Successful Info Pros in a Data-Driven Enterprise

Interview with
Mary Ellen Bates

For podcast release
Monday, May 15, 2023

KENNEALLY: In 2012 in the *Harvard Business Review*, Thomas Davenport, an authority on data analytics, and mathematician DJ Patil, who served as first US chief data scientist, declared that data scientist would prove to be the sexiest job of the 21st century. The demand for data scientists is indeed strong and is even accelerating, with the US Bureau of Labor Statistics expecting employment of data scientists to grow 36% from 2021 to 2031, much faster than the average for all occupations.

Welcome to CCC's podcast series. I'm Christopher Kenneally for Velocity of Content.

Data scientists are found working in fields where data-driven decision-making dominates, from financial services and information technology to healthcare and biotech. They often work closely with librarians and others trained in information science. The two roles are complementary, and organizations can benefit from aligning the positions strategically.

Mary Ellen Bates advises clients in research-intensive industries on their information needs. This week, she will report on findings from her recent study of best practices for info pros when working with data pros at the Medical Library Association/Special Libraries Association joint conference in Detroit. Mary Ellen Bates joins me now. Welcome to CCC's Velocity of Content, Mary Ellen.

BATES: Thank you, Chris. It's great to be here.

KENNEALLY: Looking forward to giving everyone a preview of your program at the conference later this week. Let's start by sorting out the research-related responsibilities of data scientists on the one hand and librarians and information specialists on the other.

BATES: From the interviews that I conducted, the epiphany that I got was that data scientists and information specialists are looking at data from completely different points of view. Info pros and librarians – they're looking at the data before it gets used. They're busy validating it. They're negotiating use licenses so that they have the rights to use text and data mining, for example. They're making sure that the data scientists can do what they

want to do with the data that they've identified to figure out whether the data needs cleaning up and made more consistent, what other datasets might need to be brought in to add additional insight or perspective or depth to the dataset that they'd identified. That's all the stuff that the info pros do with the data before it goes to the data scientists.

Then the data scientists – they're assuming that they have problem-free data, that when they acquire this dataset, it is beautiful and all set up and it works and that they can use their skills and tools to figure out what they want to do, figure out what questions they want to ask, what stories the data can tell, what insights they can glean that they couldn't have just gotten from doing manual research.

So I think that librarians – this is kind of a strange analogy, but I think librarians are like food inspectors. They make sure that the cut of beef that you buy at the supermarket is safe to eat. The chef or any cook who's going to the store just assumes that the food is safe. They're counting on and they take for granted the safety that comes from having food inspectors. And I think the same thing – data scientists are trusting that there's been data inspectors who have ensured that this information or that the data that they're acquiring with the assistance of info pros is clean and healthy and not contaminated with viruses.

KENNEALLY: Well, I liked that analogy up to the last part, Mary Ellen Bates. (laughter) But it really works very well, I think. Data inspection is so important. Data quality is such an issue with all manner of data today. And you suggest that info pros should learn to speak the language of those counterparts, those colleagues of theirs, the data scientists. So how does someone go about acquiring the vocabulary and the syntax of data science and data analytics?

BATES: I think that one of the challenges is that information professionals and librarians may not realize that they're not speaking the same language as data scientists. So I think what's important is for the librarians to get much more comfortable working on project teams with data scientists. Get in the room with them. Be the one person – the fly on the wall, almost – listening to them talk about their project, about what they're going to do with the data, so that we have a context of the workflow that they're bringing this data into.

I think just being on the teams, asking lots of questions – and I'm just going to pause here and say one of the challenges for info pros is we like having answers. That's probably why we went into this field is we like giving people answers. So it's hard for us sometimes to come to data scientists with that beginner's mind and say, I don't know what you mean when you say such-and-such. I'm assuming you mean this, but I'm thinking maybe you don't. So that ability to be with people who – just like info pros, they also look at data and value it and see its potential. But they're looking at it from different points of

view. They're looking at it from what kinds of questions can I ask it, and what kind of insights can I glean from it?

KENNEALLY: What are some examples of how and when info pros should participate in a data-driven research project?

