

MAKING RESEARCH SOFTWARE FAIR AND CITABLE

Stephan Druskat

German Aerospace Center (DLR), Institute for Software Technology

DOI [10.5281/zenodo.7835318](https://doi.org/10.5281/zenodo.7835318)

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

netherlands
eScience center



Software
Sustainability
Institute



Making research software FAIR and citable



1. Motivation
2. FAIR research software
3. Software publication
4. Software citation
5. Making research software FAIR and citable
6. Conclusion & outlook

Making research software FAIR and citable



- Do you write code that you use to conduct your research?
- Do you write code that *others* use (or could use) in their research?
- Can others find, assess and reuse the code that you write?

A satellite with large solar panels is shown in orbit above Earth. The satellite is gold-colored with two long arms of solar panels extending outwards. The Earth below is covered in green land and white clouds, with a blue atmosphere visible at the horizon.

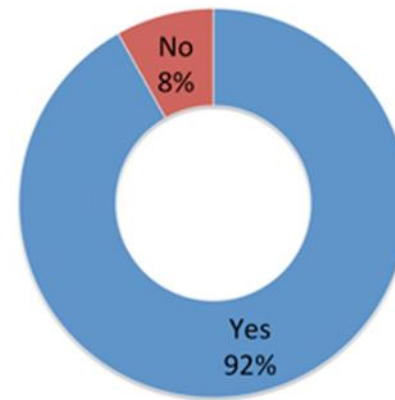
FAIR RESEARCH SOFTWARE

Software as a valid research output

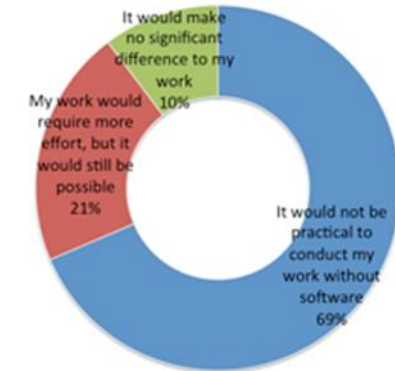


1. Importance for research [1]
2. Acknowledgment as a valid research output [2]

Do you use research software?



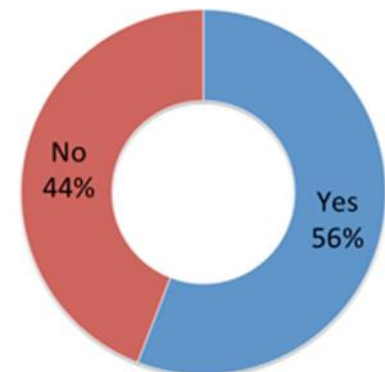
What would happen if you could no longer use research software?



[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>.

Do you develop your own research software?



FAIR research software



- Aim: increase transparency, reproducibility and reusability of research
- FAIR Principles for Research Software (FAIR4RS) [3]:
 - **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).

[3] Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A.-L., Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., Honeyman, T., et al. (2021). FAIR Principles for Research Software (FAIR4RS Principles). *Research Data Alliance*. DOI: [10.15497/RDA00065](https://doi.org/10.15497/RDA00065)

FAIR research software implementation



- **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).

Publication of software with metadata (Research) Software Engineering

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 solar cells. The satellite's main body is centrally located between the panels. 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 state of the practice



- 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

zenodo Search Upload Communities Log in Sign up

August 30, 2022 Software Open Access

Hexatomic

Druskat, Stephan; Krause, Thomas; Lachenmaier, Clara; Bunzeck, Bastian

Hexatomic is an extensible, OS-independent platform for deep multi-layer linguistic annotation of corpora. It is being developed for sustainability, in order to support research software re-use rather than new development of software with each new research project. Using Hexatomic, linguistic research projects can implement what they need on top of an existing platform. To safeguard compatibility, Hexatomic works on instances of Salt projects. Salt is a generic metamodel for linguistic data.

