Open Access, Open Data, Open Archiving: Liberating Metadata Flows across the OA Books Landscape

Javier Arias[1,2,*,†], Rupert Gatti[1,2,4,*,†], Vincent W. J. van Gerven Oei[1,3,*,†], Ross Higman[1,2,*,†], Hannah Hillen[1,*,†], Brendan O'Connell[1,2,*,†], Amanda Ramalho[1,*,†] and Toby Steiner[1,*,†]

[1]*Thoth Open Metadata, https://ror.org/00dyx6f45*

[2]*Open Book Publishers, https://ror.org/04gp2qg91*

[4]*Trinity College, University of Cambridge, https://ror.org/013meh722*

[3]*punctum books, Earth, Milky Way*

Abstract

This poster highlights the importance of integrating open metadata management in open access book publishing. It showcases publishers from a variety of regions across the globe using Thoth Open Metadata to manage and distribute FAIR and open metadata in multiple formats. Doing so, Thoth Open Metadata enables publishers to establish good metadata practice, while also fostering discoverability, sustainability, and accessibility of publishers' valuable contributions to the scholarly record. The implementation of open practices also promotes collaboration with open infrastructures supporting open scholarship.

Keywords

Open metadata,, open access,, open infrastructures,

1. Introduction

Now that open access is rapidly becoming the mainstream mode of publishing scholarly publications, including monographs and edited collections, integrating open metadata management into book production workflows, library cataloging systems, and long-term archiving solutions is gaining importance and urgency.

The poster will showcase the metadata workflows of a diverse sample of independent scholarled and university- and library-based institutional publishers of open access books from across the globe – including presses from the Netherlands, the UK, the US, Latin America, and Africa – through the usage of Thoth Open Metadata, a non-profit open source and community-led platform providing innovative metadata management and distribution solutions tailored to

tackle the problems of getting open access books and chapters into the book supply chain, ensuring their long-term discoverability, sustainability and accessibility.

2. Methods

As part of the Community-Led Open Publishing Infrastructures for Monographs (COPIM) project[1], the contributors have worked to develop a dissemination system tailored to tackle the problems of getting long-form open access works into the book supply chain[2, 3], ensuring their long-term sustainability and accessibility.

Following dedicated research into the underlying topics of dissemination workflows, stakeholders, and metadata formats relevant to open access book publishing[4], an open-source web platform was developed. This was followed by the formal inception of Thoth Open Metadata, a not-for-profit community interest company to ensure a community- and values-led implementation and stewardship of platforms and services that have subsequently been established.

3. Results

With the poster, we showcase how the platform now empowers publishers to create fully open and FAIR metadata (released under a CC0 public domain dedication to enable re-use and remix), which are automatically exported to a variety of different metadata formats required to participate in the wider book supply chain, incl. industry standards such as MARC21[5], MARCXML, ONIX[6] in versions 2.1, 3.0, and 3.1, KBART[7], Crossref XML, BibTeX, json, and csv via Thoth's open APIs (with BIBFRAME[8] and OPDS[9] as additional export formats on the development radar). Additional options provided by Thoth will include the provision of OAI PMH[10] to facilitate discovery by the global repository community.

Utilising the open metadata made available via Thoth, We will also show how metadata and content can then be directly disseminated to a variety of platforms and aggregators including, but not limited to, OAPEN, the Directory of Open Access Books DOAB, the Open Book Collective, JSTOR, Project MUSE, EBSCOhost, ProQuest ExLibris, and OCLC; how DOIs are automatically registered with Crossref for books and chapters alike; how fully open catalogue data can easily be integrated with libraries; and how publications are automatically archived in open repositories such as the Internet Archive and Zenodo through the Thoth Open Archiving Network[11] - a novel, open, transparent and auditable alternative to existing long-term preservation mechanisms.

