WHAT RSEs SHOULD KNOW ABOUT SOFTWARE CITATION

Stephan Druskat
German Aerospace Center (DLR), Institute for Software Technology

DOI 10.5281/zenodo.7560153

Software
Sustainability
Institute

DLR

Who / what am I?

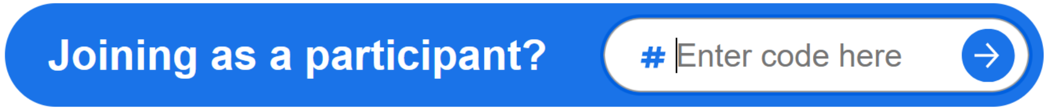
- MA in English, modern German literature, linguistics
- RSE in linguistics (2009-2020)
- Fellow of the Software Sustainability Institute (since 2018)
- Doctoral researcher in computer science (since 2019)

- Project lead Citation File Format
- Co-founder / inaugural board member of de-RSE
- PI HERMES project

Your experience with software citation

Please share your experience with software citation:

1. Go to slido.com
2. Enter the code **4123 315**



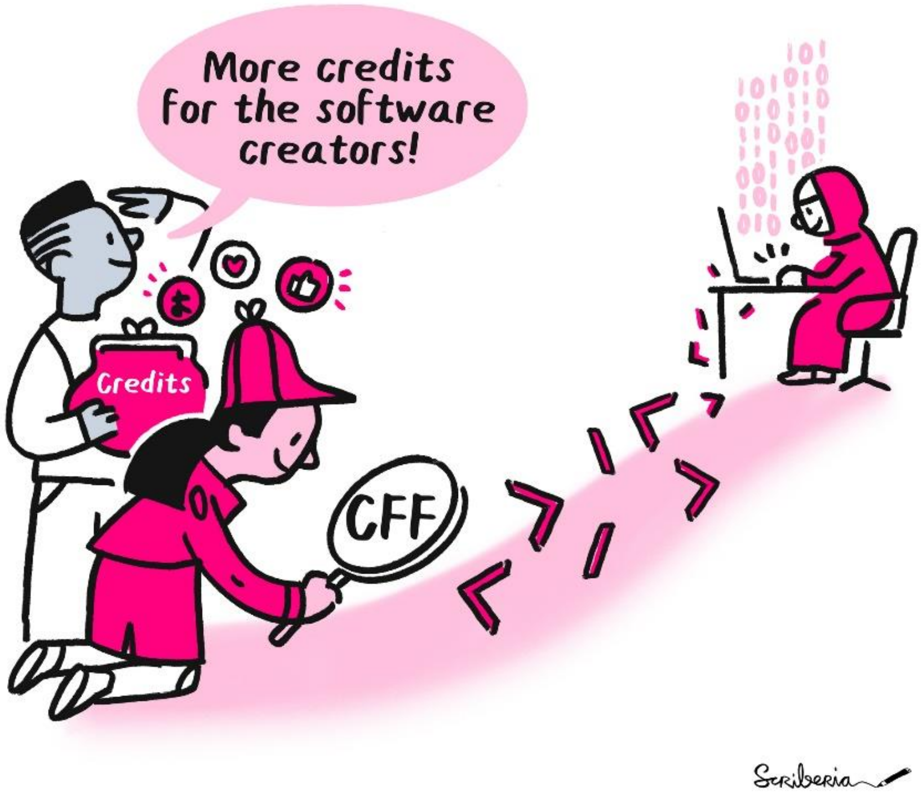
Joining as a participant?

Enter code here

Why should RSEs care about software citation?

- **Importance:** research software as part of the scholarly record

- **Credit:** it‘s fair, for RSEs
- **Findability, accessibility, reusability:** it‘s FAIR, for research software
- **Reproducibility**



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 citation works

PeerJ Computer Science

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@ieee.org

Academic editor
Silvio Peroni

DOI 10.7717/peerj-cs.86

How to cite this article Smith et al. (2016), Software citation principles. **PeerJ Comput. Sci. 2:e86; DOI 10.7717/peerj-cs.86**

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.

1. **Importance**

   Software is cited like papers are cited.

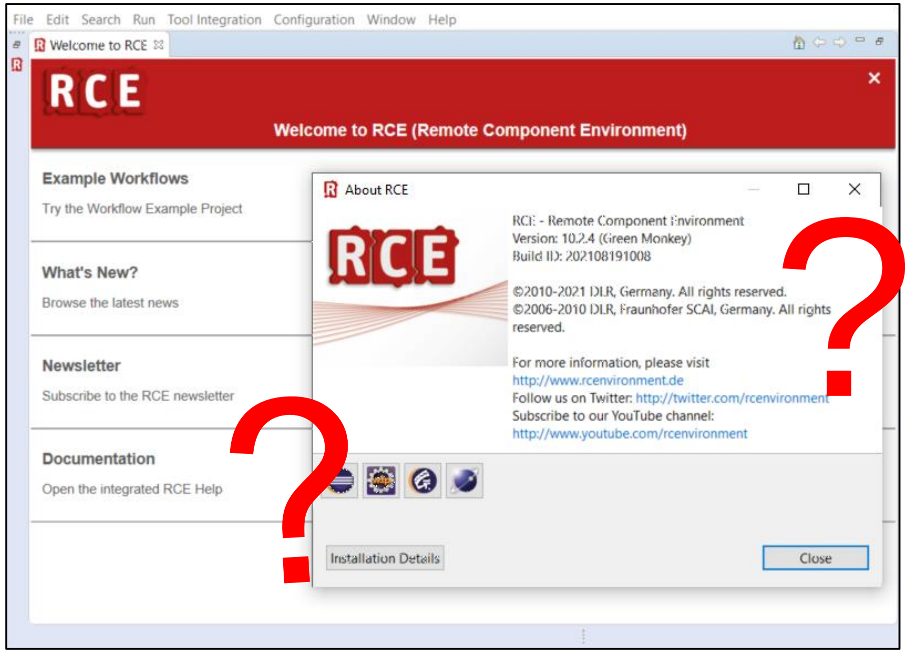
2. **Credit and attribution**
3. **Unique identification**
4. **Persistence**
5. **Accessibility**

   Citation allows access to software and metadata.

6. **Specificity**

   Citation identifies the software version used in research.

Why software isn‘t easily citable


File Edit Search Run Tool Integration Configuration Window Help
Welcome to RCE
RCE

Welcome to RCE (Remote Component Environment)

Example Workflows

Try the Workflow Example Project

What's New?

Browse the latest news

Newsletter

Subscribe to the RCE newsletter

Documentation

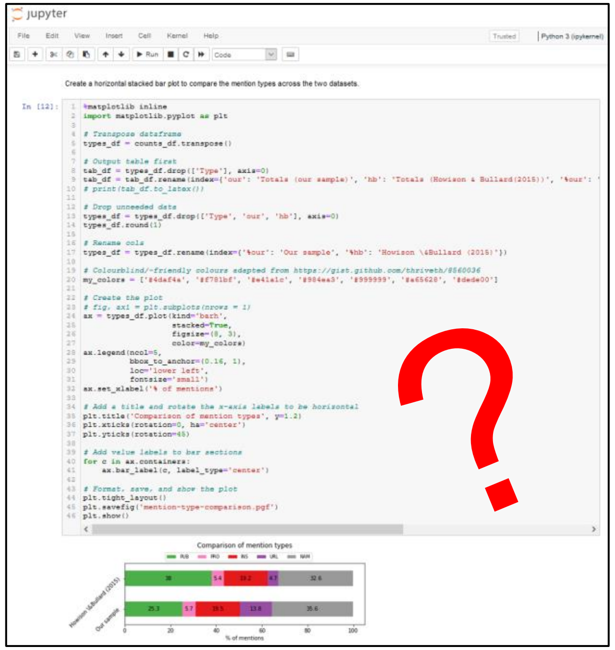
Open the integrated RCE Help

About RCE

RCE - Remote Component Environment
Version: 10.2.4 (Green Monkey)
Build ID: 202108191008

©2010-2021 DLR, Germany. All rights reserved.
©2006-2010 DLR, Fraunhofer SCAI, Germany. All rights reserved.