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

Preview

hexatomic-v1.0.1.zip

The preview is not showing all the files

- hexatomic-hexatomic-4fa5704
 - all-contributorsrc 1.7 kB
 - checkstyle 505 Bytes
 - github
 - ISSUE_TEMPLATE
 - bug_report.md 742 Bytes
 - feature_request.md 928 Bytes
 - PULL_REQUEST_TEMPLATE.md 2.3 kB
 - workflows
 - archive.yml 317 Bytes
 - release.yml 5.1 kB
 - test.yml 1.8 kB
 - gltignore 76 Bytes
 - myn
 - extensions.xml 193 Bytes

Files (8.7 MB)

Name	Size	Preview	Download
hexatomic/hexatomic-v1.0.1.zip	8.7 MB		

md5:54ca6b45d34222149f5d4431a7036e8e4

Citations 0

Show only: Literature (0) Dataset (0) Software (0) Unknown (0) Citations to this version

No citations.

Available in: GitHub, OpenAIRE

Publication date: August 30, 2022

DOI: 10.5281/zenodo.7034163

Related identifiers: Supplement to <https://github.com/hexatomic/tree/v1.0.1>

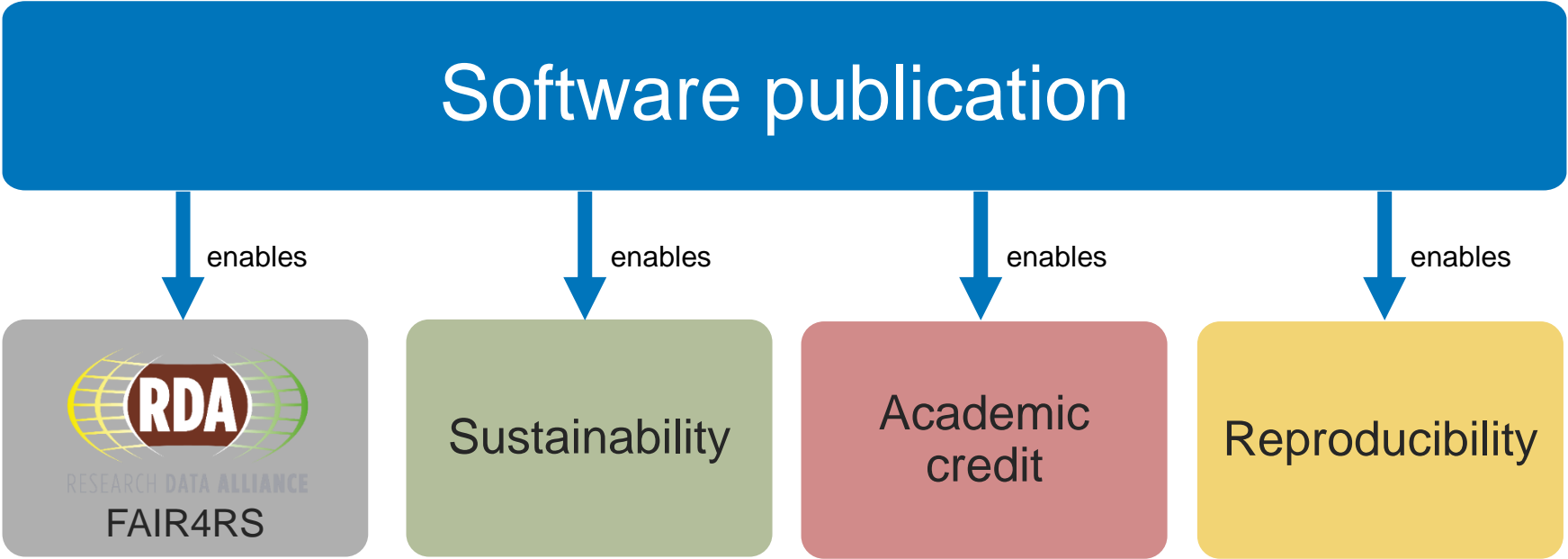
License (for files): Apache License 2.0

Versions

Version	Date
Version 1.0.1 10.5281/zenodo.7034163	Aug 30, 2022
Version 1.0.0 10.5281/zenodo.7016810	Aug 23, 2022
Version 0.14.0 10.5281/zenodo.7016685	Aug 23, 2022
Version 0.13.0-SNAPSHOT 10.5281/zenodo.6900690	Jul 25, 2022

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.6900690. This DOI represents all versions, and will always resolve to the latest one. Read more.

