



**Interview with Dr. Leslie McIntosh
Co-founder and CEO, Ripeta**

**For podcast release
Monday, August 5, 2019**

KENNEALLY: In 2016, *Nature* magazine asked more than 1,500 scientists a simple, yet critical question – have you ever failed to reproduce an experiment? Reproducibility of research results lies at the heart of the scientific method, which relies on testing a hypothesis based on keen observations of results from an experiment. In the *Nature* poll, more than 70% of researchers reported they had tried and failed to reproduce another scientist’s experiments. More than half even admitted they had failed to reproduce their own experiments.

Welcome to Copyright Clearance Center’s podcast series. I’m Christopher Kenneally for Beyond the Book. The so-called reproducibility crisis raises questions that can undermine public confidence in science, from the efficacy of drugs and vaccines to the reality of climate change. In St. Louis, Missouri, researchers at a startup founded by experts in biomedical informatics have developed what they call a credit report for scientific publications. Just like a financial credit report, which shows the fiscal health of a person, this automated analysis shows the hygiene of the scientific paper. The company, called Ripeta, aims to make science easier by highlighting verifiability and reproducibility. Dr. Leslie McIntosh is a co-founder and CEO at Ripeta. She joins me on the line from her St. Louis office. Welcome to Beyond the Book, Dr. McIntosh.

McINTOSH: Hi, Chris. Thanks for having me.

KENNEALLY: Well, we’re delighted to speak with you. We want to congratulate you first. In June, the Association of Learned and Professional Society Publishers nominated Ripeta for its shortlist to the ALPSP Award for Innovation in Publishing 2019. So congratulations and best of luck in September, when the award is announced. My eye was caught by the announcement, because reproducibility is such an essential area of science. Tell us why you chose to focus on it at your startup.

McINTOSH: So it came from my work and a little bit of frustration when I worked at Washington University School of Medicine. There were two things that were happening. Obviously, as you stated, I was in the biomedical field, but I ran a service center there. What I would see is, one, that I worked with really good



researchers who worked really hard, and when I would look at their publications, I would notice that they had described their research quite well, but really hadn't articulated the data that they used. As that was my job, to provide them – not everybody, but a lot of the research with data from, say, the electronic medical record or a clinical study, it was important for me to get credit for that. So it was kind of a frustration and also an acknowledgment about how research was really hard.

The other thing I looked at was just our own processes to see, you know, what were we doing that weren't reproducible, and how could we make that better? So that started me down the rabbit hole of reproducible research to see, how could we improve science? And one of the things that became clear is that if we were going to do this, we had to both make science better, but make science easier as well, because all of the burden couldn't be put onto researchers in order to keep up with all of the guidelines and everything that needs to be reported in research. So that's how I got started.

KENNEALLY: It's a really important topic, and if I follow correctly – I'm not a scientist, you can help me and our listeners to be sure we understand – you're separating the scientific method, the reporting and the robustness of all this data, from the actual quality of the science.

McINTOSH: Correct. One of the things that I think we have to acknowledge is that just because things are not recorded robustly within a scientific manuscript doesn't mean that the science was done or that the conclusions were wrong. It's just that science became very complicated, particularly with the advent of the modern-day computing environment, and it's hard to capture everything that we do. So yes, I'm trying to separate out sort of the scientific hygiene, if you will, and the reporting practices from the science.

KENNEALLY: So can we say that technology is at fault or in part?

McINTOSH: I would say it's in part. You know, technology has made science and scientific calculations, algorithms, so much easier, and yet describing everything has become so much more complicated because of that.

KENNEALLY: Right. So what do you do there at Ripeta? How do you make possible this scientific hygiene report?

McINTOSH: So what we're looking for – we still use the scientific manuscript or the paper that scientists publish as the basis for what we're looking at. And we've gone through hundreds of guidelines that exist and created many variables that are



absolutely needed to be reported within a scientific manuscript, and then what we're doing and where our sweet spot tends to be is in looking at before a manuscript gets published to use machine learning, actually, to read through it and see if we can detect certain things within the manuscript.

For instance, a hypothesis is a really good thing to have in a research study, and these days, we actually need it to be human-readable, and we need it to be machine-readable. So what we're doing is we're trying to detect it from a machine-readable fashion if that exists. Or if they've cited their data source or their analytics, things like that.

KENNEALLY: This is also something that must be welcomed by scientists themselves, because they're into the research. They have to write the articles. They want to document their work. They want to report back to funders or try to get tenure at a university. But really, they're about being in the lab.

