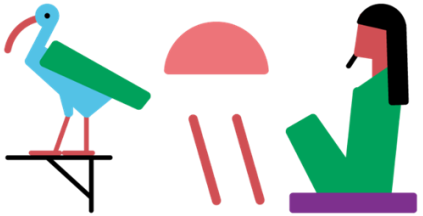
# Thoth Archiving Network

## Supporting Small and Scholar-led Publishers with Repository-Led Preservation of OA Books


Dr Miranda Barnes
Loughborough University
Research Associate in Archiving & Preserving Open Access Books

Dr Gareth Cole
Loughborough University

Toby Steiner
Product Manager & COO
Thoth Open Metadata

THOTH
Open Metadata.

# A bit of context ...

## Open Book Futures & the COPIM Project

# Open Book Futures & COPIM

- Two projects funded by Arcadia and Research England
  - COPIM - 2019-2023; OBF - 2023-2026
- Remove hurdles preventing new and existing open access book initiatives from adopting open access workflows
  - E.g. by building open-source, community-based infrastructures that support the publication of open access books, such as Thoth.
- Showcase alternative (non-BPC) business models that incorporate infrastructural innovations and/or cost-reductions through streamlined operating processes, production workflows and economic efficiencies
- OBF will deepen and accelerate the work completed on COPIM
  - Main outputs: Opening the Future; Open Book Collective; Thoth

# Work Package 7: Archiving & Preservation

- Scholarship needs to be preserved in order to be available long term
- Preservation was and is fully embedded in COPIM and OBF
- Don’t want to reinvent the wheel
  - BUT - needs to be straightforward and where possible use existing technology and workflows
- We saw preservation and archiving as another form of dissemination
  - Publishers send material to dissemination platforms so why not see preservation as the same?
  - Keep it simpler for publishers and systems
  - Automate where possible
- Opportunity to use Thoth
  - Had metadata and links to files
  - Had processes to send these to places in different formats

# Thoth

## Open Metadata Management System

# Issues in OA book publishing: Dissemination

**WP5 Scoping Report: Building an Open Dissemination System**

*by Graham Stone, Rupert Gatti, Vincent W.J. van Gerven Oei, Javier Arias, Tobias Steiner, and Eelco Ferwerda*

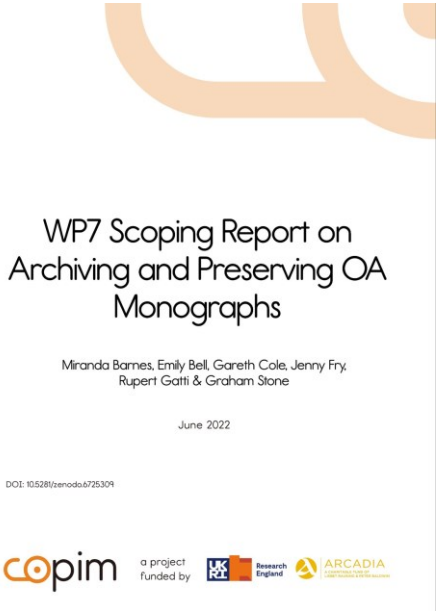
*Published: Apr 21, 2021*

doi: 10.21428/785a6451.939caeab

- 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 > limits downstream reuse & remix

