

Publication status: This preprint has not been published elsewhere.

ris-enrich: An automated Python tool for enriching bibliographic RIS files with full abstracts

Eduardo Santos

<https://doi.org/10.1590/SciELOPreprints.15571>

Submitted on: 2026-07-17

Posted on: 2026-07-22 (version 1)
(YYYY-MM-DD)

The moderation of this preprint received the endorsement of:

- Shinichi Nakagawa (ORCID: <https://orcid.org/0000-0002-7765-5182>)

ris-enrich: An automated Python tool for enriching bibliographic RIS files with full abstracts

Eduardo S. A. Santos

COSSEE, Collaboration for Open Science and Synthesis in Ecology and Evolution,
Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada
Email: esantos2@ualberta.ca | ORCID: <https://orcid.org/0000-0002-0434-3655>

2026-07-17

Summary

Systematic reviews and meta-analyses require rigorous literature screening, a process highly dependent on the availability of accurate abstracts for each bibliographic record. However, researchers often encounter a bottleneck during the discovery phase: some academic search engines, such as Google Scholar, routinely truncate abstracts in their bibliographic records. **ris-enrich** is an open-source Python package designed to resolve this data loss. It parses **.ris** files and cascades searches across four major open academic databases (Semantic Scholar, OpenAlex, Europe PMC, and Crossref) to retrieve and append the full-text abstracts, utilizing strict NFKC Unicode normalization and fuzzy string matching to prevent false-positive data contamination. Our case study demonstrates that **ris-enrich** can recover complete abstracts for 21% of records across six languages, significantly enhancing the efficiency and accuracy of title-and-abstract screening in evidence synthesis workflows. By automating this enrichment process, **ris-enrich** empowers researchers to conduct more comprehensive and inclusive literature reviews, ultimately advancing the quality of systematic reviews and meta-analyses in ecology, evolution, psychology, and beyond.

Statement of need

The preparation of a preregistration protocol and the subsequent screening workload in evidence synthesis studies (e.g., in fields like ecology, evolution, and psychology) are reliant on the quality and integrity of the initial bibliographic dataset. While tools exist to optimize search string generation ([Grames et al. 2019](#)) or facilitate the literature screening process, for example, Rayyan or SysRev ([Ouzzani et al. 2016](#); [Bozada et al. 2021](#)), there is a distinct lack of

lightweight tools focused strictly on bibliographic data enrichment post-export.

When researchers export search results, they are frequently left with snippets rather than full abstracts, rendering title-and-abstract screening virtually impossible without manual intervention. While manual intervention can be conducted when the number of records is manageable, it becomes infeasible at scale (e.g., when you have thousands of records to process). This is especially common from search records exported from Google Scholar, because records in Google Scholar present only truncated abstracts. Google Scholar is one of the few search engines available that is capable of retrieving academic records in non-English languages, thus being an important search engine in attempts to minimize bias in data collection for evidence synthesis studies. **ris-enrich** automates the recovery of this missing metadata. By prioritizing a sequential API fallback architecture and enforcing an 80% title-similarity threshold using the **difflib** and **unicodedata** libraries, the software ensures high-fidelity data retrieval even across international, diacritic-heavy, and logographic languages.

ris-enrich was developed to directly support the screening workflows of evidence synthesis studies, seamlessly preparing enriched **.ris** datasets of bibliographic records that can be incorporated in the record screening process of the synthesis studies.

In practice

To install the current version of the **ris-enrich** package, run the command:

```
pip install git+https://github.com/esantos2ua/ris-enrich.git
```

After installation, you can use the **ris-enrich** command line tool:

```
ris-enrich data/GoogleScholarPortugueseMateChoiceExample.ris
```

```
Reading 'GoogleScholarPortugueseMateChoiceExample.ris'...
Found 68 references. Starting enrichment...
```

```
No email provided for Crossref API. Rate limits may be lower.
Use --email or set RIS_ENRICH_EMAIL.
[1/68] Searching: Caracterização da reprodução e ensaios de crescime...
-> Found abstract via OpenAlex (3887 chars)
[2/68] Searching: The role of olfaction in sexual interactions of ba...
-> No exact matching abstract found.
[3/68] Searching: Marcadores genéticos de previsão de fenótipos em c...
-> No exact matching abstract found.
[4/68] Searching: Benchmarking na gestão de unidades de saúde: relev...
-> Found abstract via OpenAlex (466 chars)
[5/68] Searching: O impacto da terceirização sobre os custos de mão-
...
```

```
-> Found abstract via OpenAlex (1232 chars)
...
```

Done! Safely updated 22 abstracts.

Using the **ris-enrich** command above will output to screen information about the **.ris** file being enriched. It will display the file name, followed by the number of references found in the file and the command will print a run-time statement of each record that is being enriched, with a message to the user on whether or not a complete abstract was found. At the end of the file, the output will display how many records had abstract information updated.

