

# Open Data, Open Access, Open Source: Liberating Metadata for Books and Chapters

Toby Steiner, Thoth Open Metadata

HMC Conference: Metadata in Action

April 28, 2025



## Open Science & book publishing

Within the context of Open Science, book publishing has traditionally lagged behind in the adoption of open data and corresponding practices.





## Dissemination workflows for journals ...



Journals



<JATS>  
<TEI>

xml metadata

*[slight oversimplification]*



## Dissemination workflows for books ...



Books



# FORMAT

# SPECIFICATION

# PLATFORM

MARC21 (Record, Markup, XML)

LMS & union catalogues (e.g. Jisc NBK)

CSV

Thoth

BibDesk, Zotero

BibTeX

Zenodo

ONIX 2.1

EBSCOHost

EBSCOHost

ONIX 3.0

Project MUSE

Project MUSE

OAPEN

OAPEN

Google Books

Google Books

JSTOR

JSTOR

ONIX 3.1

Thoth

SciELO Livros

KBART

OCLC

Amazon

OCLC

ProQuest KB + ExLibris

XML for DOI Deposit

Crossref

EBSCO KB

Crossref



# INTERNATIONAL METADATA RECOMMENDATIONS AND PLATFORM-SPECIFIC REQUIREMENTS FOR OPEN ACCESS BOOKS AND CHAPTERS

Report, Jan 2026



- ★ Review of policy recommendations
- ★ Overview of platforms' requirements
- ★ Proposition of a two-tiered framework of essential and desirable metadata for OA books and chapters.

✓ **Publishers:** offers a well-defined way to prioritise metadata work and retain full control over their data and operations.

✓ **Libraries:** provides a clearer basis for understanding what good metadata for OA books looks like, while also highlighting prevailing issues with the existing commercial supply chain.

✓ **Funders & Policymakers:** outlines why book publishing needs community-led approaches distinct from journals, and why metadata quality and openness are prerequisites for meaningful policy implementation, monitoring, and evaluation.



*Read the report*

Steiner, T., Arias, J., Bennett, M., Booth, E., Edmunds, J., Gatti, R., Higman, R., Hillen, H., Laakso, M., Nason, M., O'Connell, B., Pogačnik, A., Rabar, U., Ramalho, A., Stone, G., van Gerven Oei, V. W. J., & Wake Hyde, Z. (2026). International Metadata Recommendations, and Platform-Specific Requirements for Open Access Books and Chapters (1.0). Thoth Open Metadata. <https://doi.org/10.5281/zenodo.18173982>



# Key findings

1

Metadata **openness** and **interoperability** are critical for inclusion of books (open access and non-open access) in research assessment, policy development and monitoring, and equitable dissemination.

2

**Legal and accessibility frameworks** (e.g. EU GPSR, Americans With Disabilities Act) increasingly **necessitate the provision of richer metadata**, including funder data, licensing, and accessibility fields, flowing through degradation-free metadata supply chains.

3

**Fragmentation of metadata standards and formats** (ONIX, MARC, KBART, etc) in use across stakeholders and intermediaries active in the supply chain impedes visibility and reuse of open access books, while also impinging on the reuse of the metadata itself

4

The highly fragmented nature of the book supply chain – with an existing multiplicity of distribution channels and correspondingly fragmented metadata requirements that is substantially different to that of journal publishing – means that **policy and technical maturity for open access books has traditionally lagged behind that of journals.**

# Recommendation: Two-Tiered Metadata Framework for Books & Chapters

1

**Essential:** key bibliographic and access data including information on title and subtitle, contributors, copyright holders, subjects classifications, landing page and full-text URLs and/or DOIs at book and chapter level, licence, publisher details, publication date. Releasing metadata into the public domain and making this explicit in the metadata record (e.g. via a CC0 dedication) to facilitate easy re-use e.g. across libraries. [relevant in the context of e.g. the [Barcelona Declaration on Open Research Information](#)]

2

**Desirable:** includes elements such as a fuller set of Persistent Identifiers (ORCID, ROR), usage of a broader set of controlled vocabularies, multilingual metadata, abstracts, cover images, tables of contents, and funder details.

