THE CITATION FILE FORMAT

Growing up to enable better software citation

Stephan Druskat[1,2], Jurriaan H. Spaaks[3]
[1] German Aerospace Center (DLR), Germany. [2] Humboldt-Universität zu Berlin, Germany. [3] Netherlands eScience Center, Netherlands

DLR
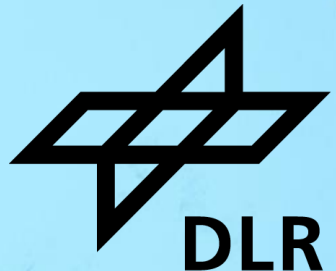
Overview

- The Citation File Format (CFF)

- Growing up: new integrations with CFF

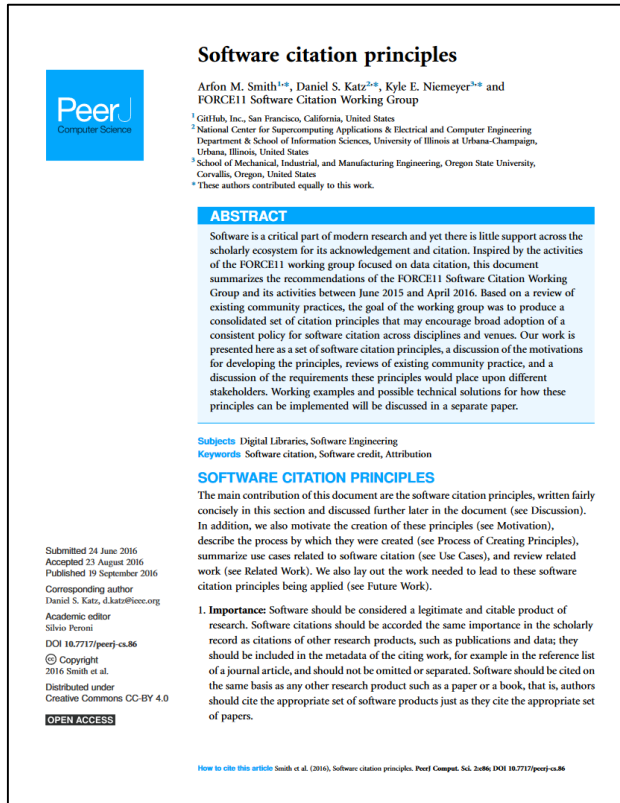- "In the wild": state of the CFF practice

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

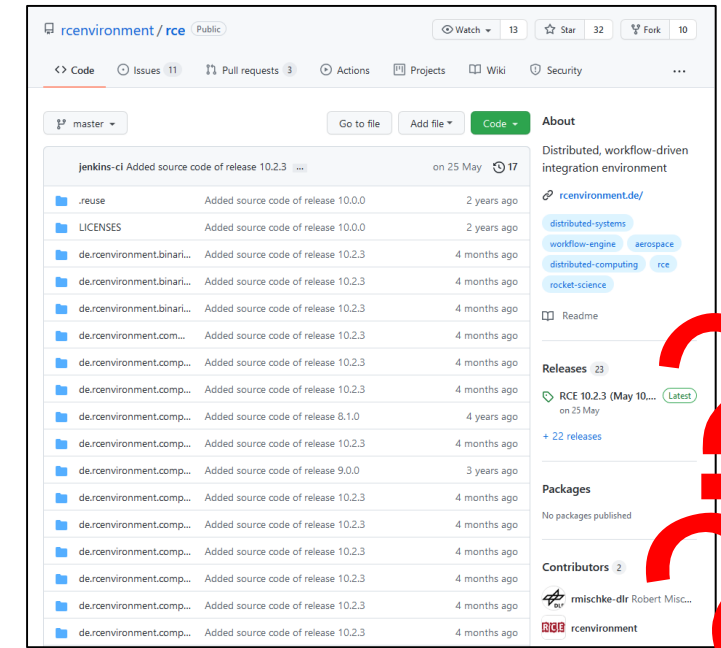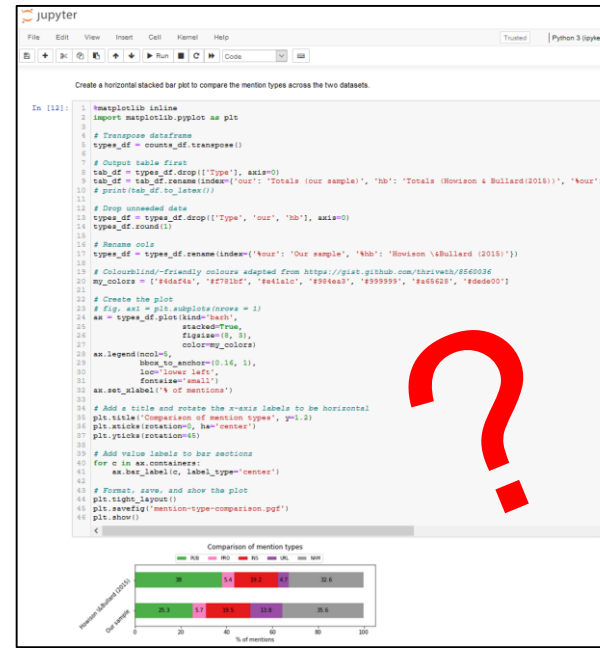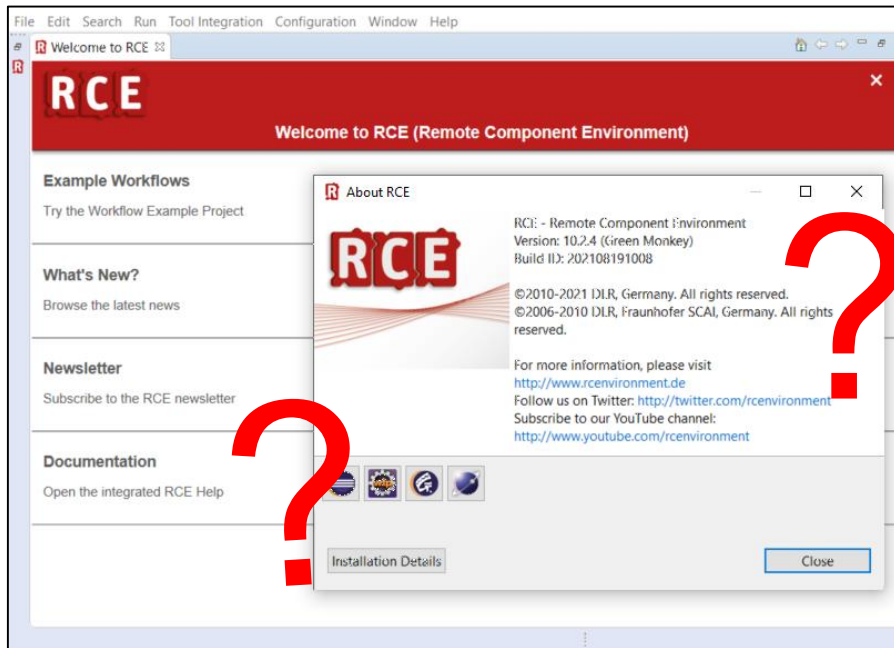4. Persistence

