COPIM

Open Metadata in Thoth

Vincent W.J. van Gerven Oei

Published on: Nov 24, 2020

DOI: 10.21428/785a6451.eb0d86e8

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Based on the recommendations set out in the [WP5 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 easily expose and transfer their metadata to other platforms, such as digital repositories, libraries, and vendors.

At the moment that metadata imported, edited, and stored inside Thoth leave the system, we encounter the question of the licensing of the book and chapter metadata that are exported. More than often, publication metadata records do not come with an explicit license, which creates an undesirable uncertainty as to the precise terms under which metadata can be (re)used and altered. This is only one of the many challenges faced by the scholarly communications community with regard to the processing of metadata ([Gregg et al. 2019](#)).

The Open Knowledge Foundation's [Principles on Open Bibliographic Data](#) recommend "the use of the Public Domain Dedication and Licence or Creative Commons Zero Waiver" ([Open Bibliographic Data Working Group of the Open Knowledge Foundation, 2010](#)); The [Open Metadata Principles](#) ([Discovery 2011](#)), which have been [signed](#) by COPIM partners [Lancaster University](#), [Jisc](#), and the [British Library](#), recommend the usage of Open Data Commons Public Domain Dedication & Licence ([ODC-PDDL](#)), [Creative Commons CCO licence](#), or the [UK Open Government Licence](#) (OGL); [Plan S](#) mandates the release of "high-quality [...] metadata in standard interoperable non-proprietary format, under a CCO public domain dedication" ([cOAlition S n.d.](#)); and the [Data Exchange Agreement](#) of Europeana, an EU-funded project to open up the digital collections of European cultural institutions, has a policy of requiring CCO-licensed metadata ([Europeana Foundation 2015](#)). The [Metadata 2020](#) project, a collaboration of commercial publishers, service providers, and libraries has so far issued no guidance with regard to metadata licensing, but states as one of its [principles](#) that "metadata must be as open, interoperable, parsable, machine actionable, human readable as possible."

These proposals to release metadata in the public domain have, understandably, encountered resistance from commercial vendors who benefit from a closed metadata ecosystem. For example, [OCLC](#) adopted a more restrictive Open Data Commons Attribution License ([ODC-By](#)) license for its metadata records, [stating](#) that the obligatory attribution "represents a significant investment of time and resources both from OCLC and from each of its member institutions that contribute records" ([OCLC 2012](#)).

There is no doubt that many of the stakeholders in the scholarly communications chain make significant investments into creating metadata records, and it is perhaps a feature of the current

closed and compartmentalized book metadata ecosystem that much of these investments are actually unnecessary double work. Implementing an ODC-By or CC-BY license for metadata would imply the creation of a potentially sizeable parallel administration of *meta*-metadata, recording the different authors, editors, and contributors to individual metadata records, who may license their contributions in different ways, thus creating eventually the necessity of licensing *those* meta-records. This phenomenon is known as “attribution stacking” ([Korn and Oppenheim 2011, 5](#)).

The only way to halt such undesirable recursion is for metadata records to be fundamentally open, that is, released under one of the licenses suggested by the Open Metadata Principles. As all member presses of [ScholarLed](#) ([Barnes 2018](#), [Deville et al. 2019](#)) already release their books under one of the available Create Commons licenses, it makes sense to stay within the same conceptual framework and license Thoth’s dataset as [CC0](#). This implies that the present and future users of Thoth, which is [freely available](#) as open source software, commit to making their metadata open as well.

[Flynn \(2013\)](#) envisions an “Open Collaborative Catalog (OCC),” which “would allow libraries to combine their cataloging efforts and achieve maximum benefit with OA metadata housed in the cloud, and then pulled into local OPACs [Online Public Access Catalogs].” These “would house the cataloging records released as OA metadata [...]. It would be ideal for the OCC to be editable as a wiki in order to maintain the best records and share them with ease, so libraries could download exactly the ones they need” (30).

There is no reason to restrict this vision solely to libraries. The pioneering usage of Thoth within the COPIM project and the ScholarLed consortium, represents only one, rather limited use case: the management of a relatively small set of open book metadata. Connected to multiple metadata feeds, such as those listed on Archive.org’s [Bulk Bibliographic Metadata](#) page, Thoth has the potential to become the functional core of the Open Collaborative Catalog that Flynn envisions, providing a collaborative platform for the creation, mashup ([Stephens 2011](#)), management, and export of high-quality open metadata without restrictions.

