

WHAT RSEs SHOULD KNOW ABOUT SOFTWARE CITATION

Stephan Druskat

German Aerospace Center (DLR), Institute for Software Technology

DOI [10.5281/zenodo.10650644](https://doi.org/10.5281/zenodo.10650644)

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



Software
Sustainability
Institute






Who / what am I?



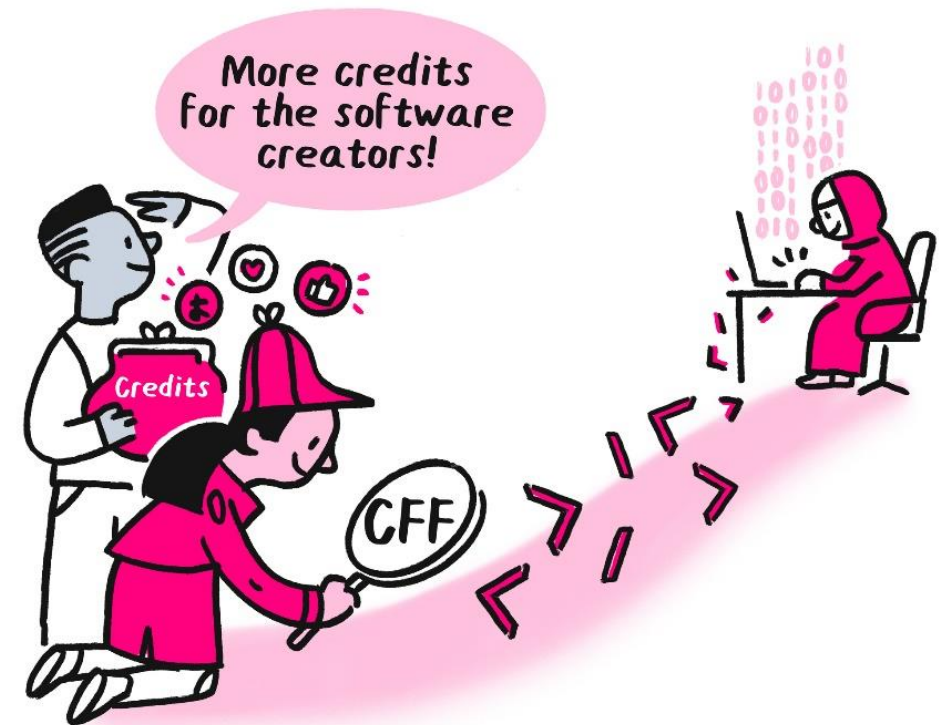
- MA in English, modern German literature, linguistics
- RSE in linguistics (2009-2020)
- Fellow of the [Software Sustainability Institute](#) (since 2018)
- Doctoral researcher/RSE in computer science (since 2019)



- Project lead [Citation File Format](#) 
- Co-founder / inaugural board member of [de-RSE](#) 
- PI [HERMES](#) project (2021-2023) 

Why RSEs should know about software citation

- Principles of research apply: disclosure of research process
- Accountability: taking credit
- **Formal citation** as established method:
 - **Importance** of software in research
 - Findability (**FAIR**)
 - **Reproducibility**
 - **Credit**



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

How software should be cited



PeerJ
Computer Science

Software citation principles

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

¹GitHub, Inc., San Francisco, California, United States
²National Center for Supercomputing Applications & Electrical and Computer Engineering Department & School of Information Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States
³School 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@force11.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(086), 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.

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).

Why software as such isn't easily citable



METADATA: Name? Authors? Version? Publisher? Publication date?

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 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, with a thin white layer of clouds. The curvature of the Earth is visible at the bottom of the frame.

MAKING SOFTWARE CITABLE

Making software citation easier as an RSE



1. Provide citation metadata for your software!

Only you/your project knows the correct metadata:

- What's the **name** of the software?
- Who are the **authors**? [*Credit and attribution*]
 - Is the person who fixed that typo / designed the UI / wrote unit tests a "[software author](#)"?
- What is the current **version** of the software? [*Specificity*]
 - Version strings
 - What versions should people use? (Releases only? Any commit in `main`?)
- Where can people find the source code? [*Accessibility*]
- ...

