Community-led Open Publication Infrastructures for Monographs (COPIM)

Thoth’s Impact: A Series of Interviews with Open Access Publishers

Livy Onalee Snyder\textsuperscript{1} Tobias Steiner\textsuperscript{2}

\textsuperscript{1}punctum books, \textsuperscript{2}Community-Led Open Publication Infrastructures for Monographs (COPIM)

Community-led Open Publication Infrastructures for Monographs (COPIM)

Published on: Apr 07, 2023

URL: https://copim.pubpub.org/pub/cwk9aks2

License: Creative Commons Attribution 4.0 International License (CC-BY 4.0)
The major challenges faced by small and/or new OA publishers in the established supply chain for scholarly books differ notably from those faced by larger publishers. One of these challenges includes specialized title management systems. Thoth, a metadata management system developed by Work Package 5 of the COPIM project, aims to alleviate this challenge and seeks to level the playing field between large and small publishers in the scholarly book market.

To learn more about Thoth’s impact, we conducted a series of interviews with representatives from small, scholar-led OA presses including Jeff Pooley at mediastudies.press, Laura Rodríguez and Luca Baffa at Open Book Publishers, and Vincent W.J. van Gerven Oei at punctum books. Interviewees spoke about their decision to switch their metadata management system to Thoth, how it impacted each press’s workflow, and which of Thoth’s specialized features have been the most useful in their particular setting.

Before migrating to Thoth, many of these smaller OA presses compiled their metadata using spreadsheets and quickly found it limiting as an organization tool. One major limitation included the laborious task of generating multiple output formats, which makes it difficult to cater to the needs of a wider variety of platforms to increase discoverability.

Jeff Pooley, from mediastudies.press, described his experience data distribution before migrating to Thoth, he explains:

“...we were able to at least not lose track of those various numbers and other identifiers. So, it worked fine, but it was hard to get stuff out, especially hard to get it out into formats that were important to various places that these books would land, which include like OAPEN and Project MUSE to take to examples where they wanted this like messy XML that had to be hand crafted basically.”

Yet another obstacle with using spreadsheets for small publishers was the amount of manual labor involved, which in turn proved prone to errors due to the variety of inputs required over time.

At Open Book Publishers, Laura accounted for the time difference it took her to work with spreadsheets versus Thoth. She recalled that processing data and distributing it took her around three to five hours if there weren’t any problems and everything had to be entered manually. Now, Laura’s distribution work is completed in less than half that time using Thoth, so less than two hours. She also notes that the quality of OBP’s data outputs has massively improved.

In addition to expanding publishers’ discoverability, minimizing manual labor and human error, and streamlining workflows, Thoth inspired new ways of thinking about metadata.

As Jeff Pooley described,
“as it [Thoth’s interface] evolved a little bit over time, we started using it and it generated ideas for us to use things like linked open data for location, so you can associate publications with locations and record all of these in granular detail, that was something that we hadn't thought about previously. Even the idea of a dropdown menu with different pre-formatted locations already gave us ideas about reaching out to JSTOR, and things like that….”

For Laura and Luca at Open Book Publishers, Thoth’s API proved to be an incredibly helpful feature. Thoth’s open-source GraphQL API is open to any user to construct and manipulate queries to retrieve existing data from Thoth’s database. This feature is unlike most databases which restrict what users can query to a limited set of retrieval options. Thoth’s GraphQL allows users the agency to retrieve the data they seek, not the server. Laura shares that

“...I always say like I'm a philologist, so I don't know much about technology in that sense, but the one thing that proved to be really useful was the open API. I'm not someone who knows how to code or how to do much technical stuff – but after attending the workshop that was put together by the team, I guess the inner workings of the open API made sense to me, and I now kind of learn more and more about how it works.

Another good thing I got out of working with Thoth is that it helps me to retrieve data about a given book quite easily. At OBP, we have a partnership with the Royal National Institute of Blind People in the UK [...], and they request mass data updates from books – but they requested that information in a very particular set of formats. And with Thoth, now that I know how to move in that open API to retrieve that data from any other output I need, such as other partners that we have that haven't been included in the previously-queried dataset – and even though Thoth doesn't currently have a specific output for that, something that you can just click and export, I can actually work with that data.”

As Laura notes, Open Book Publishers have also made extensive use of Thoth to automate the process of feeding publication-related data into their freshly-launched website. For example, their catalog of books presented on the publisher website is populated directly from Thoth data, and as Javier Arias, Lead Developer at OBP, notes in a recent OBP blog post,

"Thoth has allowed us to reduce the number of databases we had containing book metadata to just one, which we now use for all processes in which metadata was needed, including the actual catalogue of our website."

