

# SOFTWARE ZITIEREN – CITATION FILE FORMAT (CFF)

**Stephan Druskat**

German Aerospace Center (DLR), Institute for Software Technology

DOI [10.5281/zenodo.7346192](https://doi.org/10.5281/zenodo.7346192)

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



Software  
Sustainability  
Institute



Stephan Druskat: "Software zitieren - Citation File Format (CFF)", RSE Express 2022, Westfälische Wilhelms-Universität Münster, 2022-11-22

# Overview



- The Citation File Format (CFF)
- CFF integrations
- „In the wild“: state of the CFF practice



The background of the slide is a high-resolution image of a satellite in orbit above Earth. The satellite has a central body with various instruments and two long, rectangular solar panel arrays extending outwards. Below the satellite, the Earth's surface is visible, showing a mix of green landmasses and blue oceans, partially covered by white clouds. The curvature of the Earth is visible on the right side of the image.

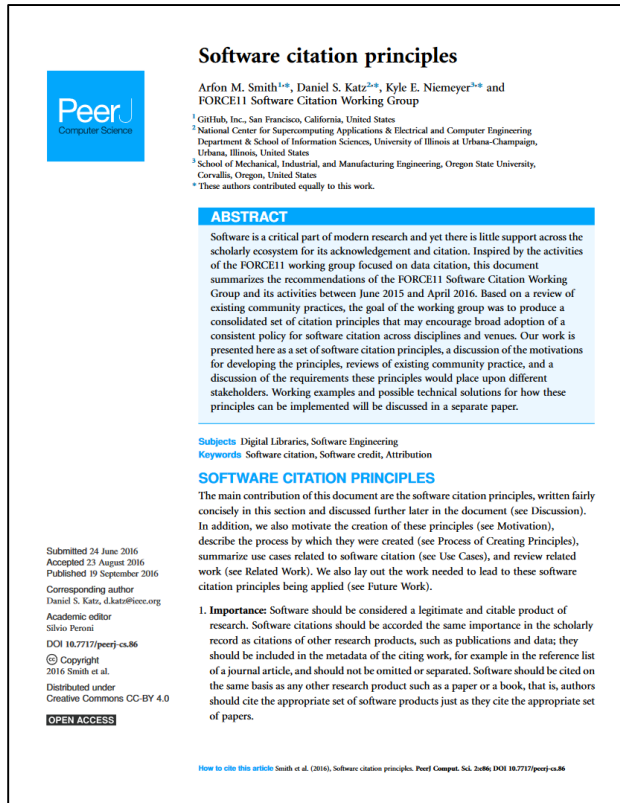
# THE CITATION FILE FORMAT (CFF)

# Why make research software citable?



- **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**

# Software citation principles



## 1. Importance

Software is cited like a paper.

## 2. Credit and attribution

## 3. Unique identification

## 4. Persistence

## 5. Accessibility

Citation enables access to software and metadata.

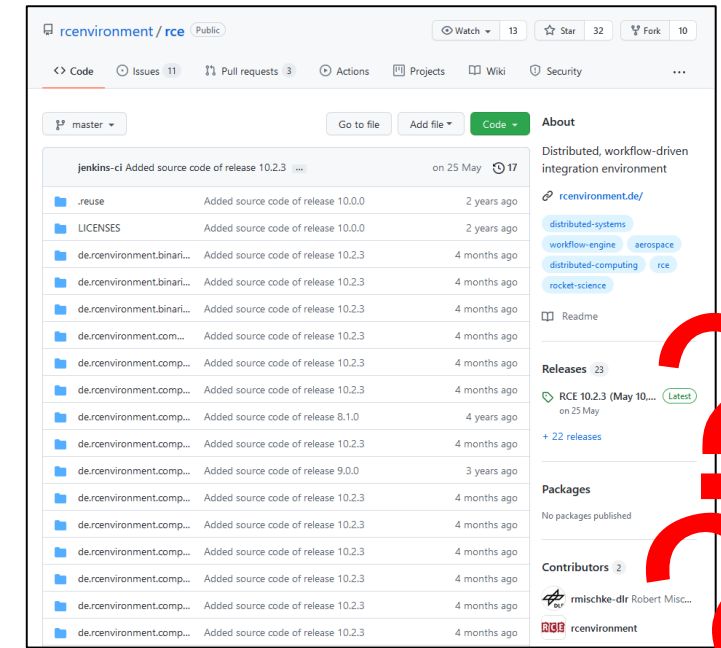
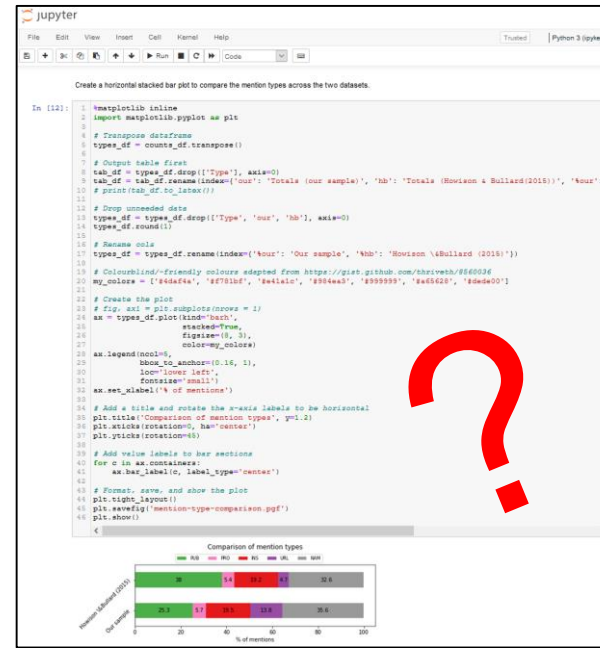
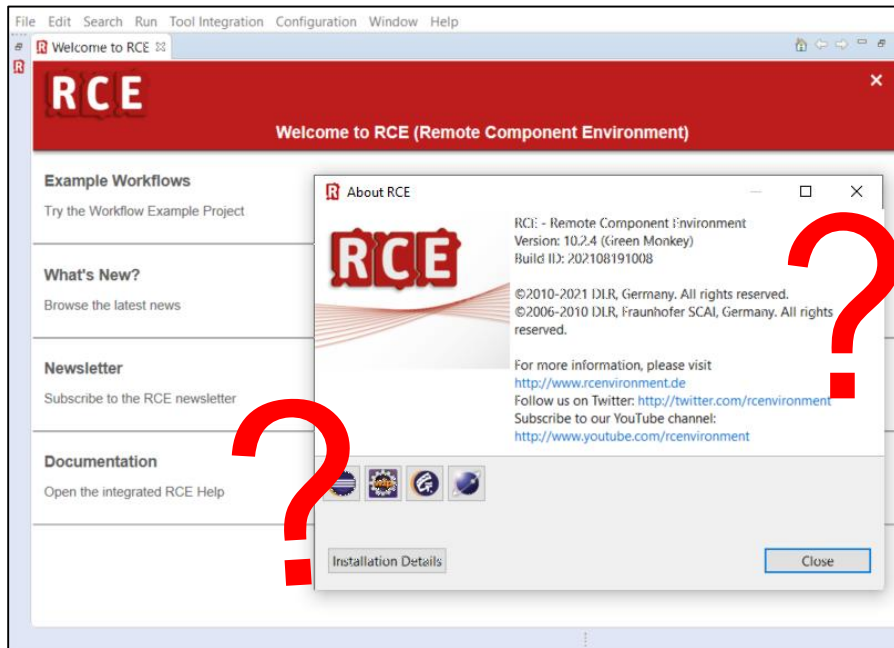
## 6. Specificity

Citation identifies the version of the software that was used.

