VELOCITY OF
content
A Copyright Clearance Center Podcast

The State of Scholarly Metadata in 2023 – Industry Insights from Around the Globe

A CCC webinar
Recorded Thursday, July 20, 2023

For podcast release
Wednesday, August 2, 2023

CARMICHAEL: I'm Jamie Carmichael with CCC, and I'd like to welcome you all to today's program, *The State of Scholarly Metadata in 2023: Insights From Around the Globe,* which is an opportunity today for us to assess and confront the challenges around low-quality metadata and underutilization of persistent identifiers that disrupts various stages of the research lifecycle – including, but definitely not limited to, the transition to open access publication.

It's clear from conferences, working group initiatives, and technology investments that there's a renewed focus on metadata and persistent identifiers, or PIDs, about people, about places, about objects as an essential component of a vibrant industry. And in embarking on our study of metadata management, it's become clear, at least to me, that an eco-wide commitment to improving data quality from the policy level down to editorial system configurations will help facilitate the transition to open while also helping to preserve research integrity, enhance findability of research, and improve impact measurement.

Joining me now on the program is Deni Auclair of Media Growth Strategies, who worked with us to create the State of Scholarly Metadata interactive report. Welcome, Deni.

AUCLAIR: Thank you.

CARMICHAEL: Deni, we engaged you to help us map metadata management across the research lifecycle, because we saw as an intermediary firsthand how publishers and institutions were challenged in disambiguating author affiliations in order to determine OA funding entitlements. And we thought, hmm, there seems like an opportunity here to learn more about the chain of events that leads to this challenge and then to share our findings to help the industry overcome these and other hurdles related to this topic. You spoke with dozens of community members to map the complexities, the breakages, as well as the value of metadata across all research stages. Can you tell us a little bit about the kinds of questions you asked folks?

AUCLAIR: Sure. We asked questions around implementation and use of quality metadata with the goal of figuring out how we can improve. So we asked questions like who should create and maintain metadata? Where should it originate? What resources do these various stakeholders invest to create, curate, or maintain various types of metadata? What are their biggest challenges when it comes to metadata management or the use of persistent identifiers? What are the most critical metadata elements? What's at stake if those elements don't persist through the scholarly communication process? And who should own metadata quality and control?

CARMICHAEL: Can you tell us a little bit about who you spoke with for the survey and what might have stood out among the responses you got?

AUCLAIR: We spoke to representatives of the various stakeholder groups. So on the institutional side, we spoke to librarians, repository managers, research offices, and grant managers. We also spoke to publishers, researchers and authors, and funders. As I think most people know, there's been a lot of really excellent research and analysis in this space that's been done. We didn't uncover anything startlingly new in terms of what the major issues are, but we got some really powerful insights as to where the pain points are and who's feeling them the most.

Basically, the industry is leaving money on the table, because the lack of standards is hindering search and discovery. Research is repeated or it's slowed down, because content isn't discoverable, especially in underrepresented areas of the world. And there's a massive amount of manual effort involved in managing metadata.

For funders, especially as it relates to open access, it's difficult to track compliance with mandates as well as measure the impact of funded research. Also, some of the ethical issues this industry is having could be addressed with quality metadata. There are just so many ways that effective implementation of metadata standards could support all industry stakeholders as well as the general public. The question is how to do that. Not can we do it, but how do we do it, and who takes the lead in those efforts?

I'd like to turn our attention to our panel now and introduce our first panelist, Randy Townsend, editor-in-chief of *George Washington Journal of Ethics in Publishing* and president of SSP. Welcome, Randy.

TOWNSEND: Thank you, Jamie. Just a clarification – I'm the inaugural editor-in-chief. I'm no longer editor-in-chief. I'm just an editor. I'm a humble editor of the journal.

CARMICHAEL: Thank you for that. Randy, what's your take on the overall state of scholarly metadata, and do you have any suggestions for how we can resolve some of the pain by potentially moving metadata management or better metadata stewardship upstream in the research lifecycle?

