

11. Metadata management

Strategies for managing core information about your books and authors

- [Key issues](#)
- [Case studies](#)
- [Questions to guide good practice](#)
- [Resources](#)

Key issues

Metadata consists of descriptive information about a publisher's book titles. This ranges from information about the book or parts of the book (e.g. book title, chapter title, year of publication, subject classifications, ISBN, DOI) to information about the author (e.g. name, institutional affiliation) as well as other information (e.g. whether the book is part of a series, the book format).

As for conventional publishers, for Open Access publishers selling hard copies, this metadata is important for book sales – to enable books to be found by vendors and distributors, for example. However, metadata becomes particularly important to enable a book's digital discoverability. The richer and more accurate the data, the easier it is for interested readers to find a book on different platforms and in library catalogues.

Preparing high-quality metadata can be a labour intensive task, as at least some manual input will likely be required. This can pose challenges for smaller and scholar-led presses. However, the importance of metadata for discoverability means that presses should seriously consider how to include a sustainable metadata management strategy into their workflow. Metadata management is also an area where publishers associated with universities can reasonably demand support from their institutions, with most university libraries having staff already appointed who have a good understanding of the processes involved.

Once the outlines of a metadata management strategy are in place, a publisher may wish to consider a 'metadata model'. This is the core metadata that a publisher will maintain on every book. An example of a metadata model is provided by [Jisc](#) and [OAPEN](#), in a 2016 [resource](#). It is a short four-page document that includes tables for each of the proposed categories of data: book, creator format, and collection, detailing the main characteristics of books that are important for discovery, distribution and tracking metrics. A strength of the guide is that it recognises that a balance can sometimes need to be struck between the richness of metadata and the overheads involved in maintaining rich metadata.

Publishers may also need a working understanding of different metadata "formats" and the role of "persistent identifiers" (PIDs). The [NUP Toolkit](#) provides an excellent overview of each, including formats such as ONIX, MARC 21, and KBART and PIDs such as DOIs (Digital object identifiers) and ORCID (Open Researcher and Contributor Identifier) records. More detailed discussions of these formats can be found in [a report from 2021](#), by Graham Stone and colleagues.

Metadata management platforms

A number of third-party platforms exist that are designed to make metadata management more straightforward.

Of the options available, we would strongly recommend that Open Access publishers consider [Thoth](#). Thoth is a UK-based non-profit that has built an Open Source metadata management platform designed specifically around the needs of smaller and scholar-led Open Access publishers. It provides a single, simple-to-use repository for a publisher's metadata and a wide range of export functionality aimed at reducing the workload involved in sending book metadata to a wide range of users, from distributors, through catalogues, to libraries and repositories. This includes different metadata formats. Within the [Open Book Collective](#), we use Thoth's open API to populate our collective catalogue; we also use Thoth's ability to collect institutional affiliation metadata in our outreach work with universities.

Other metadata management platforms may, however, better suit the needs of individual publishers. Here are some alternatives to consider:

- [Ubiquity](#), purchased by [De Gruyter](#) in 2022, specialises in providing a full publishing service for publishers wishing to publish Open Access. As part of this, it provides metadata management services, including for example preparing MARC records for scholar-led and university presses that use its infrastructure (Ubiquity website).
- [Bibliographic Data Services \(BDS\)](#) is a UK-based commercial supplier that creates metadata for the library supply chain. Presses should note that to do so will mean handing the copyright of the metadata to BDS, who may charge the press for the use of their own metadata ([NUP Toolkit](#)).

It is also worth noting that most print-on-demand services provide integrated book metadata dissemination to the global book trade ([NUP Toolkit](#)).

Case studies

- The [Thoth](#) metadata manager is presently being used in production by a number of small/scholar-led publishers, including [African Minds](#), [Mattering Press](#), [mediastudies.press](#), [meson press](#), [Open Book Publishers](#), and [punctum books](#)
- [White Horse Press](#) currently submits metadata to Nielsen and BDS but is working towards making book metadata available via Thoth and [ScienceOpen's BookMetaHub](#) (direct correspondence)
- [Stockholm University Press](#) used [Ubiquity](#) to manage its metadata, as well as [Lightning Source](#), its print-on-demand service, to distribute metadata to vendors ([Business Models for OA books](#)).

Questions to guide good practice

- Is metadata management built into the production workflow? Who is responsible for managing metadata?
- How will metadata be generated and managed? Have different metadata management platforms been assessed?
- If using a third party metadata management platform, does the press retain the copyright to its own metadata?

Resources

- [NUP Toolkit](#)
- Graham Stone, Rupert Gatti, Vincent W.J. van Gerven Oei, Javier Arias, Tobias Steiner, and Eelco Ferwerda (2021). WP5 Scoping Report: Building an Open Dissemination System. COPIM. <https://doi.org/10.21428/785a6451.939caea>
- Jisc and OAPEN (2016). Metadata for open access monographs. <https://oopen.fra1.digitaloceanspaces.com/55111e5ee86a4653a0f9708034d5f9ae.p>