

The Citation File Format (CFF): Why, what, how?

Scientific Software Registry Collaboration Workshop,
University of Maryland, College Park, MD, 2019-11-13

Stephan Druskat (German Aerospace Center (DLR),
Friedrich Schiller University Jena, Humboldt-Universität zu Berlin)



seit 1558



Citation File Format: Why?

Software citation has a metadata problem

- Growing awareness of the central role of software in research
- Consequence: treat software as a first-class citizen in research
- Ergo: cite software as you would cite a paper

Software citation has a metadata problem

- Growing awareness of the central role of software in research
- Consequence: treat software as a first-class citizen in research
- Ergo: cite software as you would cite a paper
- Citation-relevant metadata not readily available
- Available metadata not reliable
(e.g., “correct” authors & contributors != GitHub committers)

CITATION files

- “Put a file called CITATION in your source code repository, and include a BibTeX snippet” [1]
- Ambiguous file name
- BibTeX itself may be ambiguous, and not fit for software citation based on the Principles [2]

CITATION files

- “Put a file called CITATION in your source code repository, and include a BibTeX snippet” [1]
- Ambiguous file name
- BibTeX itself may be ambiguous, and not fit for software citation based on the Principles [2]
- **Idea:** Develop a standard format for CITATION files? (WSSSPE 5.1) [3],[4]
- **Outcome:** Citation File Format (CFF) [5]

What is the Citation File Format?
(Hint: “What’s in a name?”)

Features

- YAML 1.2 dictionary
- Human-/machine- -readable/-writable
- Self-descriptive (`CITATION.cff`, required `message` field)
- Modeled after the Software Citation Principles paper [2]
- Compatible with CodeMeta, BibTeX, etc.
- Tooling available

What does it look like?

```
cff-version: 1.1.0
message: "If you use MRT in your research, please cite it as below."
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: https://orcid.org/0000-0003-4925-7248
title: "My Research Tool"
version: 1.0.4
doi: 10.5281/zenodo.1234
date-released: 2017-12-18

# New in v1.1.0
identifiers:
  - type: "swh"
    value: "swh:1:rel:99f6850374dc6597af01bd0ee1d3fc0699301b9f"
```

What tooling is available?

- A schema for validation
- A validation tool
- A simple web form for manual completion (also as web service)
- A conversion tool (from/to CodeMeta, BibTeX, zenodo.json, ...)
- Others, e.g.:
 - A Maven plugin that creates CFF files from metadata
 - Object models in Java & Ruby
- Development supported by:
German Aerospace Center (DLR) & Netherlands eScience Center

How to use CFF?

Target groups

CFF has three main target groups:

1. Developers / Research Software Engineers
2. Researchers & other human “citors”
3. Downstream users, e.g., software registries

... can make their software citable in a simple manner:

1. Create `CITATION.cff` file in source code repository
2. Complete file with desired detailedness
(manually/semi-/automatically)
3. Update as required (manually/semi-/automatically)
4. Profit!

Researchers ...

... can easily identify software citation metadata:

1. Find `CITATION.cff` in a repository
2. Read file & understand its purpose (from `message` field)
3. Cite software (version) appropriately
4. Profit!

Registries & other downstream users ...

... can use CFF as source for their own metadata model/format:

1. Check for `CITATION.cff` in ingested data
2. (Optionally: validate `CITATION.cff` + citation metadata)
3. Convert metadata to own model/format?
(Alternatively supplied by user before ingestion)
4. Profit!

Thank you!

Thanks for listening, and thanks to ...

... CFF contributors; Alice Allen, Tom Morrell & Michael Hucka for the invitation; the Software Sustainability Institute (UK) for funding activities around CFF.

citation-file-format.github.io

github.com/citation-file-format/citation-file-format

stephan.druskat@dlr.de

Twitter: [@stdruskat](https://twitter.com/stdruskat)

ORCID [0000-0003-4925-7248](https://orcid.org/0000-0003-4925-7248)

Slides: [doi:10.6084/m9.figshare.10296917](https://doi.org/10.6084/m9.figshare.10296917)

Appendix

References

- [1] R. Wilson, “Encouraging citation of software – introducing CITATION files.” 2013 [Online]. Available: <https://goo.gl/7gZuSY>.
- [2] Arfon M. Smith, Daniel S. Katz, Kyle E. Niemeyer, and FORCE11 Software Citation Working Group, “Software citation principles,” *PeerJ Computer Science*, vol. 2, no. e86, 2016 [Online]. Available: <https://doi.org/10.7717/peerj-cs.86>.
- [3] Stephan Druskat, “Should CITATION files be standardized?” in *Proceedings of the Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE5.1)*, 2017 [Online]. Available: <https://doi.org/10.6084/m9.figshare.3827058.v4>.
- [4] Stephan Druskat, Radovan Bast, Neil Chue Hong, Alexander Konovalov, Andrew Rowley, and Raniera Silva, “A standard format for CITATION files,” *Software and research*. Dec-2017 [Online]. Available: <https://www.software.ac.uk/index.php/blog/2017-12-12-standard-format-citation-files>.
- [5] Stephan Druskat, Jurriaan H. Spaaks, Neil Chue Hong, Robert Haines, and James Baker, “Citation File Format (CFF),” Aug. 2018 [Online]. Available: <https://doi.org/10.5281/zenodo.1003149>.