Velocity of Content
A Copyright Clearance Center Podcast

Interview with
Michele Avissar-Whiting, editor in chief, Research Square

For podcast release
Monday, May 16, 2022

KENNEALLY: Science traditionally advances in two steps. The first is taken in a laboratory or library, the second when sharing the results. Online, digital publishing of so-called preprints has quickened the beat tremendously, closing the gap between researcher and reader. For the most part, we are better off, with the accelerated development of treatments and vaccines for COVID-19 as the model example. Yet when research is retracted, the record may not reflect the reversal.

Welcome to Velocity of Content. I'm Christopher Kenneally for CCC. A peer-reviewed article published this month in *PLOS ONE* has examined what happens in the scientific record when journal-published versions of articles are retracted in cases of research previously available on preprint servers.

Michele Avissar-Whiting, editor in chief of Research Square, a leading preprint provider, found a reassuringly small number of such retractions. However, she also writes that inconsistencies in publisher responses pose a threat to the scholarly record and to scientific integrity. Michele Avissar-Whiting joins me now from Raleigh, North Carolina, where Research Square is based. Welcome to the program, Michele.

AVISSAR-WHITING: Hi, Chris. Thanks for having me.

KENNEALLY: We are looking forward to speaking with you, Michele. You reviewed the record of retractions for preprint articles appearing on Research Square as well as bioRxiv and medRxiv. Just 30 retractions turned up, representing 0.01% of all content posted on those servers. So that's the good news, isn't it?

AVISSAR-WHITING: Yeah, I was certainly relieved that I didn't find a lot more. Actually, it's something I thought of only after I wrote and published the paper, but I was thinking about that range of 0.01% to 0.05% and wondering is it really small? That seems like an inherently small proportion, but how does it compare to the overall rate of retractions? Is it commensurate? And actually, from what I can tell, it's right in the range. I found one study that claimed about four out of every 10,000 papers are retracted, which is 0.04%. And if you consider that there are 32,000 entries in the Retraction Watch database, that's relative to the 50 million-some odd papers that have been published all time. That also

ends up around 0.06%. So we're dealing with the same kind of order of magnitude as what I found, which both – it puts me at ease a little bit that we're looking at probably just a continuation of the same process that's been happening for a long time, and there's not a disproportionate number of retractions among papers that were previously preprinted.

But I'm not going to be totally at peace about it until I have a better handle of some of the open questions that remain. The big one is how many preprints are we failing to link to downstream publications, and as a result, just missing the opportunity to update them with downstream retractions and events like that?

KENNEALLY: So we better understand, though, what retractions are about, what are the kinds of problems we're talking about, Michele? And why does it seem, do you think, that we hear more about retractions of scientific literature than in the past?

AVISSAR-WHITING: Retractions are typically carried out by journals after what is often a protracted investigation process, either by the journal's integrity committee or the institution of the author that published the paper. And there's a variety of reasons. I have them categorized in my paper as well. For simplicity, I think I only have three categories that fall into broadly the category of misconduct. So things like fabrication or data falsification, image manipulation, things like that are one thing. There's piracy and plagiarism, of course. And then there are also just retractions that happen because of errors. The author might even come forward, realizing that there was some critical error in their data and that they can't stand by the result anymore, and they'll reach out to the journal. This is pretty rare, but it happens. The journal will then retract the paper. So those are the kinds of problems we're talking about.

And the question of why we hear about retractions now more than we used to – I was interested in this, too. I talk about it a little bit in the discussion of the paper. It's been studied. Indeed, retractions are much more common now than they used to be. In fact, there was roughly a doubling of retracted papers between 2003 and 2009. And the time to retraction, meaning the time it takes a journal to retract a paper, has also gotten shorter significantly in the last couple of decades.

What the meta-research on this seems to suggest is that it's not due to a wild increase in misconduct. It's not that people are fabricating data at a much higher clip necessarily now. It's just that journals are more likely to retract papers with errors or other problems than they used to be. It's become more normalized, which is a great thing. And a lot more discrete journals actually retract papers now. It's not just the high-impact-factor journals that are taking responsibility and retracting problematic papers. So the retractions are now distributed over many different journals. That also helps to hear about them more often.

They get reported on. We have Retraction Watch specifically dedicated to talk about high-profile retractions.

I also suspect that the shift to digital, like with everything else, has made things more easy to hear about and discussed more readily in public. And then the open access movement – these are the spaces that we've created on the internet to discuss research publicly that have helped this along as well. So it's much harder to hide serious problems with a study now, which is, I think, overall a really good thing.