Software publication

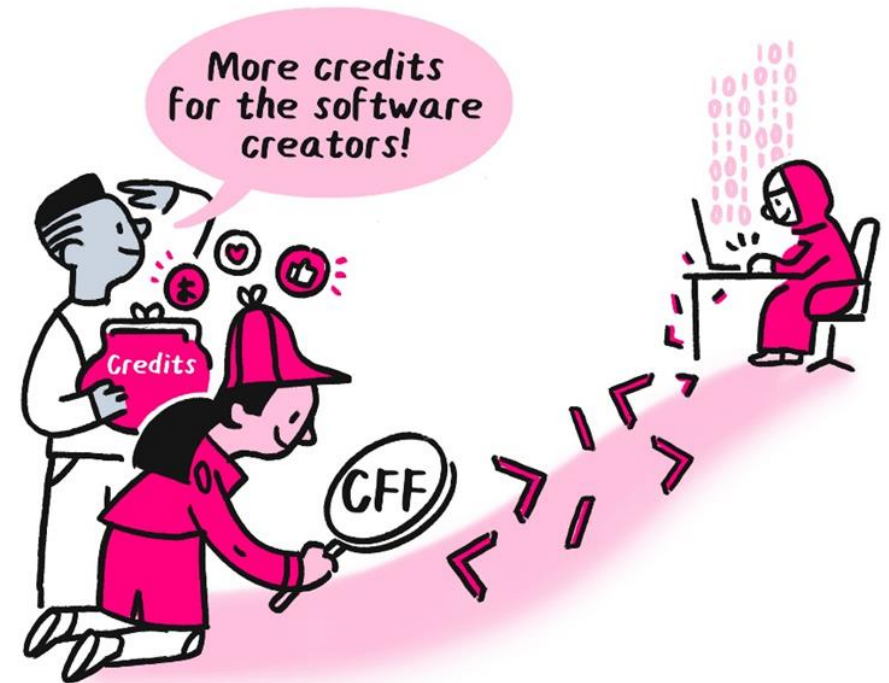


The background of the slide is a high-resolution photograph of a satellite in orbit above Earth. The satellite is a rectangular platform with two long, parallel solar panel arrays extending outwards. The panels are covered in a grid of small solar cells. The satellite's main body is centrally located between the panels. Below the satellite, the Earth's surface is visible, showing a mix of green landmasses and blue oceans, with white clouds scattered across the scene. The curvature of the Earth is visible at the bottom of the frame.

SOFTWARE CITATION

Software citation

- Good research practice [4]
- Attribution and credit for research software creators
- Enables reproducibility
- (for other functions see [5])



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

[4] Deutsche Forschungsgemeinschaft, "Guidelines for Safeguarding Good Research Practice. Code of Conduct," Apr. 2022, doi: [10.5281/zenodo.6472827](https://doi.org/10.5281/zenodo.6472827).

[5] S. Druskat, "Software and Dependencies in Research Citation Graphs," in Computing in Science & Engineering, vol. 22, no. 2, pp. 8-21, March-April 2020, doi: [10.1109/MCSE.2019.2952840](https://doi.org/10.1109/MCSE.2019.2952840).

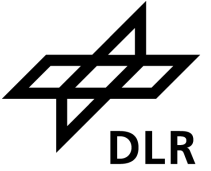
Software citation principles



- Software citation principles [6]
 - **Importance:** Software itself is cited like papers are cited.
 - **Credit and attribution**
 - **Unique identification**
 - **Persistence**
 - **Accessibility:** Citation allows access to software and metadata.
 - **Specificity:** Citation identifies the software version used in research
- Challenges: **citation metadata** (authors, title, version, publication date, ...)