2. Publish your software!

Making software citation easier as an RSE in practice: Provide citation metadata with the Citation File Format (CFF)



citation-file-format.github.io

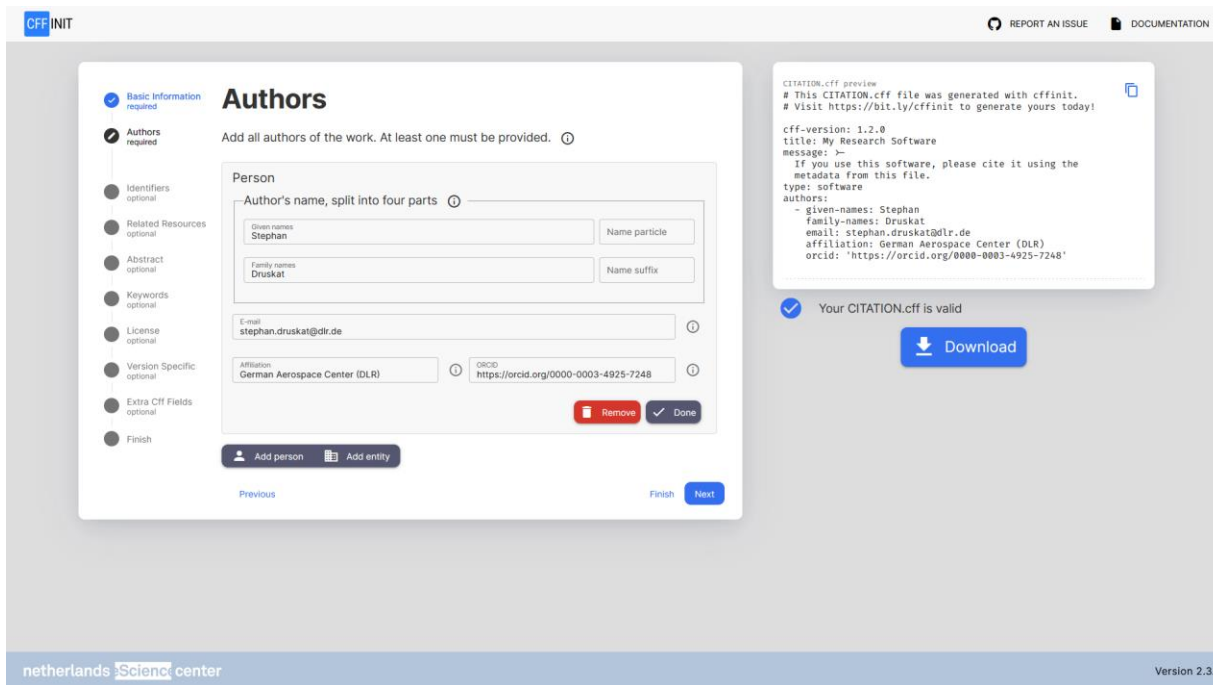
```
cff-version: 1.2.0
message: If you use this software, please cite it using these metadata.
title: My Research Software
abstract: This is my awesome research software. It does many things.
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: "https://orcid.org/0000-0003-4925-7248"
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"
```

+ references, preferred citation, experimental dataset support

METADATA: In a **CITATION.cff** file in the source code repository

Why use CFF?

1. Tooling: Create, edit & update, validate, convert



cffinit: web tool for CFF file creation - <https://bit.ly/cffinit>

Overview: <https://bit.ly/cff-tools-list>

Why use CFF?

2. GitHub support



10 days ago

10 days ago

10 days ago

Readme

CC-BY-4.0 License

Cite this repository ▾

Cite this repository

If you use this software in your work, please cite it using the following metadata. [Learn more](#)

APA BibTeX

Druskat, S., Spaaks, J. H., Chue Hong, N., H...

View citation file

- Detects **CITATION.cff** files
- Renders citation metadata (APA, BibTeX)
- Points to **CITATION.cff** instance
- [Documents](#) and supports creation

cfftracker / CITATION.cff in main Cancel changes

Adding a **CITATION.cff** file helps users to easily cite your software from the repository overview. [Learn more](#) Insert example

<> Edit new file Preview Spaces 2 No wrap

```
1 cff-version: 1.2.0
2 message: "If you use this software, please cite it as below."
3 authors:
4 - family-names: "YOUR_NAME_HERE"
5   given-names: "YOUR_NAME_HERE"
6   orcid: "https://orcid.org/0000-0000-0000-0000"
7 - family-names: "Lisa"
8   given-names: "Mona"
9   orcid: "https://orcid.org/0000-0000-0000-0000"
10 title: "cfftracker"
11 version: 1.0.0
12 doi: 10.5281/zenodo.1234
13 date-released: 2021-08-19
14 url: "https://github.com/sdruskat/cfftracker"
15
```

Why use CFF?