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



# The problem: metadata Software has no title page



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

# Providing citation metadata with the Citation File Format



- authoritative
- controllable
- "principled":

1. Importance ✓
2. Credit and attribution ✓
3. Unique identification\* ✓
4. Persistence\* ✓
5. Accessibility\*\* ✓
6. Specificity\*\* ✓

```
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, dataset support (experimental)

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

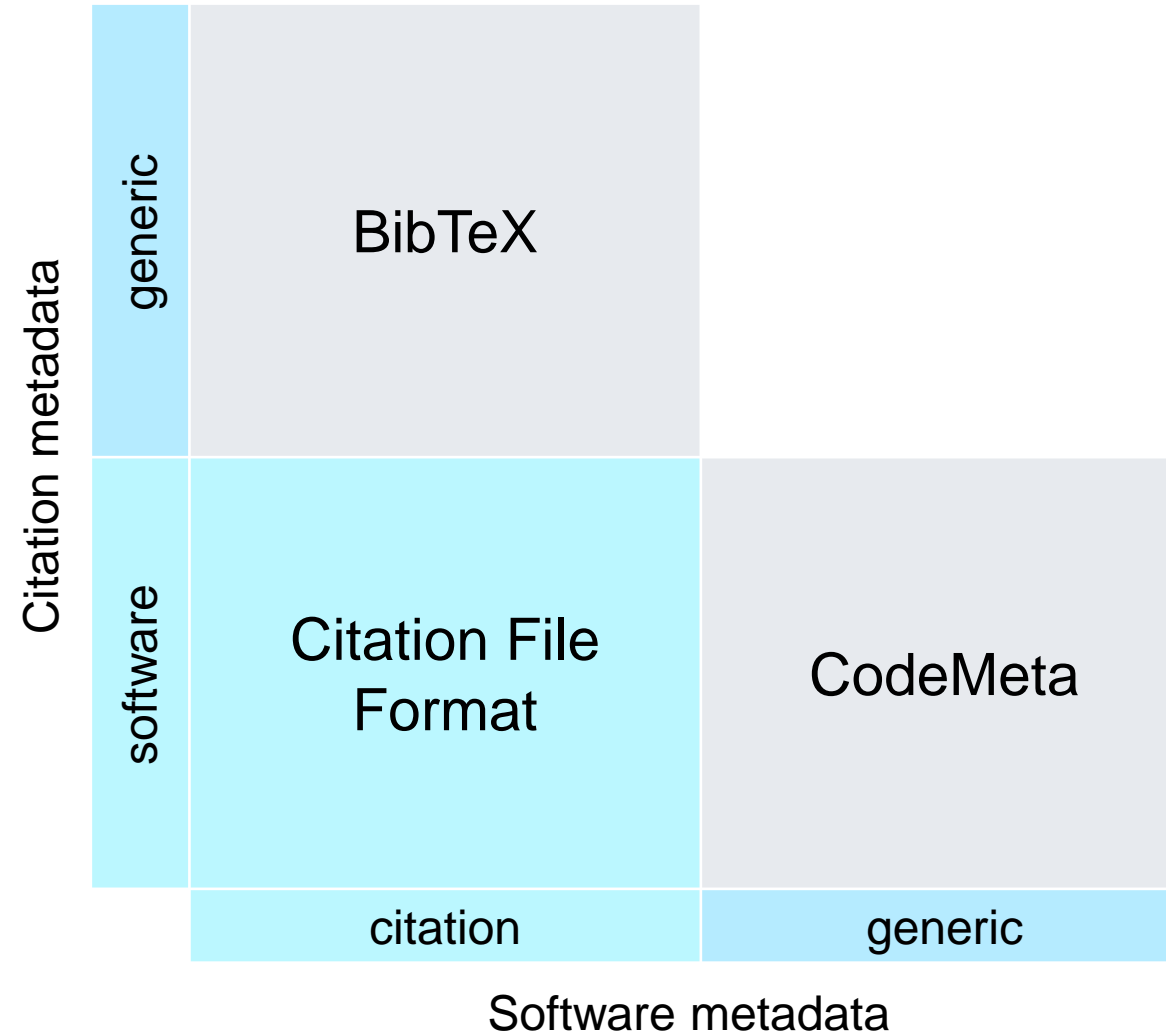
# Citation File Format as a project



- Community project (since WSSSPE5.1, 2017)
- Main development work at DLR, NLeSC, U Manchester
- Openly developed:  
<https://github.com/citation-file-format/>
- Open licenses:  
CC0, CC-BY, MIT, Apache 2.0
- Collaboration with external projects/groups
- Dedicated section in [The Turing Way](#)
- Governance development as part of the CS&S [Digital Infrastructure Incubator](#)
- [CampusSource awardee](#) 2022, Helmholtz [Software Spotlight](#)