TOWNSEND: That’s a big, big question. I’m a practical optimist, and I think that we’ve come a long way to make a case about the importance of scholarly metadata and the potential that it presents for business models for the content and then for the service to the authors. I think that we’re still flirting with the interoperable part of FAIR. Generally speaking, the publishing industry’s just teasing out the potential value of a committed relationship with interoperability. Most publishers understand the value of persistent identifiers, and then we go to the authors, and we say we’re going to need you to start categorizing everything through keywords and index terms respective to a particular research community. The author says, well, why? And we say because it will help people find your research. Great. We have the F in FAIR, which is the first ask.

Then accessible follows shortly after. Authors, we need you to put your data in a repository. Why? So that people can access the data, test it, verify it, and build on your results, which also jumps into the R, reproducibility.

But the fun – I say fun, but some could say the challenge – is in the interoperability piece, making sure that publishers are delivering in that value stream and it makes sense to the researcher. Why are publishers asking for ORCIDs and funder data if we’re unable to connect that data to the benefits that they relate to? Why are we asking for keywords and index terms if they’re not being used to identify qualified peer reviewers? It has to make sense, and we as publishers need to be confident when explaining the benefits of the ask to a particular stakeholder. If we’re asking so much of the researcher but not delivering, we’re doing ourselves a disservice, especially to our respective missions.

So when scholarly publishers voluntarily explore the benefits for the stakeholders, it can be fun and rewarding. But it’s less fun when you find yourself being forced to do it, and even worse when you put a timer on it, and you know the buzzer’s about to hit, and that tick-tock grows louder and louder, like we have to open this up. We have to open this up. It changes the tone from exploring to surviving. If publishers haven’t been investing their resources into exploring, then the regulations and then the mandates put their business at risk, and the decisions that they’ll have to make will be reactions.

I’ll just give you a quick example. The Department of Education recently announced their public access plans that includes the elimination of any embargo period before the public gains free access to journal articles resulting from federal funding. That research space has lost the convenience of time to explore the potential use case, value, and benefits of the metadata, and the publishers that operate in that space are going to be forced to accelerate their plans if they haven’t been working on them already.

I’ll wrap it up, because I can keep going on and on, clearly. (laughter) The case for rich metadata is pretty clear, but the way in which we integrate that metadata into our processes is still a challenge. And when metadata is an afterthought, because we’re laser-focused on the post-publication value, then we’ve already missed opportunities in the after-submission and peer review processes to continually enhance and enrich that metadata. So without solutioning this, the question of who does what when is really fundamental in these discussions, and we need to include all of the stakeholders.

From my perspective – and just my perspective representing the publishing space – metadata should begin to be captured upstream in the idea development phase, which is from the research that you all presented a few minutes ago. Theoretically, that puts the onus on the researcher. And the challenge there is we’re putting more and more requirements and pressure on the researcher, who may not be qualified or have the bandwidth to meet those expectations. Different publishers have different taxonomies, and they have different questions, and they have to use the collected data in different ways for different reasons.

All that to say that the author experience can be confusing, and they may have different thoughts that are actually opposite from what I presented, and the funders may have a different opinion from the authors and the publishers. So it really needs to be an inclusive discussion. Again, I say that I’m a practical optimist, and I love the journey. We still have a ways to go, but I’m really confident that we’ll get there.

CARMICHAEL: Thank you for that, Randy. Deni, did you have any follow-up questions for Randy?

AUCLAIR: No, but it just hit me when you were saying about putting the onus on the researcher, because in the survey, there were quite a few people who responded that researchers have enough on their plates. They have enough to do. And to put this on top of them – one of the questions that we asked was would you provide training? And the same answer – it’s too much. There’s too much that we’re asking researchers to do already. So it would be too much to ask them to be trained in originating metadata. We just have to make it easy for them – as automatic and as easy as possible for them to generate that metadata.

TOWNSEND: Yeah, I wonder if there’s an opportunity – again, being the practical optimist, right – is there an opportunity for a third-party solution to work in that space to provide the support for the authors? Again, I say that different publishers have different requirements. So I’m asking for X, Y, Z. Let’s say a paper is submitted and rejected, and you want to submit somewhere else. If they have a different set of requirements, then already it’s confusing and frustrating for the authors. So I'm wondering if there's an opportunity for some industrious group to come in and say, well, how can we provide the support for this particular area?