References

Barnes, Lucy. 2018. “ScholarLed Collaboration: A Powerful Engine to Grow Open Access Publishing.” *Impact of Social Sciences* (blog). October 26, 2018. <https://blogs.lse.ac.uk/impactofsocialsciences/2018/10/26/scholarled-collaboration-a-powerful-engine-to-grow-open-access-publishing/>.

cOAlition S. n.d. “Accelerating the Transition to Full and Immediate Open Access to Scientific Publications.” Accessed November 22, 2020. <https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/>.

Deville, Joe, Jeroen Sondervan, Graham Stone, and Sofie Wennström. 2019. “Rebels with a Cause? Supporting Library and Academic-Led Open Access Publishing.” *LIBER Quarterly* 29 (1): 1. <https://doi.org/10.18352/lq.10277>.

Discovery. 2011. “Open Metadata Principles.” <http://discovery.ac.uk/businesscase/principles/>.

Europeana Foundation. 2015. “Europeana Data Exchange Agreement.” https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Data%20Exchange%20Agreement.pdf.

Flynn, Emily Alinder. 2013. “Open Access Metadata, Catalogers, and Vendors: The Future of Cataloging Records.” *The Journal of Academic Librarianship* 39 (1): 29–31. <https://doi.org/10.1016/j.acalib.2012.11.010>.

Gregg, Will James, Christopher Erdmann, Laura A. D. Paglione, Julianne Schneider, and Clare Dean. 2019. “A Literature Review of Scholarly Communications Metadata.” *Research Ideas and Outcomes* 5: e38698. <https://doi.org/10.3897/rio.5.e38698>.

Korn, Naomi, and Charles Oppenheim. 2011. “Licensing Open Data: A Practical Guide.” http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf.

OCLC. 2012. “WorldCat Data Licensing.” <https://www.oclc.org/content/dam/oclc/worldcat/documents/worldcat-data-licensing.pdf>.

Open Bibliographic Data Working Group of the Open Knowledge Foundation. 2010. “Principles for Open Bibliographic Data.” *Open Bibliography and Open Bibliographic Data* (blog). October 15, 2010. <http://openbiblio.net/2010/10/15/principles-for-open-bibliographic-data/index.html>.

Stephens, Owen. 2011. “Mashups and Open Data in Libraries.” *Serials* 24 (3): 245–50. <https://doi.org/10.1629/24245>.

Citations

1. Gregg, W. J., Erdmann, C., Paglione, L. A. D., Schneider, J., & Dean, C. (2019). A literature review of scholarly communications metadata. *Research Ideas and Outcomes*, 5, e38698. <https://doi.org/10.3897/rio.5.e38698> ↵
2. of the Open Knowledge Foundation, O. B. D. W. G. (2010). *Principles for Open Bibliographic Data*. *Open Bibliography and Open Bibliographic Data*. Retrieved from <http://openbiblio.net/2010/10/15/principles-for-open-bibliographic-data/index.html> ↵
3. Discovery. (2011). *Open Metadata Principles*. Retrieved from <http://discovery.ac.uk/businesscase/principles/> ↵

4. S. (n.d.). *Accelerating the Transition to Full and Immediate Open Access to Scientific Publications*. Retrieved from <https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/> [↵](#)
5. Foundation, E. (2015). *Europeana Data Exchange Agreement*. Retrieved from https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Data%20Exchange%20Agreement.pdf [↵](#)
6. OCLC. (2012). *WorldCat Data Licensing*. Retrieved from <https://www.oclc.org/content/dam/oclc/worldcat/documents/worldcat-data-licensing.pdf> [↵](#)
7. Korn, N., & Oppenheim, C. (2011). *Licensing Open Data: A Practical Guide*. Retrieved from http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf [↵](#)
8. Barnes, L. (2018). *ScholarLed collaboration: a powerful engine to grow open access publishing. Impact of Social Sciences*. Retrieved from <https://blogs.lse.ac.uk/impactofsocialsciences/2018/10/26/scholarled-collaboration-a-powerful-engine-to-grow-open-access-publishing/> [↵](#)
9. Deville, J., Sondervan, J., Stone, G., & Wennström, S. (2019). Rebels with a Cause? Supporting Library and Academic-led Open Access Publishing. *LIBER Quarterly*, 29(1), 1. <https://doi.org/10.18352/lq.10277> [↵](#)
10. Flynn, E. A. (2013). Open Access Metadata, Catalogers, and Vendors: The Future of Cataloging Records. *The Journal of Academic Librarianship*, 39(1), 29–31. <https://doi.org/https://doi.org/10.1016/j.acalib.2012.11.010> [↵](#)
11. Stephens, O. (2011). Mashups and open data in libraries. *Serials*, 24(3), 245–250. <https://doi.org/10.1629/24245> [↵](#)