

Journée francophone OpenAlex

Questions/Answers – Jason Priem (CEO)

2 DÉCEMBRE 2025



Questions/Answers

Question : Would you talk a little bit about abstracts? They are never CC0, aren't they (considered as work of mind).

Jason Priem : *This is a better question for a lawyer, but I can say that lawyers I have talked to have uniformly told me that this is correct. Abstracts are considered intellectual products and covered by copyright, at least in the United States. Of course, the copyright holder can revoke their copyright and declare it CC0, or they can use a copyright license like Creative Commons. Again, however, I'd like to emphasize that I'm not a lawyer, and this is not legal advice :)*

Q : Are there any plans to introduce a dedicated 'Event' entity in OpenAlex? Conferences, workshops, and similar scholarly events are currently not well represented in most open scholarly indexes, which makes it hard to properly connect conference papers with their venues and the communities that produce them. Is this something OpenAlex is considering for future development?

JP : *No, we don't have any such plans. I think an event as an event would be out of scope for OpenAlex because we're really covering scholarly information. An event which produces written material is in scope, but we would call that a source. We do have many sources which are conference proceedings. In order to find these, you can search for sources and filter by type equals conference. Like this: <https://openalex.org/sources?page=1&filter=type:source-types/conference> We have more than 2 million works that have a conference source: https://openalex.org/works?page=1&filter=primary_location.source.type:source-types/conference However, conference proceedings are, as you note, more challenging to track than other types of sources like journals. So I would say our coverage there is still a work in progress.*

Questions/Answers

Q : How do you deal with the ongoing big wave of predatory publishing and AI slop nonsense that geos into « legitimate outputs »? (hijacked journals, peer review mills, etc.)

JP : *For now, the process is very manual. People tell us about problem journals or problem outputs, and we remove them. However, very soon, we're going to have to escalate to automated detection. I think it'll be very similar to how email started being overrun with spam until spam filters showed up. The only way to fight bots is with other bots.*

Q : Could you talk more about the topics of OpenAlex ? I've been starting to work with it, and I'm facing a lot of problem with publications in french that the topics doesn't seem right at all. Are your team working o this issue ?

JP : *I'm sorry to hear that. The topics are trained on a multilingual corpus, and so we have heard people tell us that they work fairly well in non-English languages. It sounds like you're certainly not having that experience, and I'm sorry to hear that. We'll be looking at revising the topic model next year, but it's probably not our top priority right now. We are deeply committed to support across multiple languages, and I believe that right now OpenAlex has the best support for works in languages other than English than any other system in the world. Indeed, probably compared to any other system in history. But I know that's a small consolation when you're facing this specific problem in your own workflow, and I'm sorry about that. All I can say is we'll keep working to improve the system. It's still early days.*

Questions/Answers

Q: very curious to see how you envision curating openalex data "in the style of Wikipedia", do you already have any mockup? How long before production?

JP : *We have a few mockups, but nothing that's ready to share. The basic conceit is that in the user interface, next to data that can be edited, there will be a pencil icon. Qualified users, and these will have to be pre-vetted by us, can click the pencil icon, and then they'll get a window where they can edit the data. Perhaps a better example is actually the Wikidata interface rather than Wikipedia, but I often use Wikipedia as an example because people are not as familiar with Wikidata.*

We are planning to have this launched in Q1 of next year.

Q : Do you have an idea of how much OpenAlex is used in institutions in Africa specially in francophone countries

JP : *I'm sorry, no, I don't have much idea of this. Our server logs show considerable use in Africa, along with all other continents, but I don't think that's a very useful answer. I think what would be useful is to understand how much African institutions are using us compared to other systems, and I don't have that data.*

Questions/Answers

Q : Unpaywall is closely tied to OpenAlex for computing OA status. With the Walden release, we noticed unexpected classifications: many of our repository records without full text ('metadata-only') are labelled as Green OA.

JP : *I'm sorry to hear that. It definitely sounds like a bug. As you note, Unpaywall and OpenAlex now share the same codebase, so the good thing is, if we can fix it in one, it will be automatically fixed in the other. Can you please submit a bug report? Thank you.*

Overall, our data show that since the Walden release, our open access data has become more accurate in both OpenAlex and Unpaywall. We've published some of this data on our blog. However, I understand that's not very helpful to hear when you are dealing with a regression in accuracy for your own use case. We'll try to get that fixed for you!

Q : The documentation doesn't clearly explain the decision logic. Could you clarify how OA status is determined?

JP : *I think this is what you're looking for? It has a decision tree which articulates the steps of the algorithm*
<https://support.unpaywall.org/support/solutions/articles/44001777288-what-do-the-types-of-oa-status-green-gold-hybrid-and-bronze-mean->

Q : Have you explored providing the OpenAlex data in a platform that supports direct SQL queries, like Google BigQuery?

JP : *Yes, we have. But so far, interest from the user community has been limited. We will definitely continue to listen, and if there's enough interest, we'll make this happen.*

MERCI !