The **ris-enrich** command will also create a **enrichment_log.txt** log file with the record of the enrichment information. Finally, the command will create a new **.ris** file with the suffix **_Enriched.ris** that contains the original inputted records with the updated abstract field containing the new complete abstract, when available.

Arguments

The **ris-enrich** command can be implemented with the following arguments:

- **input_file**: Path to the original **.ris** file (Required).
- **-o, --output**: Path to save the enriched **.ris** file. Defaults to ***_Enriched.ris**.
- **-l, --log**: Path to save the execution log. Defaults to **enrichment_log.txt**.
- **-e, --email**: **(optional)** Your email address, which will be included in the User-Agent header when querying the Crossref API. Providing an email helps Crossref identify polite usage and may increase rate limits. You can also set this value by exporting **RIS_ENRICH_EMAIL** in your environment before running the tool:

```
export RIS_ENRICH_EMAIL=you@example.com
ris-enrich input_file.ris -o output_file.ris
```

or pass it directly on the command line:

```
ris-enrich input_file.ris -o output_file.ris --email you@example.com
```

Current applications

ris-enrich was developed as part of an evidence synthesis study that is updating two meta-analyses in Ecology and Evolution. **ris-enrich** was used to enrich bibliographic records that were retrieved from literature searches conducted in Japanese, Polish, Portuguese, Russian, Simplified/Traditional Chinese, and Spanish on Google Scholar using the application Publish or Perish software

(Harzing 2007). With **ris-enrich**, I was able to enrich a set of 425 bibliographic records with 89 complete abstracts, a 21% average improvement considering the six languages. This improvement has clear benefits for screeners that are reading titles and abstracts of these records to make decisions on whether a record should be included for full-text screening in the data pipeline of the study.

Our open source Python tool, **ris-enrich**, has proven to be an effective and easy to use tool for automating the enrichment of bibliographic records, and can be used in a variety of evidence synthesis workflows.

Author contributions

Eduardo S. A. Santos: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Project administration.

Conflict of interest disclosure

The author declares that no competing interests exist.

Data availability

All the datasets supporting the results of this study are available in the GitHub repository and can be accessed at <https://github.com/esantos2ua/ris-enrich>. The source code for **ris-enrich** is distributed under the MIT License. The example **.ris** dataset used in this study is included in the **data/** directory of the repository.

Declaração de disponibilidade de dados da pesquisa

Todo o conjunto de dados de apoio aos resultados deste estudo foi disponibilizado no repositório GitHub e pode ser acessado em <https://github.com/esantos2ua/ris-enrich>. O código-fonte do **ris-enrich** é distribuído sob a Licença MIT. O conjunto de dados de exemplo no formato **.ris** utilizado neste estudo está incluído no diretório **data/** do repositório.

Acknowledgements

ESAS, as well as this research work, are supported by the Canada Excellence Research Chairs (CERC) program (grant number CERC-2022-00074). ESAS would like to thank Ayumi Mizuno and Santiago Ortega for testing the software and providing feedback on the user experience.

AI usage disclosure

- **Tool use:** For this paper, I used Google Gemini 3.1 Pro to write the main code of this open-source Python package, and to prepare the initial versions of the documents of the github repository.
- **The nature and scope of assistance:** Assistance was used for code generation, refactoring, and test scaffolding.
- **Confirmation of review:** ESAS reviewed, edited, validated all AI-assisted outputs and made the core design decisions in this project.

References

- Bozada, Thomas, James Borden, Jeffrey Workman, Mardo Del Cid, Jennifer Malinowski, and Thomas Luechtefeld. 2021. “Sysrev: A FAIR Platform for Data Curation and Systematic Evidence Review.” *Frontiers in Artificial Intelligence* 4 (August). <https://doi.org/10.3389/frai.2021.685298>.
- Grames, Eliza M., Andrew N. Stillman, Morgan W. Tingley, and Chris S. Elphick. 2019. “An Automated Approach to Identifying Search Terms for Systematic Reviews Using Keyword Co-Occurrence Networks.” *Methods in Ecology and Evolution* 10 (10): 1645–54. <https://doi.org/10.1111/2041-210X.13268>.
- Harzing, A. W. 2007. *Publish or Perish*. Released. <https://harzing.com/resources/publish-or-perish>.
- Ouzzani, Mourad, Hossam Hammady, Zbys Fedorowicz, and Ahmed Elmagarmid. 2016. “Rayyan—a Web and Mobile App for Systematic Reviews.” *Systematic Reviews* 5 (1): 210. <https://doi.org/10.1186/s13643-016-0384-4>.

This preprint was submitted under the following conditions:

- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The deposited manuscript is in PDF format.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place.
- The authors agree that the approved manuscript will be made available under a [Creative Commons CC-BY](#) license.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.