

SOFTWARE PUBLICATION AND CITATION

Stephan Druskat

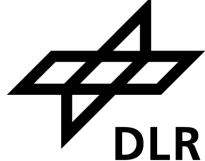
Institute for Software Technology (SC-IVS), Sustainable Software Engineering

WAW Software Engineering X, Berlin, 2023-09-13

License CC-BY-4.0 International

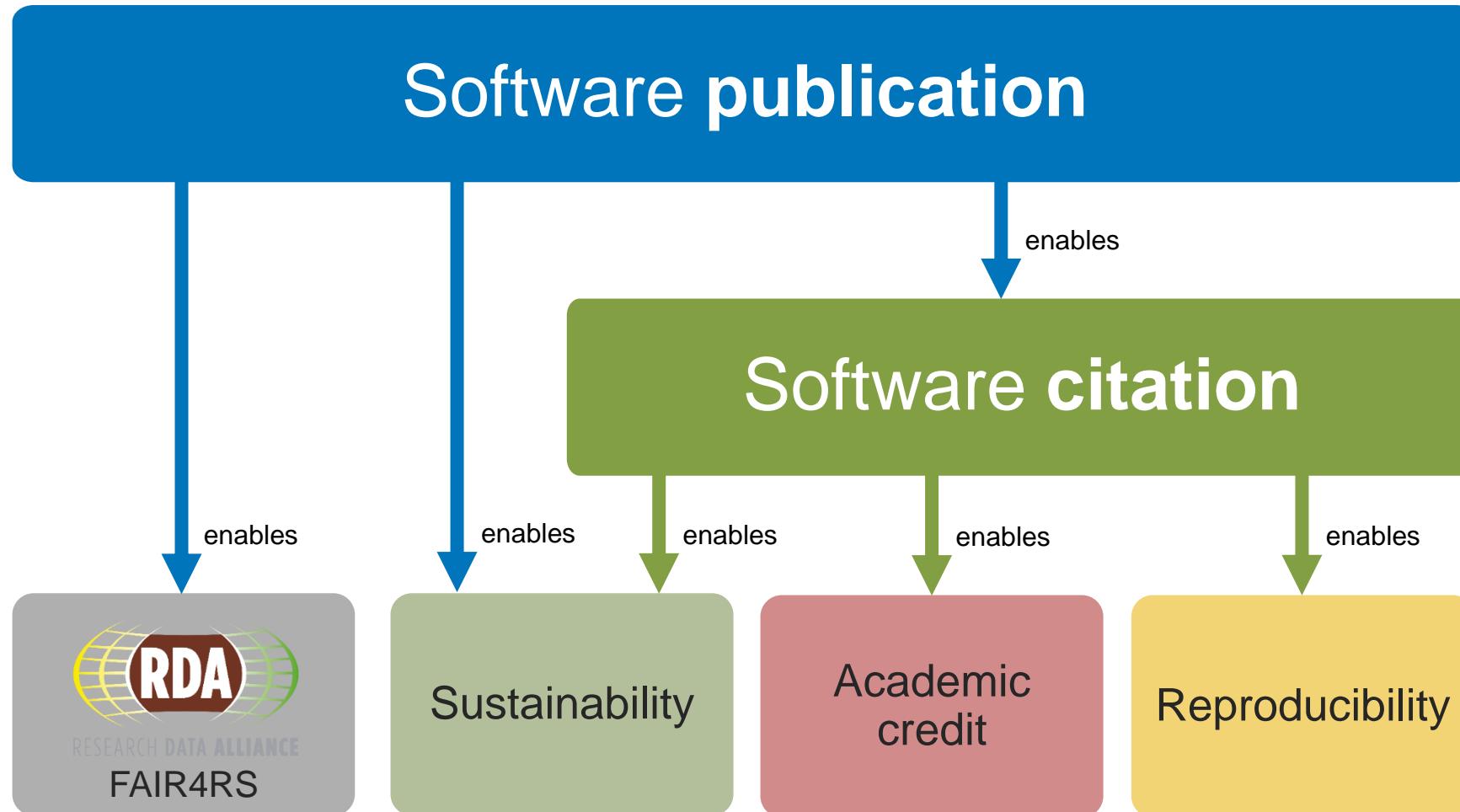


Overview



- Why are software publication and software citation important?
- What is software publication?
- What is software citation?
- How to make software citable?
- How to publish software?

