cffr: Generate Citation File Format Metadata for R Packages

Diego Hernangómez[1]

1 Independent Researcher

Summary

The Citation File Format project (Druskat et al., 2021) defines a standardized format for providing software or datasets citation metadata in plaintext files that are easy to read by both humans and machines.

This metadata format is being adopted by GitHub as the primary format for its built-in citation support (GitHub, 2021). Other leading archives for scientific software, including Zenodo and Zotero (Druskat, 2021), have included as well support for CITATION.cff files in their GitHub integration.

The cffr package provides utilities to generate and validate these CITATION.cff files automatically for R (R Core Team, 2021) packages by parsing the DESCRIPTION file and the native R citation file. The package also includes utilities and examples for parsing components as persons and additional citations, as well as several vignettes which illustrate both the basic usage of the package as well as some more technical details about the metadata extraction process.

Statement of need

Citation of research software on research project is often omitted (Salmon et al., 2021) . Among many reasons why software is not cited, one is the lack of a clear citation information from package developers.

Some of the main reasons for citing software used on research are:

1. **Reproducibility**: Software and their versions are important information to include in any research project. It helps peers to understand and reproduce effectively the results of any work. Including versions is also crucial as a way of recording the context of your manuscript when software changes.
2. **Developer Credit:** On the context of Free and Open Source Software (FOSS), many of the software developers themselves are also researches. Receive credit for software development shouldn't be different from the credit received on other formats, as books or articles.

CITATION.cff files provides a clear citation rules for software. The format is easily readable by humans and also can be parsed by appropriate software. The adoption of GitHub of this format sends a strong message that research software is something worthy of citation, and therefore deserves credit.

The cffr package allow R software developers to create CITATION.cff files from the metadata already included on the package. Additionally, the package also include validation tools via the jsonvalidate package (FitzJohn et al., 2021), that allow developers to assess the validity of the file created using the latest CFF schema.json.

Acknowledgements

References

Boettiger, C., & Salmon, M. (2021). *codemetar: Generate 'CodeMeta' metadata for R packages*. https://doi.org/10.5281/zenodo.4748266

Druskat, S. (2021). *Making software citation easi(er) - The Citation File Format and its integrations*. https://doi.org/10.5281/zenodo.5529914

Druskat, S., Spaaks, J. H., Chue Hong, N., Haines, R., Baker, J., Bliven, S., Willighagen, E., Pérez-Suárez, D., & Konovalov, A. (2021). *Citation File Format*. https://doi.org/10.5281/zenodo.5171937

FitzJohn, R., Ashton, R., Hill, A., Schep, A., Lyttle, I., Woo, K., library), M. B. (Author. of bundled imjv, & library), E. P. (Author. of bundled A. (2021). *jsonvalidate: Validate 'JSON' schema*. https://CRAN.R-project.org/package=jsonvalidate

GitHub. (2021). About CITATION files. In *GitHub Docs*. https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/about-citation-files

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Salmon, M., Chamberlain, S., & Ram, K. (2021). *Make Your R Package Easier to Cite*. https://ropensci.org/blog/2021/02/16/package-citation/