Springer Nature TDM Podcast Transcript

Mary Ellen Bates:

Welcome to the Springer Nature Podcast, focused on the text and data mining projects taking place at Springer Nature and the ways that info pros and KM professionals can contribute strategically to TDM projects in their organizations.

Today, we're talking with Robin Padilla of Springer Nature, a global research publishing and service provider. Rob is a director of product management with Springer Nature's data and analytics solutions team. Some of Rob's special focus areas include building tools and services to improve R&D in the life sciences and supporting the digitization of research labs.

I'm Mary Ellen Bates, principal of Bates Information Services. And I'm working with Jinfo Limited, a global industry analyst group on this podcast project.

So, Rob, I've been watching the TDM initiatives that Springer Nature has been developing for clients, but now you're also using TDM for internal projects. Could we start out by you're telling me a little bit about those internal TDM initiatives?

Robin Padilla:

There are a lot of different applications of TDM within Springer Nature. Many parts of the company are doing a lot of really interesting things with TDM. In the time that we have, we'll talk about a couple of cases that are kind of close to what my team is doing and what my department is doing. The first involves the use and the construction of discovery tools. And then the second case for TDM involves what I'm calling expanded delivery mechanisms.

In the first case, discovery tools, this is basically looking at how we can build specialized platforms that cater to specific fields basically.

For example, the platform that I'm looking after is Springer Nature Experiments. It's a platform for researchers to discover experimental procedures in the biological sciences. And in this case, you can imagine researchers would use it for things like trying to find protocols, lab protocols that are containing rabbit antibodies for western blotting, or they would want to look up different ways of visualizing proteomics data for example.

A more timely use-case would be how can you use PCR to detect Corona viruses, for example, things like that. And so that's one of the platforms that we're looking at.

Another platform is the Springer Materials platform, which is a platform that's dedicated to data in the material sciences. And there, we've been embarking on this kind of multi-year journey to really digitize a lot of high-quality material science data.

Two areas in particular are areas relating to semiconductor materials. So electronics materials, and also things relating to chemical engineering applications, for example, thermal physical properties of different types of organic compounds. Those two areas are quite different. But one thing that they do have in common is that we're using TDM basically to create all of this different - to not only digitize data, but also to be able to do things with data.

And so in both cases, what we're aiming at with these specialized discovery platforms is to have a place where you are able to not only find the data, but also be able to do things with that, to be able to process it, to be able to visualize it, to be able to connect it to other parts of the workflow that you are looking at.

There are a lot of different use cases. One example would be thinking about how you can extract that data. And then essentially, what you're doing with that extracted data that you are feeding into these platforms.

So that was talking quite a bit about kind of this first case for TDM, these discovery platforms. The second case is involving what I'm calling expanded delivery mechanisms. It focuses on how we can deliver different types of content for TDM purposes.

And on a really high level, you can understand this as having a shift from human-readable content to machine-readable content. And in the not-so-distant past, when we say things like electronic format, that basically meant that you have content available in an XML or a PDF format. However, as TDM becomes more widespread, it's very clear that although the PDF is very nice from a human-readable perspective, it's often just not up to task.

It's basically an electronic reproduction of a human-centric way to consume data, which is great for humans. But then, now, if you're talking about having data or having content that machines could use and do things with, then you have to think about things in different ways. That's where things like APIs start to become really, really powerful. And that is definitely a big area of development for us because, with APIs, you could do all sorts of things with it.

You can have APIs where you are essentially querying the full text of the articles, but we're also really keen to build APIs where you could do more specific types of queries.

So going back to the types of data that I was mentioning a few minutes ago with, let's say, comparing genomics data with material science data, semiconductor data, for example. Can you have different types of APIs that you could call that will allow you to query for those types of data? So it's a very, very rich area for us - the area of APIs.

And then the other area in this kind of expanded discovery mechanism space is data feeds basically. And in this case, essentially, we've seen a lot of success and we've seen quite a fair amount of development in this space with regards to having data feeds that can be customized specifically according to the parameters that our users need them.

Mary Ellen Bates:

Can you say a little about what the driving force was behind these TDM initiatives? I mean, was it something like we've got this amazing technology, so let's build a sandbox and see what we can do? Or was it more that we've been waiting to do this and finally, we can apply TDM and accomplish this?