Q: Why are software publication and citation important?
A: Because software is important for research!



Adapted from Druskat et al. (2023)

FAIR for Research Software (FAIR4RS)



- **F:** Software, and its associated metadata, is **easy for both humans and machines to find**.
- **A:** Software, and its metadata, is **retrievable via standardized protocols**.
- **I:** Software **interoperates with other software** by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.
- **R:** Software is both **usable (can be executed)** and **reusable (can be understood, modified, built upon, or incorporated into other software)**.

Chue Hong et al. (2022)

Publication of software with metadata (Research) Software Engineering

What is software publication?



Publication of metadata and artifacts* for software **versions** in publication repositories

Persistent identifier for each version

Not software publication:

- Software available on a source code platform (*GitHub*, *GitLab*, etc.)
- A paper *about* the software

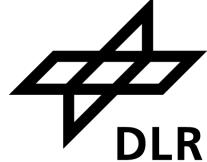
Interim solution:

- software journals ([JOSS](#), [JORS](#), etc.)

Challenge: quality assurance

A screenshot of a Zenodo software publication page for the project "Hexatomic". The page header includes the Zenodo logo, search bar, upload button, and community links. Below the header, it shows the date "August 30, 2022", the title "Hexatomic", and the authors "Druskat, Stephan; Krause, Thomas; Lachenmaier, Clara; Bunzeck, Bastian". It highlights that Hexatomic is an extensible, OS-independent platform for deep multi-layer linguistic annotation of corpora. The page features a "Software Open Access" section with download statistics (44 views, 0 downloads), a "Preview" section showing the contents of "hexatomic-v1.0.1.zip" (including ISSUE_TEMPLATE, PULL_REQUEST_TEMPLATE.md, workflows, archive.yml, release.yml, test.yml, .gitignore, .mvn, and extensions.xml), a "Files" section (8.7 MB total), a "Citations" section (0 citations), and a "Versions" section listing multiple versions (1.0.1, 1.0.0, 0.14.0, 0.13.0-SNAPSHOT, 0.13.0) with their respective DOIs and dates. A note at the bottom states: "Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.6900699. This DOI represents all versions, and will always resolve to the latest one. Read more."

What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic* version 1.4.2, Mar. 2023.
Zenodo. DOI: [10.5281/zenodo.7778709](https://doi.org/10.5281/zenodo.7778709).

Software must be cited like a paper!

Smith et al. (2016)

What is software citation?

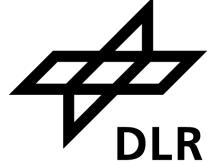


S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic* version 1.4.2, Mar. 2023.
Zenodo. DOI: [10.5281/zenodo.7778709](https://doi.org/10.5281/zenodo.7778709).

A software **version** must be cited like a paper!

A software **project** is cited like a paper!

What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic*-version-1.4.2, Mar. 2023.
Zenodo. DOI: [10.5281/zenodo.6900689](https://doi.org/10.5281/zenodo.6900689).

A software **version** must be cited like a paper!

A software **project** must be cited like a paper!

What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck,
Hexatomic version 1.4.2, Mar. 2023. Software Heritage.
SWHID: [swh:1:rel:2bc06dfdd14699b00c5acecc37213734988a15d6](https://doi.org/10.5281/zenodo.734988a).

A software **version** must be cited like a paper, even if it isn't published!

A software **project** must be cited like a paper, even if it isn't published!

What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck,
Hexatomic version 1.4.2, Mar. 2023. GitHub. URL:
<https://github.com/hexatomic/hexatomic/releases/tag/v1.4.2>.

A software **version** must be cited like a paper, even if it isn't published!

A software **project** must be cited like a paper, even if it isn't published!

What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck,
Hexatomic version a0c59ed, Mar. 2023. GitHub. URL:
<https://github.com/hexatomic/hexatomic/commit/a0c59edf8be8c1d7c045745e4bd043f7369c7d1b>.

A software **version** must be cited like a paper, even if it isn't published!

A software **project** must be cited like a paper, even if it isn't published!

What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck,
Hexatomic version a0c59ed, Mar. 2023. GitHub. URL:
<https://github.com/hexatomic/hexatomic/commit/a0c59edf8be8c1d7c045745e4bd043f7369c7d1b>.

A software **version** must be cited like a paper, even if it isn't published!

A software **project** must be cited like a paper, even if it isn't published!