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

DLR

- 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"
```

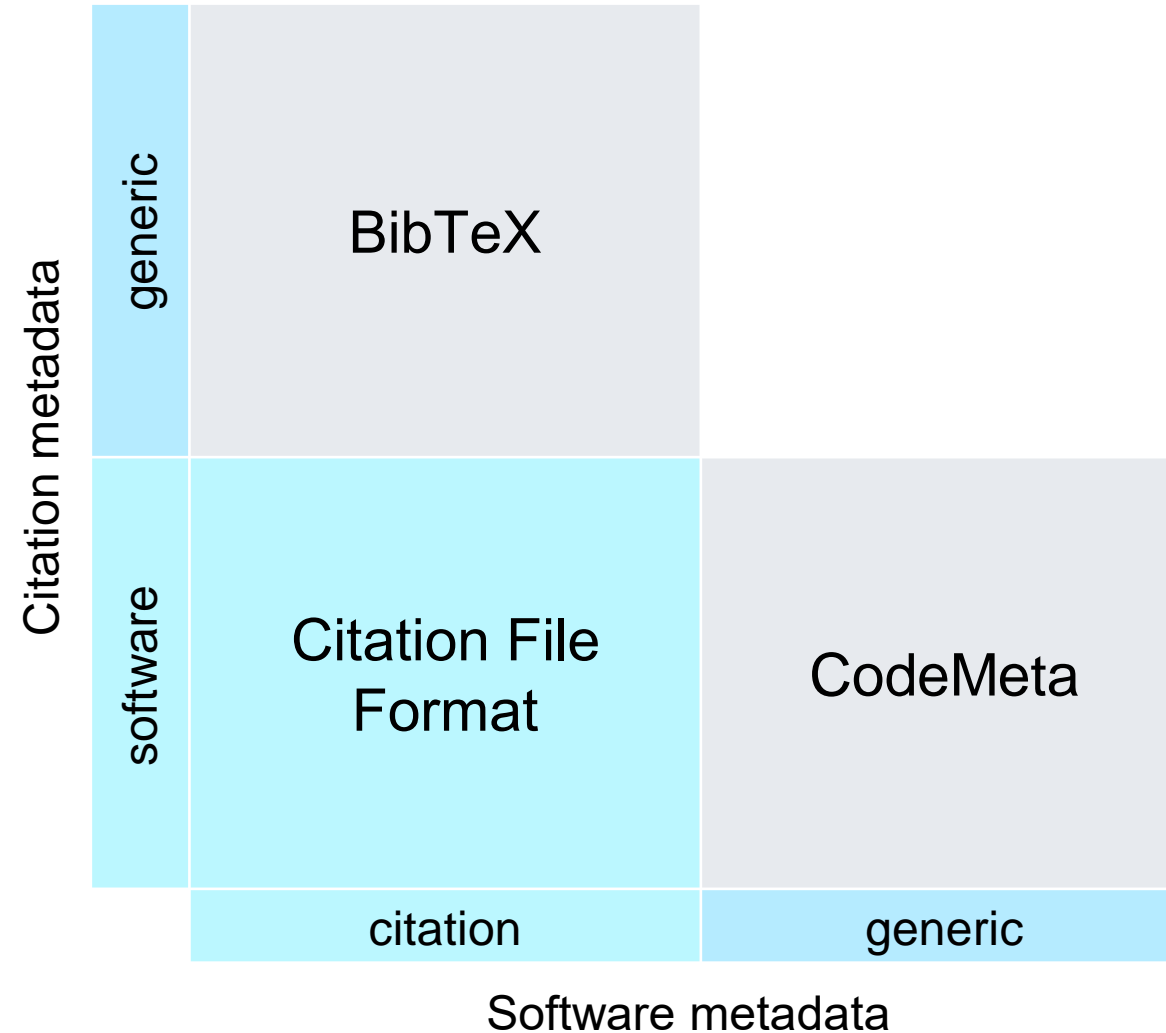+ references, preferred citation, dataset support (experimental)

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

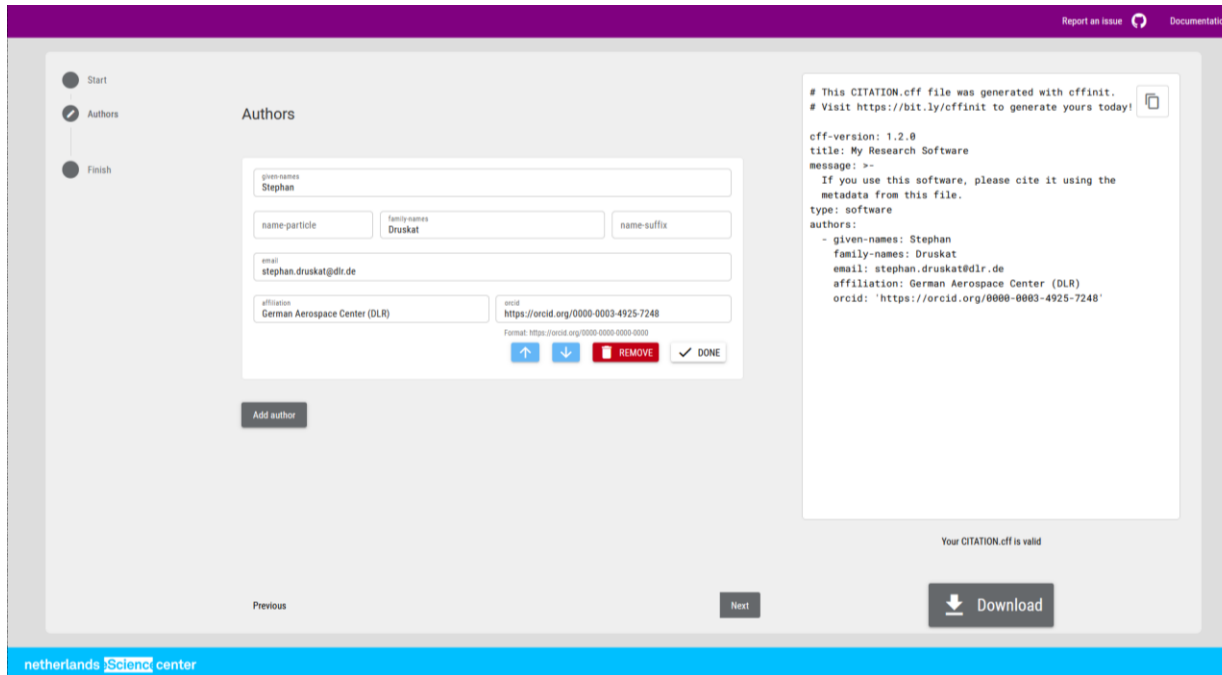# Citation File Format as a project

- 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

Citation File Format in the metadata ecosystem

Citation File Format: more than "just" a format
Create, edit & update, validate, convert

cffinit: web tool for CFF file creation - https://bit.ly/cffinit
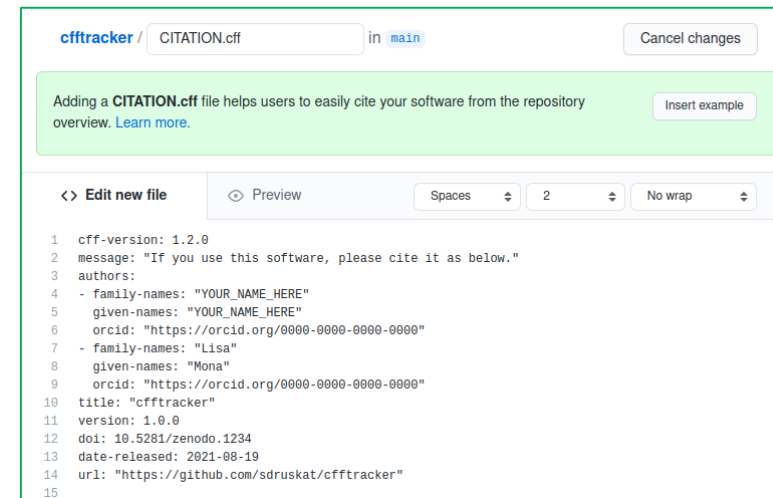
Overview: https://bit.ly/cff-tools-list

CFF INTEGRATIONS