BATES: Often what happens is the data scientists – they have a question or they have a business problem that needs to be solved. Often in pharmaceuticals, they have a condition that needs to be addressed, and they need to find an answer – what's a biologic that can address this? They don't know what data to use. So info pros may find the data, but then they'll see that, oh, this data has limitations to the usage license. Or it'll require a whole lot of cleanup before it can be used. Sometimes, the info pro team will suggest to the data scientists maybe they want to shift their resources, or maybe they need to use a different resource instead of the one that they thought they would be using.

What happens then is that the info pros ensure that the organization isn't spending a whole lot of money on data that can't be used or that can't be used the way they thought it could, that can't be reused, and it causes more work downstream that isn't worth the cost to the organization. Often, it's making a strategic decision about not acquiring one kind of data, because while it appears good, the info pros see the bigger picture and see that the ramifications of acquiring this data does not actually serve the organization as well as the data scientists may have thought of it.

That's one just general example that happens – whenever I was talking to an info pro when they're involved in a data project, they'll consistently say I keep my organization from wasting money, either from acquiring a source that's not appropriate, acquiring a source that requires too much cleanup to be worthwhile, or one that can't be reused, and that's the way that they justify buying it is that they can get multiple uses out of it.

I think the other thing that info pros really bring to the table is that they break down data silos. Any project team within a larger organization – they're focused on their mission, their questions, what kinds of insights they're trying to glean, what problem they're trying to solve, and they sometimes kind of get tunnel vision in a good way to make sure that they're pulling in the resources and getting what they need and delivering the answer. The info pros, when they're brought in to evaluate a data source that is being considered, they've seen other project teams and all those other people who are looking at data content. So they can break down those silos and say, did you know this other team over here is working on something similar? Maybe you want to talk to them as well.

KENNEALLY: What can interfere to make collaboration between data scientists and info pros difficult?

BATES: It's interesting. I asked that question of a number of the info pros that I talked to, and I was expecting to have big jurisdictional conflicts or departmental priorities or anything like that. And it turned out it was much simpler than that, and it really boiled down to – especially the info pros – they would look at a data scientist and say we're both info nerds. We both love data. We love talking about data. We like thinking about data. We like thinking about what it can do. So we just assume that we speak the same language, and in fact, we don't. Even though we're both really excited about and get all nerded out about this collection of data, we're looking at it from completely different points of view. I think info pros assume that they understand the workflow of data scientists. So they make assumptions about the utility or the issues involved in a dataset based on those assumptions, and they may not realize that those are unquestioned assumptions that they need to challenge themselves.

And on the other side, there tends to be an assumption on data scientists' part that info pros are not equally expert in their area. So the data scientists may not realize the expertise that info pros are bringing into the acquisition and licensing of content to make sure that they're getting material from publishers that will enable them to do what they want. Info pros can contextualize data in a way that data scientists may not realize the information needs to be contextualized. It's these unquestioned assumptions on both sides about what the other side values, needs, and is able to provide.

KENNEALLY: It sounds like, professionally speaking at least, it's all about empathy and respect for each other.

BATES: When you say it like that, it sounds so simple. I think both sides kind of overthink it in terms of trying to figure out where the disconnect is and how to work more closely together. And it is. It's really much more of a people issues thing and for both sides to just be willing to say, I don't know how you operate. You work with the same tools that we do. We're working in the same organization. But gosh, we really see the world from a different point of view – to just have that humility to say this isn't the only way to look at this dataset or this data environment.

KENNEALLY: Mary Ellen Bates with Bates Information Services, best wishes this week at the MLA/SLA conference, and thank you for speaking with me today on Velocity of Content.

BATES: Thank you, Chris. It was a pleasure.

KENNEALLY: Mary Ellen Bates presents, "Successful Info Pros in a Data-Driven Enterprise," Thursday, May 18th at the Huntington Place Convention Center. On Friday, May 19th at 9:00, Christine McCarty represents CCC in a panel discussion, Standards Update: What's

New With Standards? Both programs are part of the Medical Library Association/Special Libraries Association joint conference in Detroit May 16th to the 19th.

That's all for now. Our producer is Jeremy Brieske of Burst Marketing. You can subscribe to this program wherever you go for podcasts, and please do follow us on Twitter and on Facebook. You can also find Velocity of Content on YouTube as part of the CCC channel. I'm Christopher Kenneally. Thanks for listening.

END OF FILE