Reading the Data Compass
London Book Fair 2023 Preview

For podcast release
Monday, April 10, 2023

KENNEALLY: To read a map of the scholarly research ecosystem, stakeholders should keep a data compass handy.

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

The guiding role of metadata in an increasingly complex scholarly research ecosystem will be the focus of a CCC presentation at the London Book Fair that I will moderate.

As part of the Research & Scholarly Publishing Forum on Thursday, April 20, my panel will share insights why a data compass is essential when scaling the heights of Open Access.

Daniel Shanahan Publishing Director, PLOS, and Michaela Atherley, Director of Editorial Operations - Open Research, Taylor & Francis will join me along with Dr. Jose Salm of the Federal University of Santa Catarina, Brazil.

In February at the Researcher to Reader Conference in London, Dr. Jose Salm shared with me his experiences on the guiding role of metadata from the Brazilian perspective.

When we place quality data at the center of our organizations and our work, what are we trying to achieve? That's the question we'll ask.

British economist and Nobel laureate Ronald Coase once said that if you torture your data long enough, it'll tell you anything. My panel, thankfully, will advise us to treat our data well, because then it will return the favor. They each have to share ambitious goals and some early-stage success stories for how a focus on metadata management opens a 360-degree view of the research ecosystem, and they will explain why data-driven decision-making has helped them to address concerns of sustainability, compliance, and access.

KENNEALLY: Dr. José Salm. I want to welcome you. Thank you for joining us today. You're a professor at Santa Catarina State University in Brazil, but you have an extensive knowledge – or background, rather – in knowledge engineering. You've worked for a

variety of government and science organizations – the Pan American Health Organization, NIH, NSF, and so forth.

But you want to tell us about a database – a data management platform that's been around for quite some time that Brazil developed in the late 1970s, the Lattes platform. It's named for Cesar Lattes. He discovered the pion, which is a subatomic particle, and he even has a song written by Gilberto Gil about his work. So he's quite the scientist to put a foot into the world of the arts as well. But this platform, this Lattes platform, is something that really has changed the way you do science in Brazil. Tell us about that.

SALM: Well, it was developed in the end of the '90s. At that time, we had a list of different CV systems regarding researchers and the data about their past activities, research, and projects. For this specific federal agency to do evaluation on proposals of grants and the evaluation process, it was very hard. What we did was we unified this. We studied some metadata that would be necessary to integrate some of these different data sources. And we consulted from a list of 600 researchers in different fields on what would be interesting for them regarding the CV structure.

So we built that into the system – and we're talking 1998. If here, the internet was kind of not a main thing yet, imagine in Brazil. So what we did was we developed this simple service that brought researchers into this space in using the platform. We would give them different graphs regarding his or her CV information. Who have you published more? What areas do you work more? And then we just built their web page. That was a big thing at that time, right?

Today, Lattes has almost 8 million users in Brazil and from different countries. And we developed other systems that were connected to it. But it really changed the evaluation – the funding that was granted in Brazil – because it took out of some main cities and started to spread funding money into different regions that at that time weren't getting that approved.

KENNEALLY: We've been hearing a lot throughout the conference of finding ways to establish greater equity and to be more equitable in research and science and the entire workflow. Here, what you're saying is because of the data you were able to collect, because of the vision you had into the system, you were able to drive funding, to drive research, into places where they might have been deprived or on the margins in the past.

SALM: Yes, exactly. And we did that because once there was a platform in place and the evaluation cycle had to consider publicly available data on the researcher or his profile or her profile, people that were part of the evaluation panel would have to justify why they

didn't have any conflict of interest. So decision-making here with data and with this platform was bringing some form of equity and some form of transparency.

Then there were these data warehouse structures built on top of research groups and other data sources, and this helped build not just the evaluation phase, but also monitoring. So yes, this is –

KENNEALLY: And as you say, though, in 1997, it was very early days for all of this, and there were none of the now internationally accepted standards in existence. So how are you leveraging the standards that have come along? That must make a great difference in your work.

SALM: Yes. We're partnering with Portugal, and we have learned from them how to bring all these persistent identifiers in a way that you can add outside data sources and combine data, and we've learned from their lessons on using this data.

And we are finding some challenges regarding data quality in some of these data sources. This is something that we have to address. Data isn't referenced – say you have a publication. There's no DOI. There's no coauthors' identifications. In Lattes today, once you add a publication to your profile, you have to reference the DOI and also reference coauthors. Once we extract the data, if the coauthor is not identified, it won't let you add to your CV. And there's other validations. This is not rocket science, but it should benefit the quality of data.

KENNEALLY: In Brazil, though, your focus is, you told me, on building bridges, that there's various elements of the research ecosystem that you want to enable to participate and collaborate. Tell us a little bit about that.

SALM: We are trying to connect with different data sources, mainly the theses and dissertations that are published in Brazil and outside, and now we're starting to work and reengineer the whole Lattes model to add SERIF (sp?) and VIVO as references. So we're working together with a group in Germany and discussing the ontology alignment between SERIF structure and VIVO structure, the issues between these alignments, and after this is done, also align with the Lattes ontology or the Lattes structure.

KENNEALLY: So I think we're hearing that you've come a long way with Lattes. There's still a long way to go.

SALM: For sure.

KENNEALLY: Well, Dr. José Salm, thank you very much.

Velocity of Content
A Copyright Clearance Center Podcast