

RESEARCH SOFTWARE AND CITATION

Stephan Druskat

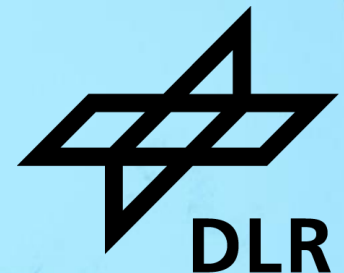
German Aerospace Center (DLR), Institute for Software Technology

DOI [10.5281/zenodo.7620412](https://doi.org/10.5281/zenodo.7620412)

License [CC-BY-4.0 International](https://creativecommons.org/licenses/by/4.0/)



Software
Sustainability
Institute





Stephan Druskat: "Research software and citation",
Forschungsseminar "Aktuelle Themen des Research Software Engineering", Inst. f. Informatik u. Computational Science, Universität Potsdam, 2023-02-08

Who / what am I?

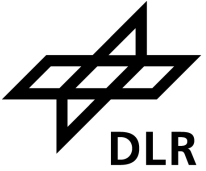


- MA in English, modern German literature, linguistics
- RSE in linguistics (2009-2020)
- Fellow of the [Software Sustainability Institute](#) (since 2018)
- Doctoral researcher in computer science (since 2019)



- Project lead [Citation File Format](#) 
- Co-founder / inaugural board member of [de-RSE](#) 
- PI [HERMES](#) project 

German Aerospace Center (DLR) Institute for Software Technology



DLR institute for **Software Research**, **Software Engineering**, **Artificial Intelligence** and **Scientific Computing**

Ca. 140 staff across our sites:

- Köln-Porz
- Berlin-Adlershof
- Braunschweig
- Oberpfaffenhofen
- Bremen (ECOMAT)
- Hamburg

<https://www.DLR.de/sc>



Institute for Software Technology Sustainable Software Engineering



We improve the quality of software at DLR, within Helmholtz, and worldwide.

We research, develop and implement software engineering solutions.

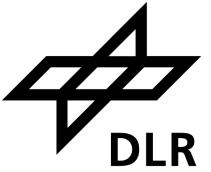
We advocate for software to be acknowledged as a valid research output.



HELMHOLTZ
SPITZENFORSCHUNG FÜR
GROSSE HERAUSFORDERUNGEN



Research software and citation



1. Research software in the academic ecosystem
2. Software citation
3. Software publication as a prerequisite for citation

The background of the slide is a high-resolution photograph of a satellite in orbit. The satellite is a rectangular platform with two long, multi-panel solar arrays extending horizontally from its central body. It is positioned in the center-right of the frame, appearing to fly over the Earth. The Earth's surface is visible, showing a mix of green landmasses and blue oceans, with a thin white layer of clouds. The curvature of the Earth is visible on the right side, where the atmosphere transitions into the blackness of space.

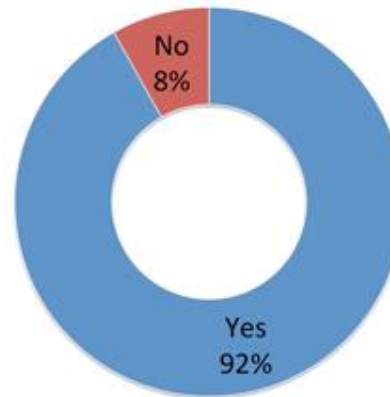
RESEARCH SOFTWARE IN THE ACADEMIC ECOSYSTEM

Research software in the academic ecosystem: The status of research software

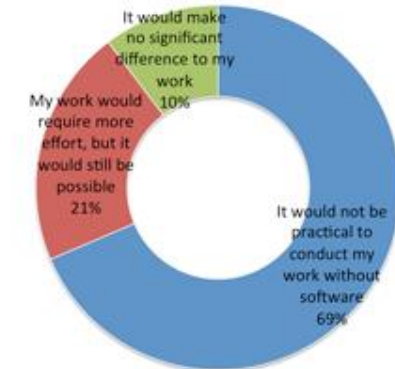


1. Importance for research [1]
2. Acknowledgment as a valid research output [2]

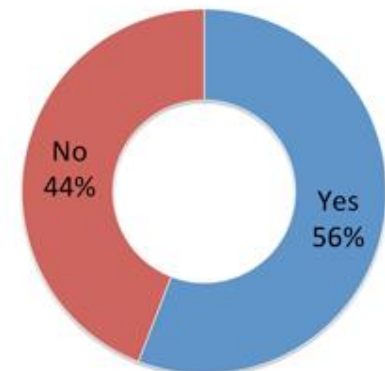
Do you use research software?



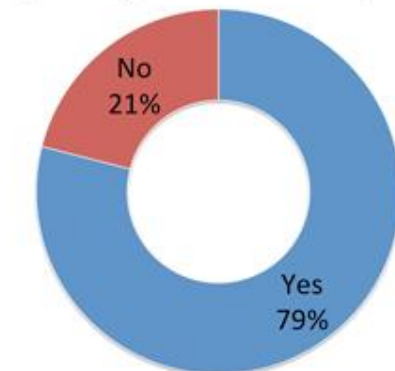
What would happen if you could no longer use research software?