Because like I mentioned, we are adding more and more pressure to the researchers, whose fundamental job is to do the research, right? They want to explore. They want to present the results. Publishing is a part of their ecosystem, but not necessarily what they are there to do. So really having that conversation to figure out how to help them with this, because helping them helps us, and it ultimately helps the entire enterprise. But I agree. Pushing it upstream adds that kind of pressure, and I hate to say that, but it makes my life easier if they did it (audio cuts out; inaudible) and we don't have to chase them down and try to fix it after the fact.

CARMICHAEL: They're excellent points, Randy. We find ourselves having a platform to facilitate open access workflows, working very closely with publishers and their submission systems to go back and rethink their configuration settings, which were set up a very long time ago, before they started embracing open access programs, and to go back and figure out what tweaks might be necessary to pull the right data in from the manuscript persisted throughout the editorial and peer review process so that we're not asking researchers to assert or reassert fundamental identifiers that the system should already have in place.

I want to now introduce Ana Heredia. Ana is a PhD and affiliate senior associate at Maverick Publishing Specialists. Welcome, Ana. Ana, can you tell us a little bit about the metadata landscape in Brazil and how the challenges around metadata or persistent identifiers are different or similar to what we see in North America or Europe or other parts of the world?

HEREDIA: Thank you, Jamie. Thank you, Deni and Randy for setting up the ground. Yeah, I think that what I'm going to say resonates a lot with what's just been said. I think there's quite a lot of expectations currently set around the role of researchers in metadata sharing.

As you were mentioning at the beginning, there are several articles, blog posts, talks, and guidelines or tool kits being shared and being prepared on how to raise awareness and engage researchers around the importance of metadata. I myself am involved in some of them, advocating for researchers to take a more active role in metadata sharing. So although there is indeed metadata information that can only be delivered by researchers themselves, because it's about – you know, no one knows better the data themselves – I'm reconsidering, as I think we all are, the relative responsibility of researchers on metadata sharing.

I think there are several linked parameters to consider, depending on the context we're talking about. They were addressed previously by Deni and Randy. It resumes to three questions – is someone mandating the deposit of metadata, providing clear standards, guidelines, and resources? I don't know if there is a difference in the Northern Hemisphere with the Global South around that. But at least from my perspective here, it's not that clear what funding agencies or the research offices of the institutions want and how they would like this to be.

The second question would be who indeed needs the metadata to be complete and accurate? Of course, researchers benefit from accessing contextual information around other researchers' data, but only very engaged individuals will have the necessary knowledge, as was mentioned before, or would take the necessary time to enter the data properly for it to be reused.

And the third question would be – and it also was addressed before – is there knowledge inside the organizations that can help to streamline the process of metadata sharing? Librarians and publishers typically are the ones who have this knowledge, because they are used to indexes. They are used to use taxonomies and standards for research information.

So if I had to summarize it, it's quite a lot in line of what Deni and Randy just said. If I had to summarize, I would say librarians and publishers have the how. Funders have the why. And researchers have the what. So I'd just reinforce here – let's let researchers do what they do, and there is already a lot that they are doing. It's research.

CARMICHAEL: When we get down to it, do researchers care about metadata?

HEREDIA: They can care, right? Not necessarily they will care. There is also something important to know – is that if you are a researcher, not necessarily a team leader or a lab leader who has more an overview of what's going on in your field – if you're a researcher, when you finish a paper, when you finish a subject, you are already looking at the next one. You don't care anymore. You have to be a very special person with a very special profile really caring about all these little intricacies of the research information. But otherwise, I'm happy by publishing my paper, and I'm bothered they are asking me to deposit the data, and the metadata is something that is from the subjective world somehow for the majority of the researchers, I would say. Of course, it depends on the field, right? I'm a biologist. Maybe someone who is on the information side of the research – maybe they care more. But a biologist? I'm not sure.

CARMICHAEL: Thank you. I would now like to introduce our next panelist, Wolfgang Mayer, head of e-resource management at the University of Vienna. Welcome, Wolfgang. From the institutional perspective, Wolfgang, what could stakeholders do better to improve the quality of scholarly metadata in terms of standardization or raising awareness? Any thoughts on that?

