In 2019, It's the Data

Special guests include

- **[Ripeta](#) co-founder Dr. Leslie McIntosh**
  - **CCC's [Ian Synge](#)**
- **Springer Nature OA Books Director [Ros Pyne](#)**
  - **[NewsGuard](#) co-founder Gordon Crovitz.**

For podcast release

Monday, December 23, 2019

KENNEALLY: Welcome to Copyright Clearance Center's podcast series. I'm Christopher Kenneally for Beyond the Book.

In the final weeks of the year, Beyond the Book is looking back at the past twelve months of our programs.

In this edition of our three-part review for 2019, innovative publishers and their partners describe the role of data and information – good, bad and open – in shaping the future of scientific research and journalism.

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. 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, explains co-founder and CEO Dr. Leslie McIntosh.

McINTOSH: 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.

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: How do you make possible this scientific hygiene report?

McINTOSH: 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: 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.

KENNEALLY: "The next Darwin is more likely to be a data wonk than a naturalist," says David Weinberger, a Harvard researcher who focuses on the essential elements of the information age and what he has called the new digital disorder.

For publishers looking to organize their complex digital worlds, metadata – information about information, data that describes data – can be a friend or a fiend.

With metadata, the information and descriptions provided conceivably might go on forever. A shortcut to success, however, can begin with defining Minimum Viable Metadata. The MVM itself will reflect a mix of internal and external factors, from IT systems to compliance requirements.

Ian Synge, a principal consultant in Copyright Clearance Center's UK office, told me at the London Book Fair in April that publishers and others with media content often have more metadata than they think they do.

SYNGE: Absolutely. It's almost invariably the case. So when I sit down with a publisher who's trying to do something, and it's usually, 'I can't find my stuff. I can't control my stuff. I don't know where my stuff is. Help me.'

Last year, was working with a publisher who – their main content was pretty well controlled.

But a lot of the ancillary material that was around it – so, in particular, their video content and their imagery – was out of control. And they described a situation like they had a warehouse where the lights were turned off, so they had brilliant stuff in there, but they couldn't find it. So a lot of the time they just recreate it.

And we sat down with our media manager and said, 'OK, what do you have?' And he goes, 'we're hopeless. We have nothing.'

And that's always really scary. You present someone with a blank sheet of paper, and they're always going to – you know, it's terrifying.

So we said, 'well, actually, we think we've probably got more than you think, so show us some of your files.' And he'd bring up a file and say, 'look, there's no metadata. It's just a picture.'

And I said, 'well, OK, right-click on it and see view properties. What do you see? Well, you see a data created, so that tells you something. There's a geocode on it because you took it with a camera that's put a latitude and longitude onto it, so you can tell where it was taken, and that tells you something. With video, you can tell how long it is, and that tells you something. And from those, you can start to infer things. And that started to get him towards this principle of minimally viable.

Now, it's not earth-changing. It's not something that unlocks all of the value there, but it gets you started, and it meant that they could start actually start actually surfacing this content rather than just endlessly reinventing the wheel.

KENNEALLY: Open access is much more than just a publishing business model. When a work of scholarly research is made freely available to readers across the globe, the impact is dramatic.

Earlier this year, Springer Nature asked scholarly authors to share their views on the quality and impact of OA books. A white paper on the future of open access books details the survey findings. Co-author Ros Pyne, director, open access books at Springer Nature, shared with me highlights. Pyne noted that while OA books are a popular idea, authors question whether "open" is also "trustworthy."

PYNE: Perhaps our most exciting finding – and I was a little bit surprised by this, to be honest – was that the majority of all book authors support the idea that all future scholarly books should be open access. Perhaps I shouldn't be surprised. But because OA has relatively low take-up levels at the moment, I was perhaps expecting a little bit more skepticism. But we had more than half of all authors, both those who had published open access before and those who had not published OA before, say yes, we want this for the future.

We did find that attitudes varied. Pro-open access attitudes were stronger in Europe and Asia compared to, for example, North America. We also found that junior researchers, so those with under 10 years' experience, tended to be more pro-open access, and that senior researchers, so those with more than 25 years of experience, were more skeptical or more cautious.

We also found some really interesting findings about what was stopping people from publishing open access. Some of it was just lack of awareness. People said I didn't know this was an option

for me publishing a book. I know about open access. I think it's great. But I didn't know I could publish a book OA. I think people were also concerned about how it would be perceived. So even if they themselves think publishing a book open access is a great thing, they're concerned about the perceptions of the quality of open access books. They might know that it's peer-reviewed, but does everybody else know that? How will their tenure committee review that? What will people think if they're looking for a new position or looking for a promotion? Will publishing OA affect that?

