Open Metadata Management and Dissemination with Thoth

Vincent W.J. van Gerven Oei @ontakragoueke
No open access without open infrastructure.
Open Metadata for OA Publishing

- Smaller, non-profit, and library-based OA publishers often struggle with metadata management and dissemination
- Commercial solutions are expensive/not tailored to OA books > metadata often managed in custom spreadsheets
- Dissemination channels require different output formats according to different specifications > technical knowledge required is often absent, leading to lower discoverability
- OA book metadata from commercial vendors often of low(er) quality
- Metadata often not CC0 licensed
Open Metadata for OA Publishing

- Thoth: Open Metadata Management & Dissemination
  - Open source, open documentation on wiki
<table>
<thead>
<tr>
<th>Folder</th>
<th>Description</th>
<th>Last Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>.github</td>
<td>Update build dependencies</td>
<td>last month</td>
</tr>
<tr>
<td>src</td>
<td>Allow setting actix keep alive via cli</td>
<td>2 months ago</td>
</tr>
<tr>
<td>thoth-api-server</td>
<td>Bump v0.11.0</td>
<td>last week</td>
</tr>
<tr>
<td>thoth-api</td>
<td>Bump v0.11.0</td>
<td>last week</td>
</tr>
<tr>
<td>thoth-app-server</td>
<td>Bump v0.11.0</td>
<td>last week</td>
</tr>
<tr>
<td>thoth-app</td>
<td>Bump v0.11.0</td>
<td>last week</td>
</tr>
<tr>
<td>thoth-client</td>
<td>Bump v0.11.0</td>
<td>last week</td>
</tr>
<tr>
<td>thoth-errors</td>
<td>Bump v0.11.0</td>
<td>last week</td>
</tr>
<tr>
<td>thoth-export-server</td>
<td>Bump v0.11.0</td>
<td>last week</td>
</tr>
<tr>
<td>.dockerignore</td>
<td>Ignore dev files</td>
<td>9 months ago</td>
</tr>
<tr>
<td>.env.example</td>
<td>Set docker environment</td>
<td>2 years ago</td>
</tr>
<tr>
<td>.gitignore</td>
<td>Ignore dev files</td>
<td>9 months ago</td>
</tr>
<tr>
<td>CHANGELOG.md</td>
<td>Update changelog</td>
<td>last week</td>
</tr>
<tr>
<td>CODE_OF_CONDUCT.md</td>
<td>Use COPIM code of conduct</td>
<td>2 years ago</td>
</tr>
<tr>
<td>Cargo.lock</td>
<td>Bump v0.11.0</td>
<td>last week</td>
</tr>
<tr>
<td>Cargo.toml</td>
<td>Bump v0.11.0</td>
<td>last week</td>
</tr>
<tr>
<td>Dockerfile</td>
<td>Update build dependencies</td>
<td>last month</td>
</tr>
<tr>
<td>Dockerfile.dev</td>
<td>Update build dependencies</td>
<td>last month</td>
</tr>
</tbody>
</table>

**About**

Metadata management and dissemination system for Open Access books

- `thoth.pub`
- `graphql`
- `rust`
- `metadata`
- `books`
- `webassembly`
- `openaccess`
- `bibliographic-data`

**Releases**

- **79**
- **Latest**
- **v0.11.0** last week

**Contributors**

[Profile Picture] [Profile Picture] [Profile Picture] [Profile Picture] [Profile Picture]
The following wiki lists all output formats, platforms, and other stakeholders that relate to the development of open source metadata management system Thoth. Those wikis marked with an asterisk (*) are identified in the WP5 Scoping Report and are indexed as an issue in the Thoth Project. Those wikis marked with a caret (**) have been successfully implemented in Thoth.

The transfer protocols and file specifications listed in this wiki are those which have current or future support by Thoth. Any updates are welcome at wp5@copim.ac.uk.

The Thoth User Manual, for publishers and other creators of metadata records in Thoth, can be found here.

**Data and Metadata**

**Data Formats**

Data formats are the formats in which the textual content of an object are commonly digitally encoded. The Thoth project is agnostic as to which data format is used.

- AZW3
- EPUB
- HTML
- KFX
- MOBI
- PDF
- XML
Open Metadata for OA Publishing

- Thoth: Open Metadata Management & Dissemination
  - Open source
  - Tailor-made for Open Access book metadata + persistent identifiers
<table>
<thead>
<tr>
<th>Locations</th>
<th>Full Text URL</th>
<th>Platform</th>
<th>Canonical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landing Page</td>
<td><a href="https://muse.jhu.edu/book/76457">https://muse.jhu.edu/book/76457</a></td>
<td>Project MUSE</td>
<td>No</td>
</tr>
<tr>
<td>Landing Page</td>
<td><a href="https://library.oapen.org/handle/20.500.12657/25553">https://library.oapen.org/handle/20.500.12657/25553</a></td>
<td>OAPEN</td>
<td>No</td>
</tr>
<tr>
<td>Landing Page</td>
<td><a href="https://punctumbooks.com/titles/the-non-library/">https://punctumbooks.com/titles/the-non-library/</a></td>
<td>Other</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Contributed to: Love Don't Need a Reason, from: punctum books

Given Name
Matthew J.