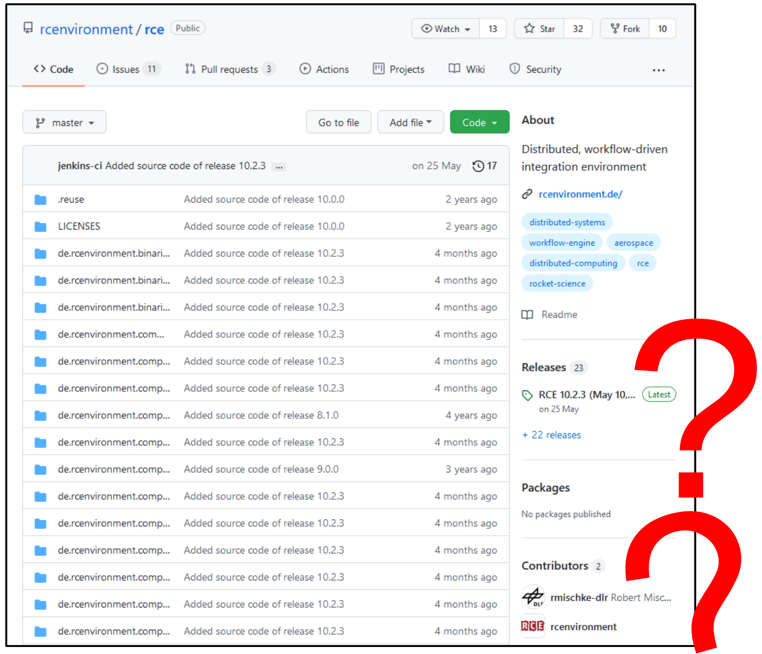
For more information, please visit
http://www.rcenvironment.de
Follow us on Twitter: http://twitter.com/rcenvironment
Subscribe to our YouTube channel:
http://www.youtube.com/rcenvironment



Create a horizontal stacked bar plot to compare the mention types across the two datasets.

In [12]:

```
%matplotlib inline
import matplotlib.pyplot as plt

# Transpose dataframe
types_df = counts_df.transpose()

# Output table first
tab_df = types_df.drop(['Type'], axis=0)
tab_df = tab_df.rename(index={'our': 'Totals (our sample)', 'hb': 'Totals (Howison & Bullard(2015))', '%our': '
# print(tab_df.to_latex())

# Drop unneeded data
types_df = types_df.drop(['Type', 'our', 'hb'], axis=0)
types_df.round(1)

# Rename cols
types_df = types_df.rename(index={'%our': 'Our sample', '%hb': 'Howison \&Bullard (2015)'})

# Colourblind/-friendly colours adapted from https://gist.github.com/thriveth/8560036
my_colors = ['#4daf4a', '#f781bf', '#e41a1c', '#984ea3', '#999999', '#a65628', '#dede00']

# Create the plot
# fig, ax1 = plt.subplots(nrows = 1)
ax = types_df.plot(kind='barh',
                   stacked=True,
                   figsize=(8, 3),
                   color=my_colors)
ax.legend(ncol=5,
          bbox_to_anchor=(0.16, 1),
          loc='lower left',
          fontsize='small')
ax.set_xlabel('% of mentions')

# Add a title and rotate the x-axis labels to be horizontal
plt.title('Comparison of mention types', y=1.2)
plt.xticks(rotation=0, ha='center')
plt.yticks(rotation=45)

# Add value labels to bar sections
for c in ax.containers:
    ax.bar_label(c, label_type='center')

# Format, save, and show the plot
plt.tight_layout()
plt.savefig('mention-type-comparison.pgf')
plt.show()
```



jenkins-ci Added source code of release 10.2.3 ... on 25 May 17

.reuse	Added source code of release 10.0.0	2 years ago
LICENSES	Added source code of release 10.0.0	2 years ago
de.rcenvironment.binari...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.binari...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.binari...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.com...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.comp...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.comp...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.comp...	Added source code of release 8.1.0	4 years ago
de.rcenvironment.comp...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.comp...	Added source code of release 9.0.0	3 years ago
de.rcenvironment.comp...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.comp...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.comp...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.comp...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.comp...	Added source code of release 10.2.3	4 months ago
de.rcenvironment.comp...	Added source code of release 10.2.3	4 months ago

About

Distributed, workflow-driven integration environment

rcenvironment.de/

distributed-systems workflow-engine aerospace distributed-computing rce rocket-science

Readme

Releases 23

RCE 10.2.3 (May 10,... Latest
on 25 May

+ 22 releases

Packages

No packages published

Contributors 2

rmischke-dlr Robert Misc...

rcenvironment

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

MAKING SOFTWARE CITABLE

Making software citation easier as an RSE

1. **Provide citation metadata for your software!**

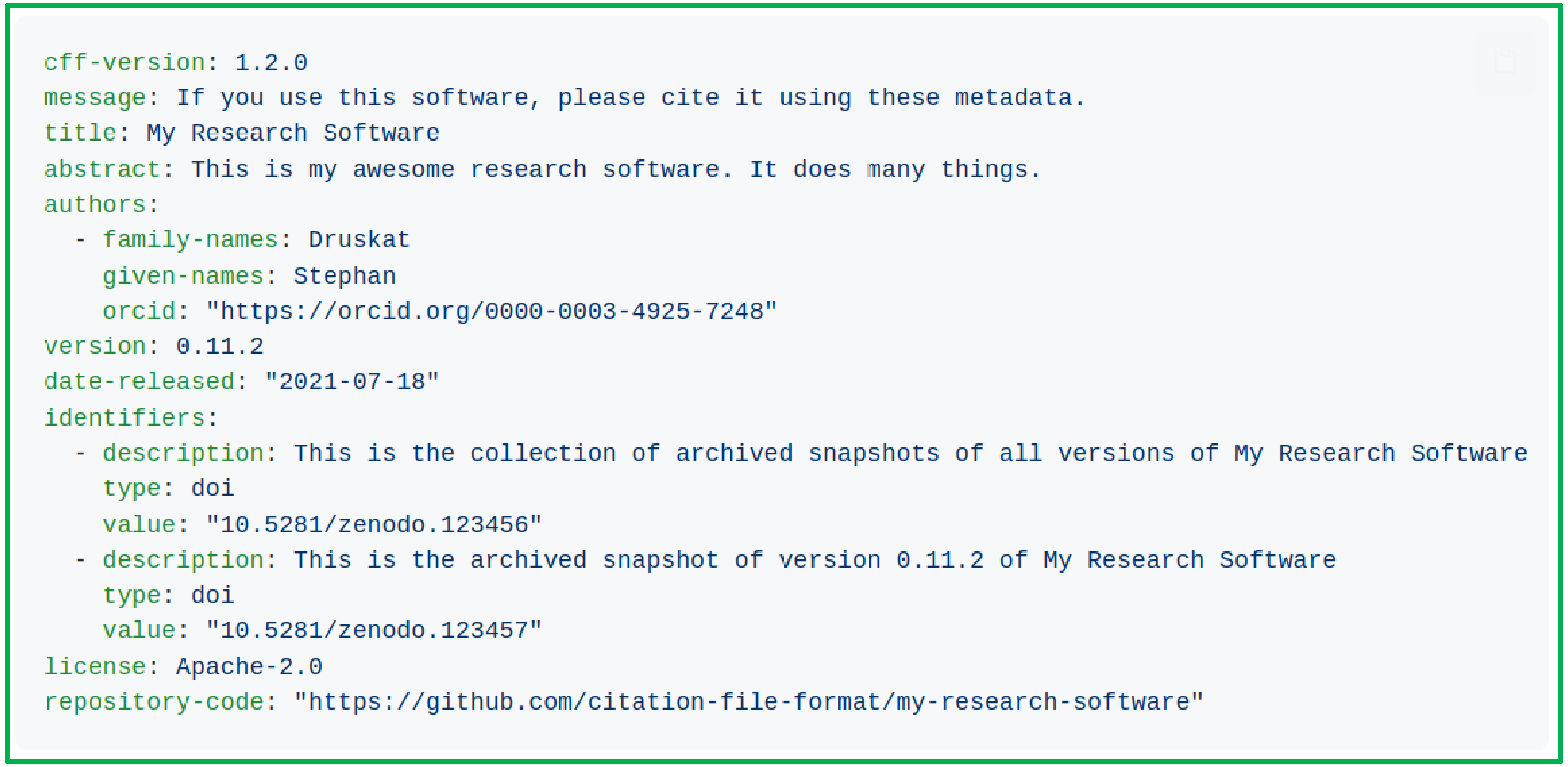
- Only you/your project knows the correct metadata:
  - What‘s the **name** of the software?
  - Who are the **authors**? [*Credit and attribution*]
    - Is the person who fixed that typo / designed the UI / wrote unit tests a "software author"?
  - What is the current **version** of the software? [*Specificity*]
    - 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