Source: International Metadata Recommendations, and Platform-Specific Requirements for Open Access Books and Chapters (1.0). Thoth Open Metadata. <https://doi.org/10.5281/zenodo.18173982>



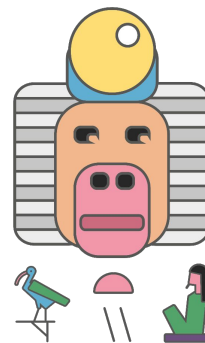
## Essential and desired metadata for OA books and chapters

Essential and desired met...			
Field name	Field desc	Field type	Status
<b>Essential</b>			
work type	Work Type (e.g. Edited Collection, Monograph)	Controlled Vocabulary, with the option to add free text	Required
work status	Work Status (e.g. published, forthcoming)	Controlled Vocabulary	Required
title	Title	Free text - separate fields for title and subtitle	Required
subtitle	Subtitle	Free text - separate fields for title and subtitle	Required
contributor names	Author/Editor/other contributor name(s) & ORCIDs (where available)	Free text, divided into Given Name(s) and Family Name(s)	Required
contribution type	role description	Controlled Vocabulary, implementing CReDIT, with the option to include additional roles not covered by the CReDIT taxonomy to accommodate needs from the Social Sciences and Humanities	Required
license	Open License (e.g. Creative Commons), All rights reserved	Controlled List, incl. machine-readable URL of the license	Required
publication date	Publication Date in ISO format	Following ISO standard	Required
place of publication	Place of Publication	Free text	Required
publication format	clear denomination of publication format (PDF, ePub, HTML, etc.)	Controlled Vocabulary, with the option to add free text	Required
publisher / imprint (if)	Publisher name, URL, ROR (encouraged)	Free text plus ROR PID	Required
DOI	DOI URL	Well-formatted DOI following best practice	Required
landing Page URL	DOI and/or Landing Page URL (where DOI is not used)	URL. Highly relevant in the context of collating usage statistics across multiple OA platforms	Required
full-text URL	URL of the full text document (link to the publication file, e.g. PDF, epub)	URL. Highly relevant in the context of collating usage statistics across multiple OA platforms	Required
page range (for chapter)		free text (to enable Roman numerals and other variations next to Arabic numbers)	Required
language	to indicate language(s) used in the publication	Controlled list following ISO standard	Required
<b>Desired</b>			
ISBN	Book-level ISBNs (ideally one ISBN per book format)	standardised input, ideally via ISBN check upon entry	Required
ISSN	for serial publications	standardised	Required
subject codes / heading	via controlled vocabularies such as Thema or VIAF, free-text keyword option also enabled	Multiple controlled vocabularies, or free-text keyword as fallback	Required
metadata record license	Explicit mention of the license applied to a given metadata record, to enable downstream re-use	Standardised CC0 or other public domain dedication in both human- and machine-readable form	Required
cover image URLs	URL where the cover image can be accessed	URL	Recommended
table of contents	in free-text, Markdown, or JATS XML	Free text, preformatted (see desc.)	Recommended
abstract	in free-text, Markdown, or JATS XML	Free text, with possibility to record in multiple languages	Recommended
references / works cited	on the level of the book and/or individual chapters	CSV (can be a list of referenced DOIs at minimum)	Recommended
Contributor PIDs	ORCID where available; ISNI as alternative	ORCID number or full URL to profile	Recommended
Affiliated institution	ROR	ROR ID or full URL to profile	Recommended
Funder information	ROR (to supersede Funder Registry)	ROR ID or full URL to profile	Recommended
Location URLs	Content URLs from the various platforms that OA content has been disseminated to	URL	Recommended
	ARK as alternative PID	ARK ID or full URL	Optional
	ISNI as alternative PID	ISNI ID or full URL	Optional
Multilingual versions of metadata	Enabling publishers from different regional contexts to record metadata in other languages than English		Recommended
Peer-review status	Indicate the type of peer review applied to a given title.	Could be using PRISM framework, or other ways to denote peer-review status.	Optional
Crossmark	(for publishers using Crossref)	See <a href="#">Crossref guidance</a> for more details	Optional

Source: Overview of International Aggregators' Metadata Requirements for Books. <https://doi.org/10.5281/zenodo.18173982>

# Implementation? Thoth can help!

- Nonprofit registered in the UK, community-owned and -managed
- Codebase available under open license on Github (<https://github.com/thoth-pub/>)
- Made for Open Access, also supports non-OA titles
- High-quality metadata on book and chapter levels
- Multiple output formats and specifications for free, including MARC, ONIX, KBART, Crossref XML, and BibTeX
- Open APIs, CC0 licensed
- Services to facilitate automated dissemination, DOI management, archiving, usage statistics and website & catalogue hosting - all via open-source solutions!