# Issues in OA book publishing: Archiving




doi: 10.5281/zenodo.6725309

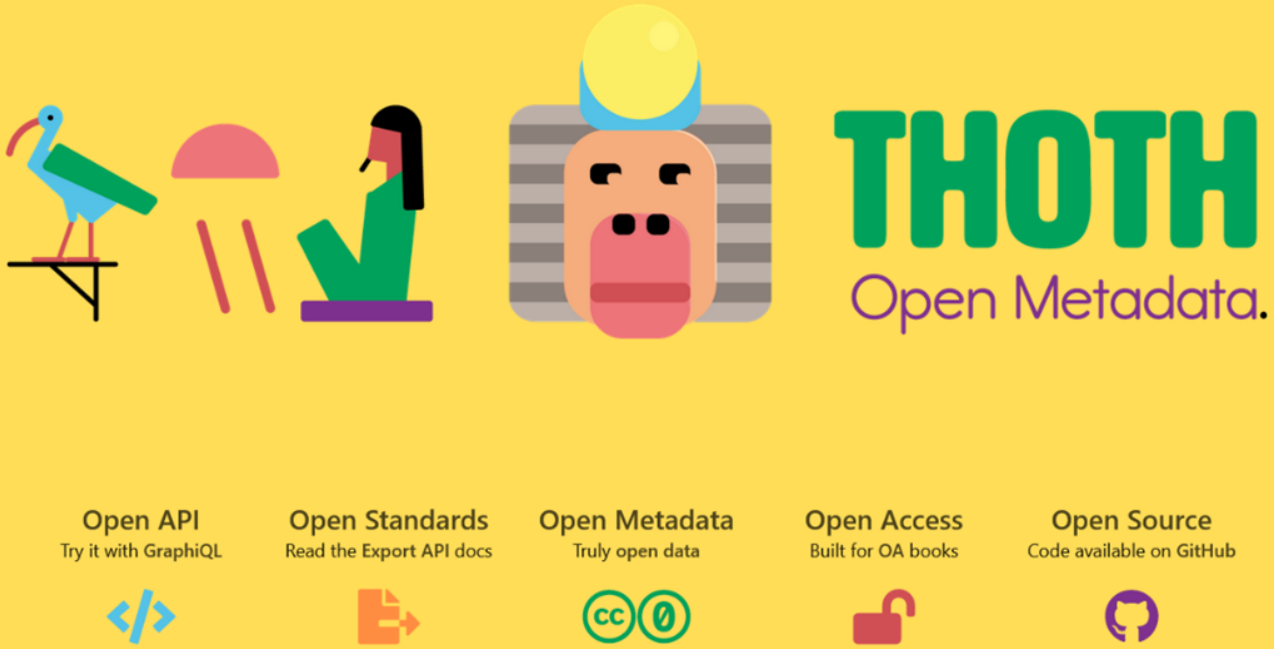
- Smaller, non-profit, and library-based OA publishers often lack expertise & resources with regards to archiving of their outputs
- Existing solutions are expensive and remain difficult to reach for small presses
- As a recent study dissecting the archiving and preservation status of Open Access books suggests, there is

*"reason for concern for the long tail of OA books distributed at thousands of different web domains as these include volatile cloud storage or sometimes no longer contained the files at all"*

