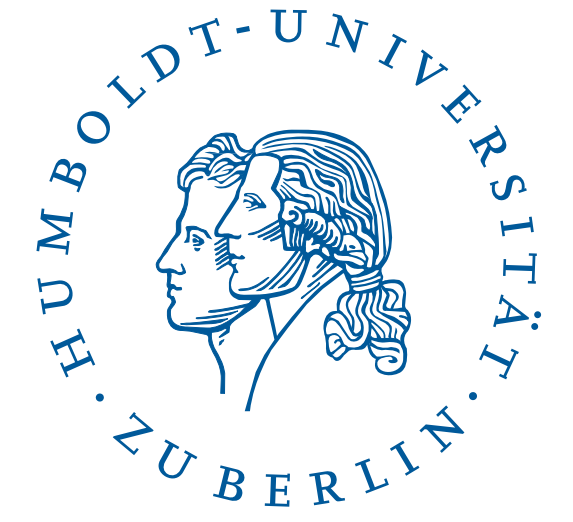


# (Forschungs-) Software zitierfähig machen

Stephan Druskat

Humboldt-Universität zu Berlin, Institut für deutsche Sprache und Linguistik



## Software als Forschungsergebnis

- Softwareentwicklung ist häufig **integraler Bestandteil** moderner Forschung.
- Forschungssoftware unterstützt Gewinnung, Verarbeitung und Auswertung empirischer Daten, Modellierung und Simulation komplexer Prozesse.
- **Software hat maßgeblichen Einfluss** auf die Qualität der erzielten Forschungsergebnisse.
- Zu selten jedoch wird Softwareentwicklung als **wissenschaftliche Leistung** anerkannt.

Die internationale Bewegung der **Research Software Engineers** ([rse.ac.uk](http://rse.ac.uk), [de-rse.org](http://de-rse.org)) arbeitet daran, dass hier ein Umdenken eintritt. Dabei wurde festgestellt, dass die **schlechte Verankerung von Software im wissenschaftlichen Reputationssystem** auch darin begründet liegt, dass Software häufig **nicht oder falsch zitiert** wird (Tabelle 1, Tabelle 2). Gründe dafür liegen unter anderem in **unzureichend aufbereiteten und schlecht auffindbaren Metadaten**, die für eine korrekte Zitierung von Software relevant sind.

- Das **Citation File Format** (CFF, [1]) bietet eine maschinenlesbare und menschenles- und schreibbare Form zur Aufbereitung und Bereitstellung von Zitationsmetadaten für Software.
- Zentrales Hinterlegen von formatierten CITATION.cff Dateien (vgl. LICENSE, README, etc.).
- Software wird für die Zitierung **sichtbarer und zugänglicher** gemacht.
- **Nachnutzung** durch Einbettung in UI, Konvertierung und Bereitstellung anderer Formate, etc.

DOI 10.5281/zenodo.1172282

## Software in Publikationen

Typ	Anzahl (n = 286)	Prozent
Zitierte Publikation (z.B. Software-Paper)	105	37
Zitiertes Handbuch	6	2
Name oder Website erwähnt	15	5
Zitiert wie Instrument	53	19
URL im Text	13	5
Name im Text	90	31
Keine namentliche Erwähnung	4	1

Tabelle 1: Inkonsistente Erwähnung von Software in Publikationen, nach [2].

Frage: "In general, when your software contributes to a paper, are you acknowledged in that paper?"

Land	"Ja"	"Nein"
DE	212	88
NL	58	17
UK	139	40
US	92	38
ZAF	8	11
Gesamt	509	194

Tabelle 2: Umfrage 2017. Quelle: <http://bit.ly/ssi-intl-survey>.

## Citation File Format

- CITATION-Dateien [3] beinhalten Metadaten zur Softwarezitierung, hinterlegt im Wurzelverzeichnis des Quellcode-Repositories.
- Die ursprünglich freie Formatierung – mit Ausnahme der Dateien für R-Pakete [4] – verhindert eine maschinelle Nachnutzung der Metadaten.
- Diese **Nachnutzung ist ausschlaggebend** für eine erhöhte Sichtbarkeit der Metadaten und die Sicherstellung korrekter Zitierungen von Software.

Das **Citation File Format (CFF)** löst dies und stellt die Umsetzung von durch die wissenschaftliche Gemeinschaft erarbeiteten **Best Practices** sicher.

- Konkret: für **konsistente, nachvollziehbare, eindeutige, persistente und zugängliche Zitierung von Software** notwendige Datenpunkte müssen in CITATION.cff vorhanden sein:
  - Erforderlich sind mindestens ("**Software Citation Principles**" [5]): Titel, Autoren, Veröffentlichungsdatum, Versionsinformation.
  - Optimal sind: ein eindeutiger Identifikator (DOI o.ä.), URL des Quellcode-Repositories; notfalls stattdessen anzugeben: Versionsindikator (Commit Hash/Revisionsnummer), Kontaktinformationen.
- Diese Informationen können in CFF hinterlegt werden, primäre vier Schlüssel sind erforderlich für erfolgreiche Validierung.
- Zusätzlich: **Anweisung** in welchen Fällen die Software wie zitiert werden soll, weitere **sekundäre Publikationsmetadaten**.

## Format

- Implementiert in YAML 1.2 [6]
- Blogpost: <http://bit.ly/ssi-citation-file-format>
- Spezifikation: <https://citation-file-format.github.io/>
- Validierung: <https://github.com/citation-file-format/schema>
- Referenzimplementierung Reader/Mapper (Java): <https://github.com/citation-file-format/cff-reader-java/>
- Portal: <https://cite.research-software.org>

## Beispiel

```
cff-version: 1.0.3
message: Please cite the Toolbox Text Modules for Pepper with the metadata given below!
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: https://orcid.org/0000-0003-4925-7248
title: pepperModules-ToolboxTextModules
version: 1.0.0
doi: 10.5281/zenodo.1162208
date-released: 2018-01-24
repository-code: https://github.com/sdruskat/pepperModules-ToolboxTextModules
```

## Infrastruktur und Community

- CFF benötigt zur Entfaltung des vollen Potenzials **technische Infrastruktur** (Validierung, Lesen, Konvertierung, etc.)
- Diese technische Infrastruktur ist im Entstehen (s. <https://github.com/citation-file-format/>): Hackathon 2017, Hackathon 2018
- CFF wird **gemeinschaftlich mitgetragen durch internationale Communities**, z.B. *FORCE11 Software Citation Implementation Working Group*, *WSSSPE (Working for sustainable software for science: practice and experiences)*

## Referenzen

- [1] S. Druskat, "Citation File Format (CFF)," Dec. 2017. [Online]. Available: <https://doi.org/10.5281/zenodo.1120389>
- [2] J. Howison and J. Bullard, "Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature," *Journal of the Association for Information Science and Technology*, vol. 67, no. 9, pp. 2137–2155, 2016. [Online]. Available: <http://dx.doi.org/10.1002/asi.23538>
- [3] R. Wilson, "Encouraging citation of software - introducing CITATION files," 2013, (accessed 17 July 2017). [Online]. Available: <http://bit.ly/2ECSdEQ>
- [4] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008, ISBN 3-900051-07-0. [Online]. Available: <http://www.R-project.org>
- [5] A. M. Smith, D. S. Katz, K. 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>
- [6] O. Ben-Kiki, C. Evans, and I. dot Net, "YAML Aint Markup Language (YAML) Version 1.2. 3rd Edition, Patched at 2009-10-01," 2009, accessed 04 October 2017. [Online]. Available: <http://yaml.org/spec/1.2/spec.html>