Family Name
Jones

Full Name
Matthew J. Jones

ORCID
https://orcid.org/0000-0002-9707-6655

Website
Website

Save
Open Metadata for OA Publishing

- Thoth: Open Metadata Management & Dissemination
  - Open source
  - Tailor-made for Open Access book metadata
  - Multiple output formats and specifications for 10+ platforms
<table>
<thead>
<tr>
<th>Format</th>
<th>Specification</th>
<th>Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARC 21 (Record, Markup, XML)</td>
<td></td>
<td>Library Management Systems</td>
</tr>
<tr>
<td>CSV</td>
<td>Thoth</td>
<td>BibDesk</td>
</tr>
<tr>
<td>BibTeX</td>
<td>EBSCO Host</td>
<td>Zenodo</td>
</tr>
<tr>
<td>ONIX 2.1</td>
<td>Project MUSE</td>
<td>EBSCO Host</td>
</tr>
<tr>
<td>ONIX 3.0</td>
<td>OAPEN</td>
<td>RNIB Bookstore</td>
</tr>
<tr>
<td></td>
<td>Google Books</td>
<td>Project MUSE</td>
</tr>
<tr>
<td></td>
<td>JSTOR</td>
<td>OAPEN</td>
</tr>
<tr>
<td>KBART</td>
<td></td>
<td>DOAB</td>
</tr>
<tr>
<td>DOI Deposit</td>
<td>JSTOR</td>
<td>Google Books</td>
</tr>
<tr>
<td>JSON</td>
<td>OCLC</td>
<td>JSTOR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BDS Live</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OCLC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ProQuest KB + ExLibris</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jisc KB</td>
</tr>
<tr>
<td></td>
<td>Crossref</td>
<td>EBSCO KB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Crossref</td>
</tr>
</tbody>
</table>
REQUEST

PATH PARAMETERS

*specification_id  string  onix_3.0:oapen

*publisher_id  uuid  9c41b13c-cecc-4f6a-a151-be4682915ef5

API Server  https://export.thoth.pub
Authentication  Not Required

Response Status: (CORS or Network Issue)

RESPONSE  RESPONSE HEADERS  CURL

`curl -X GET "https://export.thoth.pub/specifications/onix_3.0%3A%3Aoapen/publisher/9c41b13c-cecc-4f6a-a151-be4682915ef5"`

RESPONSE

200

OK
The Getty Fiend
Ken White
Earth, Milky Way: punctum books, 2024
10.53288/0534.1.00
Open Metadata for OA Publishing

- Thoth: Open Metadata Management & Dissemination
  - Open source
  - Tailor-made for Open Access book metadata
  - Multiple output formats and specifications for 10+ platforms
  - CC0 licensed
Open Metadata for OA Publishing

- Thoth: Open Metadata Management & Dissemination
  - Open source
  - Tailor-made for Open Access book metadata
  - Multiple output formats and specifications for 10+ platforms
  - CC0 licensed
  - Integration with third-party platforms
Collective Catalogue
The collective catalogue displays books that are in the Thoth system, which is an open source metadata management service used by many of the publishers supported by the Open Book Collective.
HISTORY OF THE BOOK

TOUCHING PARCHMENT: HOW MEDIEVAL USERS RUBBED, HANDLED, AND KISSED THEIR MANUSCRIPTS: VOLUME 1: OFFICIALS AND THEIR BOOKS
Kathryn M. Rudy

The Medieval book, both religious and secular, was regarded as a most precious item. The traces of its use through touching and handling during different rituals such as oath-taking, is the subject of Kathryn Rudy's research in Touching Parchment.

TEXTBOOKS AND LEARNING GUIDES • SCIENCE

INTRODUCTION TO SYSTEMS BIOLOGY: WORKBOOK FOR FLIPPED-CLASSROOM TEACHING
Thomas Sauter • Marco Albrecht

This book is an introduction to the language of systems biology, which is spoken among many disciplines, from biology to engineering. Authors Thomas Sauter and Marco Albrecht draw on a multidisciplinary background and evidence-based learning to facilitate the understanding of biochemical networks, metabolic modeling and system dynamics.
ScienceOpen and Thoth join forces on book metadata

April 13, 2023 • Author: Nina Tacheke

To increase the discoverability and visibility for open books with richer and more interoperable metadata, ScienceOpen is happy to announce a new collaboration with Thoth, an open metadata management and dissemination platform. With this collaboration, open access book publishers and the entire book publishing ecosystem will benefit from a more interconnected digital publishing environment.

When we first started working with books and aggregating publicly available Crossref records via book DOIs to the ScienceOpen citation index, we quickly learned that the potential of Crossref records as actual metadata source was hardly recognized making for rather basic publicly available records. Rich metadata was available mostly in ONIX, but while being the default format for distribution channels for book sellers and libraries, there tend to be data gaps in the ONIX information required for best practice usage in digital environments—speaking here esp. about persistent identifiers which are essential to keep records unified, connected and discoverable throughout the publication’s lifecycle.