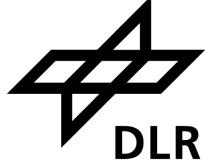
What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic* version 1.4.2, Mar. 2023.
Zenodo. DOI: [10.5281/zenodo.7778709](https://doi.org/10.5281/zenodo.7778709).

Academic credit

What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic* version 1.4.2, Mar. 2023.
Zenodo. DOI: [10.5281/zenodo.7778709](https://doi.org/10.5281/zenodo.7778709).

Reproducibility

+

Findability, Accessibility, Reusability

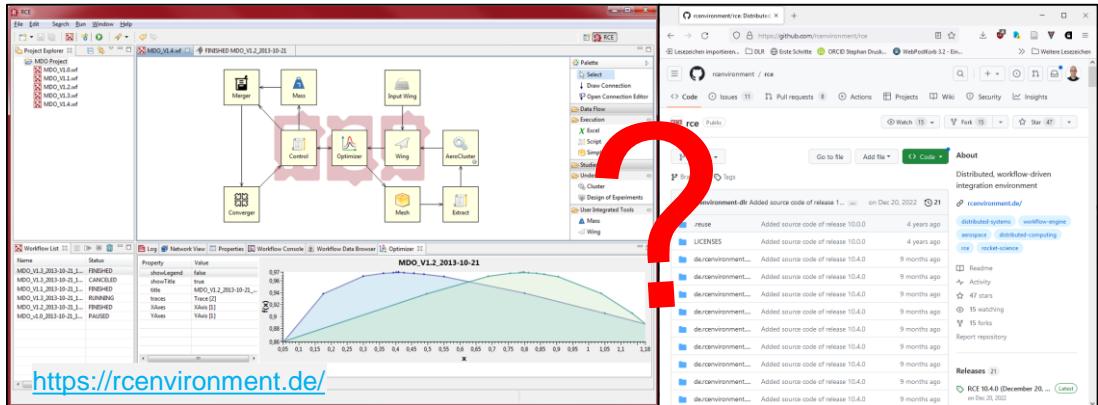
How to make software citable?



Metadata

citation-file-format.github.io

Authors? Name? Version? Publication date?



Provide metadata in CITATION.cff in the repository!

A screenshot of a GitHub repository page for 'cognition-file-format'. The main content is the 'CITATION.cff' file, which contains the following YAML metadata:

```
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"
```

A 'Cite this repository' button is at the top right, and APA and BibTeX citation options are shown below.

Druskat et al. (2021)

- YAML: citation-relevant metadata, including references (e.g., to primary dependencies)
- Easy creation/update/validation (web tool [cffinit](#)), compatibility, tool ecosystem
- Open community format (> 15,000 files on GitHub; Support: Zenodo, Zotero, JabRef, ...)

How to publish software?

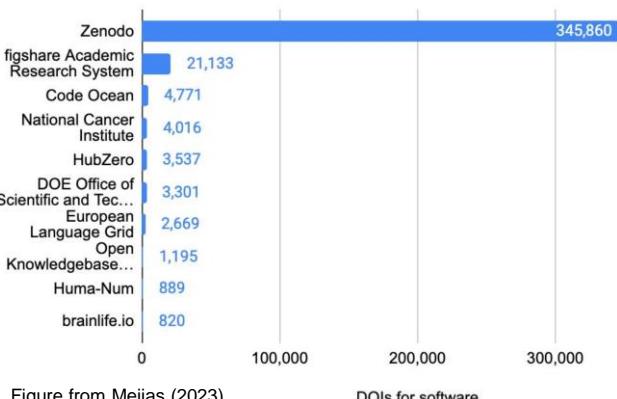


Publication repositories

DOIs for versions & projects, metadata, persistence

A screenshot of the Zenodo interface. It shows a search result for a software project. The project has a DOI of 10.5281/zenodo.1211c23. The page includes details like the author (Stephan Druskat), version (v0.1.0), and a preview of the software's contents.

Figure from Mejias (2023).



DOIs for software

Simple automation with CFF, GitHub & Zenodo!

A screenshot of a GitHub repository named "sdruskat/campussource". A specific release "v0.1.0" is shown. The release page indicates it is a "release without a CFF file". A large green callout box highlights the "campussource-0.1.0.zip" file, which is listed under the "REleases" section.