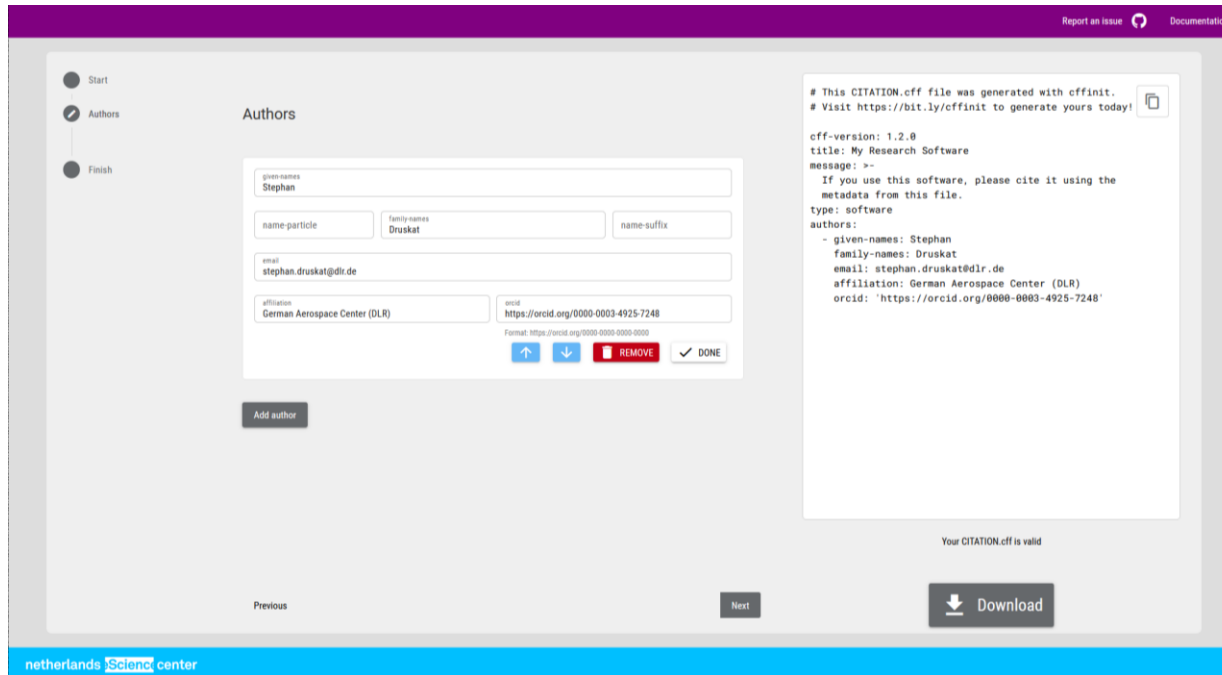


# Citation File Format in the metadata ecosystem



# Citation File Format: more than "just" a format

## Create, edit & update, validate, convert



The screenshot shows the 'Authors' section of the cffinit web tool. On the left, a sidebar has 'Start', 'Authors' (selected), and 'Finish'. The main area has a form for adding an author with 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). Below the form is an 'Add author' button. To the right, a text area shows the generated CFF file content, including the title 'My Research Software' and the author information. At the bottom, there are 'Previous', 'Next', and 'Download' buttons. The footer of the interface says 'netherlands science center'.



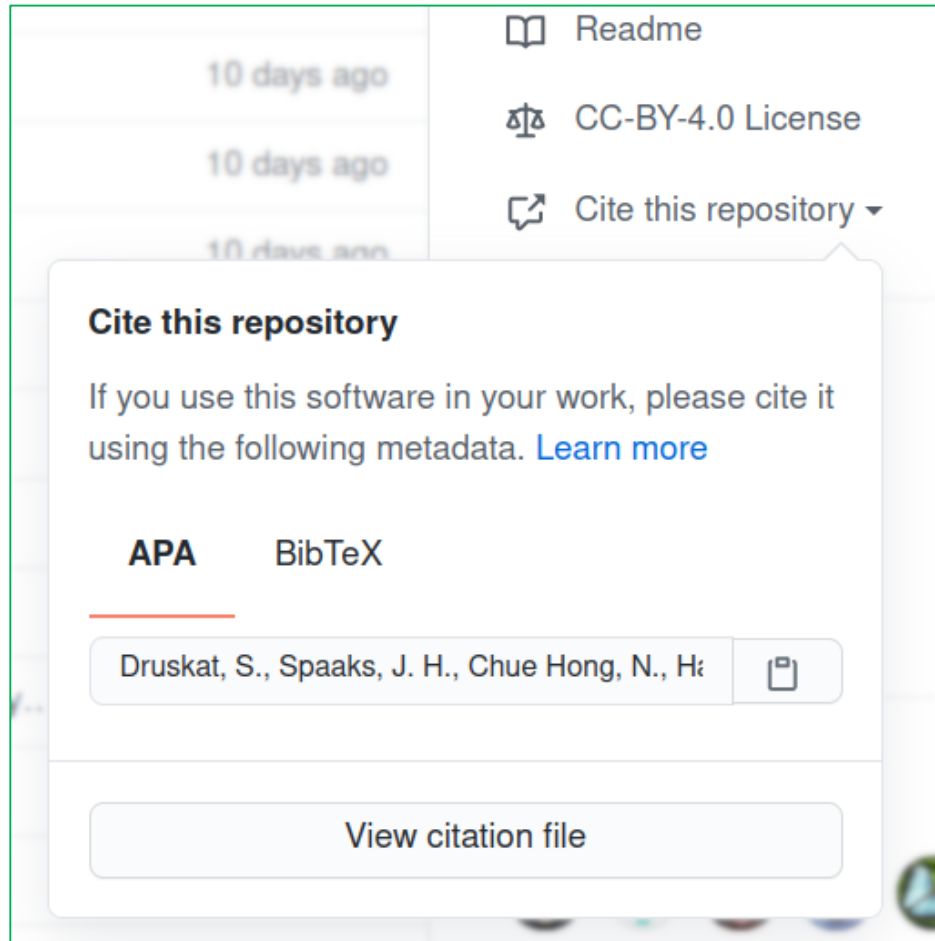
**cffinit:** web tool for CFF file creation - <https://bit.ly/cffinit>

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

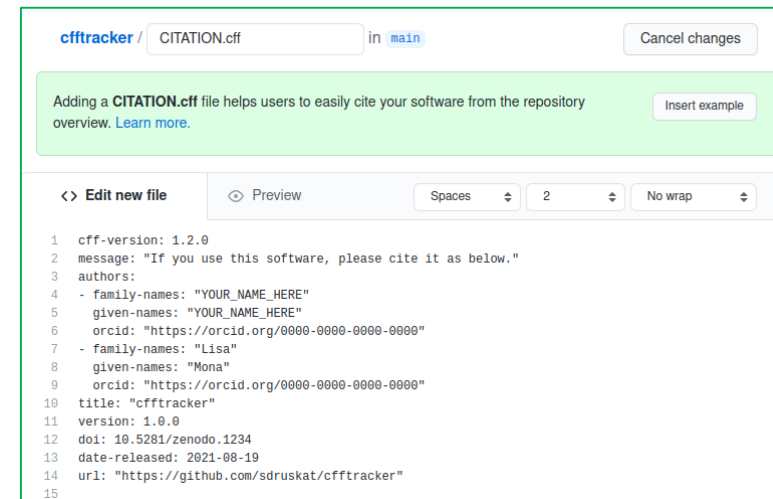
# CFF INTEGRATIONS



# Citation File Format ❤️ GitHub



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