[6] 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 metadata



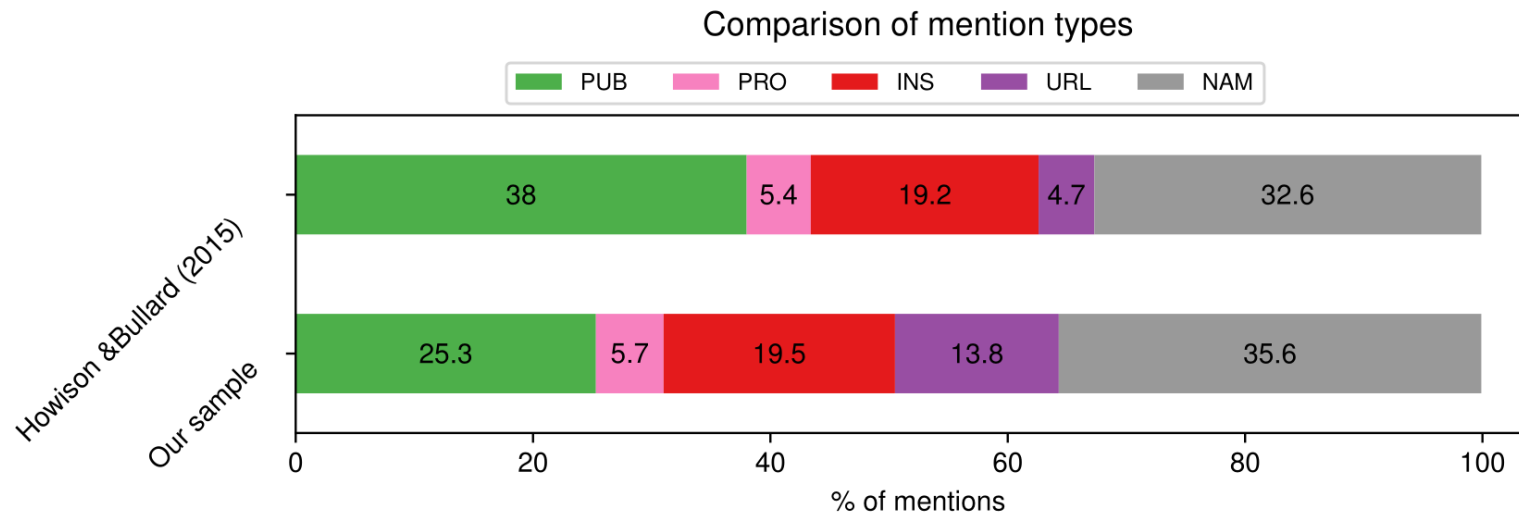
Correct and complete software citation metadata must be identified and provided by software projects

- Authors vs. contributors
- Versioning
- (Persistent) identifiers, e.g., from publication

Software citation culture change



Software is currently often mentioned, not cited!



From [8]: Mention types of software in publication in % in data from [7] and our own data.
PUB: cites publication; PRO: cites project name/website; INS: instrument-like; URL: URL in text; NAM: in-text name only.

[7] 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).

[8] 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.

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, thin solar panel arrays extending outwards. The Earth's surface below is a mix of green landmasses and blue oceans, with white clouds scattered across the scene. The curvature of the planet is visible on the right side.

MAKING RESEARCH SOFTWARE FAIR AND CITABLE

Citation File Format



Citation File Format (CFF) [9]:

- Authoritative, controllable, principled
- Open community project
- Human- and machine-readable (YAML)
- References from software
- Fallback software paper citation

CITATION.cff

```
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"
```

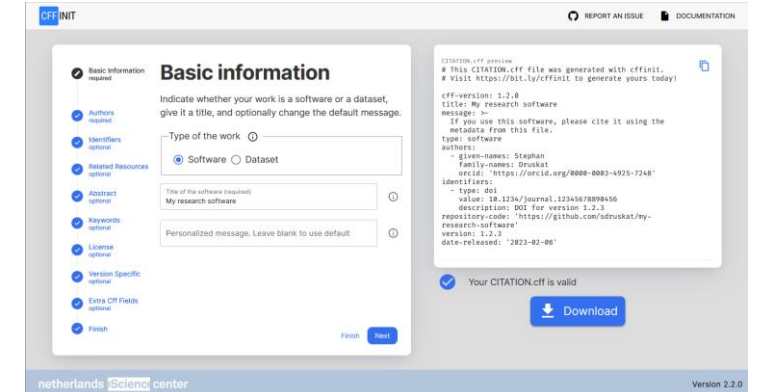
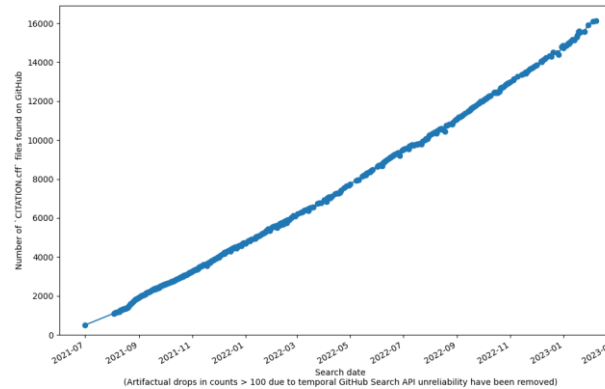
<https://citation-file.format.github.io>