A screenshot of a GitHub repository named "CampusSource Example Deposit". A specific release "v0.2.0" is shown. The release page indicates it is a "release WITH a CITATION.cff file". A large green callout box highlights the "campussource-0.2.0.zip" file, which is listed under the "REleases" section. Below the file listing, a green box contains the CFF file content.

```
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"
```

zenodo

- Most often used by far: [Zenodo](#) (developed & run at CERN), GitHub integration
- Persisted records, persistent identifiers ([DataCite](#) DOIs w/ metadata), landing pages
- Publication repository @ DLR: under construction

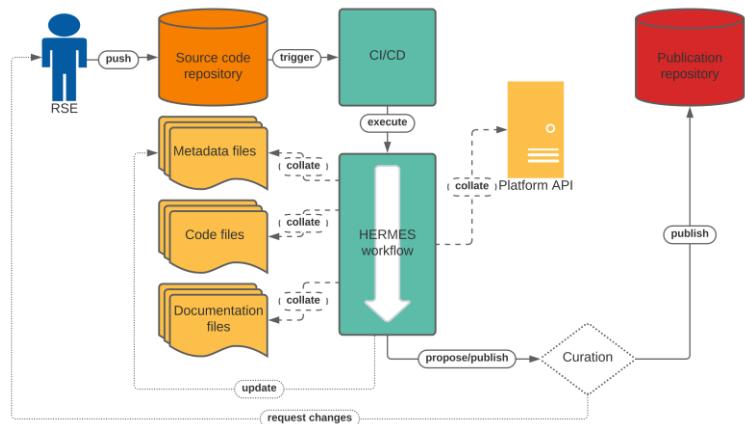
How to publish software (advanced)?



Continuous publication with HERMES

software-metadata.pub

Run publication workflow in CI



Druskat et al. (2022)

Extensible & works with DLR GitLab



- Full metadata control; includes metadata curation & updating repository metadata
- Different metadata sources (including CFF) & publication targets (including Zenodo)
- Developed as FLOSS at Helmholtz – documentation, training materials

Conclusion



- Software citation is a must ([DFG Code of Conduct](#))
 - Software metadata provided by authors
 - Cite software (versions) like papers
- Software publication is ...
 - ... highly recommended ([Helmholtz Open Research Software Indicator](#)),
 - ... possible (Zenodo),
 - ... automatable (HERMES),
 - but lacks an established step for external quality assurance!
- You can start today!

Thanks!



Projects:

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

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

Me:

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

References:

Druskat, Stephan, Bertuch, Oliver, & Struck, A. (2023). Towards Research Software-ready Libraries. ABI Technik, 43(3), 168-178. doi: [10.1515/abitech-2023-0031](https://doi.org/10.1515/abitech-2023-0031).

Smith AM, Katz DS, Niemeyer KE, & FORCE11 Software Citation Working Group. 2016. Software citation principles. PeerJ Computer Science 2:e86. doi: [10.7717/peerj-cs.86](https://doi.org/10.7717/peerj-cs.86).

Chue Hong, Neil P., Daniel S. Katz, Michelle Barker, Anna-Lena Lamprecht, Carlos Martinez, Fotis E. Psomopoulos, Jen Harrow, et al. (2022). FAIR Principles for Research Software (FAIR4RS Principles). RDA. doi: [10.15497/RDA00068](https://doi.org/10.15497/RDA00068).

Druskat, Stephan, Bertuch, Oliver, Juckeland, Guido, Knodel, Oliver, & Schlauch, Tobias. (2022). Software publications with rich metadata: State of the art, automated workflows and HERMES concept. arXiv. doi: [10.48550/arXiv.2201.09015](https://doi.org/10.48550/arXiv.2201.09015).

Druskat, Stephan, Spaaks, Jurriaan H., Chue Hong, Neil, Haines, Robert, Baker, James, Bliven, Spencer, Willighagen, Egon, Pérez-Suárez, David, & Konovalov, Alexander. (2021). Citation File Format (1.2.0). Zenodo. doi: [10.5281/zenodo.5171937](https://doi.org/10.5281/zenodo.5171937).

Mejias, Gabi. (2023, May 30). DOIs for research software: Increasing Visibility, Connectivity, Citability. DOIs for Research Software: Increasing Visibility, Connectivity, Citability, online. Zenodo. doi: [10.5281/zenodo.7985199](https://doi.org/10.5281/zenodo.7985199).