# 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

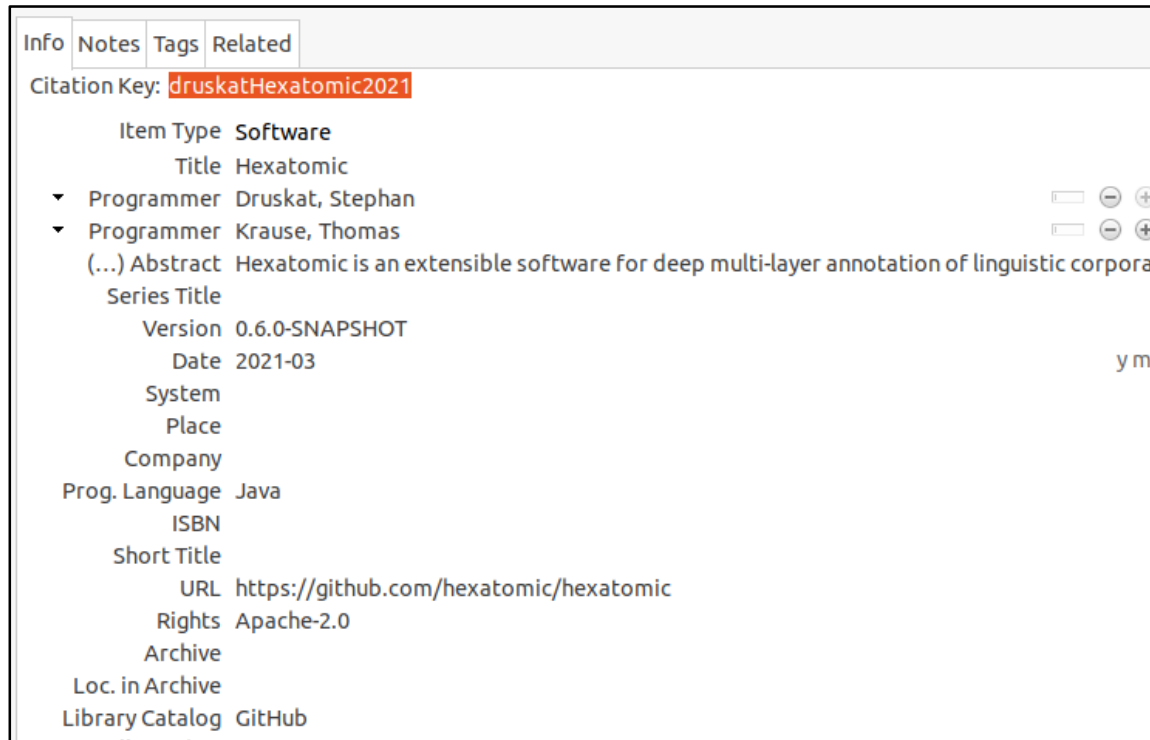
Version 0.2.0	Mar 16, 2022
10.5072/zenodo.1035737	
Version 0.1.0	Mar 16, 2022
10.5072/zenodo.1035711	

**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.](#)

# Citation File Format reference managers



Import in Zotero from GitHub via browser plugin



Info Notes Tags Related

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

System

Place

Company

Prog. Language Java

ISBN

Short Title

URL <https://github.com/hexatomic/hexatomic>

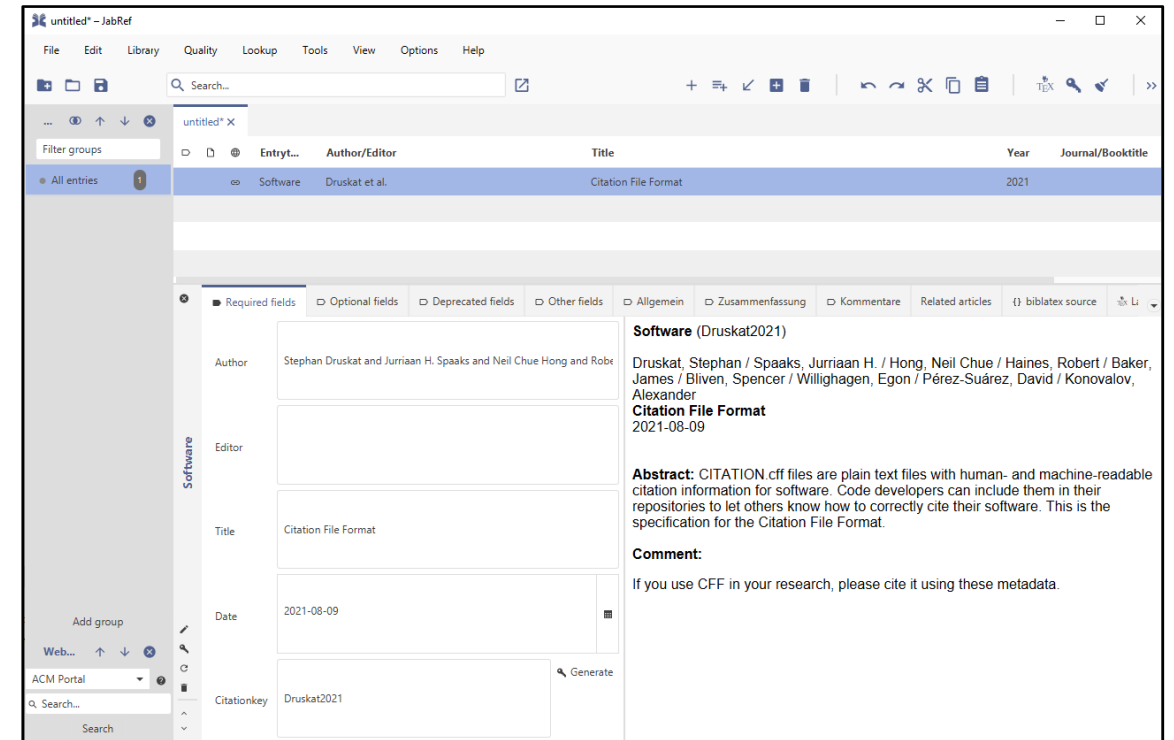
Rights Apache-2.0

Archive

Loc. in Archive

Library Catalog GitHub

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



untitled\* - JabRef

File Edit Library Quality Lookup Tools View Options Help

Search...

Filter groups

untitled\* X

Entry... Author/Editor Title Year Journal/Booktitle

All entries 1

Software Druskat et al. Citation File Format 2021

