

# Citation File Format – status and current challenges

Stephan Druskat

German Aerospace Center (DLR)

Institute for Software Technology – Sustainable Software Engineering Group

Berlin, Germany

Software Citation Workshop 2022

Harvard Center for Astrophysics, Cambridge, MA

15 August 2022

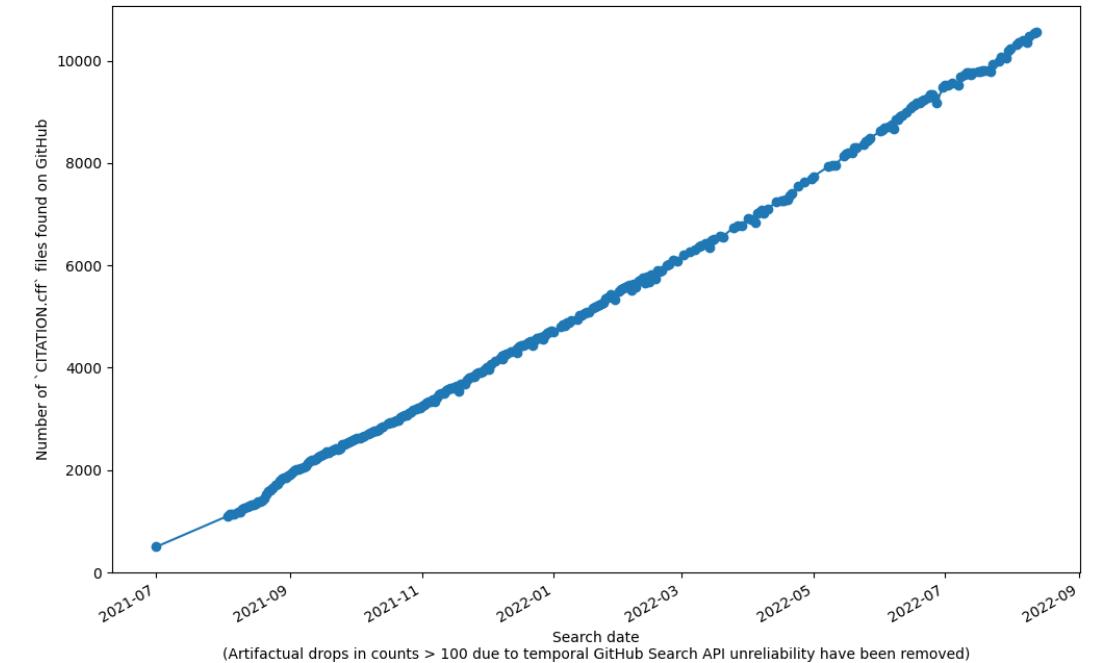
License CC-BY-4.0 International

DOI [10.5281/zenodo.6990773](https://doi.org/10.5281/zenodo.6990773)



# Citation File Format (CFF): Status

- CFF: Human- and machine-readable and –writable file format for software citation metadata (YAML)
- > 10k CITATION.cff files on GitHub
- Open source project: schema, docs, tooling
  - Creation, editing/updating, validation, conversion
- Integrations: GitHub, Zenodo, Zotero, JabRef
- Project leads at *DLR, Netherlands eScience Center* + independent projects under CFF umbrella
- Formal governance setup in progress



# CFF – Challenges I: Funding / development time

- Volunteer-driven community project
- Growth in user community and subprojects, but not funding
- "Small" funds for:
  - Events / travel ([\*Software Sustainability Institute\* fellowship](#) (S. Druskat))
  - Governance development (Code for Science and Society [\*Digital Infrastructure Incubator\*](#))
  - Development ([\*CampusSOURCE\* award](#) price money)
- "Goodwill funding" as part of day jobs

## Potential solutions

- Short term: Institutional funding for development via contract work (in planning)
- Long term: Joining a foundation / charitable org?



## CFF - Challenges II: Instance validity

- **Files invalid** as per the CFF [schema](#)
- **Metadata invalid** (e.g., leftover template values, files not updated)

### Potential solutions

- Improved documentation
- More / improved tooling (e.g., validation website, feature development for [initialization website](#))
- Automation (e.g., automated validation via CI)



## CFF – Challenges III: Metadata exchange

- **Conversion** from / to CFF implemented for several formats (e.g. [CodeMeta](#))
- Lack of **interfaces** with platforms (e.g., publication repositories, journal submission systems)
- Lack of usable **automation**

### Potential solutions

- Develop further direct interfaces ([Dataverse](#) in progress)
- Develop automation solutions for end users (e.g., [HERMES](#))
  - Harvesting, collation, depositing of metadata (and software) in InvenioRDM & Dataverse repositories



## CFF – Challenges IV: Culture and policy

- General **cultural issues** apply (citation metadata provision and usage)
- No strong, ubiquitous **incentives / mandate** for provision of software citation metadata
- Citations not to software (preferred-citation)

### Potential solutions

- Mandate with stakeholders (funders, institutions, research councils, etc.)
- Publicize
- Provide easy-to-use documentation including rationale (e.g., [cite.research-software.org](https://cite.research-software.org), [citesoftware.org](https://citesoftware.org))



# Thank you

## Citation File Format

- [citation-file-format.github.io](https://citation-file-format.github.io)
- [github.com/citation-file-format](https://github.com/citation-file-format)

## HERMES

- [software-metadata.pub](https://software-metadata.pub)

## Me

- [stephan.druskat@dlr.de](mailto:stephan.druskat@dlr.de), [sdruskat.net](http://sdruskat.net)
- Twitter: [@stdruskat](https://twitter.com/stdruskat)
- iD <https://orcid.org/0000-0003-4925-7248>

**Slides:** <https://doi.org/10.5281/zenodo.6990773>, CC BY 4.0