Citation File Format ♡ GitHub
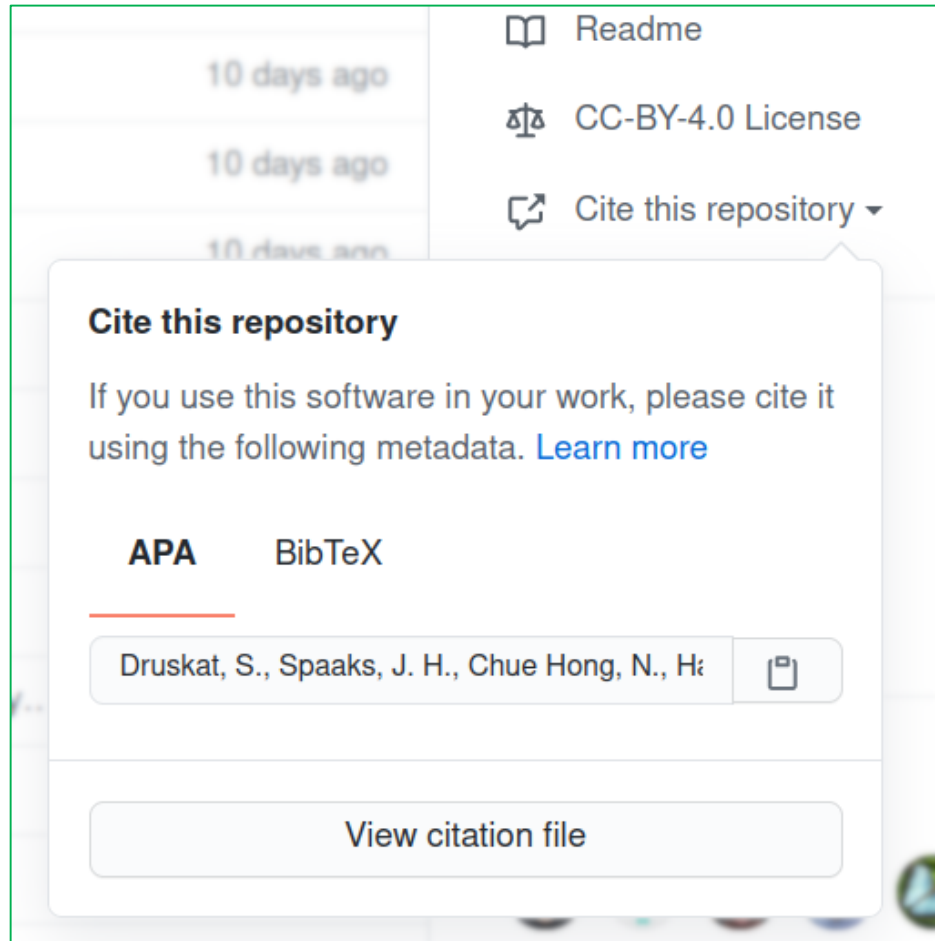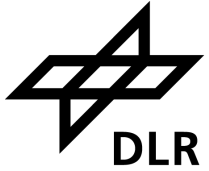
- Detects **CITATION.cff** files

- Renders citation metadata (APA, BibTeX)

- Points to **CITATION.cff** instance

- [Documents](#) and supports creation

S. Druskat, J. Spaaks, "The Citation File Format", RSEConUK 2022, Newcastle-upon-Tyne, UK, 2022-09-06

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

March 16, 2022   Software   Open Access

CampusSource Example Deposit

ⓘD 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-1211c23
   ○ 📄 CITATION.cff   360 Bytes
   ○ 📄 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"
```
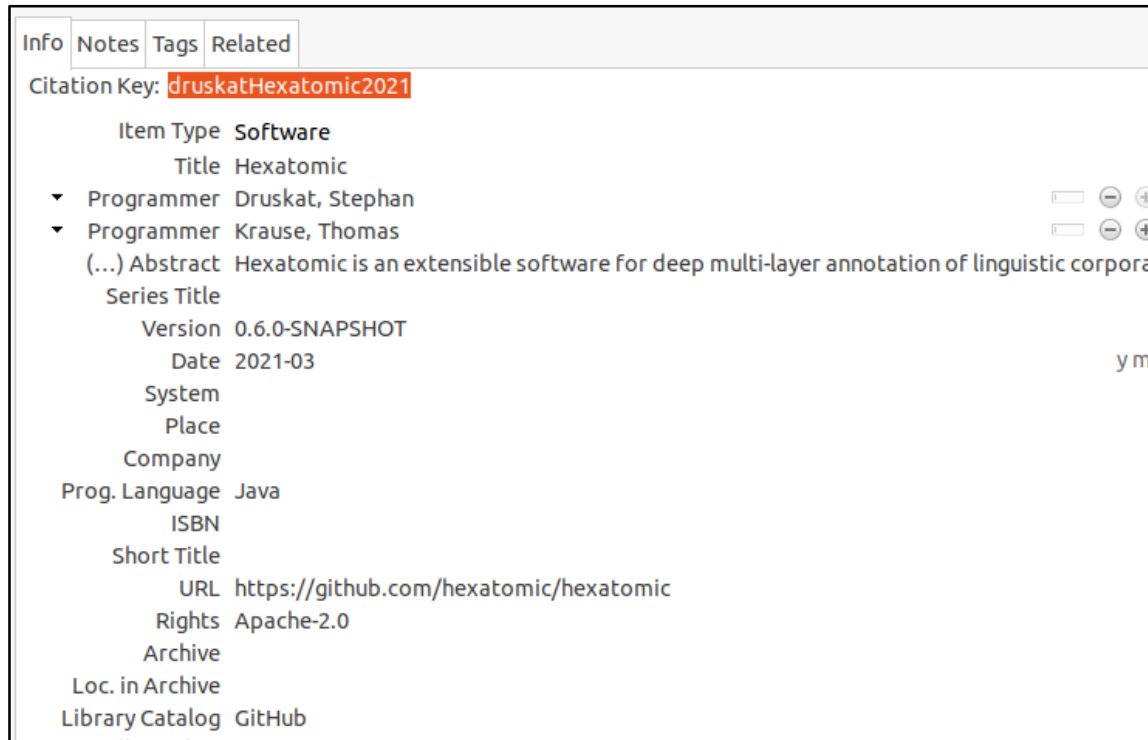
Citation File Format ♡ reference managers