3. Zenodo support



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.](#)

Why use CFF?

4. Reference manager support



Import in **Zotero** from GitHub via browser plugin

The screenshot shows the 'Info' tab of a Zotero entry. The citation key is 'druskatHexatomic2021'. The item type is 'Software' with the title 'Hexatomic'. The programmers listed are 'Druskat, Stephan' and 'Krause, Thomas'. The abstract states: 'Hexatomic is an extensible software for deep multi-layer annotation of linguistic corpora'. Other fields include 'Version: 0.6.0-SNAPSHOT', 'Date: 2021-03', 'System', 'Place', 'Company', 'Prog. Language: Java', 'ISBN', 'Short Title', 'URL: https://github.com/hexatomic/hexatomic', 'Rights: Apache-2.0', 'Archive', 'Loc. in Archive', and 'Library Catalog: GitHub'.

Import in **JabRef** from **CITATION.cff** file

The screenshot shows the 'untitled* - JabRef' window. The entry is for 'Software' with the title 'Citation File Format' and year '2021'. The author is 'Druskat et al.'. The 'Required fields' section shows: Author: 'Stephan Druskat and Jurriaan H. Spaaks and Neil Chue Hong and Robert...', Editor: (empty), Title: 'Citation File Format', Date: '2021-08-09', and Citationkey: 'Druskat2021'. The 'Software (Druskat2021)' section contains the following text: 'Druskat, Stephan / Spaaks, Jurriaan H. / Hong, Neil Chue / Haines, Robert / Baker, James / Bliven, Spencer / Willighagen, Egon / Pérez-Suárez, David / Kononov, Alexander', 'Citation File Format', '2021-08-09', 'Abstract: CITATION cff files are plain text files with human- and machine-readable citation information for software. Code developers can include them in their repositories to let others know how to correctly cite their software. This is the specification for the Citation File Format.', and 'Comment: If you use CFF in your research, please cite it using these metadata.'

Why use CFF?

5. IDE support



- Auto-completion and validation via JSON Schema / schemastore.org

The screenshot shows an IDE window with a CFF file named 'CITATION.cff'. The file content is as follows:

```
1 cff-version: 1.2.0
2 message: If you use this software, please cite it using the metadata from this file.
3 authors:
4   - family-names: Druskat
5     given-names: Stephan
6   title: Software project
7   license: UNKNOWN LICENSE
8 references:
9   - type: article
10     authors:
11       - name: anon.
12       title: Article
13
```

The 'license' field on line 7 is selected, and a dropdown menu is open, showing a list of valid SPDX license identifiers. The 'license' field is highlighted in blue in the original image.

Field	Description
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.
copyright	The copyright information pertaining to the work.

The screenshot shows the same IDE window with the CFF file. The 'license' field on line 7 is now set to 'Apa'. A dropdown menu is open, showing a list of valid SPDX license identifiers. The 'license' field is highlighted in blue in the original image.

```
1 cff-version: 1.2.0
2 message: If you use this software, please cite it using the metadata from this file.
3 authors:
4   - family-names: Druskat
5     given-names: Stephan
6   title: Software project
7   license: Apa
8 references:
9   - type: Apache-1.0
10     auth Apache-1.1
11         Apache-2.0
12         APAFML
13 titl Press Eingabe to insert, Tabulator to replace
```

The 'license' field is highlighted in blue in the original image.

Document 1/1 > license: > Apa

Problems: Current File 1 Project Errors

CITATION.cff C:\Users\drus_st\tmp\cff-in-ides 1 problem

- Schema validation: Value should be one of: "0BSD", "AAL", "Abstyles", "Adobe-2006", "Adobe-Glyph", "ADSL", "AFL-1.1", "AFL-1.2",

Why use CFF?

