

Starting November, all Rogue Scholar blog posts will be archived by the Internet Archive

Martin Fenner 

Published October 30, 2023

Citation

Fenner, M. (2023). Starting November, all Rogue Scholar blog posts will be archived by the Internet Archive. *Front Matter*. <https://doi.org/10.53731/hhtx0-wb293>

Keywords

News, Rogue Scholar



Copyright

Copyright © Martin Fenner 2023. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Front Matter

Today I am happy to announce an important milestone for the [Rogue Scholar](#) science blog archive. Starting November 1st, all blog posts from participating blogs will automatically be archived by the [Internet Archive](#).

Front Matter has signed a contract with Internet Archive to use their [Archive-It](#) service, and archiving will start in November. The archiving test runs to estimate the data volume that needs to be archived have been completed, and proper archiving will start this week. The archived blog posts will be available via [this page](#), and it will take 1-2 weeks until all blog posts from participating blogs are archived.

Some participating blogs are part of a larger website and the archiving will be restricted to the blog posts (and all associated resources such as images and other attached files). Blog posts will automatically be archived every six months. The only limitation is video which will not routinely be included because of large file sizes, which was an issue with one participating blog. The archive size is currently limited to one GB per participating blog, more than enough for the currently included blogs (only one participating blog exceeds that data limit but is included anyway).

The archived blog posts will be integrated into the Rogue Scholar service so that DOIs are automatically updated when a blog is no longer accessible via the public web. The DOIs will then resolve to the blog post archived by the Internet Archive. So far this is the case for only one blog included in Rogue Scholar ([Project THOR](#)), but science blogs that are no longer maintained and disappearing from the public web are a well-known phenomenon. Rogue Scholar is coordinating these archiving activities with Crossref and their work on [digital preservation](#).

Long-term archiving of science blog posts is not a good fit for a startup such as Front Matter, and the collaboration with the Internet Archive provides the necessary long-term perspective. This does not preclude other archiving activities, and I have started conversations with the [Zenodo](#) open repository about archiving blog posts in ePub format. This will happen later this year once [work on the Rogue Scholar API](#) is completed. Please reach out to Rogue Scholar if you want to see science blog posts archived in other formats and/or other archiving services.

References

Fenner, M. (2023). *Use cases for science blogs: Grant-funded projects*. <https://doi.org/10.53731/mh9a1-dw902>

Eve, M. (2023). *A Request for Comment: Automatic Digital Preservation and Self-Healing DOIs* [Website]. Crossref. Retrieved October 30, 2023, from <https://www.crossref.org/blog/a-request-for-comment-automatic-digital-preservation-and-self-healing-fois/>

Fenner, M. (2023). *The Rogue Scholar API now automatically indexes blog posts*. <https://doi.org/10.53731/qq4a5-6zc45>