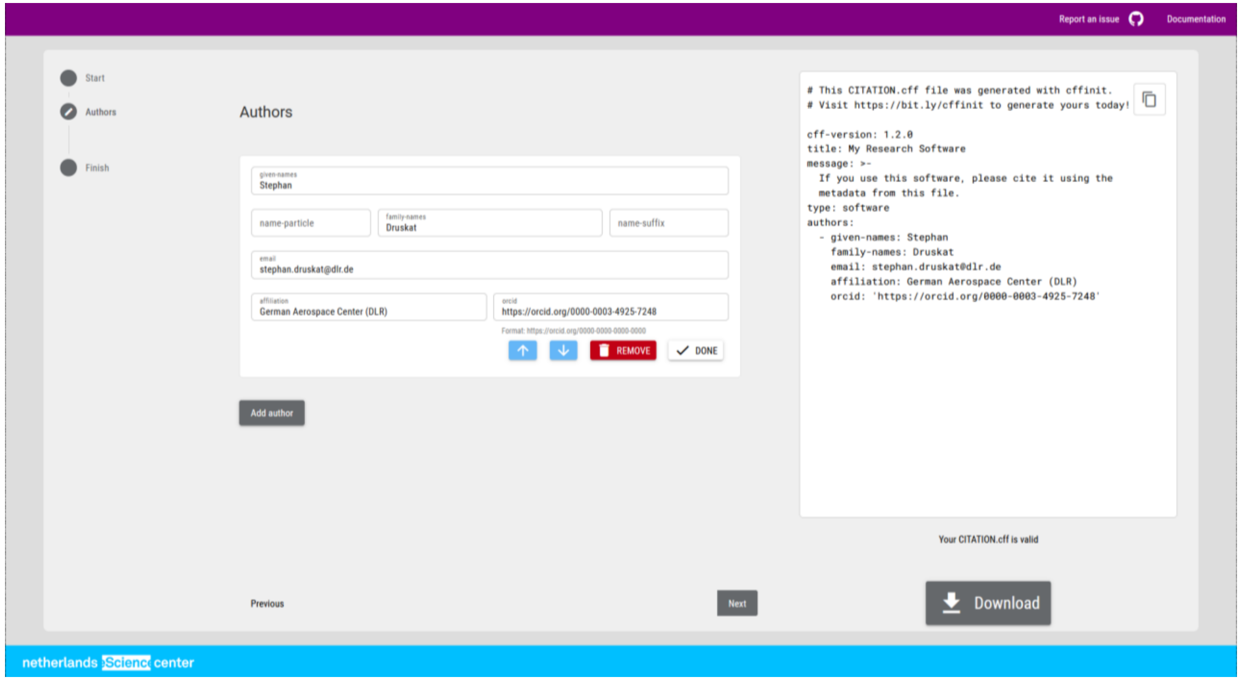
```yaml
cff-version: 1.2.0
message: If you use this software, please cite it using these metadata.
title: My Research Software
abstract: This is my awesome research software. It does many things.
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: "https://orcid.org/0000-0003-4925-7248"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



Authors

given-names: Stephan

name-particle

family-names: Druskat

name-suffix

email: stephan.druskat@dlr.de

affiliation: German Aerospace Center (DLR)

orcid: https://orcid.org/0000-0003-4925-7248

Format: https://orcid.org/0000-0000-0000-0000

```
# This CITATION.cff file was generated with cffinit.
# Visit https://bit.ly/cffinit to generate yours today!