Likewise, Laura confirms that with their previous website, we had to enter everything manually while also needing to make sure there was consistency in the way things were displayed, and things quickly got fragmentized … with Thoth, I can now devise different bits of information that will appear in different categories, and this helps us keep
things very organized, and we don't have to worry too much about how to display a blurb or how the affiliation of an author is displayed, because everything on OBP's website is done properly and that's very useful because we don't have to worry about keeping content consistent on the webpage."

Javi's blog post provides a concise overview of all the new features they have been able to implement thanks to Thoth, including a customizable, interactive catalogue of all books published by OBP, chapter-level metadata and corresponding chapter landing pages, and an open-source metrics widget building on previous development work that has been done in the context of HIRMEOS – all of which has, in true open-source spirit, been made openly available via a white-label website that other publishers are welcome to reuse.2

Reflecting on the entry barrier of learning how to use Thoth, Laura says

"it is something that I myself didn't think I would actually manage to do because I'm not very tech-savvy. So it's nice to see that Thoth can help people like me who don't have that knowledge or background – it's very user-friendly and you can just make the most of it. So all in all, information is very well organized and you can use it and display it in different places, you can download it, you can share it, you can distribute it very easily."

Vincent also agrees that the UI has benefited punctum because it affords them different ways for organizing, which “definitely [helped them] to gain a lot more reflexive insight into what [their] catalog is.”

Admittedly, and as many a librarian and publishers can surely confirm, navigating the landscape of metadata management can feel intimidating, but Thoth is available to help publishers work through these challenges. Vincent says that

“Thoth's interface may seem intimidating at first, but it certainly is far less intimidating compared to doing the work on your own.”

Vincent reflects, “For me, personally, I've only started to understand the enormous complexity of the metadata landscape because of working on Thoth.” Likewise, Jeff Pooley also acknowledges the complexity of metadata and how mediastudies.press is continuously refining their data.

There is also a Thoth wiki with all of the research conducted for this project, which is openly available to publishers interested in refining their metadata and figuring out what is needed to get one's books disseminated to a variety of channels. Included is an overview of the Thoth platform, a recording of Thoth’s API workshop, and example queries to get publishers started exporting their metadata. As Vincent notes, “Working with the Thoth wiki will also help with the ongoing refinement of your data, making them more complete and making sure that your data is as compatible as possible with as many output formats as possible.” Ultimately, Thoth provides clear information about “what do you actually need as an OA publisher to put into a metadata record in order to make it compatible with the platforms out there.”
Ultimately, through these interviews, Thoth gained key perspectives from small OA book publishers about the problems of getting OA publications into the book supply chain and how Thoth has supported them in the creation of a more sustainable open source, open access metadata management and dissemination system.

See below for the full interviews.

**punctum books | Vincent W.J. van Gerven Oei, director**

The video features an interview with Vincent van Gerven Oei, the co-director of punctum books, about their experience with Thoth, a metadata management system for Open Access presses. Prior to Thoth, punctum books used Excel and Google Sheets to manage metadata, which was labor-intensive and error-prone. Thoth helped them streamline the process and make their publications more discoverable. The migration to Thoth was organic due to the involvement of punctum books in the development of the system. The transition involved a lot of data cleanup, which was made smoother with the help of UCSB Library. There were some gaps in chapter data when they initially switched to Thoth, but they were able to generate the missing metadata through manual labor. Punctum books faced challenges in generating BISAC codes for Google Books due to the transdisciplinary nature of their publications. Overall, the interview highlights the importance of using metadata management systems like Thoth for Open Access presses to increase discoverability and reach a broader audience.

Visit the web version of this article to view interactive content.

**Open Book Publishers | Laura Rodríguez, Marketing and Library Relations and Luca Baffa, Book Production, Digital Product Development and Illustration Management**

The video is an interview with two employees from Open Book Publishers, Luca and Laura, who talk about their roles and how they use Thoth to improve their workflow. Luca is a production worker who typesets books and files metadata, while Laura handles library outreach, marketing, and distribution. They both started using Thoth in 2019, which has helped them automate their metadata input and distribution process, saving them a lot of time and reducing errors. With Thoth, they can trust that the metadata is correct and focus on other tasks. The video offers a case study of how Thoth can improve book publishers’ workflow.

Visit the web version of this article to view interactive content.
mediastudies.press | Jeff Pooley, director

The interview is with the director of mediastudies.press, a small press that publishes Open Access no fee books in the media, film, and communication studies fields. They discuss their experience with metadata and how they recorded metadata in what they thought was a proper way, including tracking things like ISBNs and DOIs using an Airtable, which worked fine but was hard to get stuff out. They then talk about their integration with Thoth, and how it has replaced the Airtable and become their version of record for all of their published books. Finally, they discuss the impact Thoth has had on their outreach and generating ideas for metadata. The interview ends with a discussion of the future possibilities for Thoth.

Visit the web version of this article to view interactive content.

Thoth Case Study Interview | mediastudies.press

Header image credit: Photo by Milad Fakurian on Unsplash

Footnotes
