundation tools to help them understand their information assets. So do the profiling analysis and relationship discovery to figure out, you know, what they've got, how is it related? And also think about the life cycle implications of that. Do we need to keep it forever, or can we look at a sunset strategy for applications or for data sources? So that baseline understanding of your information assets can go a long way towards all of your projects and all of your solutions that are designed to deliver better answers and automate business processes.

The second is around quality. I mean I can't underscore enough that organizations need to address data quality across the organization. You know, it's not just something you do once a year or once in a while. You know, the analogy that I like to use is that carnival game called Whack a Mole, where once you take a swing at once of those little moles that pops his head up, another one pops up somewhere else. So data quality is a lifestyle choice that needs to be applied across the enterprise, and one of the cornerstones of any information governance initiatives. So those are two of the best practices that I typically would recommend to just about any organization, regardless of industry, regardless of size.

Eric Green: Well thank you for that. That was really helpful. And indeed, thank you so much for all your insight today.

Patrick Connolly: Thanks a lot Eric, I enjoyed it.

Eric Green: Thanks for listening. Please do visit IBM.com/software to connect with our experts, continue the conversation, and to learn more about smarter software from IBM. Let's build a smarter planet.