Import in Zotero from GitHub via browser plugin
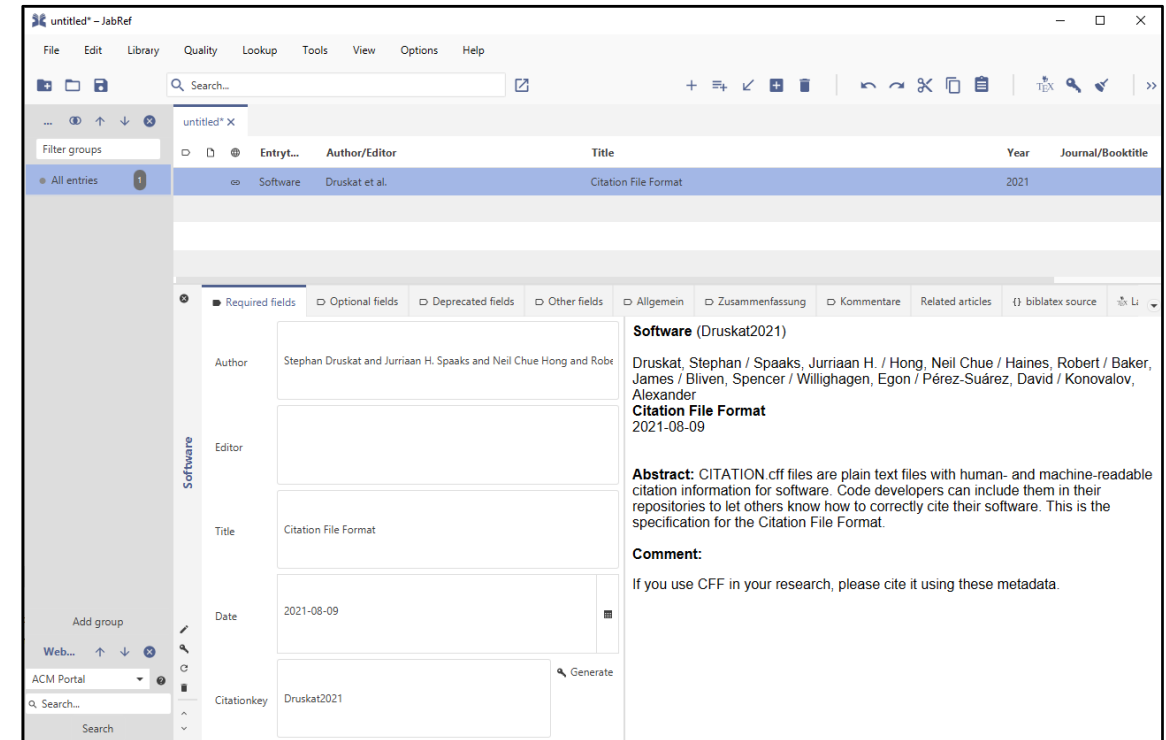
Import in JabRef from `CITATION.cff` file

