**Front Matter**

# Improving Rogue Scholar metadata conversions

**Martin Fenner** ⓘ

Published January 16, 2024

**Citation**

**Keywords**

**Copyright**

## Front Matter

Last week the Rogue Scholar science blog archive added export of blog post content in various formats (Markdown, ePub, PDF, JATS XML). This week Rogue Scholar is improving the existing metadata export, and adding metadata export in Schema.org JSON-LD format.

Until now Rogue Scholar metadata export depended on DOI content negotiation, returning the DOI metadata registered with Crossref (and in some cases DataCite) in BibTex or CSL JSON formats, or as formatted citations. While this approach worked reasonably well and allowed Rogue Scholar to offload the metadata conversion to an external service, it had major limitations:

- Delay of up to 24 hours after DOI registration until metadata are available in DOI content negotiation
- Long-standing issues with some of the metadata in BibTeX or CSL JSON
- Missing support for some formats, especially Schema.org JSON-LD
- Missing support for converting metadata in bulk, e.g. for all Rogue Scholar blog posts published last week

To address these issues, the Rogue Scholar API this week started using commonmeta-py, the metadata conversion Python library I wrote last year. This allows Rogue Scholar to show the metadata immediately after DOI registration, which now happens within 20 min after a blog post is published or updated. This also addresses some issues with the metadata returned by DOI content negotiation, e.g. the missing blog name. And this adds additional metadata formats, most notably Schema.org JSON-LD.

Going forward this change will allow us to address another limitation of the Rogue Scholar service: DOI registration and updates happen via GitHub Actions, scheduled via cron. While this workflow works well, it is currently limited to one DOI registration or update every 10 min, so updates to the metadata of all 13K Rogue Scholar DOIs would take weeks. One example is the changes last week that added links to the full-text Markdown, ePub, PDF, and JATS XML versions to all Rogue Scholar DOIs.

commonmeta-py already generates Crossref XML metadata needed for Crossref DOI registration. More work is needed to add background workers to the Rogue Scholar API that can replace the current GitHub Actions, and increase the frequency DOI metadata are updated with Crossref. As commonmeta-py can also generate DataCite metadata needed for DOI registration, Rogue Scholar could also help scholarly blogs that use DataCite DOIs (currently two of the blogs participating in Rogue Scholar).

These changes again highlight the importance of eating your own dog food: organizations should themselves use the products and services they provide.

## References

Fenner, M. (2024). Every Rogue Scholar blog post now available in Markdown, ePub, and PDF formats. *Front Matter*. https://doi.org/10.53731/1dfxr-hs665

## Front Matter

Fenner, M. (2023). Releasing commonmeta-py v0.8. *Front Matter*. https://doi.org/10.53731/xszpd-6z265

Fenner, M. (2016). Eating your own dog food. *Front Matter*. https://doi.org/10.53731/r79vxn1-97aq74v-ag58n