Required fields Optional fields Deprecated fields Other fields Allgemein Zusammenfassung Kommentare Related articles biblatex source Li

Software (Druskat2021)

Author Stephan Druskat and Jurriaan H. Spaaks and Neil Chue Hong and Robert / Baker, James / Bliven, Spencer / Willighagen, Egon / Pérez-Suárez, David / Kononov, Alexander

Editor

Title Citation File Format

Date 2021-08-09

Citationkey Druskat2021

Generate

**Citation File Format**  
2021-08-09

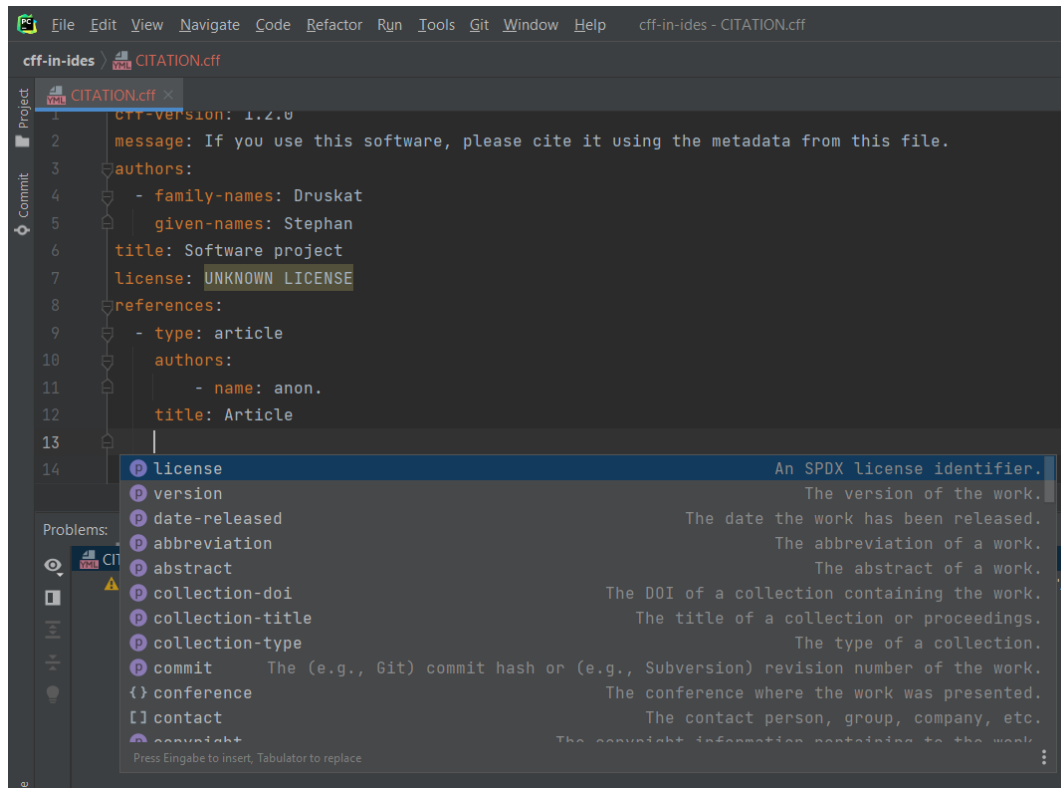
**Abstract:** CITATION.cff files are plain text files with human- and machine-readable citation information for software. Code developers can include them in their repositories to let others know how to correctly cite their software. This is the specification for the Citation File Format.

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

# Citation File Format ♥ IDEs



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

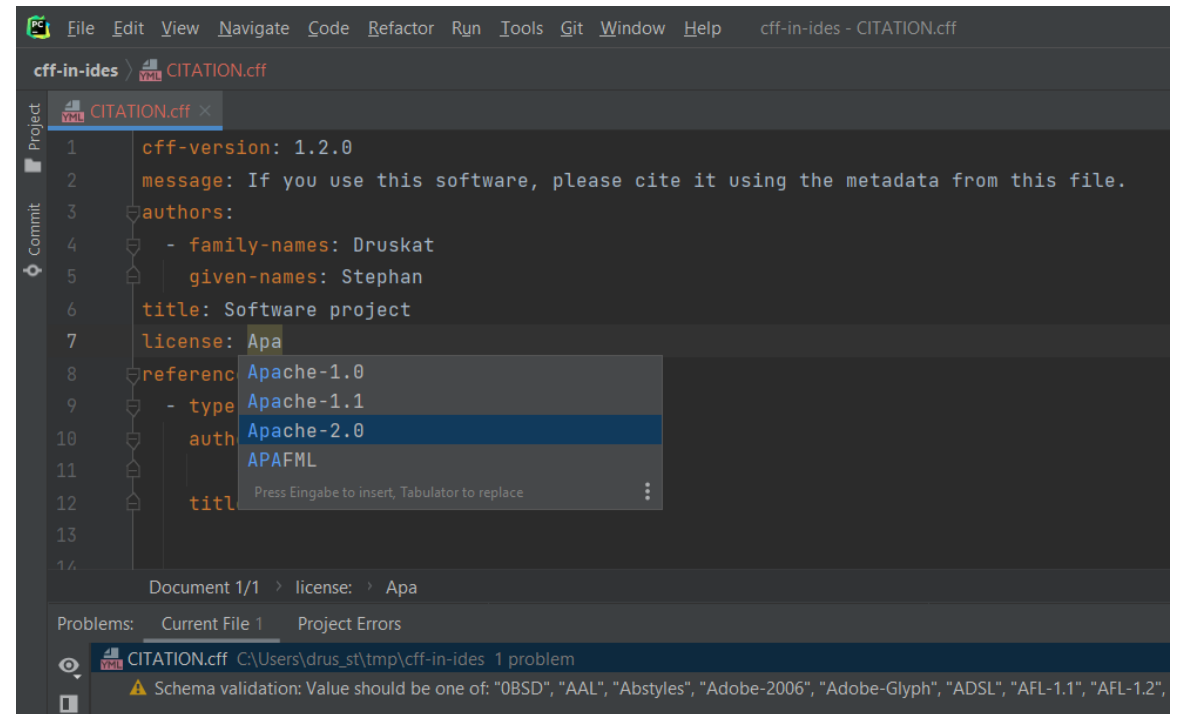


The screenshot shows an IDE window titled 'cff-in-ides - CITATION.cff'. The editor displays a CFF file with the following content:

```
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
```

A dropdown menu is open for the 'license' field, showing a list of valid SPDX license identifiers. The 'license' field is highlighted in the editor, and the dropdown shows the following options:

- 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 as the previous one, but with the 'license' field set to 'Apl'. A dropdown menu is open for the 'license' field, showing a list of valid SPDX license identifiers. The 'license' field is highlighted in the editor, and the dropdown shows the following options:

- 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 dropdown menu is currently showing 'Apl' as the selected option. The IDE also shows a 'Problems' panel at the bottom with a warning message: 'Schema validation: Value should be one of: "0BSD", "AAL", "Abstyles", "Adobe-2006", "Adobe-Glyph", "ADSL", "AFL-1.1", "AFL-1.2", ...'.

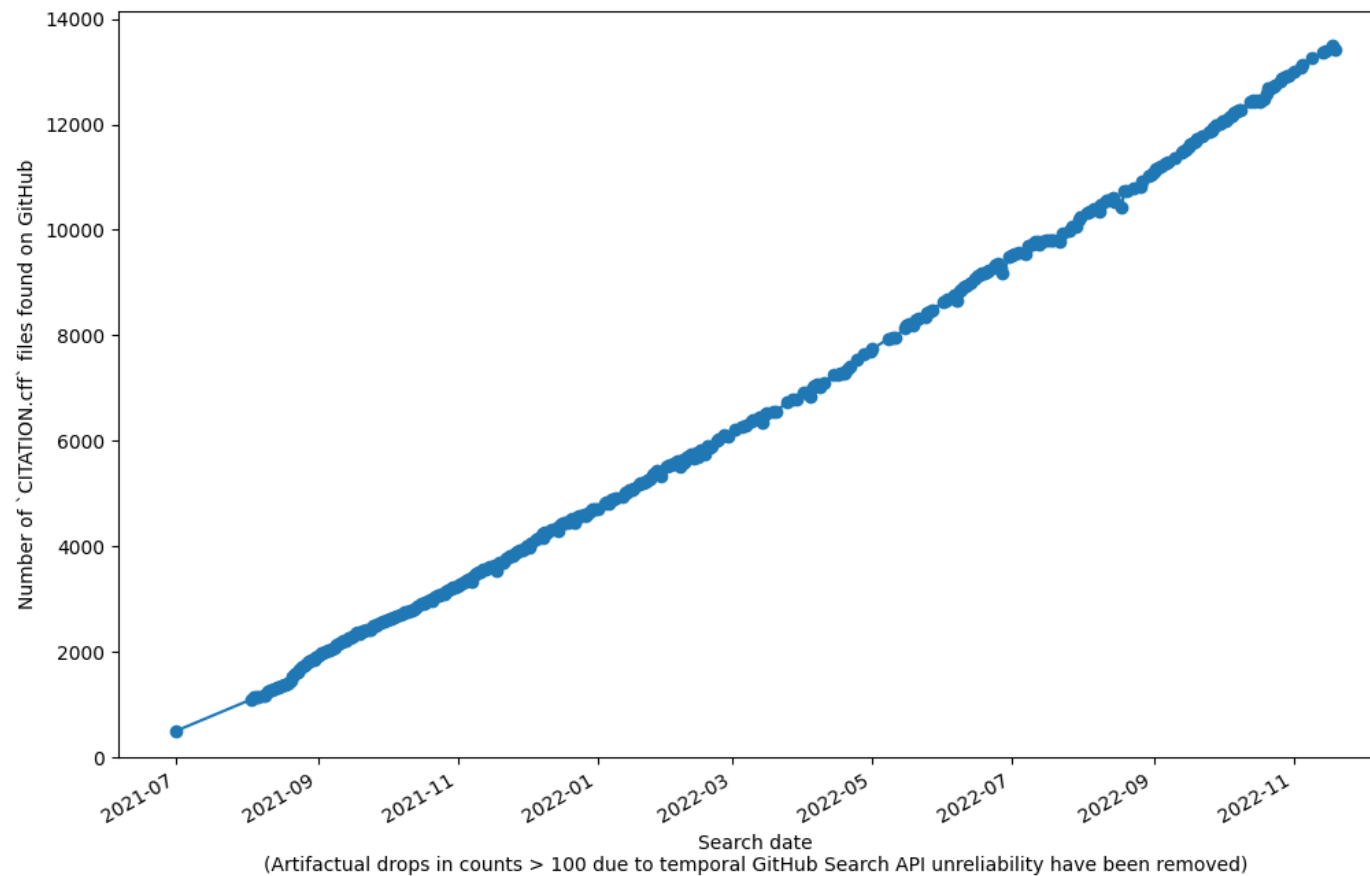


# STATE OF THE CFF PRACTICE



# How many CFF files are there?

- On GitHub, as of yesterday (if the GitHub Search API is right): **13,419**



# What do people's CFF files look like?



- **SSI Collaborations Workshop 2022 hack day:**