MAYER: It's quite difficult, because originally, I planned to start with another statement. But just playing a little bit of devil's advocate to the speakers before, for the researchers, I think there are two paradigms. Obviously, it's the task of the university, in our case, to enable them to put as much effort and energy into research and less into administration as we could. But nevertheless, administration of their publications, of their research data, of their identifiers, of the metadata is still part of the job.

Ana said before that there are different activities and different questions to stakeholders to take part in. In Europe, it's a little different than the US, I think. The funders, based upon the Coalition S initiative, are trying to increase the pressure regarding the open access part of the publishing and the traceability and visibility of these publications. To say it frank, there is some kind of pressure, and there's not the freedom that the researchers, at least if they are members of a university, are completely free to decide the time and the quality when they create the metadata.

Having said that, I think it's also a generation gap. The young researchers with new projects – when we are able to present incentives regarding the visibility, especially, they could be quite easily motivated in the beginning of the research to put thought into the metadata. I really think that older researchers will not go this way themselves. So this is a chance, like we at University of Vienna tried to do it, to place services – in our case, at the the library – with the open access office, with the repository management, with the metadata management, with doing licenses to various metadata possessors and curators, like, for example, Ringgold and other things, to have some kind of impact on the standardization of metadata and to take some of the workload off the researchers.

One way is training. But we are the second-largest university in northern Europe, with 10,000 researchers and more than 90,000 students, so we never will have the right channels to communicate all of the tools and the possibilities regarding that. The other things are simply doing some services, like creating ORCID IDs, creating sets of Ringgold IDs, helping the researchers from the beginning of their research project to create research data and the research flows, wherever possible, to help them. In some of our institutes, there are also administration staff who in reality does much of the manual submission processes of the preparation of those to help the researchers.

CARMICHAEL: So the activities that you're describing happening at the University of Vienna – is that isolated to your university? Is that something that you see other institutions doing to provide better support and standardization to their researchers for the more administrative pieces of the research process?

MAYER: Definitely others, too. But especially it's a question of the standing of the library within the university and which departments are placed at the library. In our case, bibliometrics, research (inaudible), the CRIS systems, open access office – all of them is placed at the library, which is quite different than at many European institutions.

CARMICHAEL: Deni, any questions for Wolfgang?

AUCLAIR: I have one question. You mentioned CRISes. How do you feel that CRISes support or do not support the use of metadata – implementation of metadata?

MAYER: (laughter) Perfect question. Let's try to see the other way around. The CRISes are dependent upon metadata created elsewhere. So we have now the third project funded by the ministry where all of the Austrian universities and some other research institutions take part where we try to create a comparable infrastructure for CRIS systems, even if there are many different technical solutions. But we said, OK, we try to have the same infrastructure when we harvest this data for the consortium.

Also, there's one consortium for all of the more than 60 members who have all of the open access deals. It's very, very important to evaluate the deals and to have data – which publications by which corresponding authors are published under which open access license. It's drawn from various sources – from the publishers themselves, not so good data, unfortunately, in some cases, from citation databases like Web of Science, Scopus, and so on – Dimensions – the things manually entered within our local CRIS systems, and identifiers from all sources. So in the end, we try to receive good metadata, standardized metadata, and based upon this, create a comparable output for the university and for the funders.

CARMICHAEL: Well, I just want to thank our panelists for their time and insights today. Why does this all matter? These challenges make it very difficult, as we discussed, for all the stakeholder groups in scholarly communications to advance their goals and objectives. And we heard some particularly prickly challenges when it comes to the researchers themselves.

So some of my takeaways here – definitely a dedication to metadata stewardship across each stakeholder group as a shared responsibility and the service providers supporting them is vital. New metadata strategies, inclusive policies, and a more robust framework of interoperable systems really seem essential for modernizing this element of scholarly communications.

I want to remind folks to check out our stateofmetadata.com interactive map. We are always looking to iterate on this. If you think we missed something or you have any comments or questions, there's an ability to provide us feedback through the form on that site. And we want to share this recording after the program so you can go back and reference any points of interest. Again, thank you to our speakers and our audience for taking time out of their days on a really important topic. Appreciate it.

Thank you.

END OF FILE