cff-version: 1.2.0
title: My Research Software
message: >-
  If you use this software, please cite it using the
  metadata from this file.
type: software
authors:
  - given-names: Stephan
    family-names: Druskat
    email: stephan.druskat@dlr.de
    affiliation: German Aerospace Center (DLR)
    orcid: 'https://orcid.org/0000-0003-4925-7248'
```

Your CITATION.cff is valid

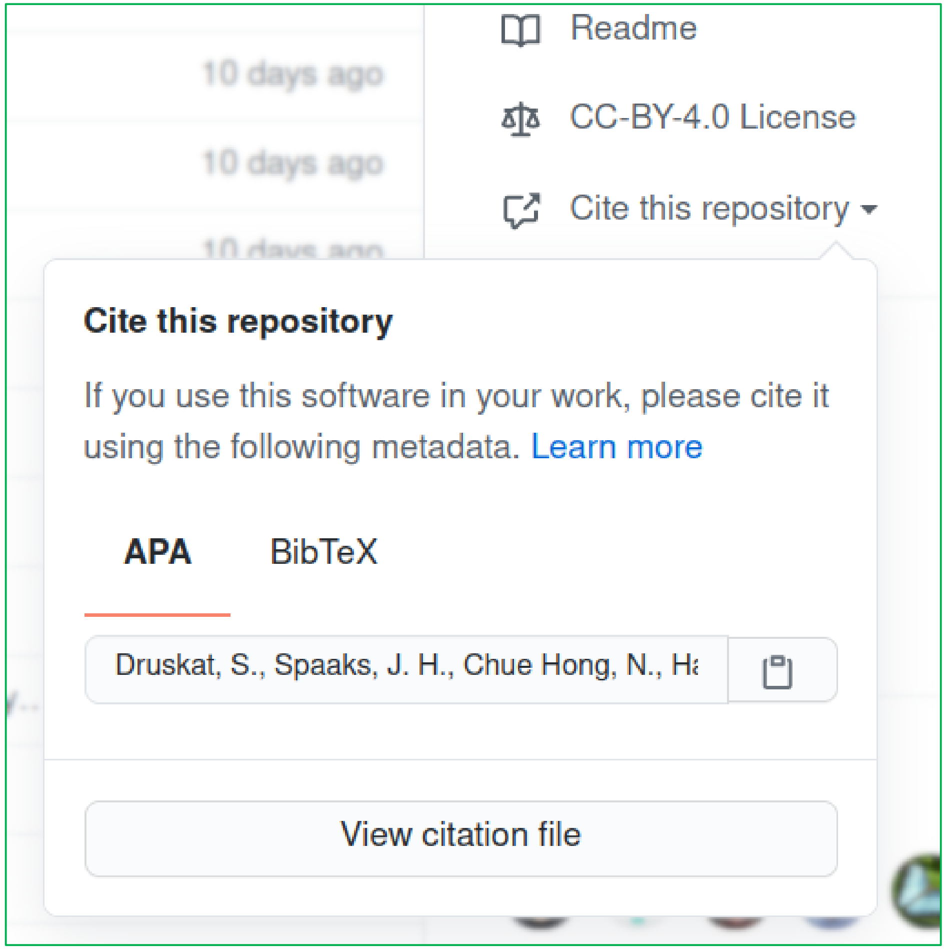
python™
julia
GO
#<_
TypeScript
JS
R
php

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

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

Why use CFF?

2. GitHub support

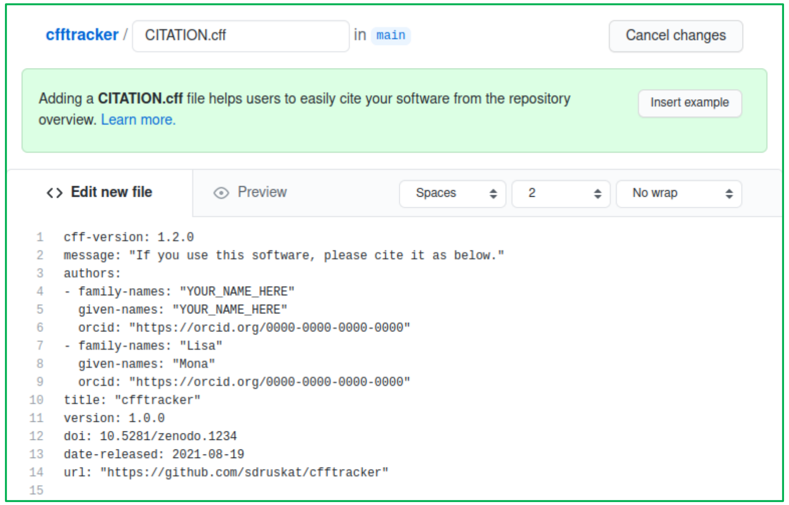


Cite this repository

If you use this software in your work, please cite it using the following metadata. Learn more

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

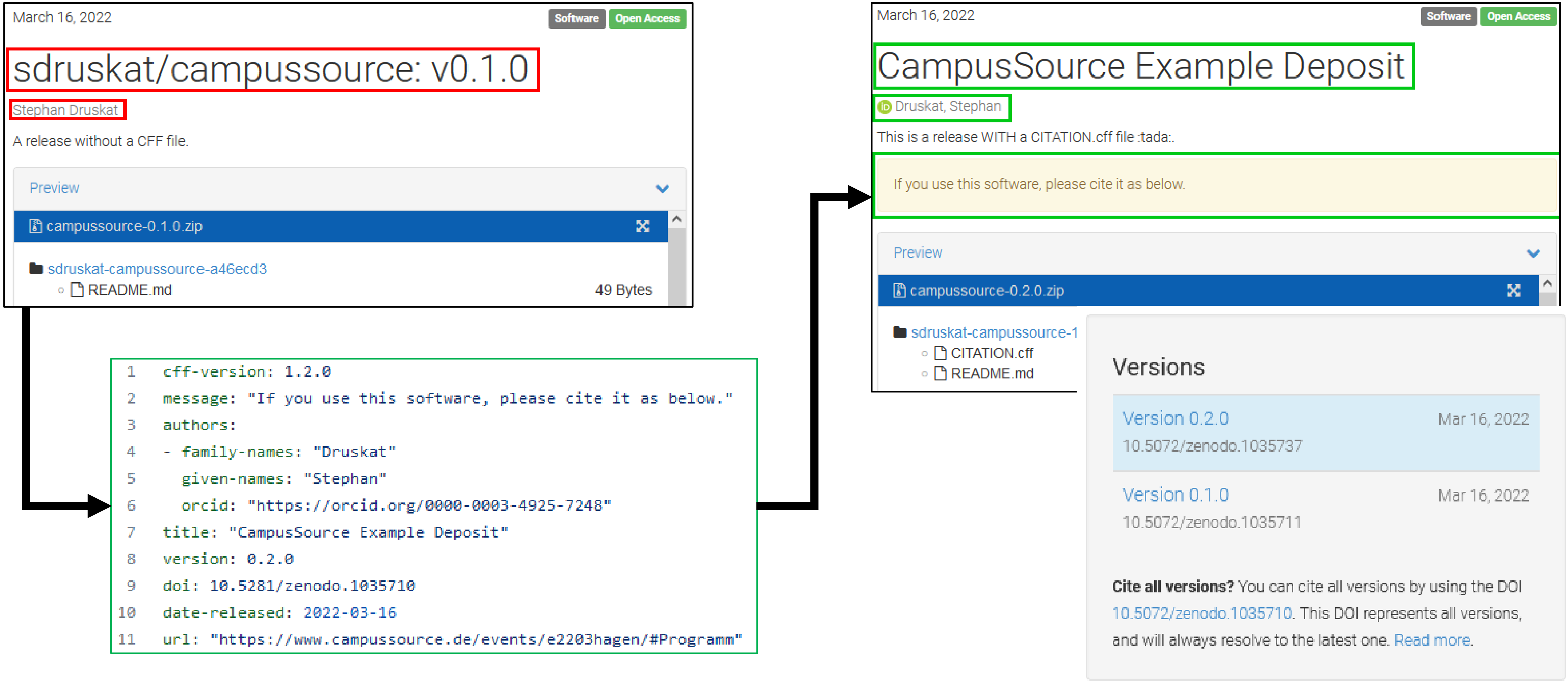
- Detects **`CITATION.cff`** files
- Renders citation metadata (APA, BibTeX)
- Points to **`CITATION.cff`** instance
- [Documents](#) and supports creation



Adding a **CITATION.cff** file helps users to easily cite your software from the repository overview. Learn more.