- Retrieved and analyzed ~ 3,000 CITATION.cff files from GitHub:
  - We seemed to find that only ~ 1.6% were adhering to the software citation principles
  - Thank you: Amal Alghamdi, Jez Cope, Sam Harrison, Mark Basham, Hugo Gruson

Sample size: 2793

Has version: 2021 (72.4%)

Has code repo URL: 785 (28.1%)

Has DOI: 1187 (42.5%)

Has preferred citation: 469 (16.8%)

*Compliant: 46 (1.6%)*

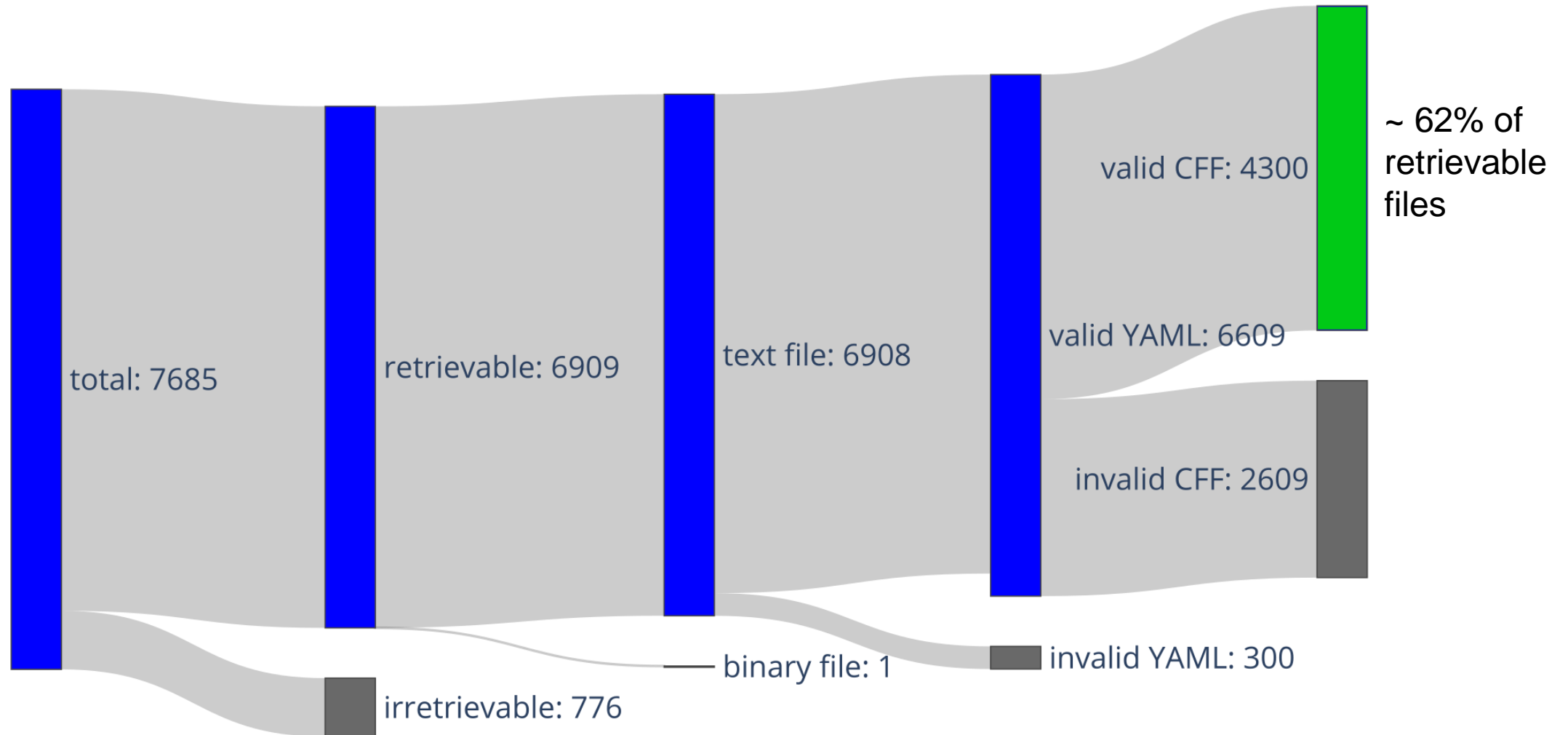
- **GitHub Policy:**

- Provided a list of ~7,000 repositories containing CITATION.cff files
  - Thank you: Peter Cihon, Mike Linksvayer, Kevin Xu, Justin Royal

- **The following results are preliminary**

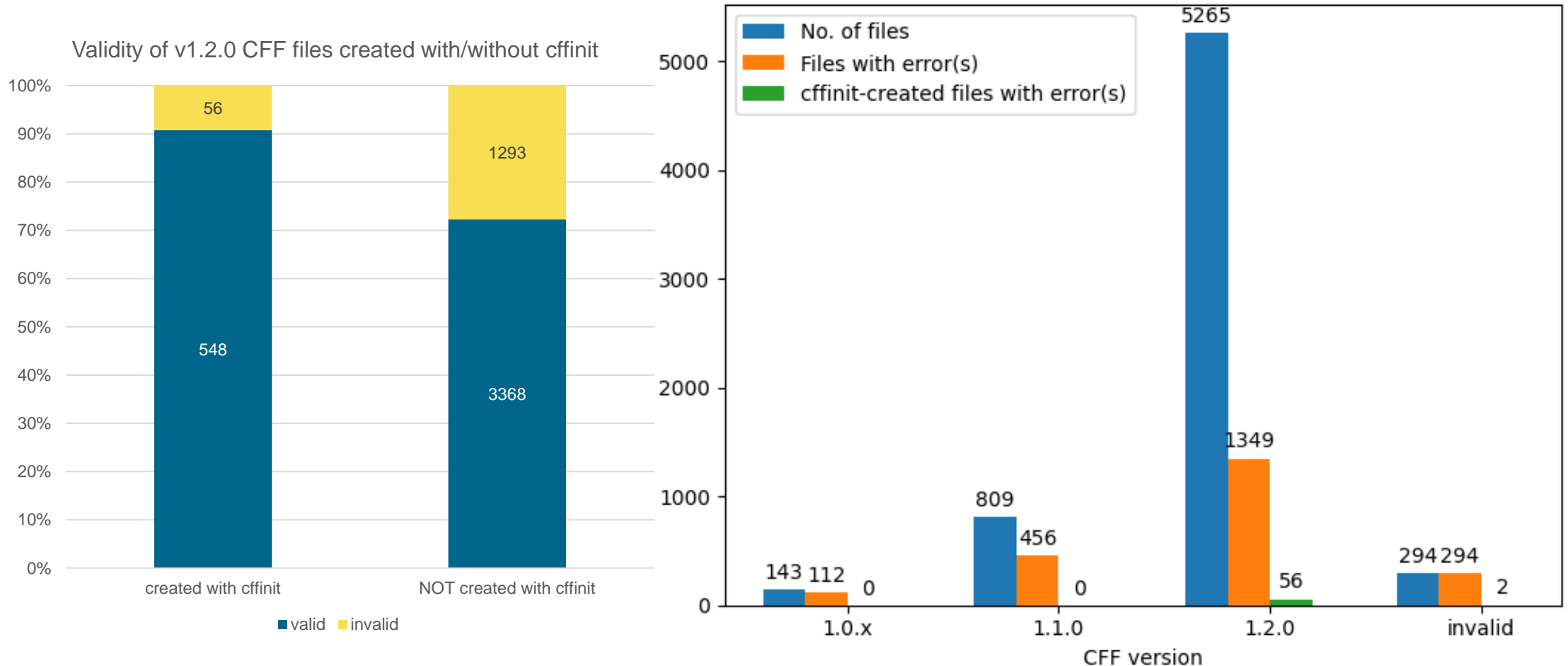
# What do people's CFF files look like?

## The road to validity



# What do people's CFF files look like?

## Does the CFF initialization website improve validity?

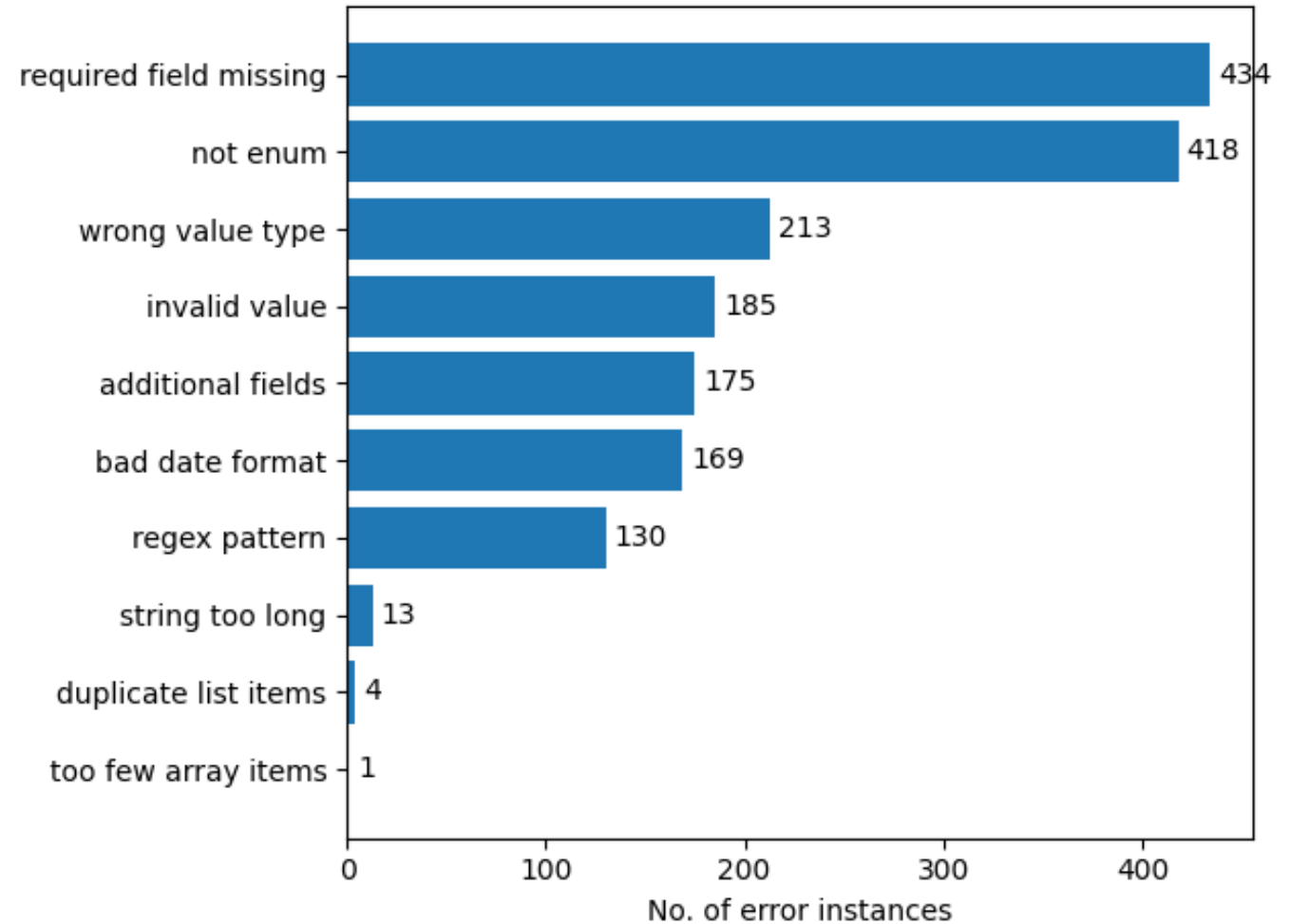




# What do people's CFF files look like?

## CFF validation errors

- Files with errors: 2,609
- Total number of errors across corpus: 4,284
  - Unmapped errors: 2,242

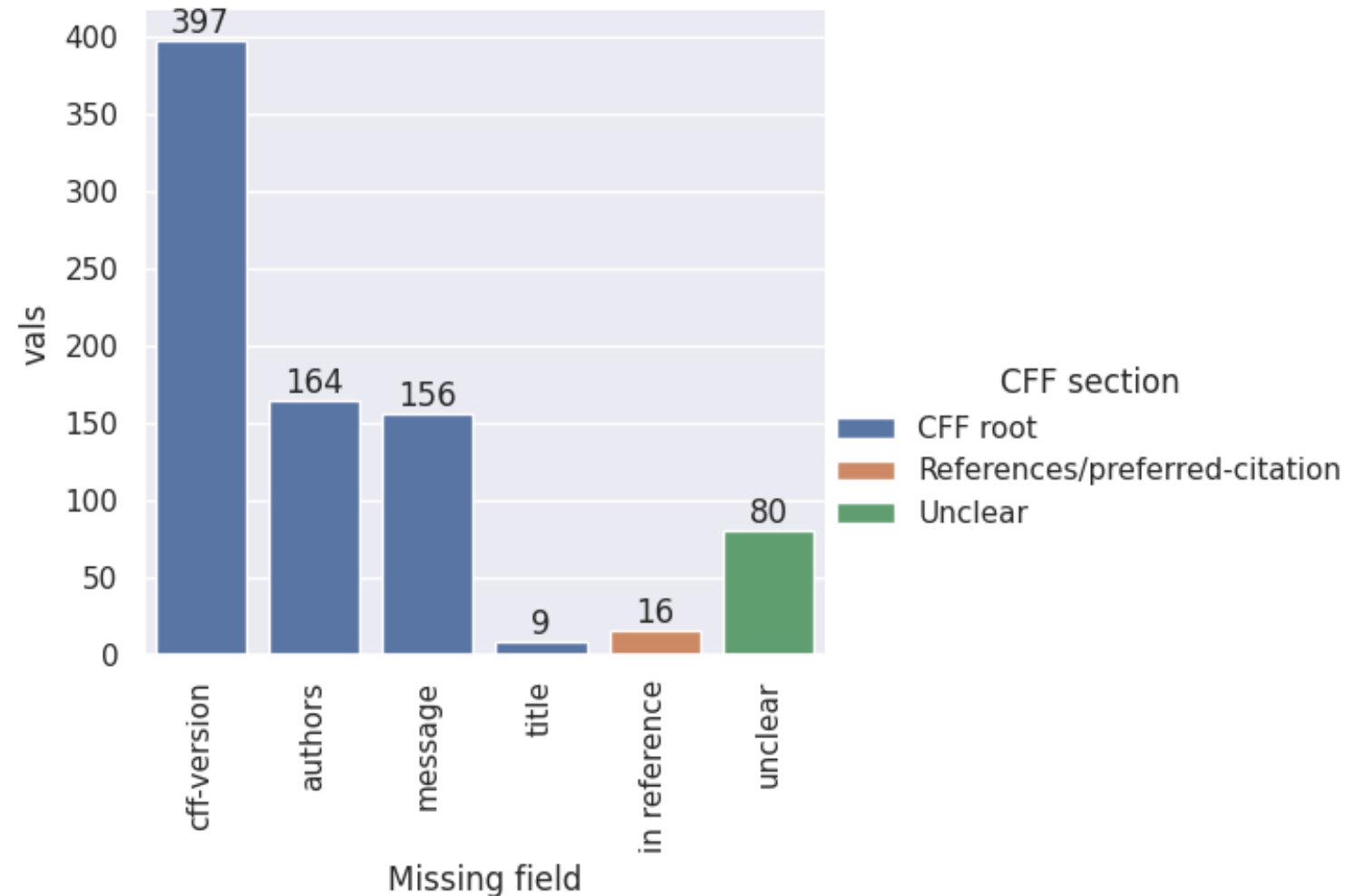


# What do people's CFF files look like?

## CFF validation errors: required fields

- Total „missing required field“ errors (one or more fields):

**852**



# What do people's CFF files look like?

## What now?



- Preliminary conclusions:
  - Tooling (e.g., `cffinit`) helps people create valid `CITATION.cff` files
  - People seem to struggle with required fields and enums
    - Can we improve documentation?
    - What new tooling can we build to help create valid files?
    - Where does tooling need improvement?
    - How can we make existing tooling more visible?
- Take home message:
  - Use <https://bit.ly/cffinit> to create a CFF file for your software

# Thanks!



**Thanks to all contributors to the CFF project!**

**Thanks to the Software Sustainability Institute, Code for Science & Society, CampusSOURCE for funding!**

**Special thanks to Jurriaan H. Spaaks and Rob Haines!**

**Get in touch ...**

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

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

**(... with me)**

[stephan.druskat@dlr.de](mailto:stephan.druskat@dlr.de)

Fediverse: [@sdruskat@scholar.social](https://sdruskat@scholar.social) |

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