Do you develop your own research software?



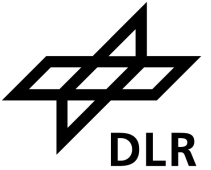
If you develop software, have you received any training in software development?



[1] Hettrick, Simon. "UK Research Software Survey 2014." Zenodo, February 23, 2018.
<https://doi.org/10.5281/zenodo.1183562>.

[2] Jay, Caroline, Robert Haines, and Daniel S. Katz. "Software Must Be Recognised as an Important Output of Scholarly Research." International Journal of Digital Curation 16, no. 1 (April 26, 2021): 6.
<https://doi.org/10.2218/ijdc.v16i1.745>.

Research software in the academic ecosystem: The status of research software



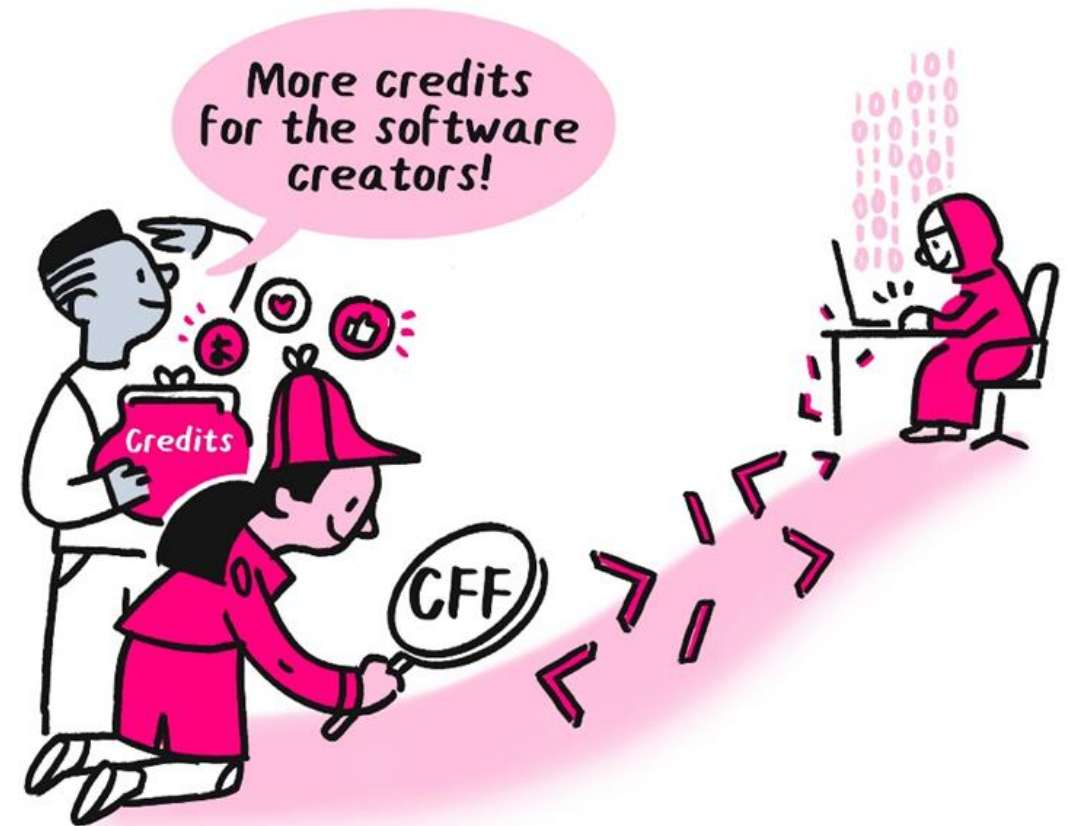
The origin of the data, organisms, materials and software used in the research process is disclosed and the reuse of data is clearly indicated; **original sources are cited.** The nature and the scope of research data generated during the research process are described. Research data are handled in accordance with the requirements of the relevant subject area. **The source code of publicly available software must be persistent, citable and documented.** Depending on the particular subject area, it is an **essential part of quality assurance that results or findings can be replicated or confirmed** by other researchers (for example with the aid of a **detailed description of materials and methods**). [3]

[3] Deutsche Forschungsgemeinschaft. "Guidelines for Safeguarding Good Research Practice. Code of Conduct," September 15, 2019. <https://doi.org/10.5281/ZENODO.3923602>. p. 14.

Research software in the academic ecosystem: Citation

Software citation serves:

1. Good research practice
2. Attribution and credit for RSEs
3. Reproducibility of research results
4. Sustainability of research software [4]



[4] Druskat, Stephan, Daniel S. Katz, and Ilian T. Todorov. "Research Software Sustainability and Citation." In *2021 IEEE/ACM International Workshop on Body of Knowledge for Software Sustainability (BoKSS)*, 1–2, 2021. <https://doi.org/10.1109/BoKSS52540.2021.00008>.