```yaml
cff-version: 1.2.0
message: "If you use this software, please cite it as below."
authors:
- family-names: "YOUR_NAME_HERE"
  given-names: "YOUR_NAME_HERE"
  orcid: "https://orcid.org/0000-0000-0000-0000"
- family-names: "Lisa"
  given-names: "Mona"
  orcid: "https://orcid.org/0000-0000-0000-0000"
title: "cfftracker"
version: 1.0.0
doi: 10.5281/zenodo.1234
date-released: 2021-08-19
url: "https://github.com/sdruskat/cfftracker"

```

Why use CFF?
3. Zenodo support



March 16, 2022

sdruskat/campussource: v0.1.0

Stephan Druskat

A release without a CFF file.

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

March 16, 2022

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.

Versions

Version 0.2.0 Mar 16, 2022
10.5072/zenodo.1035737

Version 0.1.0 Mar 16, 2022
10.5072/zenodo.1035711

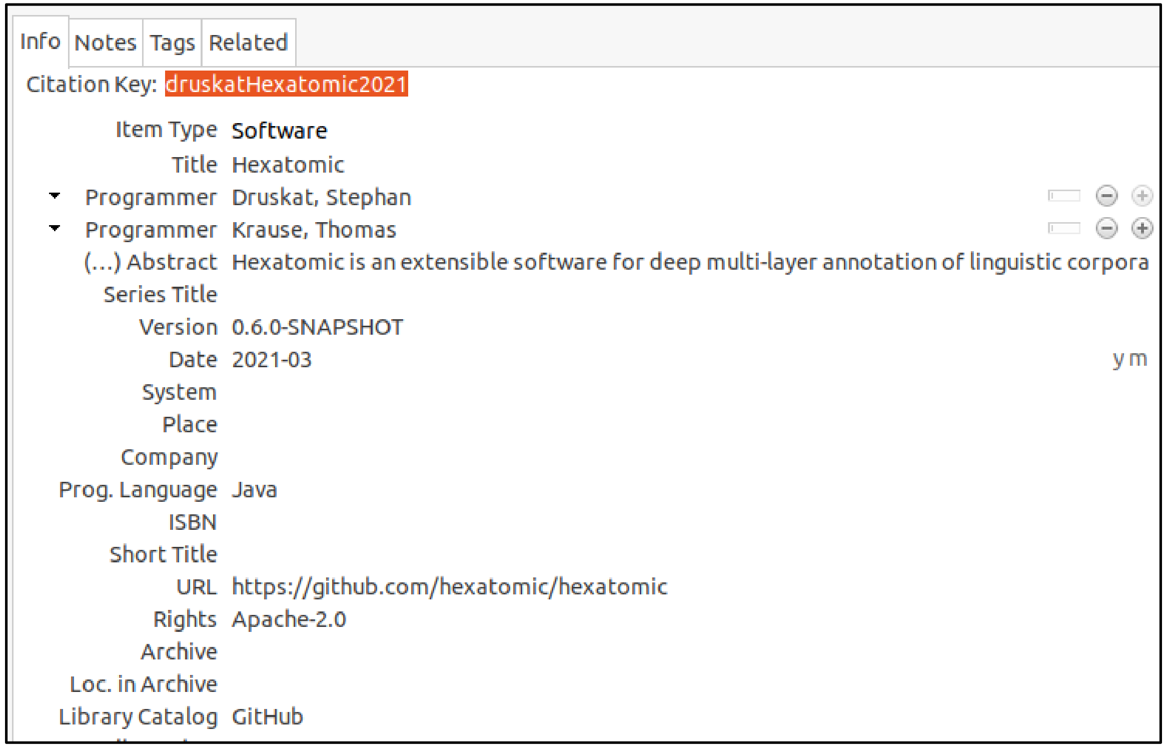
Cite all versions? You can cite all versions by using the DOI 10.5072/zenodo.1035710. This DOI represents all versions, and will always resolve to the latest one. Read more.

Stephan Druskat: “What RSEs should know about software citation", 11th HiRSE Seminar, 2023-01-23

Why use CFF?

4. Reference manager support

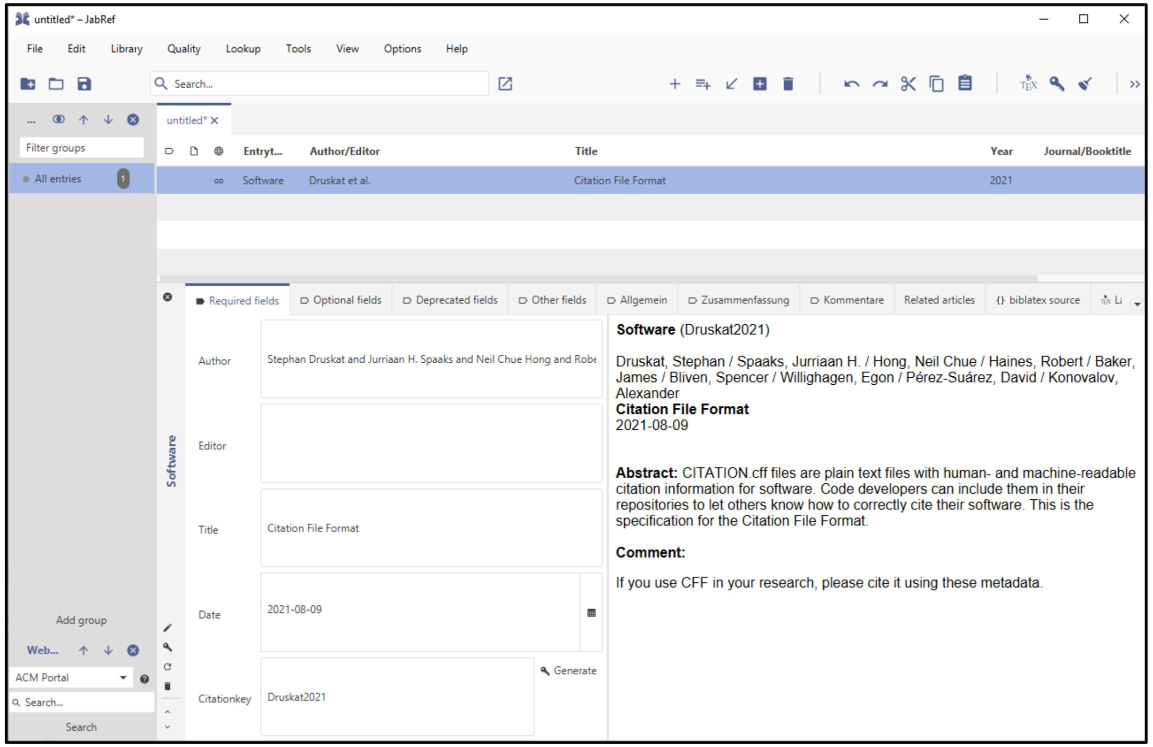
Import in Zotero from GitHub via browser plugin



Citation Key: druskatHexatomic2021

Item Type	Software
Title	Hexatomic
Programmer	Druskat, Stephan
Programmer	Krause, Thomas
(...) Abstract	Hexatomic is an extensible software for deep multi-layer annotation of linguistic corpora
Series Title	
Version	0.6.0-SNAPSHOT
Date	2021-03 y m
System	
Place	
Company	
Prog. Language	Java
ISBN	
Short Title	
URL	https://github.com/hexatomic/hexatomic
Rights	Apache-2.0
Archive	
Loc. in Archive	
Library Catalog	GitHub

Import in JabRef from `CITATION.cff` file



Entryt...	Author/Editor	Title	Year	Journal/Booktitle
Software	Druskat et al.	Citation File Format	2021	

Author	Stephan Druskat and Jurriaan H. Spaaks and Neil Chue Hong and Robe
Editor	
Title	Citation File Format
Date	2021-08-09
Citationkey	Druskat2021

Software (Druskat2021)

Druskat, Stephan / Spaaks, Jurriaan H. / Hong, Neil Chue / Haines, Robert / Baker, James / Bliven, Spencer / Willighagen, Egon / Pérez-Suárez, David / Konovalov, Alexander
Citation File Format
2021-08-09

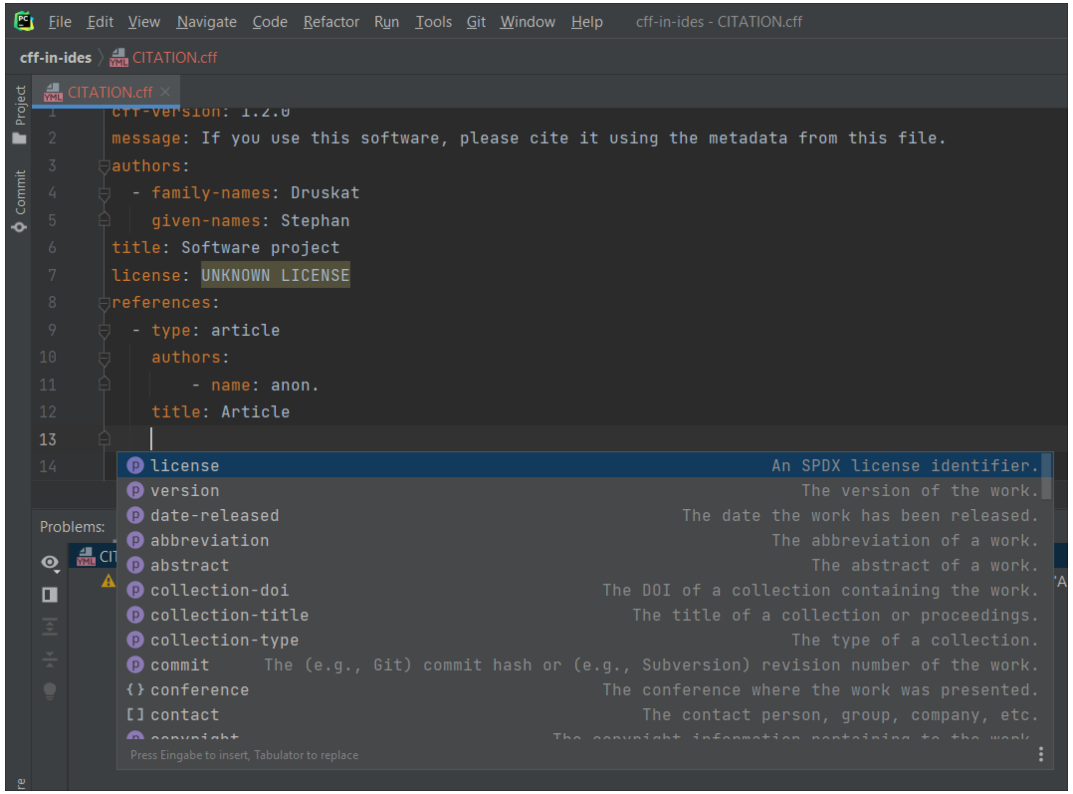
Abstract: CITATION.cff files are plain text files with human- and machine-readable citation information for software. Code developers can include them in their repositories to let others know how to correctly cite their software. This is the specification for the Citation File Format.

Comment:

If you use CFF in your research, please cite it using these metadata.

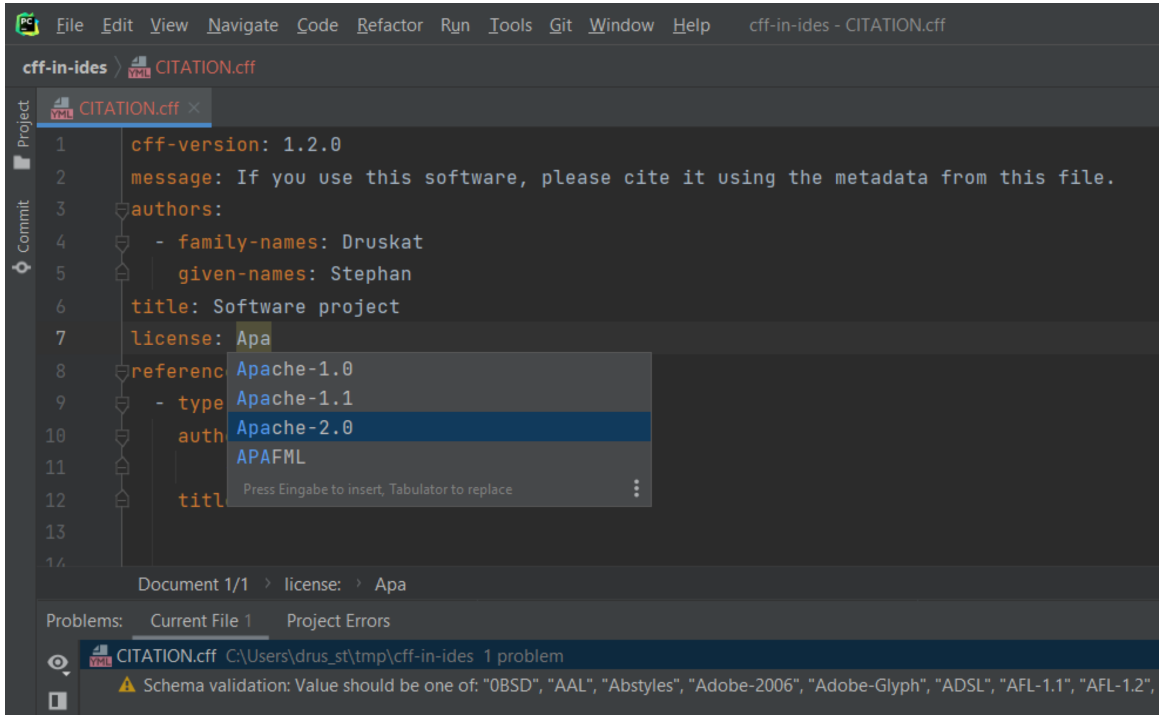
Why use CFF?
5. IDE support

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