McINTOSH: Right. That is the whole idea, so that if they write the paper, the paper can be run through our system, our software, and then it can give them a report. We couldn't find your hypothesis. Now, maybe it was there, and we couldn't find it in a machine-readable format, which is about 26% of the time, we can't find them in a machine-readable format. So we just pull it out, but it also gives them a chance to highlight what is there and kind of really look at, do they want it stated this way? It gives another viewpoint.

KENNEALLY: It's also important to point out that you're talking about these two different readers. There's the machine that's doing the reading, and then the human being, the actual reader of the content. But I would imagine today, many more machines are getting a chance to read all of these thousands of articles that are published than any human being ever could.

McINTOSH: That is very true. I do want to point out with this automation that we really are trying to make it to put checks in place so that people – whether it be the people consuming the manuscripts, whether it's the peer reviewers – can get to the science more quickly. Yes, there can be analytics done with the information after the fact. We're actually not doing this. Right now, it's trying to help scientists get back to the science and focus on that.

KENNEALLY: So in a way, you're getting back to basics. What are some of the long-term goals you have at Ripeta?

McINTOSH: Yeah, well, I guess the idealistic goal is that I want to make the reporting of science better. So that's more of, obviously, an ideal. I would like to be able to



see where science is going, to be real honest. But I want to help people like the peer reviewers be able to actually focus on the science so that they can look at the information and not have to try to keep up with a checklist on whether all the guidelines have been met or try to keep up with guidelines.

So right now, that's keeping us quite busy. There are some other long-term goals that I think if we talked in a year, I could give you. But I think I'm going to stick with that now.

KENNEALLY: All right. You mentioned that in a set of papers you've looked at, something like 26% – that'd be one-quarter of them – are lacking a hypothesis. Did I hear that right?

McINTOSH: Yeah, that's what you heard.

KENNEALLY: That's a bit of a shock.

McINTOSH: Well, again, we're focused on what can we pull out with a machine? And what we've found – and I've pored over these manually as well, because that's what you do to train a machine algorithm – is that we as scientists have gotten into a place to where we expect humans to infer what we are doing. That's a little bit different than a machine being able to detect what the purpose of your study was. Does that make sense?

KENNEALLY: I think I do get it. And again, because machines – they're really literal. Let's put it that way.

McINTOSH: Yes, that's exactly it. So I'm not saying that in all of these papers, that there is not a hypothesis or some sort of study objective, but quite a few are not machine-readable. I mean, come on. This is science. Every single paper should have its goal stated. And we've gone back and looked at those that the computer did not find, and unfortunately, it is the case that even in some very highly cited papers, the goal is not clearly stated. Let's just leave it at that. (laughter)

KENNEALLY: All right. And you mentioned the benefits of your work to the researchers – they're pretty obvious – and to the peer reviewers, as well. But there are benefits for others in that publishing ecosystem – publishers, the actual publishing society or the journal publisher, as well as funders.

McINTOSH: Yes. And one of the things that we've started doing is offering what we call a portfolio analysis. I call it a portfolio, because if you're a publisher, maybe you want to look at how your journal is doing based on a set of criteria, such as



sharing data or sharing the code or what's important to your journal. To funders, it might be the grantees – that they are interested in looking at how their grantees are publishing papers and how well they're adhering to their guidelines, as well, and supporting maybe open access – depending on what their goals are – open access or data sharing, things like that.

KENNEALLY: And the public stands to benefit, too. This notion that confidence in science is eroding because of the reproducibility crisis can't be good for science or for public policy.

McINTOSH: I would completely agree with that. And I think there's two aspects to this. One is we really need to be able to understand how to trust the science that we're working with. This is at a small scale, right? Do you trust the paper that you're reading? Is it robust enough for you to say, yes, this is science that I would like to continue reading and build on my own hypothesis, whatever that may be?

The second point is we actually need a way to think quickly, but more accurately, about the work that we're looking at. That's what we're hoping to do with Ripeta is for us to think slowly, to think deeply about what it means to have a well-written paper with good information in it, so other people can think quickly about whether they want to use that information or not.

KENNEALLY: We have been speaking today with Dr. Leslie McIntosh. She is co-founder and CEO at Ripeta in St. Louis, Missouri. She's working to take the crisis out of the reproducibility crisis. Dr. McIntosh, thanks for joining us on Beyond the Book.

McINTOSH: Thank you so much for having me, Chris. I've really enjoyed being here.

KENNEALLY: Beyond the Book is produced by Copyright Clearance Center. Our co-producer and recording engineer is Jeremy Brieske of Burst Marketing. Subscribe to the program wherever you go for podcasts and follow us on Twitter and Facebook. The complete Beyond the Book podcast archive is available at beyondthebook.com. I'm Christopher Kenneally. Thanks for listening and join us again soon on CCC's Beyond the Book.

END OF FILE