We will also present steps taken towards fostering the uptake of an emerging open ecosystem of like-minded infrastructures such as the Public Knowledge Project's Open Monographs Press (OMP), Janeway, the Open Book Collective, Crossref, OAPEN/DOAB, and the OPERAS Metrics service provided by the OPERAS network – all of which work together to promote open (meta)data for OA books and chapters, and are in the midst of establishing a network of platforms and services that embody the core tenets of open source, open data – and open scholarship.
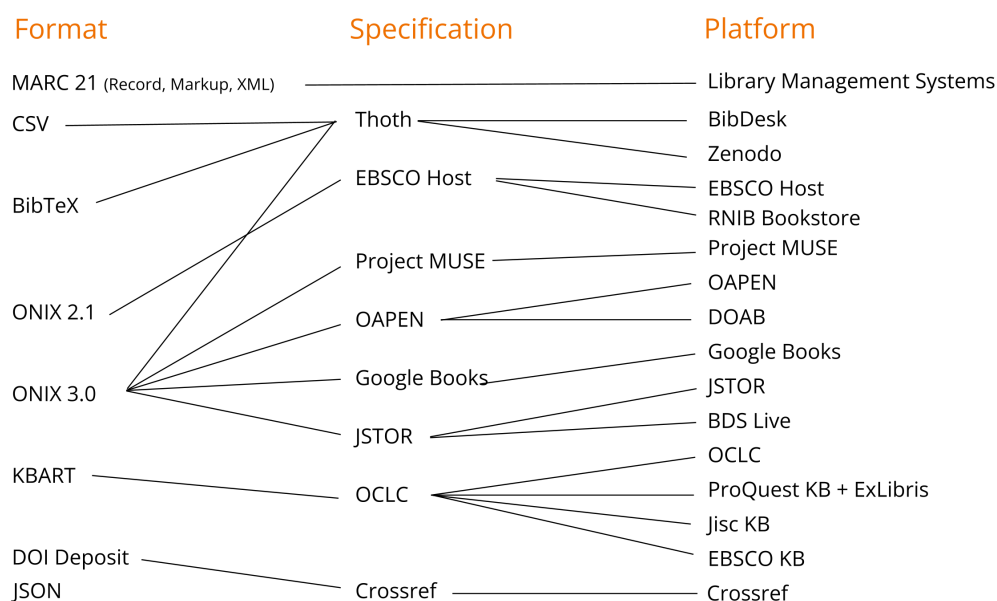
Format	Specification	Platform

Figure 1: Schematic overview of book metadata formats, different sub-specifications of those standard formats, and platforms using those specifications. Source: Thoth Open Metadata, 2025 [CC BY 4.0]

References

[1] T. Steiner, J. Adema, Community-Led Open Publication Infrastructures for Monographs: Final Report, Technical Report, Zenodo, 2023. doi:10.5281/zenodo.7961527.

[2] G. Stone, OA monographs discovery in the library supply chain: draft report and recommendations, Technical Report, Jisc, 2018. URL: https://web.archive.org/web/20240418110802/https://scholarlycommunications.jiscinvolve.org/wp/2018/10/25/oa-monographs-discovery-in-the-library-supply-chain-draft-report-and-recommendations/.

[3] M. Clarke, L. Ricci, OA Books Supply Chain Mapping Report, Technical Report, Zenodo, 2021. URL: https://zenodo.org/record/4681725. doi:10.5281/zenodo.4681725.

[4] G. Stone, R. Gatti, V. W. van Gerven Oei, J. Arias, T. Steiner, E. Ferwerda, WP5 Scoping Report: Building an Open Dissemination System, Technical Report, Community-led Open Publishing Infrastructures for Monographs (COPIM), 2021. doi:https://doi.org/10.21428/785a6451.939caeab.

[5] MARC 21 Format for Bibliographic Data: Table of Contents, 2000. URL: https://www.loc.gov/marc/bibliographic/.

[6] ONIX for books: overview, 2009. URL: https://www.editeur.org/83/Overview/.

[7] Knowledge Bases And Related Tools (KBART), 2020. URL: https://www.niso.org/standards-committees/kbart.

[8] K. Tharani, Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with BIBFRAME, Information Technology and Libraries 34 (2015) 5–19. doi:10.6017/ital.v34i1.5664.

[9] H. Gardeur, OPDS Catalog 2.0, 2017. URL: https://drafts.opds.io/opds-2.0.html.

[10] C. Lagoze, H. Van de Sompel, M. Nelson, S. Warner, Open Archives Initiative - Protocol for Metadata Harvesting - v.2.0, 2002. URL: https://www.openarchives.org/OAI/openarchivesprotocol.html.

[11] R. Higman, G. Cole, R. Gatti, J. Arias, T. Steiner, P. Stokes, P. Wheatley, M. Barnes, C. McGann, Putting the 'Open' in 'Thoth Open Archiving Network', Copim (2025). URL: https://copim.pubpub.org/pub/thoth-open-archiving-network. doi:10.21428/785a6451.76e96572.