Developing an Archiving Strategy: What metadata to archive?

Metadata is critical for the discovery and effective dissemination of books. Just putting something online doesn't mean anybody will get to know about it - and this is particularly true for future readers searching and accessing works in ways we have not yet imagined. While more metadata is always to be encouraged, it is important to use persistent identifiers and controlled vocabularies whenever possible as this will increase the likelihood of interoperability and the successful transmission of metadata to new systems in the future.

What metadata to archive?

The report *International Metadata Recommendations, and Platform-Specific Requirements for Open Access Books and Chapters* (Steiner et al. 2026) identifies metadata fields that are "Essential", and those that are "Desirable", for the effective dissemination of open access books. These criteria are appropriate for archiving purposes also.

Essential bibliographic and access metadata include:

- title and subtitle (multilingual if appropriate)
- contributors (including standardised and persistent identifiers such as ORCID or ISNI where possible)
- copyright holder and licence,
- subjects (utilising recognised schemas such as THEMA where possible)
- landing page and full-text URLs and/or DOIs at book and chapter level (ideally, for archiving purposes, a link/reference to an archived version should be included)
- publisher details, and publication date.

Desirable elements include:

- abstract (multilingual if appropriate)
- cover image
- table of contents
- contributor affiliations (using standardised and persistent identifiers such as ROR where possible)
- funder details (using standardised and persistent identifiers such as ROR where possible)

What formats to archive metadata?

Many file formats, such as PDF and EPUB, allow extended metadata to be included in the book file itself - and clearly the more metadata included this way the better.

However, we recommend that when archiving content a separate metadata file be included alongside the primary ebook file(s) in an open and standardised format that can be accessed as plain text if necessary (such as ONIX, MARC or JSON). This helps ensures that the metadata can be openly shared across systems and platforms and that engagement with specific software or formats is not required to access the metadata

This section summarises the findings of several reports created within the OBF project:

Barnes, M., Cole, G., Fry, J., Gatti, R., & Higman, R. (2023). 'Good, Better, Best': Practices in Archiving & Preserving Open Access Monographs (1.0). Zenodo.

https://doi.org/10.5281/zenodo.7876048 . This report considers the archiving of metadata specifically in Chapter 2.

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. This report identifies the most important metadata fields to provide to ensure broad discovery and dissemination of the work.

Stone, Graham, Rupert Gatti, Vincent W. J. van Gerven Oei, Javier Arias, Tobias Steiner, and Eelco Ferwerda.
‘WP5 Scoping Report: Building an Open Dissemination System’. Community-Led Open Publication Infrastructures
for Monographs (COPIM), 21 April 2021. https://doi.org/10.21428/785a6451.939caeab.

Revision #1
Created 2026-04-28 15:11:04 UTC by Rupert Gatti
Updated 2026-04-30 16:10:41 UTC by Rupert Gatti