6. Everyone does it



- GitHub (if the GitHub Search API is right) has **~15,000** CFF files

The screenshot shows a GitHub search interface with the query `path:CITATION.cff -user:sdruskat`. The search results are filtered to show 15k files. Three specific CFF files are displayed:

- gpt-engineer-org/gpt-engineer · citation.cff**

```
1 cff-version: 1.0.0
2 message: "If you use this software, please cite it as below."
3 authors:
4   - family-names: Osika
5     given-names: Anton
6 title: gpt-engineer
7 version: 0.1.0
```
- microsoft/JARVIS · CITATION.cff**

```
1 cff-version: 1.2.0
2 message: "If you find this work useful in your method, you can cite the paper as below."
3 authors:
4   - family-names: Shen
5     given-names: Yongliang
6   - family-names: Song
7     given-names: Kaitao
```
- facebookresearch/MetaCLIP · opendip_CITATION.cff**

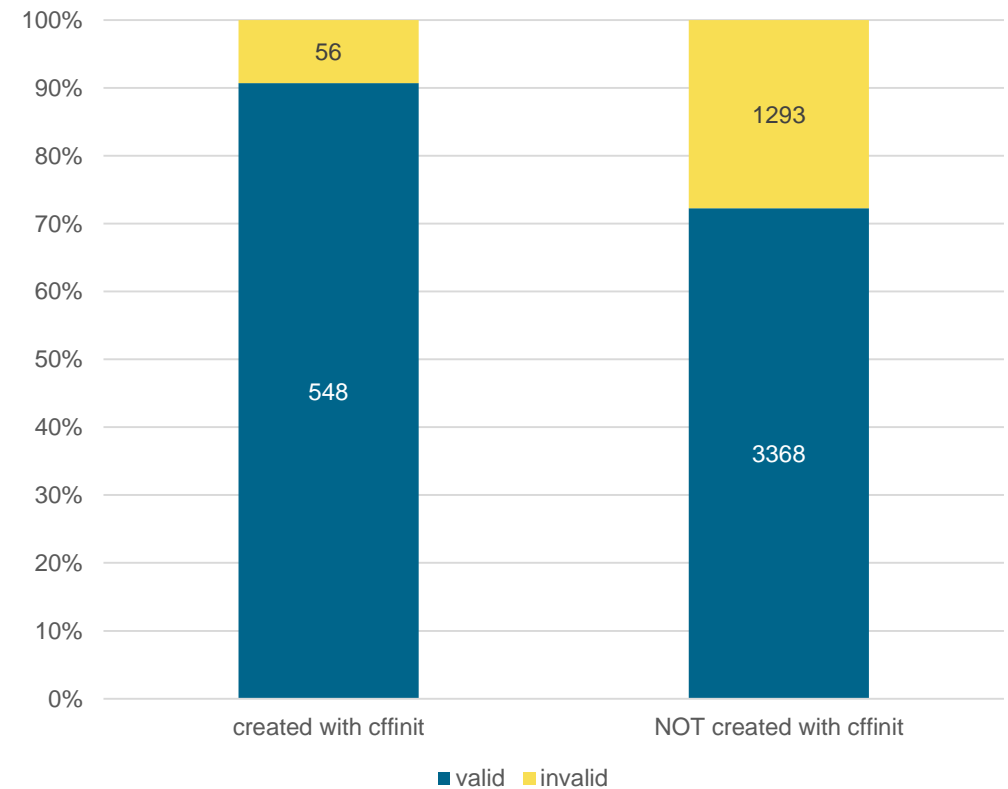
```
1 # Copyright (c) Meta Platforms, Inc. and affiliates.
2 # All rights reserved.
3
4 cff-version: 1.1.0
5 message: If you use this software, please cite it as below.
6 authors:
7   - family-names: Ilharco
```


Getting started with the Citation File Format



The screenshot shows the 'Authors' section of the CFF INIT web interface. The form is titled 'Authors' and includes a sidebar with navigation options: Basic Information (required), Authors (required), Identifiers (optional), Related Resources (optional), Abstract (optional), Keywords (optional), License (optional), Version Specific (optional), Extra CFF Fields (optional), and Finish. The main form area is titled 'Person' and contains the following fields: 'Author's name, split into four parts' (with sub-fields for Given names: Stephan, Name particle, Family names: Druskat, Name suffix), 'Email: stephan.druskat@dlr.de', 'Affiliation: German Aerospace Center (DLR)', and 'ORCID: https://orcid.org/0000-0003-4925-7248'. There are 'Remove' and 'Done' buttons for the person entry, and 'Add person' and 'Add entity' buttons at the bottom. A 'Previous' and 'Next' button are also present. On the right, a preview of the generated CFF file is shown, including the title 'My Research Software' and author information. A 'Download' button is located below the preview. The footer of the interface includes 'netherlands Science center' and 'Version 2.3.1'.