Robin Padilla:

I would say it's definitely a mix of both. If I look at this from a proper product management perspective, the starting point for us would be what are those basic user needs? What are the users trying to accomplish? What other tasks, what other goals and also what are the pain points that they have associated with that? What are their blockers? What are the things that are holding up their progress?

Once you really understand that on a deep level, then you can start to think, okay, if these are the problems that have been defined, the problems that people are requiring solutions for, what sort of solutions can we build? And when we start to think about what kind of solutions can we build, that's where things like TDM really, really come into play.

And once it's decided that TDM can in fact be quite relevant, then we essentially throw what we can at it, the best people, the best tech, the best resources that we have available to us to really address those challenges.

Of course, there's also a very strong intellectual curiosity as well in terms of let's see what can we do with things like TDM or AI or ML. We're a lot of nerds amongst my colleagues, myself included and also very creative people as well. And so there's definitely this strong desire to see well, what can we do?

One example would be can you teach a machine to read a scientific publication or a patent, let's say, to extract the ... I don't know, the structural properties of nanomaterial? Can you do this? Okay, we've done this. Okay, great. Can you scale this up so that it can now read 10 papers, a thousand papers, a million papers? And so that curiosity is also a really important part for us as well in terms of how we want to really develop the kinds of solutions that are really going to benefit our users.

Mary Ellen Bates:

Given your familiarity with the needs of info pros and KM professionals, what info skills do you think are particularly useful for TDM projects?

Robin Padilla:

One of the things that I would highlight as really important is understanding the use cases. What exactly are those use cases? What exactly are the goals? What are you trying to do? What are those success metrics really? And to really define that very clearly.

And you can say that about anything, really kind of any project. So what kind of questions, or how do we know we've made it type of questions. But I think that keeping that in mind is especially relevant for TDM. Because a lot of the tools that we use in the course of doing TDM, AI, ML, all of this other stuff,

APIs, the tools themselves are pretty flashy actually. And so I think it's quite easy to get bogged down in sort of the novelty and kind of the plain coolness of a lot of these tools.

And then it's very easy to get distracted because, oh, well, we have all kinds of really interesting algorithms that we can apply, or we can construct some kind of really nice knowledge graphs where we can uncover all kinds of really interesting information where you can correlate the number of cookies eaten with the air temperature in the Sahara Desert or some strange thing. Could you do that? Sure, but is that really a part of the goal that you're looking at?

So really understanding that becomes a really important thing. So understanding those use cases, the metrics.

On the topic of use cases, if we look at TDM, two of the major use cases that we see for our users and also for us as well within Springer Nature would be enhancing content discovery and relationship generation.

Mary Ellen Bates:

Is there anything that you wish you'd known ahead of time before getting involved in TDM initiatives or any advice you'd give someone who's embarking on a TDM project?

Robin Padilla:

I mentioned the importance of being on track, having very clearly defined goals, kind of keeping those end results in mind. I wish I had that instilled in me much earlier in my product management career precisely for some of the reasons that I mentioned. Where now, there are so many tools for TDM. There's so many different types of openly accessible data sets, algorithms, applications. There's so much stuff out there. And if you read the media, there's so much hype around things like machine learning algorithms, AI applications and all kinds of different things.

And again, this is all really cool stuff and it's great that we have this, but then what are we actually doing with this? How are these tools going to help us achieve the goals that we are trying to achieve? So I would say that the key takeaway there is to just keep those goals in mind, to always keep in mind, hey, this is what we're trying to achieve. These are the tools that we can use to help us achieve it. But also keeping in mind that as fancy and as powerful as those tools are, at the end of the day, they are just tools. They don't exist for their own sake. They exist to be used and to be applied, to help us achieve whatever objectives it is that we are looking for.

Mary Ellen Bates:

What a great answer. Thank you so much, Rob.

That was Rob Padilla of Springer Nature, speaking about the text and data mining projects within Springer Nature and the ways that info pros and KM professionals can contribute strategically to TDM projects.

Listeners, thanks for joining us for Springer Nature's podcast on key topics on information strategy and management. Learn more about this project by visiting Springer Nature's resources for information professionals.