KENNEALLY: As you say, Michele, it's not only harder to hide, but it's also important that social media plays a role. The public has access to a lot more scientific literature than in the past. And if they think they see something, they're going to say something.

AVISSAR-WHITING: That's right. Yeah, we have now PubPeer, just an entire platform basically dedicated to surfacing these issues with papers, and often participants anonymously discussing the issues with those papers and taking them to the journals and asking them what they're going to do about it.

KENNEALLY: So tell us what drives researchers to publish on preprint servers like Research Square. Are preprints also well accepted now in the scientific community?

AVISSAR-WHITING: Yeah. So there are a lot of different drivers for different people as far as why they choose to preprint. For some people, I think it's really a practical move. The work has been done, and they want to get it out and gauge the reaction of their community as quickly as possible. Some people are really treating it as an act of personal sovereignty. This work is mine to publish when I think it's ready, and then it's on my peers to decide its value to the field. And some researchers are in really fast-moving, competitive fields and looking to timestamp a finding. Preprint gets a DOI, so the primacy of the work is established at the point of posting the preprint.

During the pandemic and other emergency public health situations before that, it was really a necessity. It seemed unthinkable for a potentially really important discovery about the virus, about transmission, about masks, about vaccines to hide on an editor's desk for weeks or months. A month was a year in pandemic time. So of course, the last two years have made people think pretty hard about the existing system and whether it's fit for purpose, whether it would benefit all disciplines, really, to shift to a preprint-first model, more similar to what the physics community has embraced since the '90s. And this period just kind of, in my mind, did a lot for both the awareness and acceptance of preprints in circles that previously had only very modest uptake of the practice, like medical science.

I would say we still have a long way to go, though. There is much less awareness or acceptance in the global south, for example, relative to the US and Europe. I think that our platform, for example, is doing a lot to change that and make people aware that this is an option for anywhere that you live.

KENNEALLY: You spoke of DOIs, Digital Object Identifiers. So metadata, identifiers, help us to track the process here – the progress, if you will, of science through various stages of publication. Those DOIs, those identifiers, are going to be important with this issue around retractions. So what are the processes in place to ensure that should an article be retracted, that it rebounds back to the preprint version?

AVISSAR-WHITING: Under ideal circumstances, this process that I've been talking about of rebounding or back-propagation, as I call it, of the retraction information back to the preprint, should happen. But there needs to be a conscious effort on the part of the preprint server to ensure this happens. So first of all, there needs to be a mechanism to reliably link the preprint to a downstream publication that doesn't rely on the author actually following up, because we know this isn't a realistic expectation.

Already here with this linking, there is a caveat, in that the main mechanism used by preprint servers – our server and others – relies on a near-exact match of titles and authors in Crossref, which is the organization that issues the DOIs to all of these publications. There needs to be a near-exact match between the preprint and the journal article in order for that link to happen. So if the title or the author list deviates too much, the match isn't triggered. Already, there's some unknown number of preprints that fail to establish that link.

But if the link is established, then there is the potential, at least, to discover if something important happens with that publication down the line. That's because retractions are typically noted in the Crossref metadata for the journal article, and the preprint article link, that link between the preprint and the article, is also recorded in Crossref metadata. So both of those components work together. And to the extent that a preprint is linked to an article, a retraction is discoverable, theoretically.

Now, there are caveats there, too. For one, the Crossref data is only as good as the data that they receive. So if the journal isn't following best practices, the retraction may not be recorded. Also, there are journals that are not even Crossref members, and retractions at those journals obviously won't be discoverable via Crossref. This is where a database like the Retraction Watch database fills in a lot of the gaps. They have three or four more times the number of records of retractions than Crossref.

But moving on from those exceptions, now that you have a process in place to check against one of those databases and pull out the retractions related to your preprints, now you also need to have a process for marking the preprints appropriately and consistently, right? All these things are really resource-intensive, and this is something I discuss in the paper as well. They require development. They require engineering. And many, if not most, preprint servers are operating on shoestring budgets and really don't have the means to implement even the automated linking part to begin with.

So I don't have a specific answer on how often is this happening? My guess is it's probably pretty rare. But the good news is that at least the largest preprint servers – ours, bioRxiv, and those that are housing most of that content – they do typically have this automated linking process. So if they aren't already updating on retractions, they at least have the potential to do that.

KENNEALLY: Are there suggestions for best practices and other standard ways of handling all this that you would put forward? And are you seeing Google Scholar and ASAPbio, which is a preprint advocacy organization, stepping up and trying to help address this particular problem?