[9] 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>.

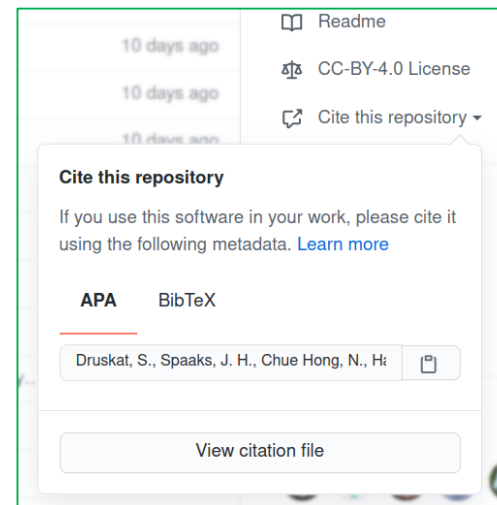
Citation File Format as community standard



- 19,000+ CFF files on GitHub
- Tool support (e.g. [cffinit](#))
- Platform support (GitHub, Zenodo)
- IDE support



```
File Edit View Navigate Code Refactor Run Tools Git Window Help cff-in-ides - CITATION.cff
cff-in-ides CITATION.cff
CITATION.cff
cff-version: 1.2.0
message: If you use this software, please cite it using the metadata from this file.
authors:
  - family-names: Druskat
    given-names: Stephan
  title: Software project
  license: UNKNOWN LICENSE
references:
  - type: article
    authors:
      - name: anon.
        title: Article
  - license: An SPDX license identifier.
  - version: The version of the work.
  - date-released: The date the work has been released.
  - abbreviation: The abbreviation of a work.
  - abstract: The abstract of a work.
  - collection-doi: The DOI of a collection containing the work.
  - collection-title: The title of a collection or proceedings.
  - collection-type: The type of a collection.
  - commit: The (e.g., Git) commit hash or (e.g., Subversion) revision number of the work.
  - conference: The conference where the work was presented.
  - contact: The contact person, group, company, etc.
  - repository: The repository identifier associated to the work.
```



FAIR software publication in practice



- Publish every stable version (“release”) in a publication repository:
 - General purpose repositories, e.g., Zenodo
 - Institutional repositories
 - Domain repositories
- Register software publication with a domain registry:
 - e.g. [swMATH](#) for mathematical software
- Submit rich metadata

Continuous software publication: GitHub + 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	