Citation File Format ♡ IDEs

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

STATE OF THE CFF PRACTICE

How many CFF files are there?

- On GitHub, as of yesterday (if the GitHub Search API is right): **11,199**

(Artifactual drops in counts > 100 due to temporal GitHub Search API unreliability have been removed)
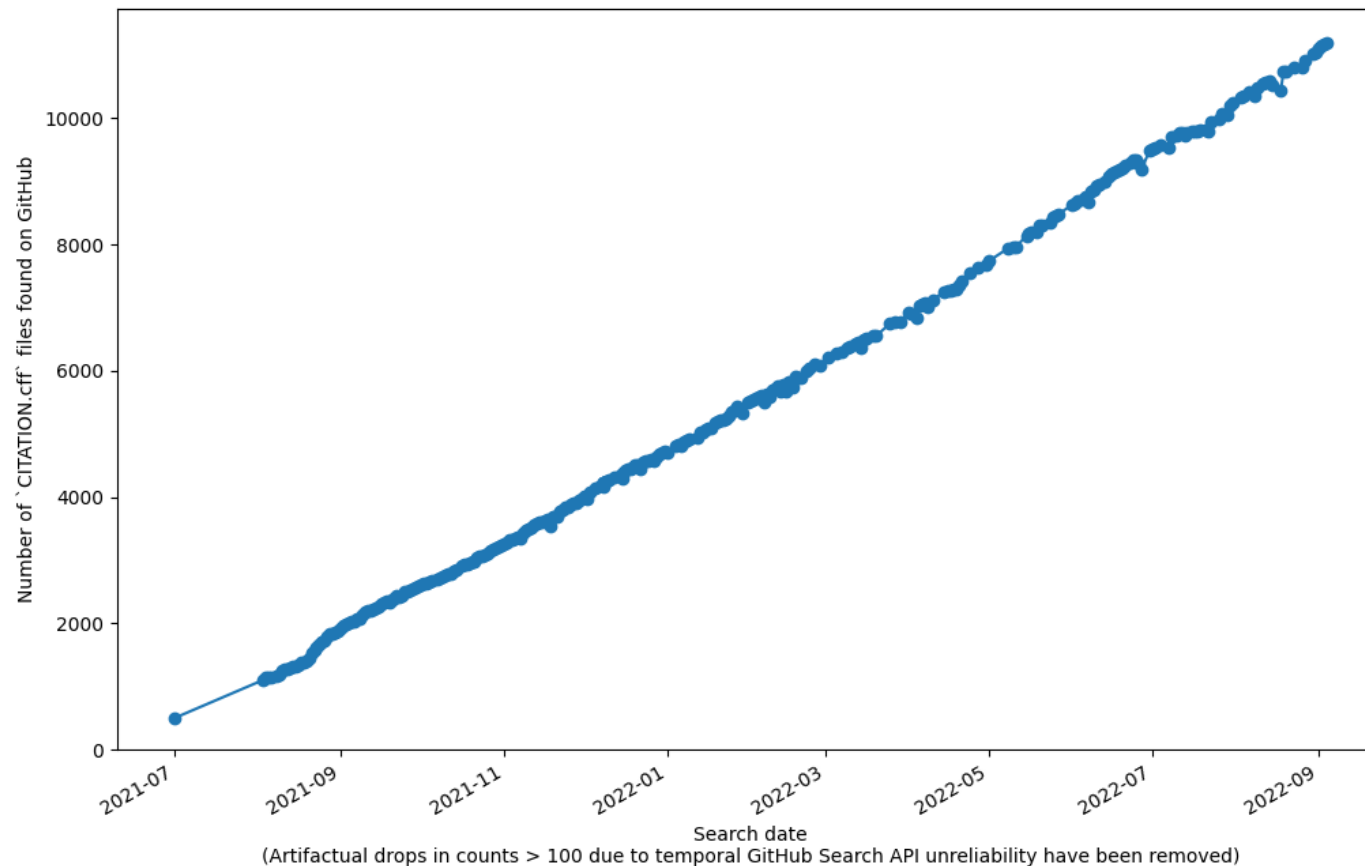
What do people's CFF files look like?