Laakso, 2023

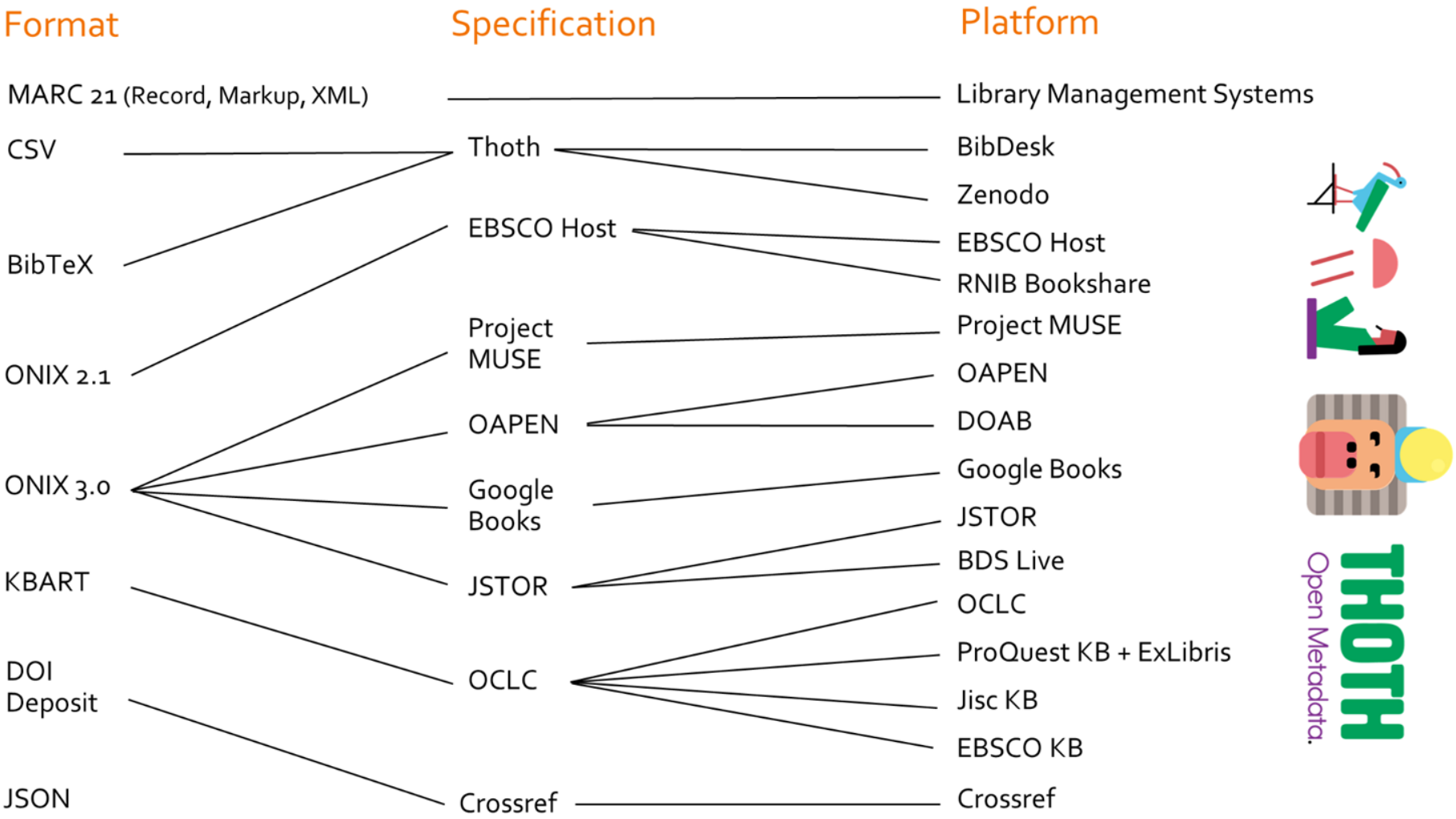
https://thoth.pub

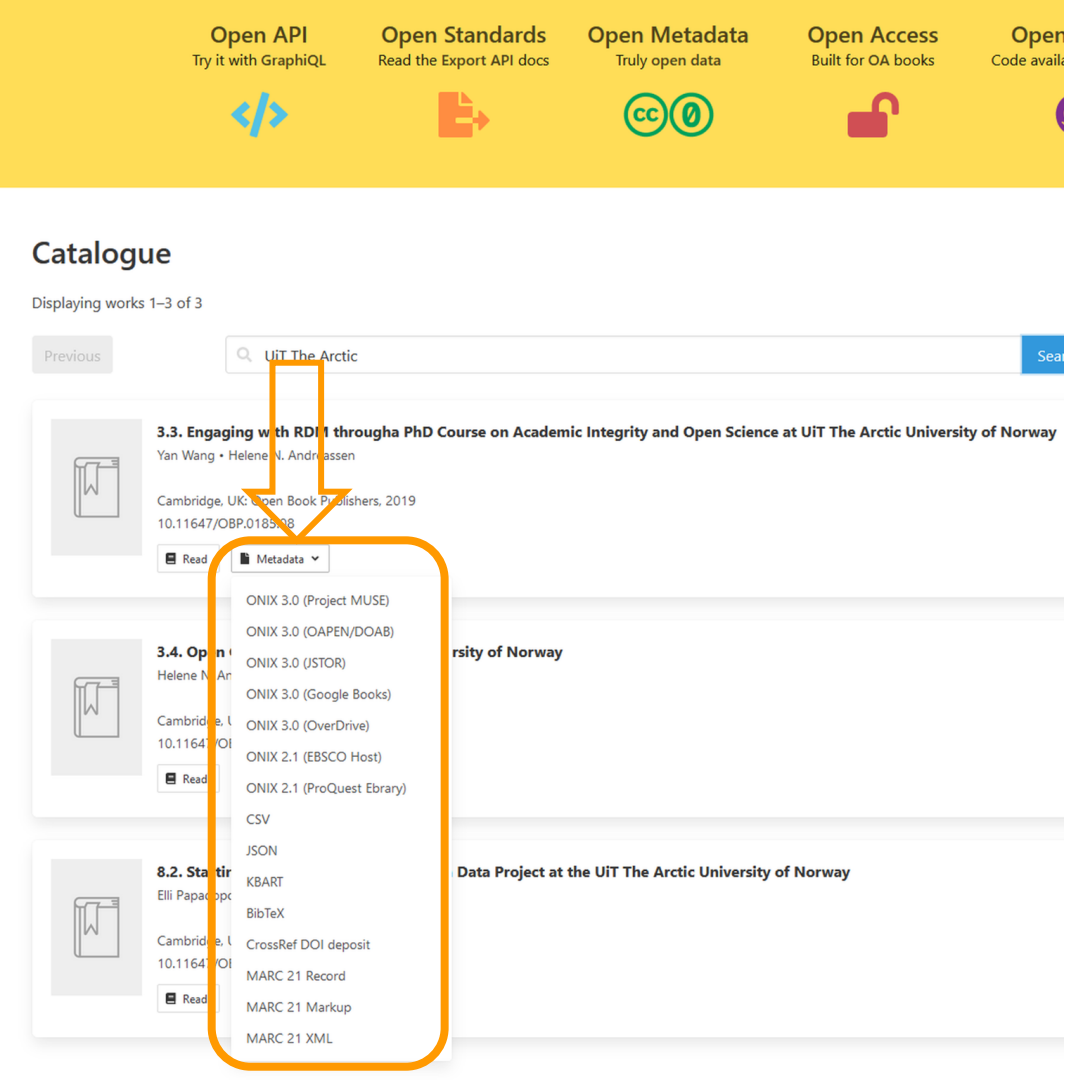
Non-profit CIC



# Metadata Management for OA books: Ingest

- Direct data entry via UI
- Import existing metadata from other platforms (e.g. ScienceOpen) & local title management systems via REST API, or structured data (e.g. ONIX)


Format
Specification
Platform
MARC 21 (Record, Markup, XML)
CSV
BibTeX
ONIX 2.1
ONIX 3.0
KBART
DOI Deposit
JSON
Thoth
EBSCO Host
Project MUSE
OAPEN
Google Books
JSTOR
OCLC
Crossref
Library Management Systems
BibDesk
Zenodo
EBSCO Host
RNIB Bookshare
Project MUSE
OAPEN
DOAB
Google Books
JSTOR
BDS Live
OCLC
ProQuest KB + ExLibris
Jisc KB
EBSCO KB
Crossref
THOTH
Open Metadata.

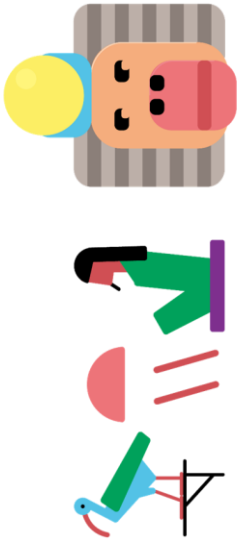



THOTH
Open Metadata.

# Data Output

automatic file generation
adhering to established standard
formats & platform specifications

THOTH
Open Metadata.



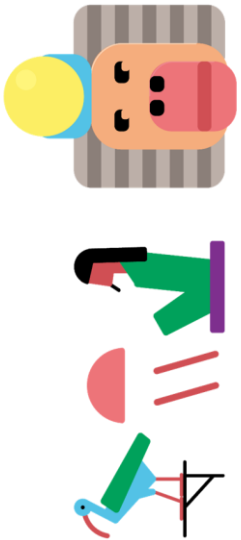
# Interoperability via open APIs

Using Thoth's open APIs, partners have already been able to integrate Thoth data in the context of …

- Joint / shared catalogues (Open Book Collective, ScholarLed)
- Platforms (Open Book Collective)
- Publisher websites (Open Book Publishers)
- Customised data searches / outputs
- Connections with other systems (e.g. ScienceOpen's BookMetaHub, PKP's Open Monograph Press)
- Repository deposits (data & content) via the Thoth Archiving Network

# Collaborations with like-minded community-led intitaitives

THOTH Open Metadata.



