Tracking references of Rogue Scholar blog posts

Martin Fenner ⓘ

Published March 27, 2024

Citation

Keywords

News, Rogue Scholar

Copyright

Front Matter

The Rogue Scholar science blog archive has been collecting the references of blog posts since June 2023, and has registered Crossref DOIs for 1,114 blog posts with references as of today. As RSS feeds – which Rogue Scholar uses to fetch content and metadata – have not standardized how metadata for references should be described, Rogue Scholar is extracting them from the full text of posts and takes advantage of the fact that there is a standard format: a section called **References** followed by a listing of metadata (author, title, publication year, etc.) that includes a DOI or URL link. Crossref Directory of Technology Geoff Bilder uses this image in presentations to make a similar point – scholarly audiences immediately understand how the references section of a publication looks:
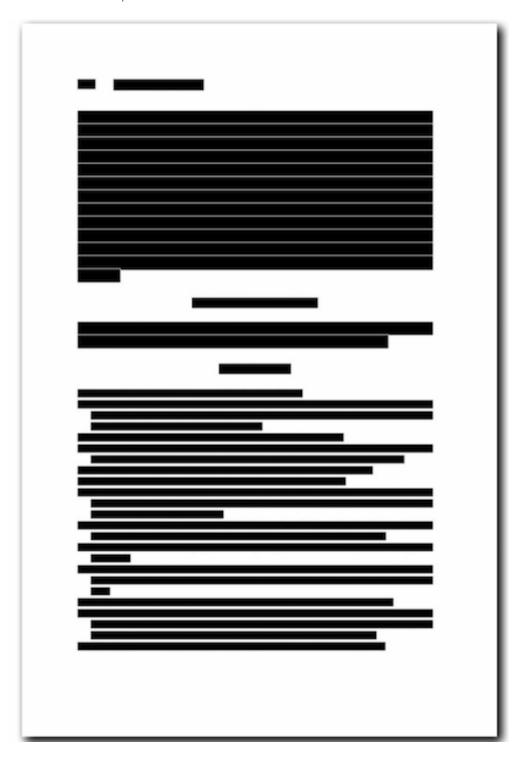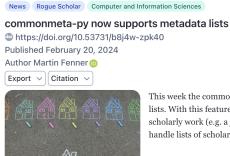
Front Matter

Image by Geoff
Bilder

An example is the references in this post. The citation style (in this case **APA**) doesn't matter as Rogue Scholar only looks for the links. If the link is a DOI, the Rogue Scholar looks up the DOI metadata via DOI content negotiation and stores DOI, title, and publication year. This information is then included in the metadata registered with Crossref.

Improvements to the references workflow

In the past few weeks I have started to improve this workflow. The first step was to store all references in the Rogue Scholar API, using a new /works API endpoint, and the commonmeta-py JSON schema. Commonmeta-py can import metadata in Crossref and DataCite formats, and the latest version has improved support for extracting metadata from web pages using schema.org and HTML meta tags. This allows Rogue Scholar to find DOI metadata for webpages, e.g. journal article landing pages (similar to what Zotero and other reference managers are doing), and to store additional metadata about web links, mainly the title and publication year. This information is now included in the references registered with Crossref.

To show the references that Rogue Scholar has found for a blog post, I changed the layout of Rogue Scholar pages which now include pages for individual posts, e.g. this page:

News   Rogue Scholar   Computer and Information Sciences

commonmeta-py now supports metadata lists

https://doi.org/10.53731/b8j4w-zpk40

Published February 20, 2024

Author Martin Fenner

Export ⌄   Citation ⌄

This week the commonmeta-py Python library adds an important new feature: metadata lists. With this feature commonmeta-py no longer only operates on metadata for a single scholarly work (e.g. a journal article, book, dataset, software, or blog post), but can handle lists of scholarly works.

References

1. Announcing Commonmeta (2023). https://doi.org/10.53731/cp7apdj-jk5f471
2. commonmeta-py (2024). https://doi.org/10.5281/zenodo.8340374

You navigate to these pages by clicking on the title in a Rogue Scholar search result. Going forward these dedicated pages will show more metadata, e.g. funding and relationship information.

The next step is adding the ability to display the references in a variety of citation styles and export them in formats such as BibTex or CSL, both straightforward as these are core functionalities of the commonmeta-py library.

By showing the references in Rogue Scholar I hope to motivate blog authors to include more references in their posts, and I will continue to work with them on how to improve reference parsing, e.g. when references don't include a DOI or URL link. I am a big supporter of the

Front Matter

Initiative for Open Citations (I4OC) and helped get it started. In the case of Rogue Scholar, the challenge is not opening up existing but closed reference metadata, but finding ways to collect and register them with Crossref that scale for science blogs.

References

Fenner, M. (2023). *Starting to include references in DOI metadata for blog posts.* https://doi.org/10.53731/6mkrk-dzh02

Fenner, M. (2024, February 20). *Commonmeta-py now supports metadata lists.* https://doi.org/10.53731/b8j4w-zpk40

Shotton, D. M. (2017). *The Initiative for Open Citations.* https://doi.org/10.59350/jdwj8-at997