Validity of v1.2.0 CFF files created with/without cffinit (preliminary data)



1. Use <https://bit.ly/cffinit> to create an initial CFF file for your software!
2. Optionally, refine the file using IDE support.

The background of the slide is a high-resolution photograph of a satellite in orbit above Earth. The satellite is the central focus, featuring a central body with various instruments and two long, rectangular 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 at the top and bottom edges of the frame.

PUBLISHING SOFTWARE


Why publish software?



- **To enable software citation**
- To make the software FAIR
 - Findable, accessible, (interoperable, reusable*)
- To let software count towards evaluation (key performance indicators, etc.)
 - Based on [Helmholtz Open Science Policy](#), number of citable software publications are counted (from 2023 onwards)

Publishing software (for software citation)



- State of the art: publish software (or at least citation metadata) to an (ideally open access) **publication repository** 
 - Repository publication > software journals
 - The source code repository is not a publication repository!

In practice:

- Publish every release on **Zenodo** (or your institutional repository).
- This gives you:
 - A DOI [*Unique identification*]
 - An archived snapshot [*Persistence*]
 - A resolvable identifier whose target provides access to the software [*Accessibility*]
 - An identifier that points to the specific version [*Specificity*]

Making software publication easier for RSEs: CFF and the GitHub Zenodo integration



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

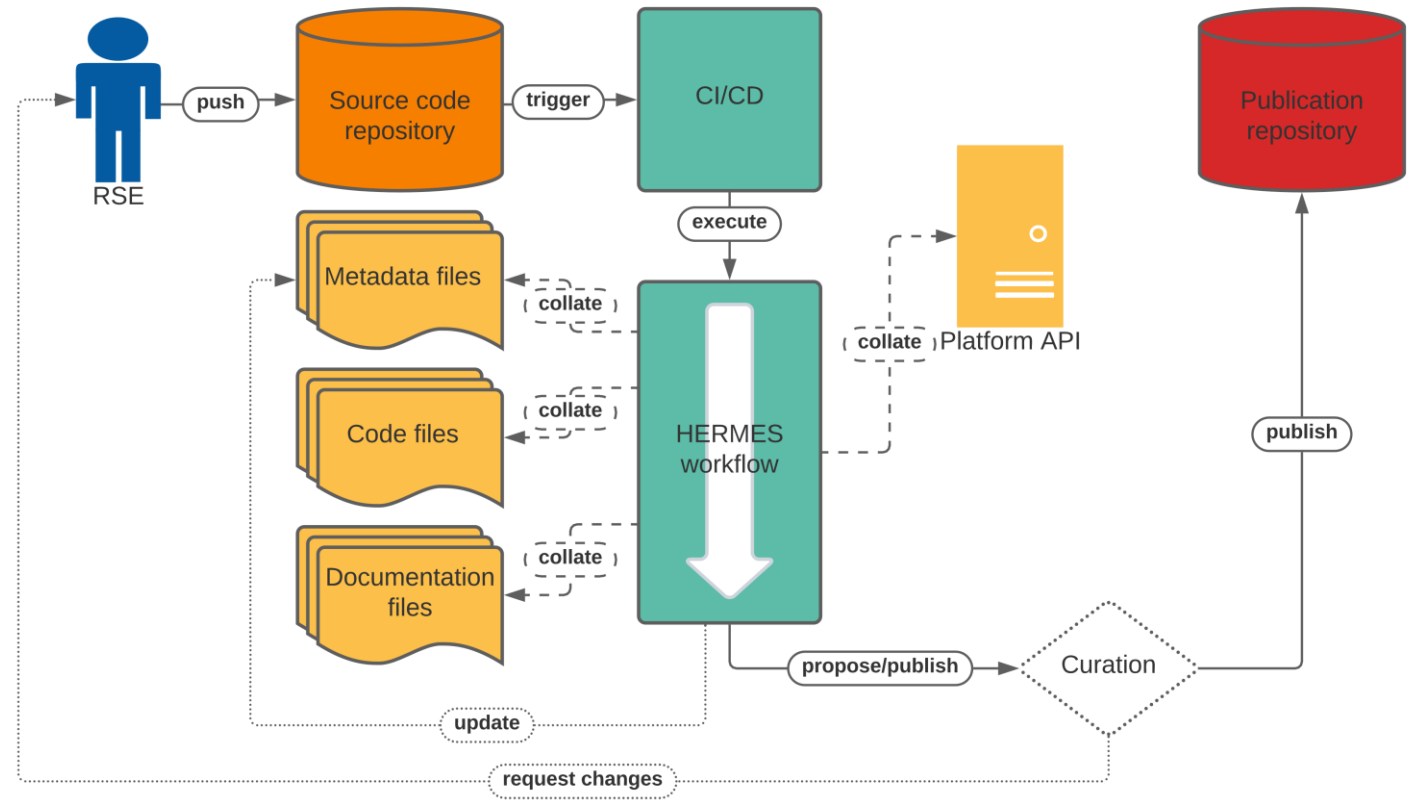
Version 0.2.0	Mar 16, 2022
10.5072/zenodo.1035737	
Version 0.1.0	Mar 16, 2022
10.5072/zenodo.1035711	