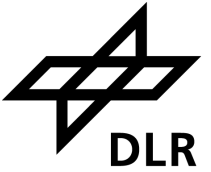
Stephan Druskat: "Research software and citation",
Forschungsseminar "Aktuelle Themen des Research Software Engineering", Inst. f. Informatik u. Computational Science, Universität Potsdam, 2023-02-08

More credits for the software creators. The Turing Way project illustration by Scriberia. Zenodo.
<https://doi.org/10.5281/zenodo.3332807>
License: CC BY-4.0

A satellite with two long solar panel arrays is shown in orbit above the Earth. The satellite is gold-colored with various instruments and antennas. The Earth below shows a coastline with green land and blue water, partially obscured by white clouds. The curvature of the Earth and the blackness of space are visible in the background.

SOFTWARE CITATION

Software citation: Requirements



Software citation requires:

1. Understanding how software citation works
2. Culture change

Software citation: Principles [5]





Software citation principles

Arfon M. Smith^{1**}, Daniel S. Katz^{2**}, Kyle E. Niemeyer^{3**} and FORCE11 Software Citation Working Group

¹ GitHub, Inc., San Francisco, California, United States
² National Center for Supercomputing Applications & Electrical and Computer Engineering Department & School of Information Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States
³ School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University, Corvallis, Oregon, United States
^{**} These authors contributed equally to this work.

ABSTRACT

Software is a critical part of modern research and yet there is little support across the scholarly ecosystem for its acknowledgement and citation. Inspired by the activities of the FORCE11 working group focused on data citation, this document summarizes the recommendations of the FORCE11 Software Citation Working Group and its activities between June 2015 and April 2016. Based on a review of existing community practices, the goal of the working group was to produce a consolidated set of citation principles that may encourage broad adoption of a consistent policy for software citation across disciplines and venues. Our work is presented here as a set of software citation principles, a discussion of the motivations for developing the principles, reviews of existing community practice, and a discussion of the requirements these principles would place upon different stakeholders. Working examples and possible technical solutions for how these principles can be implemented will be discussed in a separate paper.

Subjects Digital Libraries, Software Engineering
Keywords Software citation, Software credit, Attribution

SOFTWARE CITATION PRINCIPLES

The main contribution of this document are the software citation principles, written fairly concisely in this section and discussed further later in the document (see Discussion). In addition, we also motivate the creation of these principles (see Motivation), describe the process by which they were created (see Process of Creating Principles), summarize use cases related to software citation (see Use Cases), and review related work (see Related Work). We also lay out the work needed to lead to these software citation principles being applied (see Future Work).

1. Importance: Software should be considered a legitimate and citable product of research. Software citations should be accorded the same importance in the scholarly record as citations of other research products, such as publications and data; they should be included in the metadata of the citing work, for example in the reference list of a journal article, and should not be omitted or separated. Software should be cited on the same basis as any other research product such as a paper or a book, that is, authors should cite the appropriate set of software products just as they cite the appropriate set of papers.

Submitted 24 June 2016
Accepted 23 August 2016
Published 19 September 2016
Corresponding author
Daniel S. Katz, d.katz@ices.org
Academic editor
Silvio Peroni
DOI 10.7717/peerj-cs.86
© Copyright
2016 Smith et al.
Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

How to cite this article: Smith et al. (2016), Software citation principles, PeerJ Comput. Sci. 2(08): DOI 10.7717/peerj-cs.86

1. Importance

Software is cited like papers are cited.

2. Credit and attribution

3. Unique identification

4. Persistence

5. Accessibility

Citation allows access to software and metadata.