```
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.
copyright                              The copyright information pertaining to the work.
Press Eingabe to insert, Tabulator to replace
```



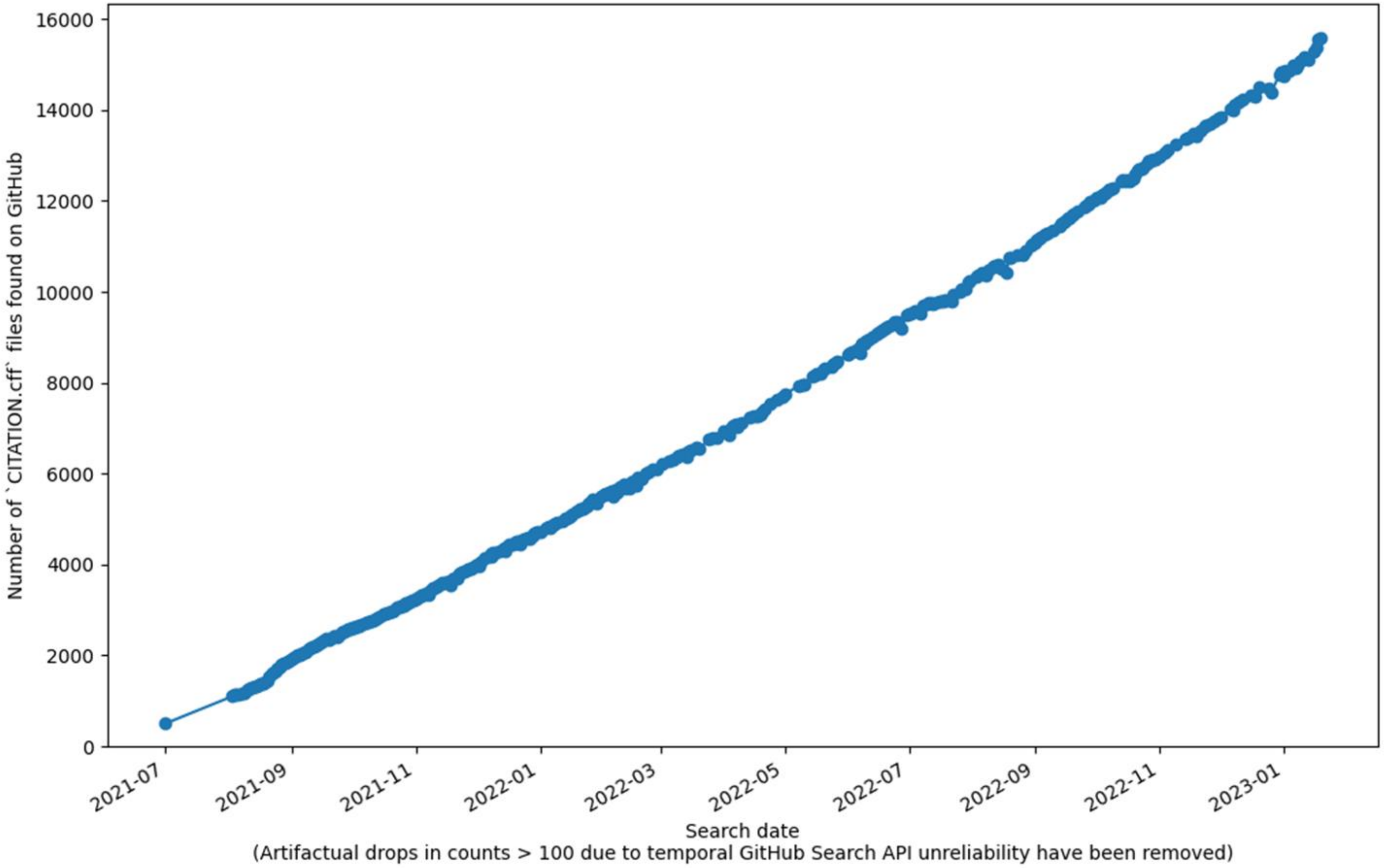
```
cff-version: 1.2.0
message: If you use this software, please cite it using the metadata from this file.
authors:
  - family-names: Druskat
    given-names: Stephan
title: Software project
license: Apa
referenc Apache-1.0
  - type Apache-1.1
    auth Apache-2.0
         APAFML
         Press Eingabe to insert, Tabulator to replace
    titl
```