# Next steps

- Improving UX, and formalising Thoth SaaS model
- Improving bibliodiversity: extension of the geographic and linguistic reach beyond anglophone countries – integrating multilingual UX, outreach to non-English publishers (e.g. Europe & SciELO Books)
- Enhancing interoperability with other open infrastructures, e.g. PKP's Open Monograph Press, COKI's Books Analytics Dashboard project; integration with OpenAlex & Wikidata, etc.
- Creating a network of National Libraries to host & archive OA books as part of the **Thoth Archiving Network**

# Thoth Archiving Network

Photo by Asal Lotfi on Unsplash

# Thoth Archiving Network - Introduction

## What is the Thoth Archiving Network?

- A network of participating institutional repositories that agree to accept open access monographs published by small, scholar-led, and library-based OA publishers that have no existing archiving or preservation arrangements

- Publishers click to deposit (through Thoth) their published OA monographs into a repository as an archiving option to protect 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

- Will also employ web-archiving platforms such as the Internet Archive & Zenodo

# Thoth Archiving Network: Outlook

- Currently have proof-of-concept completed in Figshare, Eprints, and the Internet Archive. Dspace work continues with Cambridge University's repository team.
  - PubPub by Ross Higman on programmatic archiving into the IA
  - Workflow experimentations with API, Python, and SWORD protocol

- Additional university repositories in the UK and USA have agreed to join network pilot, with further collaboration beginning within the OBF project.

- Development work continues toward archiving and preservation as part of the dissemination features in-progress within Thoth.

- Aim is to simplify this process to a single "Send to archive" button.

# The small and scholar-led OA publisher

- Publishers who are very small, independent publishers, often run by a single scholar/academic or a small team
- Are not supported by any institutional funding and are not for profit
- Usually have very limited resource: staff, financial, technological (Barnes, Bell, et al, 2022).
- Due to this, many do not have any archiving or preservation safeguards in place
- OA books therefore at risk, as part of the long tail of publishers that could 'disappear': An ongoing study investigating the preservation status of open access books (Laakso, Wise & Snijder, 2022) follows in the footsteps of an earlier study (Laakso et all, 2020) which examined journal articles.
- **Are valuable contributions to the scholarly record that need to be preserved**

# OBF WP7: National Libraries Network

Part of the work of Work Package 7 of Open Book Futures:

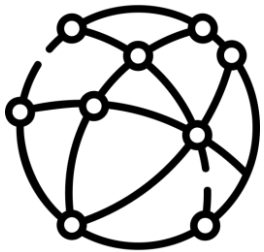
- Establish and expand an informal National Libraries Network, involving national libraries from around the world, to scope  funding, technical, and administrative needs that would be required to implement an archiving & preservation network at this scale.
- What possibilities might exist for this to created in the future?
- What barriers exist and what is already being done?

# Thoth Archiving Network - Looking to the future

## Would your organisation be interested in joining the Thoth Archiving Network?

The Thoth Archiving Network receives funding from the Research England Development Fund and Arcadia under the Open Book Futures project grant, which has been awarded to significantly expand and accelerate uptake of the open infrastructures developed in COPIM.

Our ambition is to expand the Thoth Archiving Network globally.



Earth icons created by Freepik - Flaticon

# Thank you for listening!

Dr Miranda Barnes
m.l.barnes@lboro.ac.uk
ORCiD: 0000-0001-7740-4896

Dr Gareth Cole
g.j.cole@lboro.ac.uk
ORCiD: 0000-0001-7493-0137

Toby Steiner, MA
toby@thoth.pub
ORCiD: 0000-0002-3158-3136

Photo by alleksana on Pexels