6. Specificity

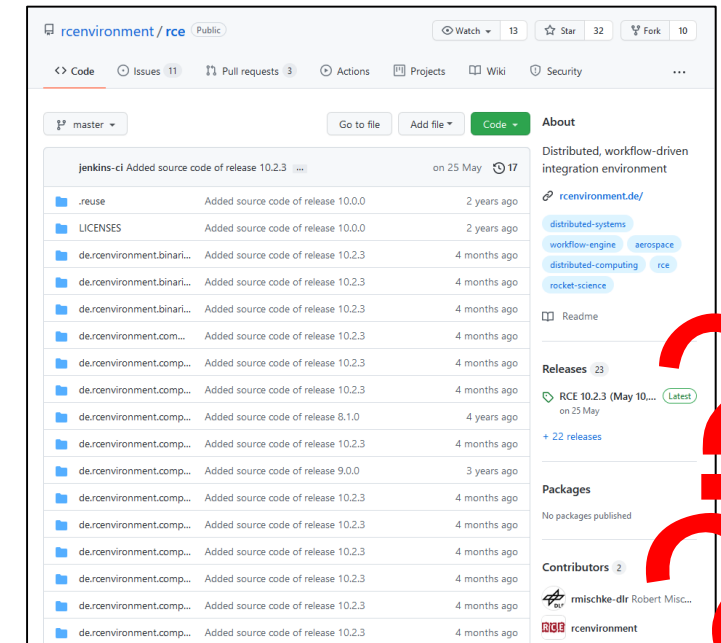
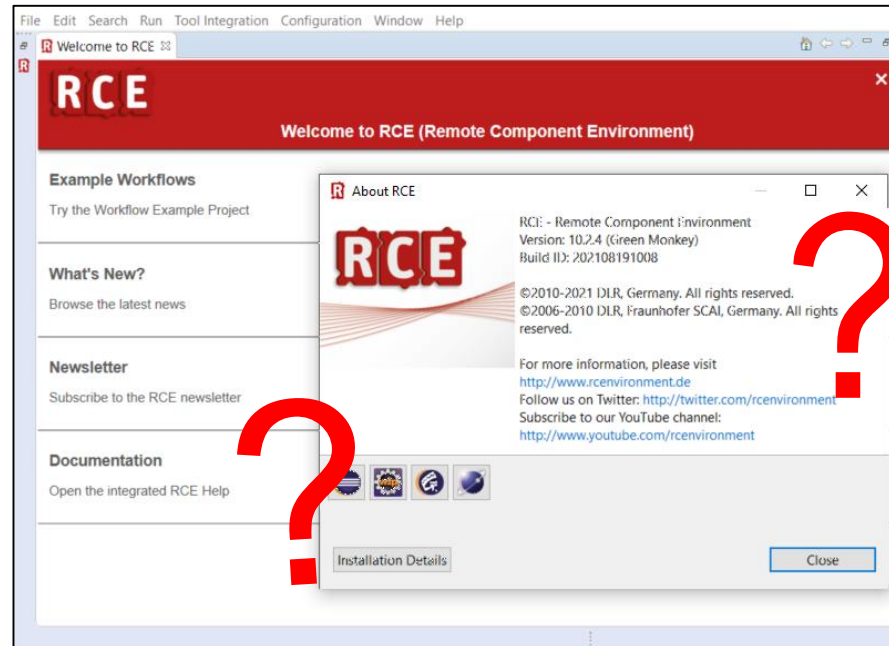
Citation identifies the software version used in research.

[5] A. M. Smith, D. S. Katz, K. E. Niemeyer, and FORCE11 Software Citation Working Group, “**Software citation principles**,” *PeerJ Comput. Sci.*, vol. 2, no. e86, 2016, doi: [10.7717/peerj-cs.86](https://doi.org/10.7717/peerj-cs.86).

Software citation: Challenges [6]

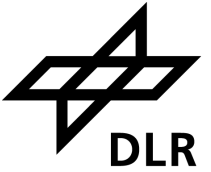
Metadata?

- Software name
- Authors
- Version
- Publisher
- Publication date
- Identifier



[6] Katz, Daniel S., Daina Bouquin, Neil P. Chue Hong, Jessica Hausman, Catherine Jones, Daniel Chivvis, Tim Clark, et al. "Software Citation Implementation Challenges," May 21, 2019. <http://arxiv.org/abs/1905.08674>.

Software citation: Metadata



Correct and complete software citation metadata must be identified by the software project!

- Authors vs. contributors
- Versioning
- (Persistent) identifiers

Software citation metadata solutions: Citation File Format



Correct and complete software citation metadata must be provided by the software project!

CITATION.cff

Citation File Format (CFF) [7]:

- authoritative
- controllable
- principled

<https://citation-file.format.github.io>

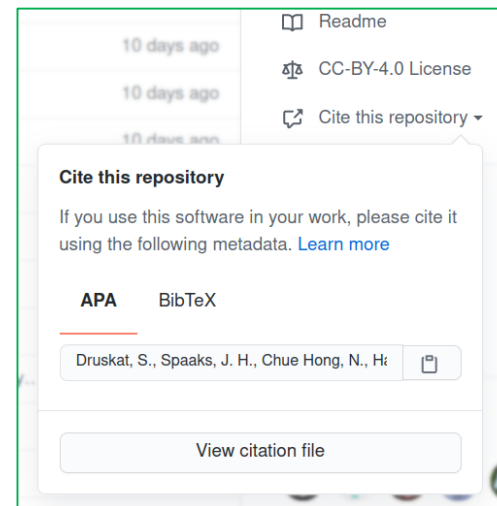
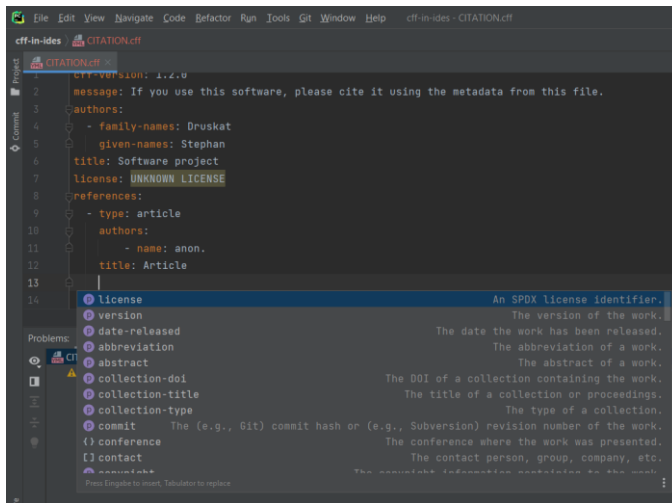
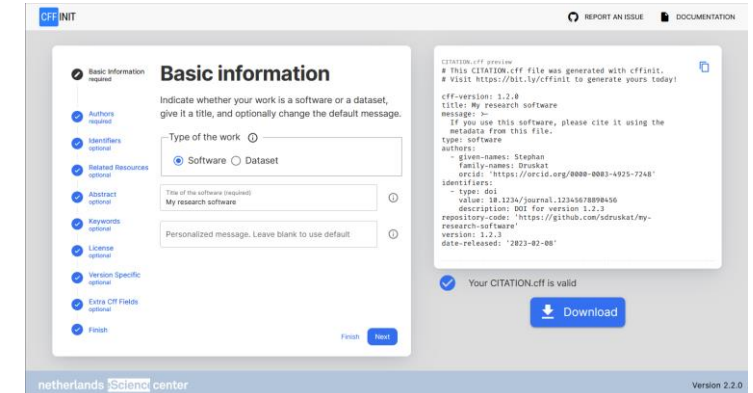
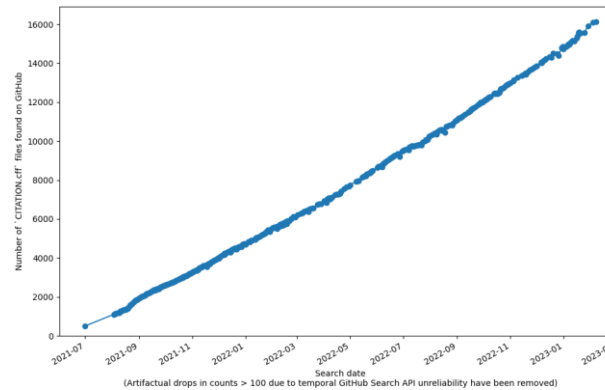
```
cff-version: 1.2.0
message: If you use this software, please cite it using these metadata.
title: My Research Software
abstract: This is my awesome research software. It does many things.
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: "https://orcid.org/0000-0003-4925-7248"
version: 0.11.2
date-released: "2021-07-18"
identifiers:
  - description: This is the collection of archived snapshots of all versions of My Research Software
    type: doi
    value: "10.5281/zenodo.123456"
  - description: This is the archived snapshot of version 0.11.2 of My Research Software
    type: doi
    value: "10.5281/zenodo.123457"
license: Apache-2.0
repository-code: "https://github.com/citation-file-format/my-research-software"
```

[7] Druskat, Stephan, Spaaks, Jurriaan H., Chue Hong, Neil, Haines, Robert, Baker, James, Bliven, Spencer, Willighagen, Egon, Pérez-Suárez, David, and Konovalov, Alexander. "Citation File Format," August 9, 2021. <https://doi.org/10.5281/ZENODO.1003149>.

Software citation metadata solutions: Citation File Format

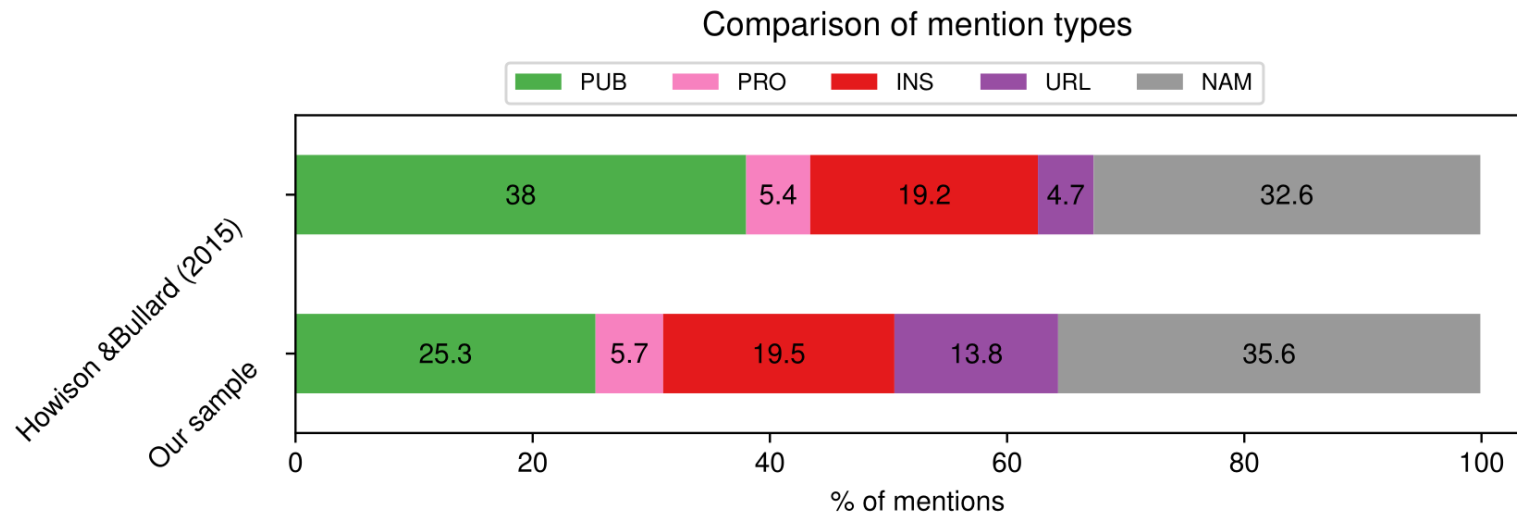


- 16,000+ CFF files on GitHub
- Tool support (e.g. [cffinit](#))
- Platform support (GitHub, Zenodo)
- IDE support



Software citation: Culture change

Software is currently mentioned, not cited!