KENNEALLY: Those findings point to one of the conclusions, one of the suggestions you make in your report, which is there's a need to educate scholarly authors about open access publishing for their books.

PYNE: Yeah, absolutely. It's something that our editors are doing day in, day out. I think one of the most powerful ways of changing minds or of communicating about new ideas is to have that one-on-one conversation. And while our editors are out there talking to academics about their latest research and how we can best help communicate that to the world, they can talk about open access and start to reset some of those perceptions. But I think there's really a role for everyone here – institutions, libraries can be getting involved in saying we support open access, and this isn't going to affect your tenure application, for example.

KENNEALLY: Misinformation and disinformation thrive best in the dark. And like fungus and mildew in a cellar, they disintegrate quickly when air and sunlight arrive on the scene.

In the emerging area of online trust technology, the startup NewsGuard has decided to disinfect the web of false reporting with the detergent of journalism.

Gordon Crovitz is a distinguished publishing veteran who co-founded NewsGuard with acclaimed journalist Steve Brill in 2018. He spoke with Beyond the Book in August about making the web a better place one website at a time, one maybe news article at a time.

KENNEALLY: This is such a very important topic, especially in 2019, Gordon. I have to start out by asking you about the state of the web today, because your team has really looked at things – you're looking at things that maybe some of us would rather not see. What have you seen? How bad is it out there?

CROVITZ: It's not good. As you were describing, the way NewsGuard operates, our analysts look at all of the news and information websites that account for at least 90% of online engagement in every country in which we operate – that's the US, UK, Germany, France, and Italy. In the US market, we've looked at 96% of all the news and information sites that people look at, and we've been a little bit surprised by what we've found. For example, more than one in 10 of those popular news websites contains health misinformation – false reporting, for example, about the dangers of vaccines. So when you think that one in 10 of the websites that Americans read for news and information contain misinformation about health issues, it's not a surprise that measles is back. It's not a surprise that people feel very anxious about the quality of the news that they're getting online, whether it's about health or about politics or other topics.

The world has become such that regular people feel very anxious about whether they're getting news from reliable sources or not. And this is the downside of the brilliance of the internet. In the print era, some of your listeners may remember, people would go to a newsstand in order to acquire newspapers, and they might say I like the Philadelphia Inquirer, I don't want the National Enquirer – knowing that one of them was a respected local newspaper and the other one was a kind of grocery store checkout gossip sheet. On the internet, where people are thumbing through their Facebook feeds or their Twitter feeds or looking at the next video that pops up on YouTube, the value of those brands has so disappeared. There are so many brands, people don't know what to make of any particular one.

And that is the problem that NewsGuard set out to help solve on behalf of news consumers – which websites are trying to do journalism? Which ones are doing something else? And for all of them, to give them a red rating or a green rating and a nutrition label writeup that explains everything that a reader would want to know about that particular website so that he or she can make up his or her mind about whether to read news from that website with an extra grain or two of salt.

KENNEALLY: And the way you go about it, Gordon Crovitz – describe how you work at NewsGuard. There is this team of experienced journalists who do the research, so it's humanly curated. That's really the very opposite of the sites that are propagating these kinds of news sites, because they use algorithms. You use human beings.

CROVITZ: Everything we do is the opposite of an algorithm. We try to be completely transparent. We use nine basic apolitical criteria of journalistic practice. For every website, our analysts look to see how they do on these nine. Among the nine are criteria such as, is there a corrections policy? And you find out who owns this website. Do they responsibly treat news and opinion differently? In other words, are they trying to follow basic journalistic practices or not?

And through that process, when it looks to one of our analysts as if a website is going to get a negative mark on any of the nine criteria, the analyst practices journalism. He or she calls that website, contacts that website, and asks to speak about those criteria or that criterion. And in many cases – more than one-quarter of all the cases so far – a news website has actually changed its journalistic practice in order to do better on our ratings, which we think is great.

KENNEALLY: Of the changes, trifling and profound, that technology has brought in the Digital Age, the most unsettling is our loss of trust. What we read, what we hear, what we see – they all may not be what they seem. Restoring trust will require not updates in computer code but a return to earlier codes, codes of decency, honor and respect for the facts.

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. Best wishes for the coming year. And join us again soon on CCC's Beyond the Book.

END OF FILE