**THOTH  
OPEN  
METADATA**

# Oasis

- Our **free** self-service product
- Manage and export metadata in multiple open formats (ONIX, JSON, MARC, KBART, Crossref XML)



# Obelisk

- Everything in Thoth Oasis
- Crossref membership
- Automated DOI deposit
- Automated distribution and archiving
- File hosting on superfast CDN



# Sphinx

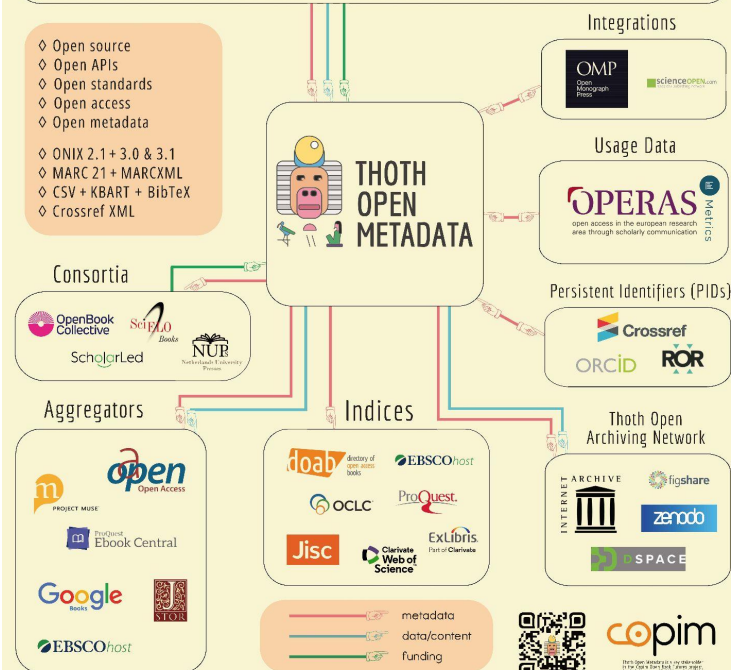
- Everything in Thoth Obelisk
- Dedicated publisher usage metrics dashboard
- Usage metrics book widgets



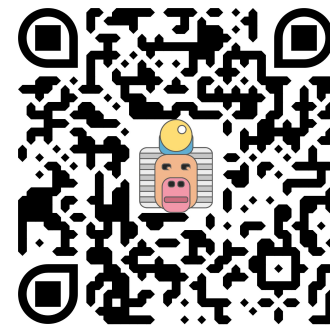
# Pyramid

- Everything in Thoth Sphinx
- Website and catalogue hosting on superfast CDN
- Customizable front-end





# A SHOWCASE OF OPEN DATA WORKFLOWS IN THE CONTEXT OF OA BOOK PUBLISHING



<https://doi.org/10.5281/zenodo.17340814>

# Enabling Good Metadata Practice



# JUST DOI IT.



THOTH  
OPEN  
METADATA

- E.g. integration of multilingual book- and chapter-level data, PIDs (e.g. ORCIDs, RORs, DOIs), controlled vocabularies (e.g. Thema), references, etc. ...
-  in line with international recommendations for ebooks such as NISO, Diamond OA Standard, AG Univerlage Quality Standard.
- **!** Enabling publishers to meet new requirements mandated by recent EU legislation (General Product Safety Regulation (GPSR), EU Deforestation Act, and Accessibility Act)



*Metadata, standards and platforms for open access books: good metadata practice with Thoth.* <https://doi.org/10.5281/zenodo.15048076>

# Building Capacity

Though as the source for high-quality open metadata that flows across open infrastructures

Example: **Crossref Participation Report**

<https://www.crossref.org/members/prep/>



Richer metadata makes content useful.  
Make sure your work can be found.

Open Book Publishers

2,573  
Total registered  
content items

Content type: Book chapters

Preprints 63 Books 158 Book chapters 2,352

Book chapters

Search by title

Current content

References

65%

Abstracts

85%

ORCID iDs

55%

Affiliations

79%

ROR iDs

78%

Crossmark enabled

100%

Text mining URLs

100%

License URLs

100%

Similarity Check URLs

100%





# Thank you.

Toby Steiner  
COO, Thoth Open Metadata

 [info@thoth.pub](mailto:info@thoth.pub)

 <https://thoth.pub>

# Like what we do?

## Here is how you can help!

1

Support our initiative via



OpenBook  
Collective

2

Adopt **open practices** to make open access books more present in your institution

3

Integrate fully **open metadata streams** within your publisher & library context

4

Engage with us to establish **open and transparent ways of archiving OA books**

5

And generally, **join us!** Because there's ..

No **open access**  
Without **open infrastructure.**