From [9]: Mention types of software in publication in % in data from [8] and our own data.
PUB: cites publication; PRO: cites project name/website; INS: instrument-like; URL: URL in text; NAM: in-text name only.

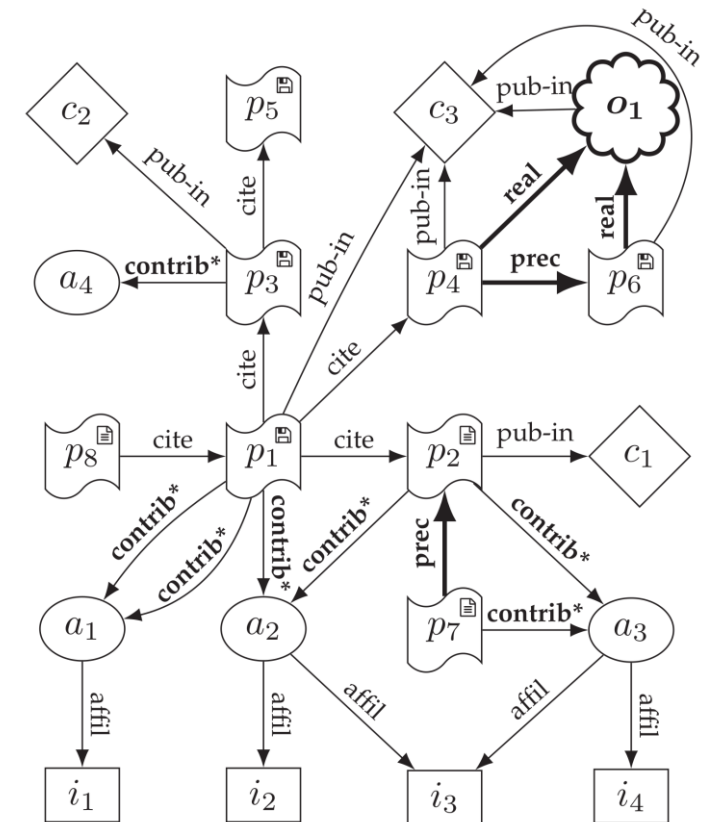
[8] J. Howison and J. Bullard, "Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature," *Journal of the Association for Information Science and Technology*, vol. 67, no. 9, pp. 2137–2155, May 2015, doi: [10.1002/asi.23538](https://doi.org/10.1002/asi.23538).

[9] S. Druskat, N. P. Chue Hong, P. Kornek, S. Buzzard, and A. Konovalov, "Don't mention it: challenges to using software mentions to investigate citation and discoverability," *PeerJ Computer Science*, forthcoming.

Software citation: Culture change

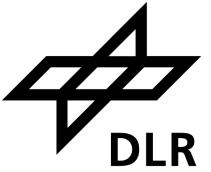
Software needs to cite its own references!

- Software as a valid research output
- Citation graphs [10]
- E.g., in `CITATION.cff` > references



[10] Druskat, Stephan. "Software and Dependencies in Research Citation Graphs." *Computing in Science & Engineering* 22, no. 2 (March 2020): 8–21. <https://doi.org/10.1109/MCSE.2019.2952840>.

Software citation and research software sustainability



Software citation supports software sustainability [11]:

1. By providing metrics and incentives
2. Enables reuse and adaptation

Software citation requires sustainability, or at least persistence.

[11] Druskat, Stephan, Daniel S. Katz, and Ilian T. Todorov. "Research Software Sustainability and Citation." In *2021 IEEE/ACM International Workshop on Body of Knowledge for Software Sustainability (BoKSS)*, 1–2, 2021.