- **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 valid
    - 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

total: 7685

retrievable: 6909

text file: 6908

valid CFF: 4300

valid YAML: 6609

~ 62% of retrievable files

invalid CFF: 2609

irretrievable: 776

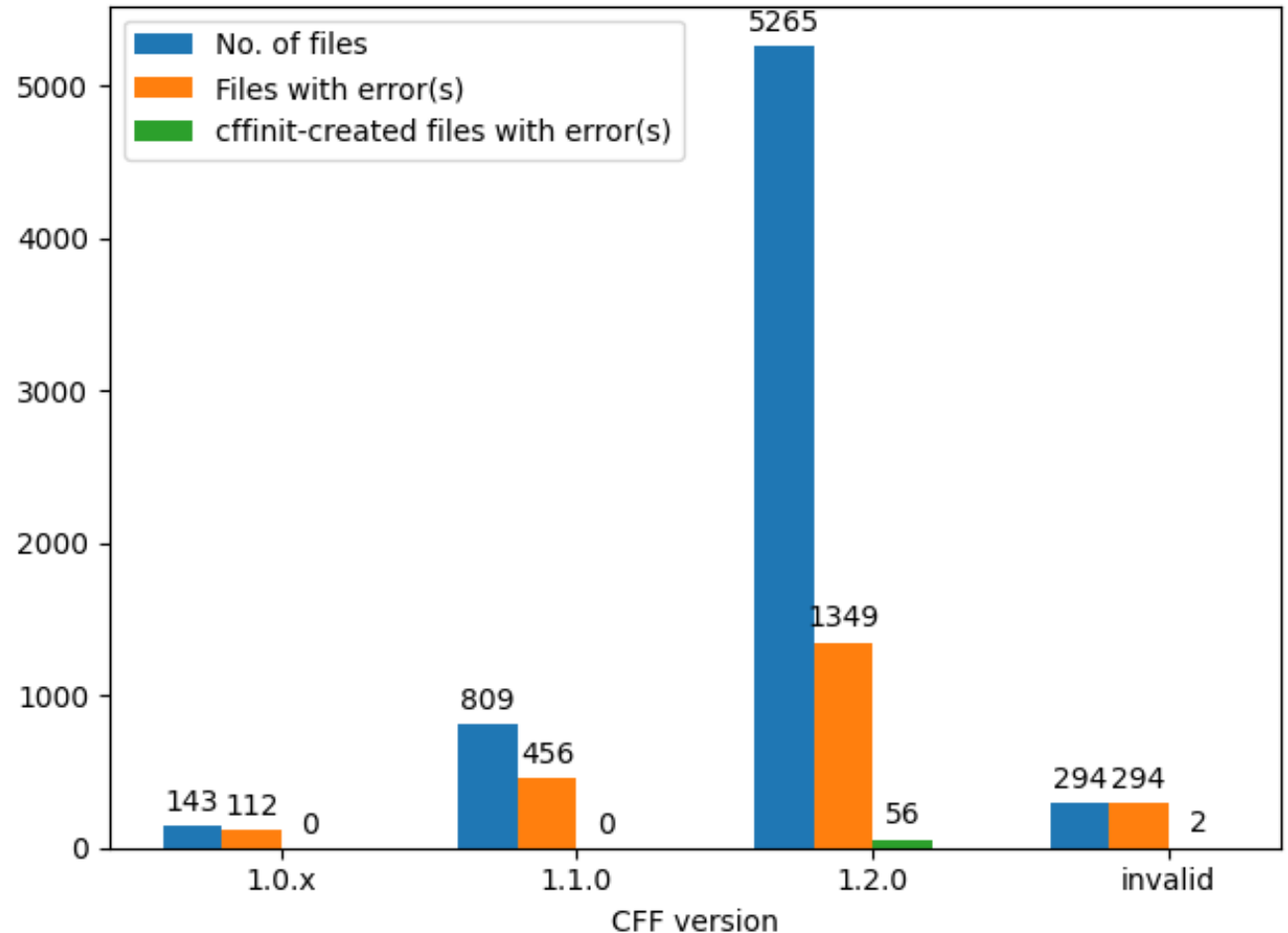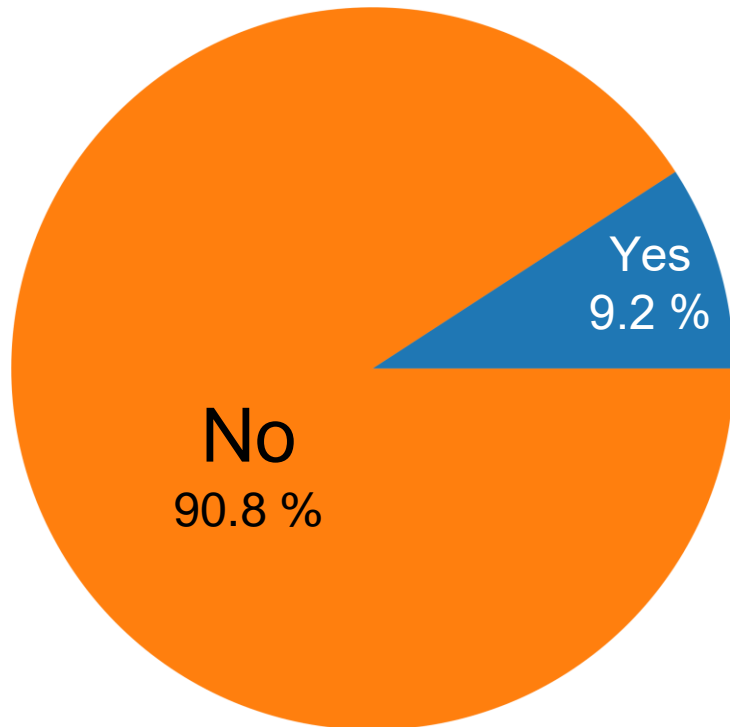binary file: 1