Thanks to this new collaboration with Thoth, publishers will now be able to create a record on BookMetaHub, send it to Thoth for further enhancement, and then gain access to additional output formats for future export. We are also working on sending data created on Thoth to the BookMetaHub, so stay tuned!

In response to the current data gap in the book publishing landscape between rich publication records used by book vendors within their sales channels and commonly available data, we are putting together two innovative and complementing digital infrastructures that will support publishers to share their works even further throughout the entire ecosystem—by offering simple DIY-tools for the creation and maintenance of metadata that
OPEN MONOGRAPH PRESS

The full, end-to-end solution for publishing books. Publish your monographs and edited volumes with full metadata for worldwide dissemination and discovery. Manage submissions and conduct internal and external reviews in a single, intuitive workflow.
OAPEN promotes and supports the transition to open access for academic books by providing open infrastructure services to stakeholders in scholarly communication. We work with publishers to build a quality-controlled collection of open access books and provide services for publishers, libraries, and research funders in the areas of hosting, deposit, quality assurance, dissemination, and digital preservation.
Open Metadata for OA Publishing

- Thoth: Open Metadata Management & Dissemination
  - Open source
  - Tailor-made for Open Access book metadata
  - Multiple output formats and specifications for 10+ platforms
  - CC0 licensed
  - Integration with third-party platforms
Thoth Archiving Network

• Network of institutional repositories hosting OA books published by small, scholar-led, and library-based OA publishers without existing archiving or preservation arrangements

• Protection against complete loss of their catalogue in case they cease to operate

• Participating institutional repositories can specify which publishers, subjects/research areas, and the amount of content deposited

• Proof of concept for Figshare, Eprints, and Internet Archive; Dspace in process
Thoth Archiving Network

Based on the recommendations set out in the WPS Scoping Report, the development of open metadata management and dissemination system Thoth has been going ahead, and we're now at the point that two scholar-led open-access publishers, punctum books and Open Book Publishers, are using Thoth for their day-to-day metadata management. The next step in our process will be to create a variety of metadata output formats, such as different flavors of ONIX, MARCXML, and BibTeX, which will allow users of Thoth to...

More...
Thoth Long-Term Sustainability

• Library subscriptions via the Open Book Collective
• Thoth Plus
<table>
<thead>
<tr>
<th>Free</th>
<th>Plus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ingest</strong></td>
<td></td>
</tr>
<tr>
<td>• API</td>
<td>• Metadata Services</td>
</tr>
<tr>
<td>• UI</td>
<td>• Formatting</td>
</tr>
<tr>
<td>• ONIX</td>
<td>• Backlist ingest</td>
</tr>
<tr>
<td></td>
<td>• Validation/Enhancement</td>
</tr>
<tr>
<td><strong>Export</strong></td>
<td></td>
</tr>
<tr>
<td>• GraphQL API</td>
<td>• Distribution Services</td>
</tr>
<tr>
<td>• Export API</td>
<td>• Curated (meta)data delivery</td>
</tr>
<tr>
<td></td>
<td>• Archiving/Preservation</td>
</tr>
<tr>
<td></td>
<td>• Support/Training</td>
</tr>
</tbody>
</table>
Thoth Long-Term Sustainability

- Library subscriptions via the Open Book Collective
- Thoth Plus
  - DOI Registration (CrossRef)
  - Content & Metadata Platforms (OAPEN or Zenodo, JSTOR*, Project MUSE*, RNIB Bookshare, Library Genesis, Google Books, EBSCOHost*, ProQuest*, Clarivate WOS Book Citation Index)
  - Metadata Platforms (DOAB, Jisc KB, ProQuest Ebook Central, EBSCO KB, OCLC KB)
  - Archive Platforms (Thoth Archiving Network)
Next Steps for Thoth

- Automated metadata dissemination to platforms
Simplifying Dissemination Pipelines

Metadata entry → XML

Cloud68 .co

PDF URL → PDF

p. punctumbooks

emails → XML

downloads → PDF
Simplifying Dissemination Pipelines

Metadata entry

XML

punctumbooks

uploads

XML
PDF
JPG

JSTOR
Simplifying Dissemination Pipelines
Next Steps for Thoth

• Automated metadata dissemination to platforms
• Update of UX
• Quality control via “traffic light” system
• Improving bibliodiversity – multilingual UX and content (SciELO)
• Integration with PKP’s OMP, COKI’s Book Analytics Dashboard, OpenAlex, WikiData
• Expansion of Thoth Archiving Network
• Other use cases: universal catalog, Metadata Broker (OABUDT)
Resources

Thoth
https://thoth.pub/

Open Metadata in Thoth
https://copim.pubpub.org/pub/open-metadata-thoth/

Introducing Thoth Plus
https://copim.pubpub.org/pub/thoth-plus-service-model/
Get in Touch

Tobias Steiner
Thoth
info@thoth.pub

Vincent W.J. van Gerven Oei
punctum books
vincent@punctumbooks.com