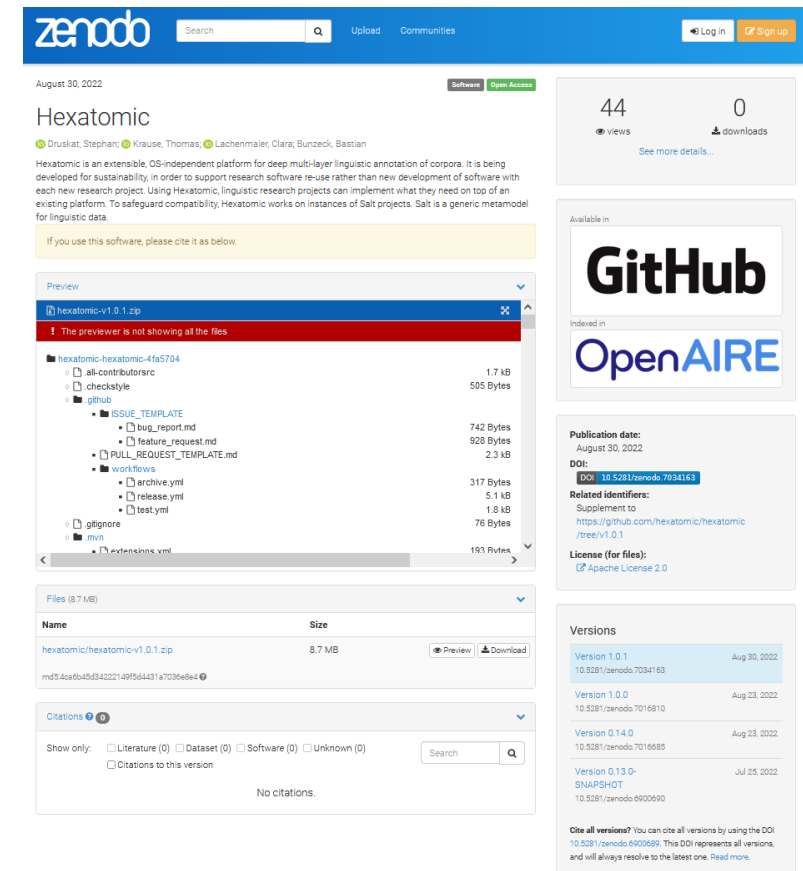
<https://doi.org/10.1109/BoKSS52540.2021.00008>.

The background of the slide is a high-resolution photograph of a satellite in orbit above Earth. The satellite is a rectangular platform with two long, parallel solar panel arrays extending outwards. The panels are covered in a grid of small, square solar cells. The satellite's central body is gold-colored and features various instruments and antennas. Below the satellite, the Earth's surface is visible, showing a mix of green landmasses and blue oceans, partially obscured by white clouds. The curvature of the Earth is visible at the bottom of the frame.

SOFTWARE PUBLICATION

Software publication as a prerequisite for citation

- **Software publication enables software citation according to principles (persistence)**
- FAIR research software through (metadata) publication
- Supports academic evaluation*
- Practice of research software publication:
 - Publication in a publication repository (Zenodo)
 - Publication in a software journal (JOSS, JORS, ...)
 - “Publication” as a paper
 - “Publication” in the methods section of a paper



The screenshot shows the Zenodo page for the software 'Hexatomic'. The page includes a search bar, navigation links for 'Upload' and 'Communities', and a 'Log in' / 'Sign up' button. The main content area displays the software title 'Hexatomic', the authors 'Druskat, Stephan; Krause, Thomas; Lachenmaier, Clara; Bunzeck, Bastian', and a brief description. It also shows a file tree for 'hexatomic-v1.0.1.zip' with subfolders like 'all-contributorsrc', 'checkstyle', 'github', 'workflows', and 'tests'. A table lists the files and their sizes. The page also features a 'Citations' section with filters for 'Literature (0)', 'Dataset (0)', 'Software (0)', and 'Unknown (0)'. On the right side, there are statistics for views (44) and downloads (0), and logos for 'Available in GitHub' and 'OpenAIRE'. The 'Publication date' is August 30, 2022, and the DOI is 10.5281/zenodo.7034163. The license is Apache License 2.0. The 'Versions' section lists several versions of the software, including 'Version 1.0.1', 'Version 1.0.0', 'Version 0.14.0', and 'Version 0.13.0-SNAPSHOT'.

Continuous software publication: Citation File Format > Zenodo

March 16, 2022 Software Open Access

sdruskat/campussource: v0.1.0

Stephan Druskat

A release without a CFF file.

Preview

- campussource-0.1.0.zip
- sdruskat-campussource-a46ecd3
 - README.md

49 Bytes

```
1 cff-version: 1.2.0
2 message: "If you use this software, please cite it as below."
3 authors:
4 - family-names: "Druskat"
5   given-names: "Stephan"
6   orcid: "https://orcid.org/0000-0003-4925-7248"
7 title: "CampusSource Example Deposit"
8 version: 0.2.0
9 doi: 10.5281/zenodo.1035710
10 date-released: 2022-03-16
11 url: "https://www.campussource.de/events/e2203hagen/#Programm"
```

March 16, 2022 Software Open Access

CampusSource Example Deposit

Druskat, Stephan

This is a release WITH a CITATION.cff file :tada:.

If you use this software, please cite it as below.