invalid YAML: 300
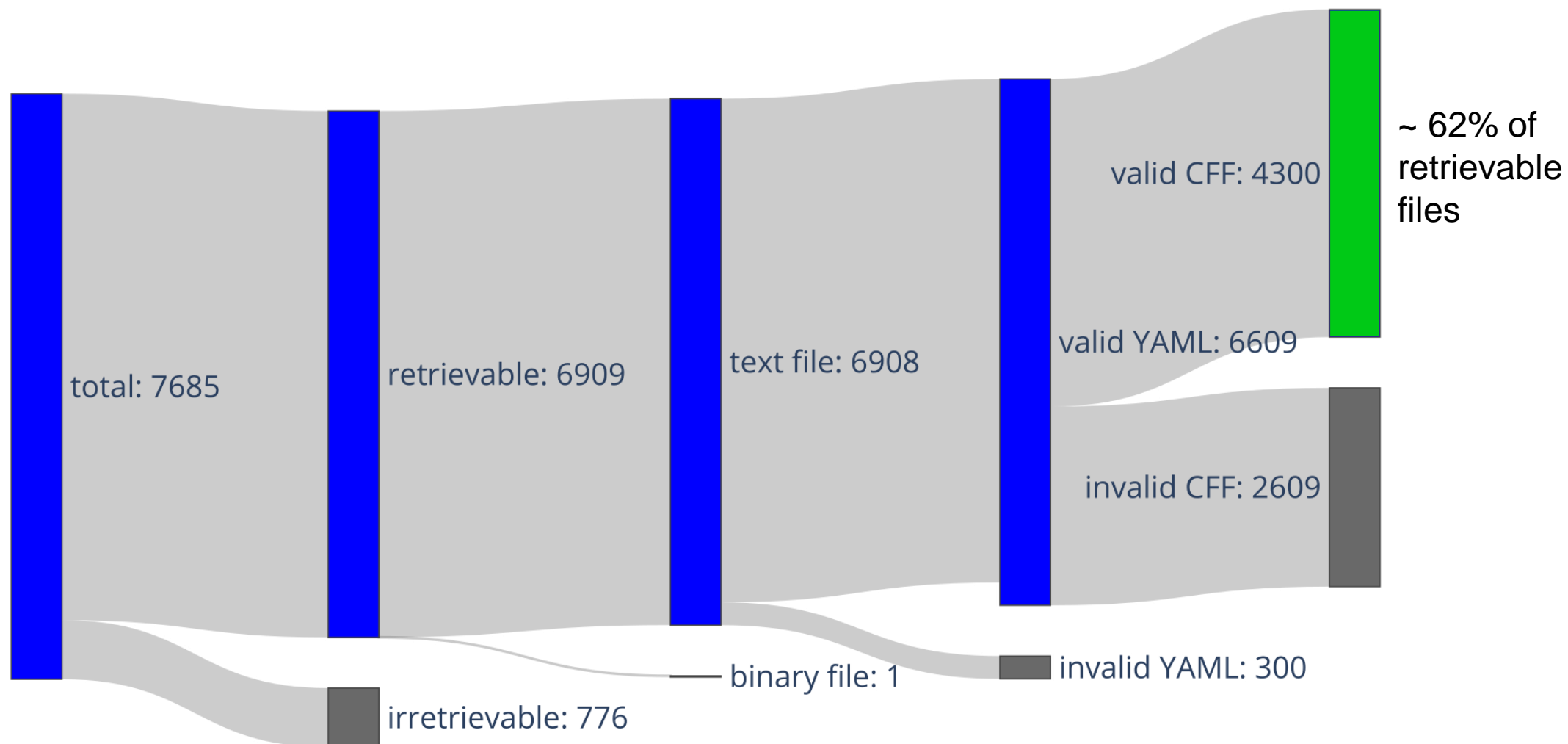
What do people's CFF files look like?
Does the CFF initialization website improve validity?

Created with the CFF initalizer web tool?

What do people's CFF files look like?
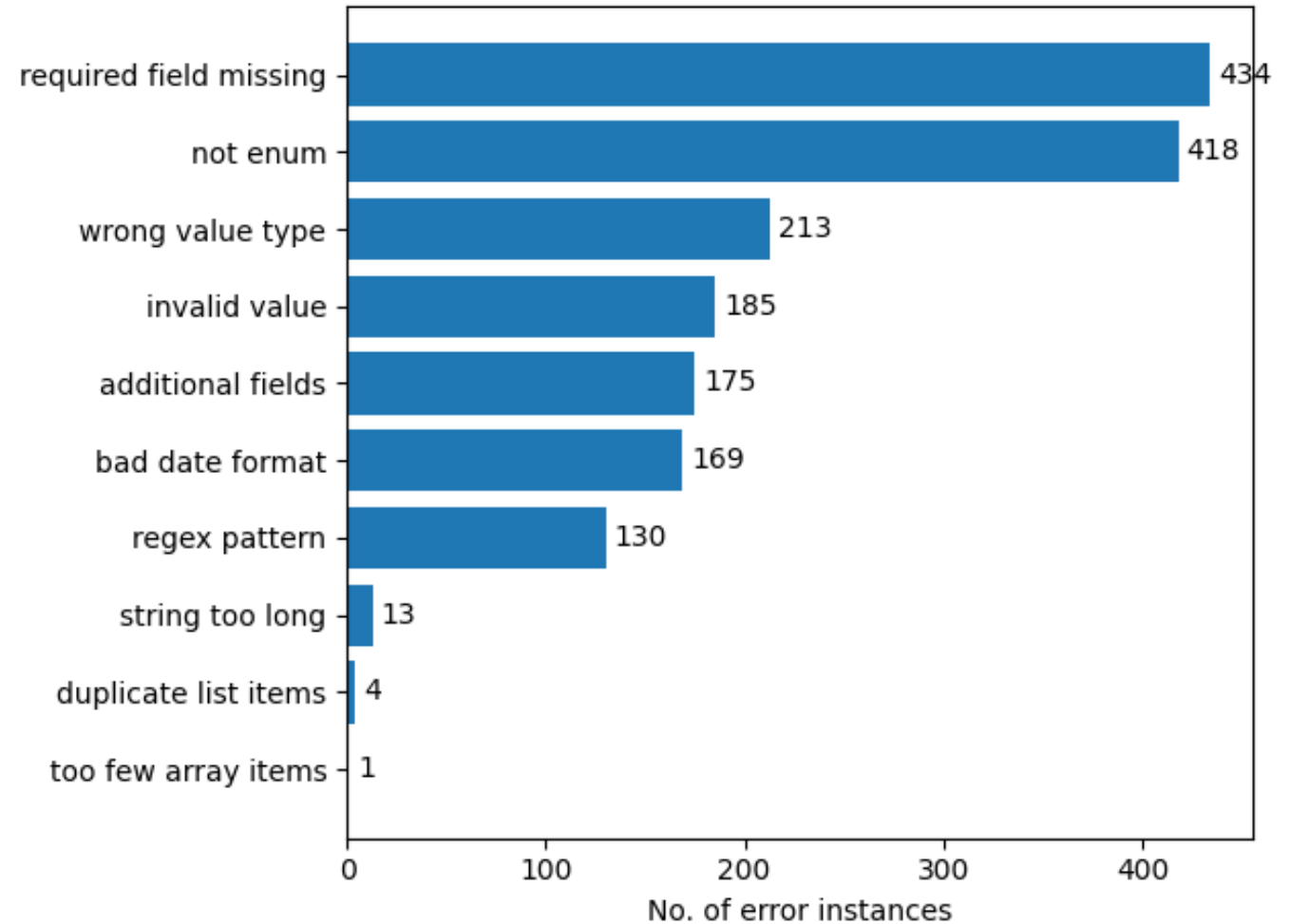# CFF validation errors