```
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",
```

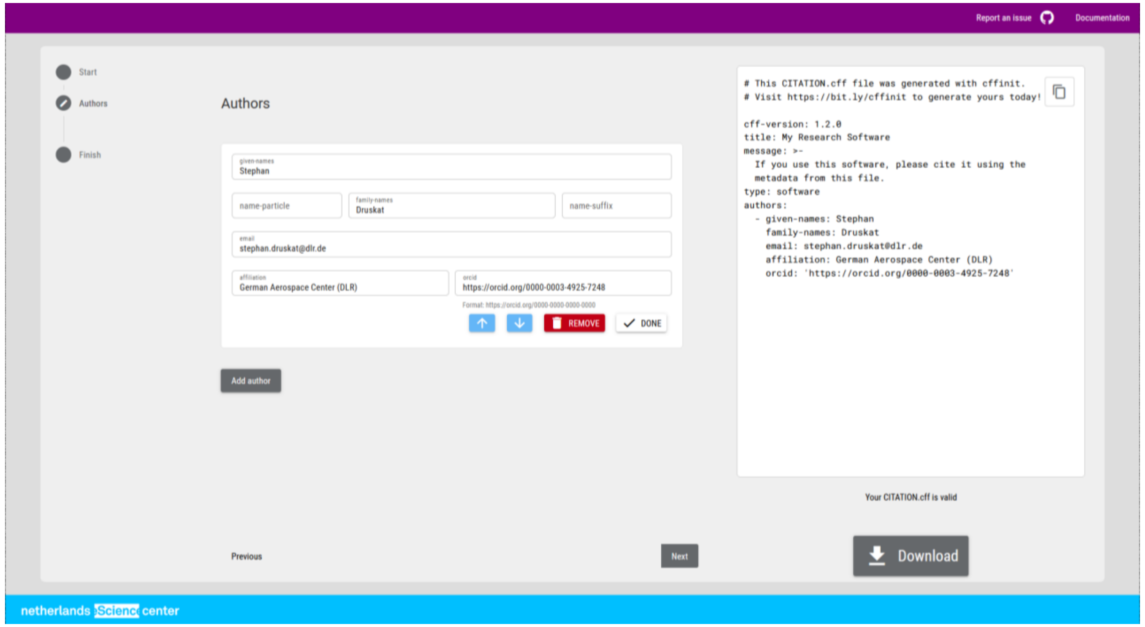
Why use CFF?

6. Everyone does it

- GitHub, as of last Friday (if the GitHub Search API is right) has > **15,000** CFF files



Getting started with the Citation File Format



Authors

given-names: Stephan

name-particle

family-names: Druskat

name-suffix

email: stephan.druskat@dlr.de

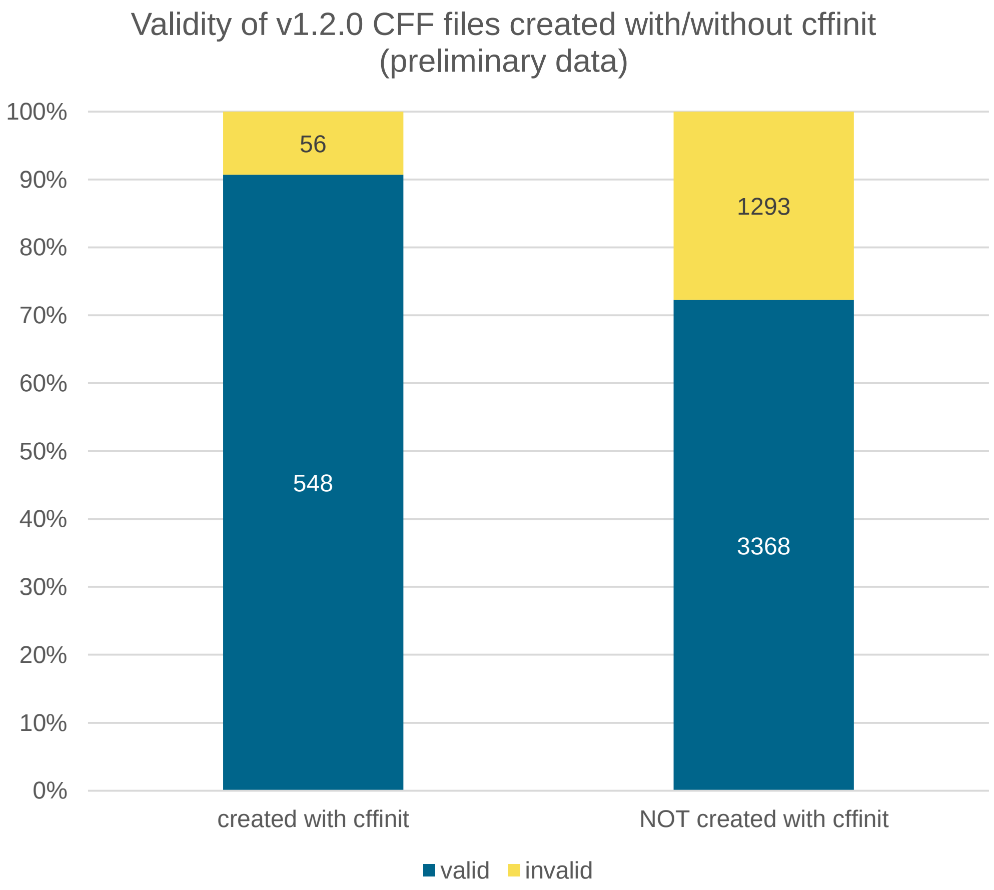
affiliation: German Aerospace Center (DLR)

orcid: https://orcid.org/0000-0003-4925-7248

Format: https://orcid.org/0000-0000-0000-0000

```
# This CITATION.cff file was generated with cffinit.
# Visit https://bit.ly/cffinit to generate yours today!