Preview

- campussource-0.2.0.zip
- sdruskat-campussource-1
 - CITATION.cff
 - README.md

Versions

Version 0.2.0	Mar 16, 2022
10.5072/zenodo.1035737	
Version 0.1.0	Mar 16, 2022
10.5072/zenodo.1035711	

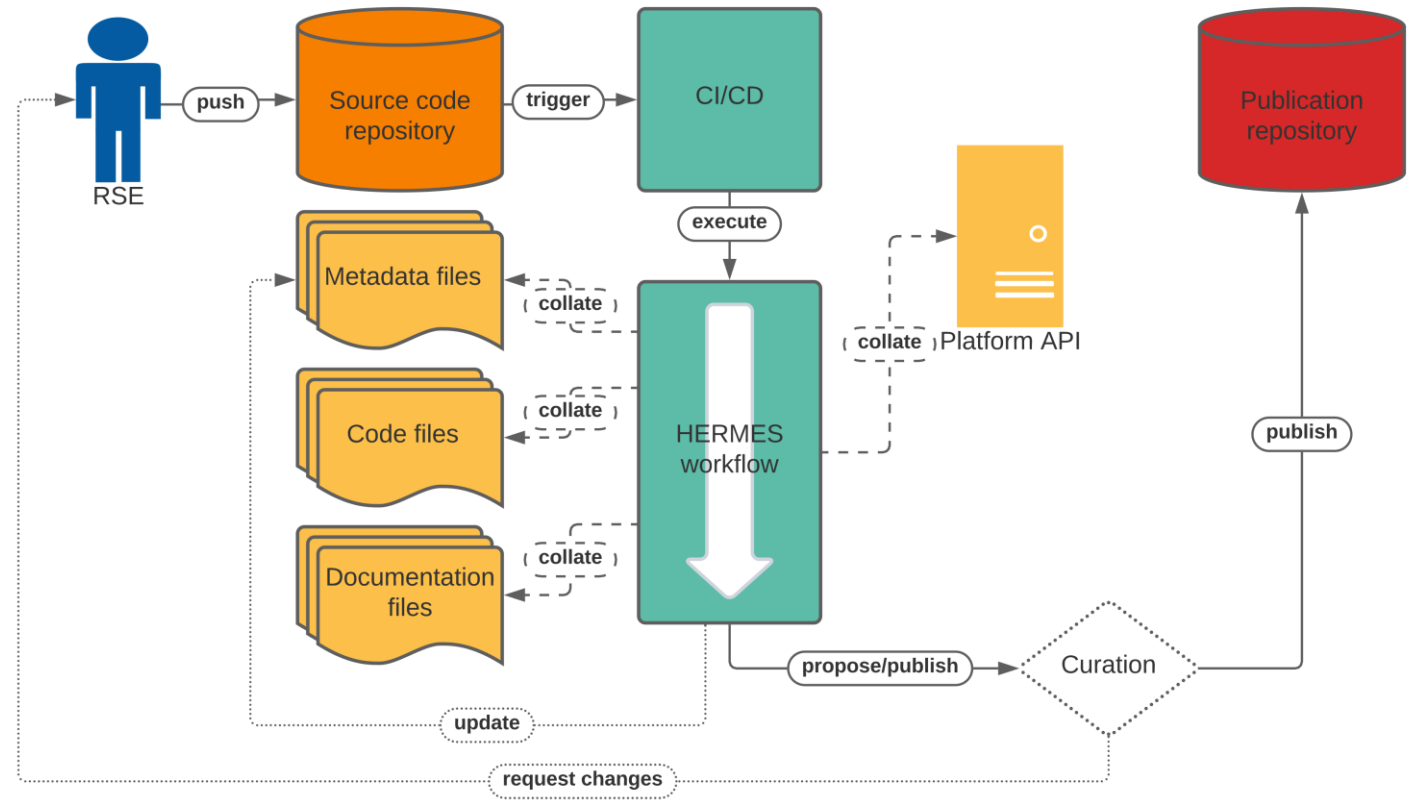
Cite all versions? You can cite all versions by using the DOI [10.5072/zenodo.1035710](https://doi.org/10.5072/zenodo.1035710). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Advanced continuous software publication

HERMES: <https://software-metadata.pub>



- Automating software publication via continuous integration
- Harvest and merge existing metadata
- Proactive pushes
- Curation & FAIR metadata-only publication (closed source) possible
- InvenioRDM and Dataverse support



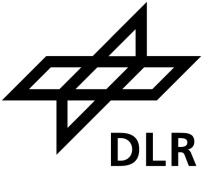
HERMES workflow, high-level view, HERMES project ([CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/))

Conclusion



1. Software citation is a requirement, not an option.
2. The principles of software citation are known, their implementation requires a culture change:
 - Researchers must cite, not mention, the software they use as they would cite a paper.
 - RSEs/software creators must provide correct and complete citation metadata.
 - RSEs/software creators should publish their software (releases).
3. Solutions exist.
 - **Software metadata:** Citation File Format and support in reference managers, publication repositories.
 - **Software publication:** Publication repositories with Citation File Format support, or automated software publication via CI.

Thanks!



Projects

<https://citation-file-format.github.io/> | <https://software-metadata.pub>

Contact

stephan.druskat@dlr.de | Fediverse: [@sdruskat@scholar.social](https://scholar.social/@sdruskat)

ORCID: <https://orcid.org/0000-0003-4925-7248>