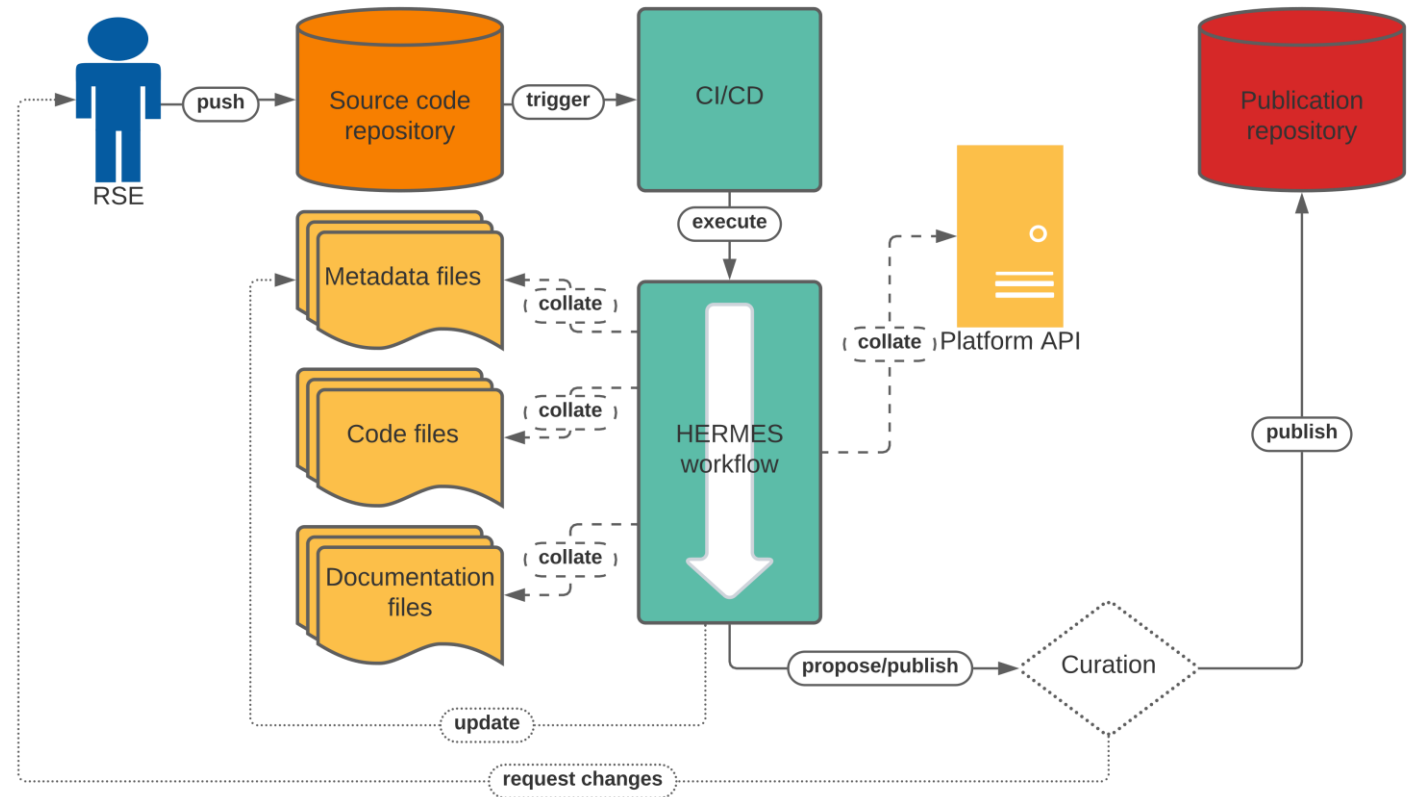
Cite all versions? You can cite all versions by using the DOI [10.5072/zenodo.1035710](https://doi.org/10.5072/zenodo.1035710). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

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



[10] S. Druskat, O. Bertuch, G. Juckeland, O. Knodel, and T. Schlauch, “Software publications with rich metadata: state of the art, automated workflows and HERMES concept,” *arXiv*, Jan. 2022, doi: [10.48550/arXiv.2201.09015](https://doi.org/10.48550/arXiv.2201.09015).

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

Conclusion & outlook



1. FAIR and citable research software can be achieved through software publication with rich metadata.
2. Software can be made more easily citable by software projects through the Citation File Format and its tooling support.
3. Initial solutions for automated software publication exist or are in development:
 1. GitHub-Zenodo integration using CITATION.cff files
 2. HERMES
4. Software publication and citation will become part of research evaluation:
 1. Helmholtz introduced basic indicator for citable software publication [11] (KPI)
 2. Quality indicator for software publication by the end of 2024

[11] Helmholtz-Gemeinschaft, “Helmholtz Open Science Policy. Version 1.0. Approved in the 119th General Assembly of the Helmholtz Association on 20-21 September 2022,” 2022, doi: [10.48440/os.helmholtz.056](https://doi.org/10.48440/os.helmholtz.056).

Thanks!



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

Funded by German Aerospace Center (DE), Netherlands eScience Center (NL), Software Sustainability Institute (UK), Code for Science & Society (US), CampusSOURCE (DE)

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

Funded by *Initiative & Networking Fund* of the Helmholtz Association (ZT-I-PF-3-006)

Contact

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

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