cff-version: 1.2.0
title: My Research Software
message: >-
  If you use this software, please cite it using the
  metadata from this file.
type: software
authors:
  - given-names: Stephan
    family-names: Druskat
    email: stephan.druskat@dlr.de
    affiliation: German Aerospace Center (DLR)
    orcid: 'https://orcid.org/0000-0003-4925-7248'
```

Your CITATION.cff is valid



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

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 will be counted from 2023

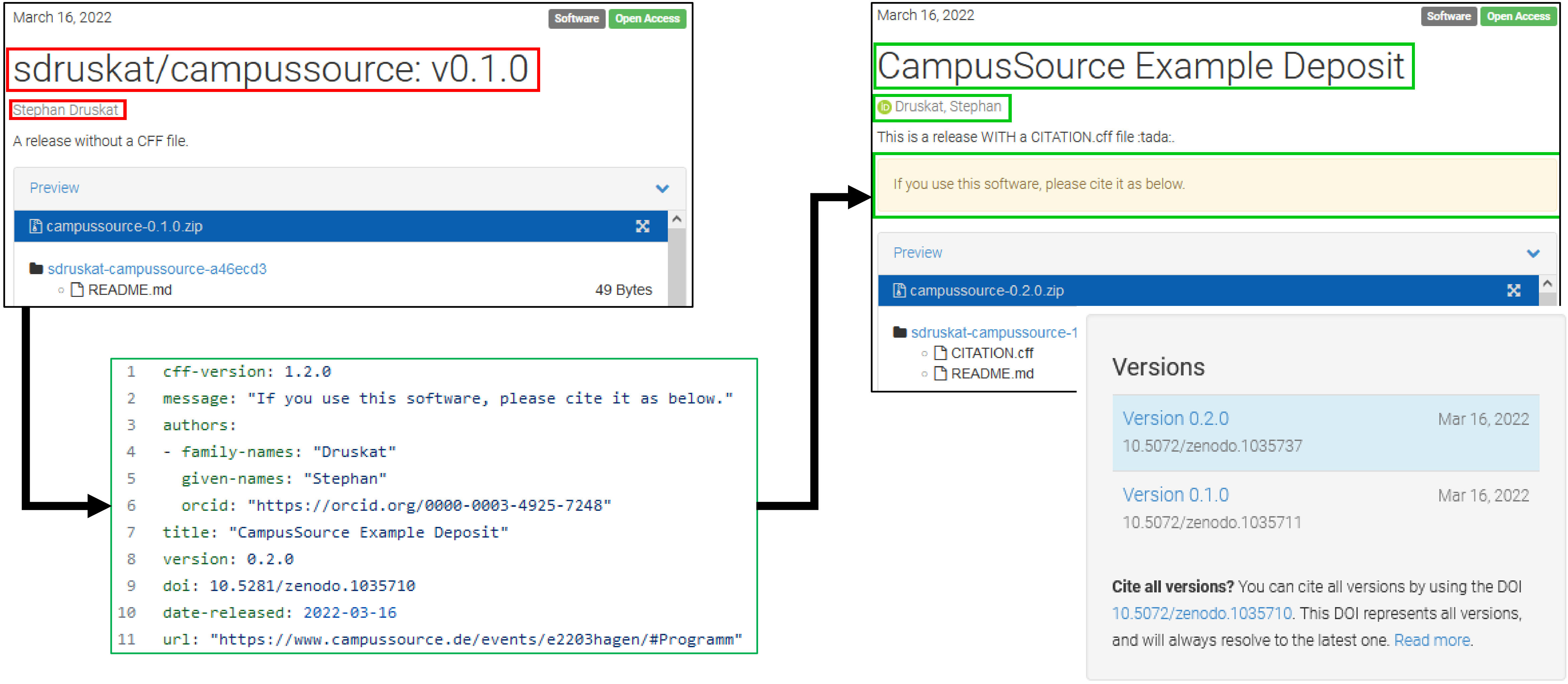
Publishing software (for software citation)

- State of the art: publish software (or at least citation metadata) to an (ideally open access) **publication repository** 🥇
  - Repository publication > software journals > papers
  - 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*]
  - An identifier for the software concept (points to latest version)

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



March 16, 2022

sdruskat/campussource: v0.1.0

Stephan Druskat

A release without a CFF file.

campussource-0.1.0.zip

sdruskat-campussource-a46ecd3
README.md 49 Bytes

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

March 16, 2022

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.

campussource-0.2.0.zip

sdruskat-campussource-1
CITATION.cff
README.md

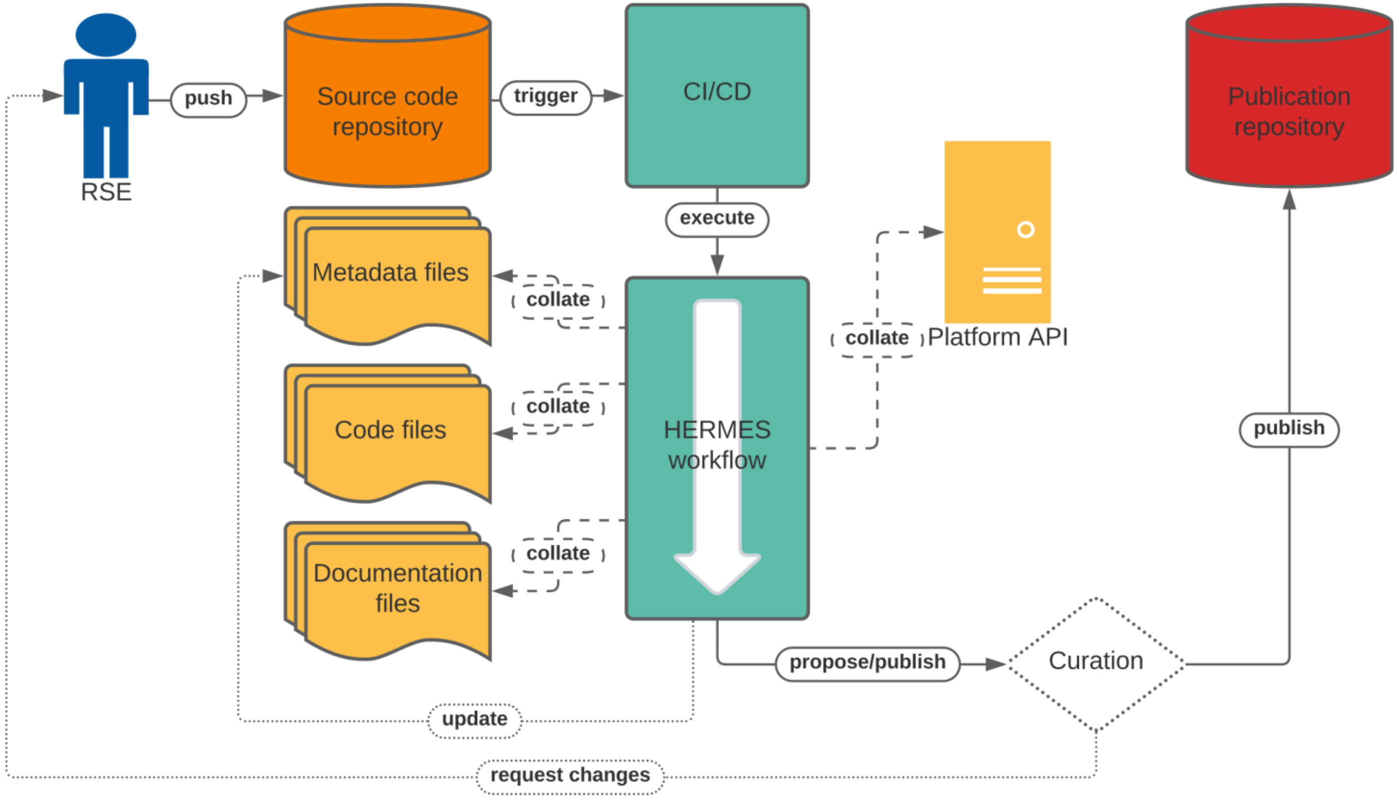
Versions

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

Cite all versions? You can cite all versions by using the DOI 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), first iteration ends 06/2023

- See the HERMES workflow in action at the **deRSE23** demo session: Michael Meinel: *The HERMES workflow for automatic software publication*

deRSE23 - Conference for Research Software Engineering in Germany

20-22 Feb 2023 Paderborn (Germany)

Conclusion

1. Software creators/RSEs should provide the **correct citation metadata** for software in a Citation File Format `CITATION.cff` file in the source code repository.

2. Software creators/RSEs should **publish every release** to Zenodo (or another publication repository).

3. 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@scholar.social

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