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.](#)

Making software publication easier for RSEs: Automation with HERMES



- Automating software publication via CI
- Harvest and merge existing metadata
- Pushing proactively to keep full control
- Curation & FAIR metadata-only publication (closed source) possible
- InvenioRDM and Dataverse support



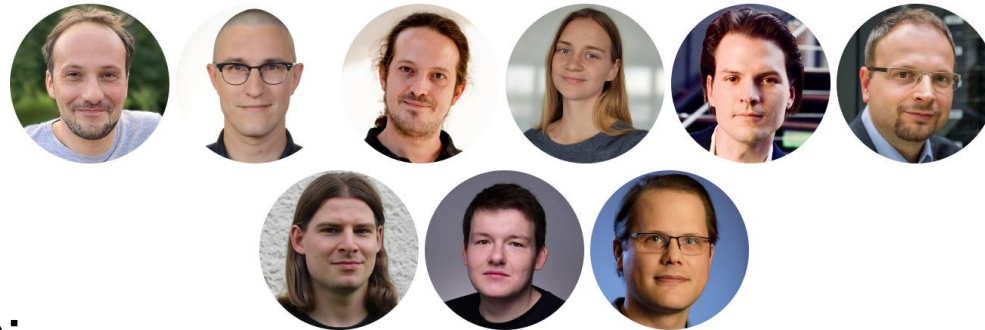
HERMES workflow, high-level view, HERMES project ([CC BY-SA 4.0](#))

Making software publication easier for RSEs: Automation with HERMES



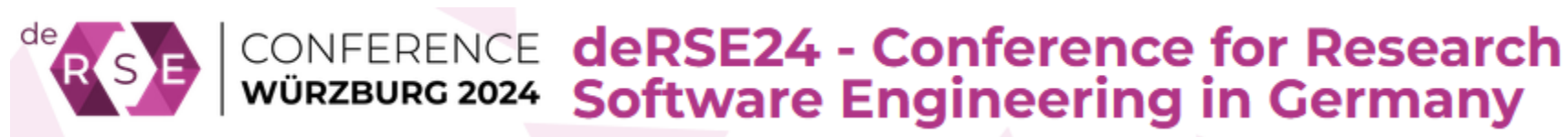
software-metadata.pub

- [Helmholtz Metadata Collaboration](#) project (DLR + FZJ + HZDR)



- **deRSE24** workshop:

S. Kernchen: [Automating your FAIR software publications with HERMES](#)



Conclusion



1. **(Enabling) software citation is part of our job as RSEs.**
2. Software creators/RSEs should provide and maintain the **correct citation metadata** for software in a Citation File Format `CITATION.cff` file in the source code repository.
3. Software creators/RSEs should **publish every release** to Zenodo (or another publication repository).
4. Researchers **must cite software** as they would cite a paper.

Thanks!



The **Citation File Format** has been funded by the *German Aerospace Center (DLR)*, the *Netherlands eScience Center*, the *Software Sustainability Institute*, *Code for Science & Society*, and *CampusSOURCE e.V.*

HERMES is being funded by the *Initiative and Networking Fund* of the Helmholtz Association as part of the [Helmholtz Metadata Collaboration](#)'s 2020 project call.. Thanks to colleagues at DLR, HZDR, FZ Jülich.

Projects

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

Contact

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

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