COPIM •

Thoth, open metadata and building structural equity: an interview for Open Access Week

Javier Arias, Lucy Barnes

Published on: Oct 27, 2021

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

The theme for Open Access Week 2021 is ‘It matters how we open knowledge: building structural equity’. Here at COPIM we spend a lot of time thinking about equitable ways to support Open Access book publishing, so, to mark the occasion, we bring you an interview with our software developer [Javier Arias](#) reflecting on a key piece of infrastructure we are building: our open dissemination system, [Thoth](#).

1. What is Thoth and why is it needed?

Thoth is a metadata management and distribution platform specifically tailored to tackle the problems of getting Open Access works into the book supply chain. It is being built with openness in mind: its source code is open, its data is exposed via open APIs and all its outputs are released under a CC-0 license.

Thoth’s main goals are:

- To lower the entry barrier to good metadata management and practices for small/medium OA publishers who are currently struggling to produce their metadata to all the various different specifications that each distributing platform requires;
- To help distribute Open Access books, which have been systematically excluded from a book supply chain that was created for closed books;
- To expose quality and first-hand metadata, using industry standards, publicly for anyone to consume.

2. Why is it important for metadata to be open?

Let’s first clarify that data cannot be owned, nor copyrighted. Databases, however, can. Therefore metadata about a book cannot be closed, but you can license an ONIX file. The book metadata supply chain does not often rely on [open APIs](#), but on file exchange. If you license those files you thus control the data flow.

Metadata aggregators collect data from various sources (mainly from the publisher), with which they produce a series of records (ONIX, MARC, etc.) that they then sell to libraries so that the libraries can include the books in their catalogues. This happens to both closed and open books, and while this process may make sense for closed books it really doesn’t for the OA ones — if a book is open, its metadata should be too.

3. Who do you think will use Thoth?

Our main users will be: publishers that want to increase their metadata outputs, archivists and distributors, and libraries and platforms wanting to gather book metadata.

Catalogue

Displaying works 1-10 of 738

A screenshot showing some of the books listed on Thoth

4. How does Thoth fit into the broader COPIM project?

Since Thoth stores book metadata it is becoming an integral part of other COPIM work packages (WPs). Thoth will be used by the Archiving WP to distribute all data needed for book preservation, as well as it will be used by Open Book Collective to provide information to libraries about the type of books their members publish.

5. Do you think Thoth will contribute to 'building structural equity'? If so, how?

Definitely - the fact that our data is exposed publicly and that our code is open source means that all our users have an insurance policy. If we ever got bought by a large commercial entity, anyone could create a new Thoth with exactly the same functionality and data in no time. There is no platform lock-in with Thoth. Everything COPIM is building is designed to be community owned and community governed, not owned by a single company or organisation that would have its own priorities -- and this openness is a safeguard against that eventuality.

So far we have been focusing on building the platform and following open data principles, and we are now in the process of discussing the governance structure we will have. We are building a community of users that can have a say in the direction we follow

and it is clear that this community will be the heart of the governance structure once it is formalised.

6. How did you come to work on Thoth -- have you worked on similar projects in the past?

The idea of Thoth sprang from ScholarLed wanting to create a collective catalogue. The first step was to gather the required data, but we soon realised that coordinating that flow had many similarities to distributing the data to a metadata aggregator, with all that comes with it. However, if we could agree on a shared, generic, data model and used open standards we would be able to achieve not just a catalogue but many other tasks. And how wonderful would it be to offer all this as a service to any other press in need?

The previous project I participated in was HIRMEOS, where we took a very similar approach to Thoth's but with book usage metrics, arguing that the first step towards ethical metrics was to open the data.

7. What is the most interesting aspect of working on Thoth?

Working with standards is a lot of fun, they tell you what you can or cannot include but they are normally subject to free interpretation, so having to provide a mechanism to output a different version of an ONIX file for each platform really tells you that standards are “more what you'd call guidelines than actual rules” .

Thoth logo

8. What is the biggest challenge?

So far the main issue we've faced has been importing back-catalogue data. Thoth's data model is very granular so we can output the data that each platform needs, and most publishers do not store that much data in a structured way but as separate files depending on the requirements of the platforms they normally distribute to.

9. What is your biggest hope for Thoth -- what would be the ideal outcome of the project?

I would love it if librarians found Thoth useful and contributed their cataloguing skills to it. It'd be great if we could build a shared environment that publishers and librarians used together to enrich metadata records and keep them alive -- something similar to Wikidata but specifically for academic books.

10. What other projects happening at the moment do you think will help to develop open access in an equitable way?

We have very good examples within COPIM: I love the pragmatism of [Opening the Future](#) as well as the idealism behind [Open Book Collective](#).

[Find out more about Thoth and the work behind its development here.](#)