- Files with errors: 2,609

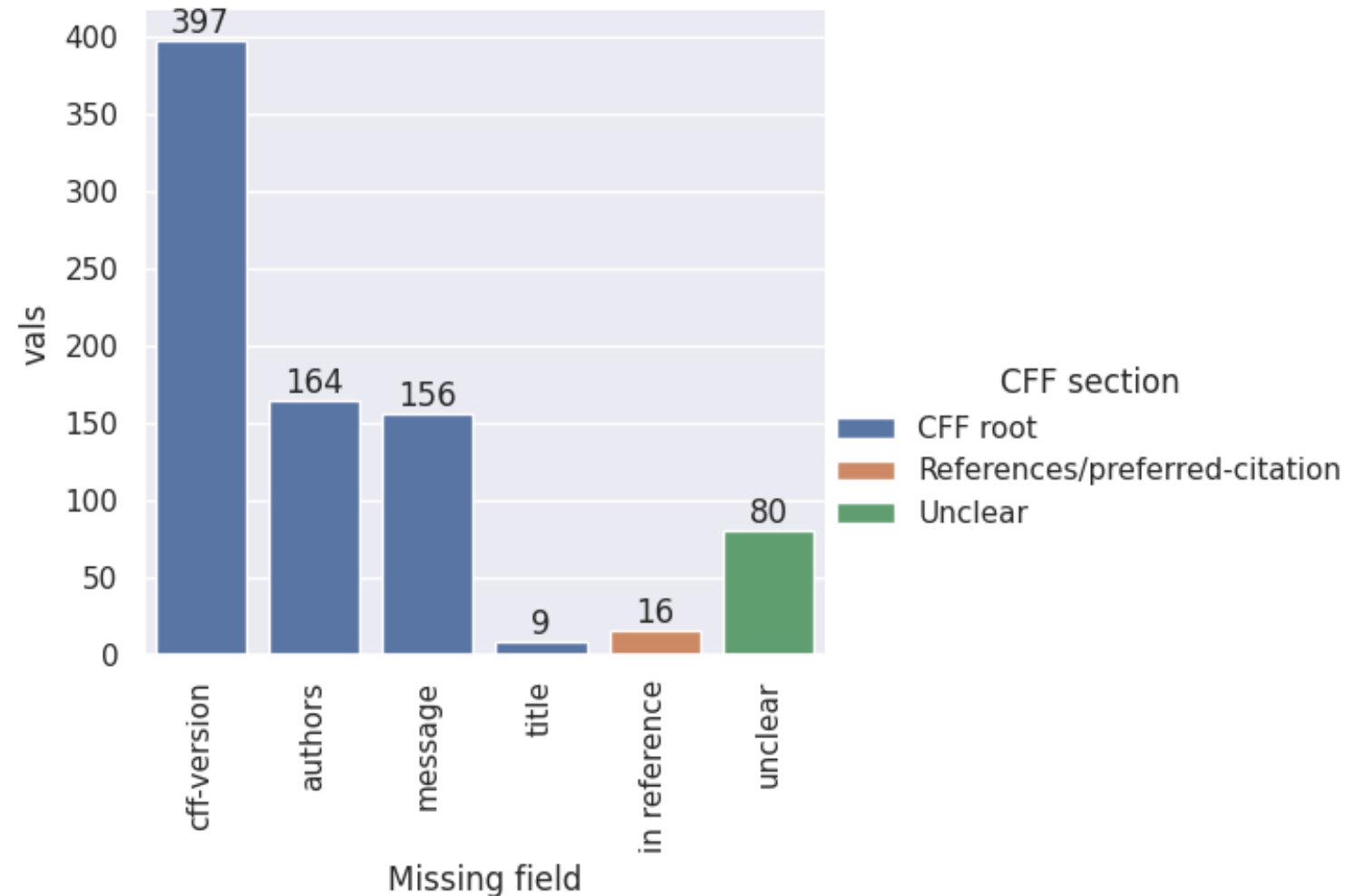- Total number of errors across corpus: 4,284
  - Unmapped errors: 2,242

What do people's CFF files look like?
CFF validation errors: required fields

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

  852

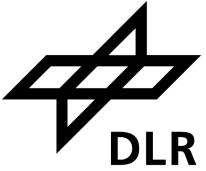S. Druskat, J. Spaaks, "The Citation File Format", RSEConUK 2022, Newcastle-upon-Tyne, UK, 2022-09-06

What do people's CFF files look like? What now?

- Preliminary conlucions:
  - 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 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

Twitter: @stdruskat | ORCID: https://orcid.org/0000-0003-4925-7248