AVISSAR-WHITING: I don't think there's been a lot of focus on this particular problem, because preprints are, at least in the spaces that I'm talking about – mine was focused specifically on the life and medical sciences – are relatively new. And this concept of a downstream retraction, and what do you do about it with the preprint – that's, I think, something that we're going to be facing more and more as preprints become normalized and as time goes on. So it hasn't been a major focus of those kinds of organizations.

What I will say is you mentioned Google Scholar. You mentioned ASAPbio. Google Scholar does a pretty fantastic job, for example, of aggregating versions. I don't even know all the mechanisms in place behind the scenes to make this work. But it does a pretty good job of putting under sort of one header all of the versions associated with a particular study and privileging the most recent version or the version of record. That's fantastic for the purposes of curation and stewardship.

And ASAPbio has done extremely important work in getting representatives from a large number of preprint servers, both commercial and nonprofit, with hugely divergent models and operating protocols – getting them all at the same table to reach consensus about the standards that we all believe should exist for this medium, and that it's not just a trivial undertaking. But at the end of the day, you're left with the realities of budgets and other constraints that mean that not all of these servers will be able to meet those standards, necessarily.

I recommend that everyone take a look if they're interested at the very extensive document that ASAPbio prepared, along with a couple of other groups a couple of years ago – I think it's like 2018, 2019 now – on recommendations for preprint servers. It's a really great document, and I think there's going to be a lot of work continuing on with this to address some of the concerns that I bring up in my paper.

KENNEALLY: And when a retraction happens for a published article, what should happen to the preprint version? Should it be withdrawn? Should it be identified as such? Do you have an idea as to what the best way forward there is?

AVISSAR-WHITING: I mention in my article that ASAPbio's recommendations kind of stop short of offering specific guidance on this. Instead, they say it's up to the preprint servers to decide what action to take when there is a downstream retraction. But in my view, there are not a lot of instances I can imagine where the reason for a retraction of a journal article wouldn't also warrant withdrawing the preprint that preceded it. It's effectively the same study. So if the issue was with the conduct of the study, like problematic treatment of data or missing ethics approval or something, that would be something that's relevant for both outputs. That would be something that if we knew it was true of the preprint itself, we would take action and withdraw it and note it. So I can't think of a lot of instances where a retraction would be warranted downstream, but not on the preprint itself.

In doing this analysis, of course, we have tidied our own house in this respect and looked at each case of a downstream retraction, decided in each case whether to withdraw the preprint. And in the end, we ultimately withdrew all of them. It was like 16 or something that we ended up withdrawing as a result.

This action just effectively tells readers that the authors or some authoritative figure has deemed the work untrustworthy, and they don't want it cited. It shouldn't be cited as valid research. So for the same reason that that would be important to do and mark very visibly on the journal article, for the same reasons, we want to do it on the preprint platform.

KENNEALLY: And when it comes to the scientific record and the confidence that we can all have in it, it seems to me that the responsibility here ultimately lies with authors. Should they be the ones coming to their preprint providers and telling them about a retraction? I suppose there are going to be exceptions to that. But it does really seem to be on their shoulders.

AVISSAR-WHITING: Yeah, in an ideal world, everyone would take this responsibility. But we already know that we can't rely on authors to even update us about the preprint being published. Richard Sever at Cold Spring Harbor, who's in charge of the bioRxiv and medRxiv platforms, has said a similar thing. They're just like, they can't rely on the

authors to come back and let them know that the preprint has been published. So it would be pretty foolish to assume that they'd update their preprint in the case of a retraction. And I would say that's true even under circumstances where the author agreed with or even initiated the retraction, because that certainly happens when errors are caught, for example, or one member of the group discovers another tinkered with the data or something like that. That's not even to speak of the many cases where authors do not agree with the retraction, which is probably the vast majority.

Look, retractions are still a contentious thing. I think some of the best work being done in the area of scientific integrity is the work to destigmatize the retraction – not just normalize it, but reward the admission of fault and the active correction of the record by authors. The more we can do in this area, I think the more we'll see that authors taking responsibility – there was even a case on Twitter, I think it was last month, where somebody came forward and said that this is one of the hardest things I've ever had to say – I think it was a PI at a lab. She said we've had to retract our paper, and then went on to list the problems with it and why they took that action. It was a hugely viral post, and there was nary a negative comment to be seen. Everybody was just lauding this behavior. I think we need to see more of that and people essentially being socially rewarded for taking responsibility in this way.

KENNEALLY: Michele Avissar-Whiting, editor in chief with Research Square, thank you so much for joining me today and sharing your thoughts on this important issue.

AVISSAR-WHITING: Thank you so much, Chris.

KENNEALLY: That's all for now. Our producer is Jeremy Brieske of Burst Marketing. I'm Christopher Kenneally for Velocity of Content from CCC.

END OF FILE