

## Metadata and FAIR Assessment for Research Software

Giacomo Peru, University of Edinburgh (*research software metadata*) Faruk Diblen, Netherlands eScience Centre (*research software FAIR assessment*) Elena Breitmoser, University of Edinburgh (*research software FAIR assessment*)



Funded by the European Union

deRSE25 Metadata for Research Software Giacomo Peru



## Metadata for Research Software

An overview



Funded by the European Union

deRSE25 Metadata for Research Software Giacomo Peru





## Outline of Research Software Metadata Session

What we are covering

An introduction on what research software metadata is and why it matters

Core concepts of research software metadata

Real-world implementations from research software projects

Hands-on exercise: creating .cff files using CFFINIT







## A library metaphor

How do you find books?





## Catalogues

#### A very ancient and current problem

#### Books ISBN (unique identifier) Title and subtitle Author(s) Publisher Publication date Edition number Keywords and categories Physical location in the library

#### Software Package identifier Version number Authors and contributors Release date Dependencies License Repository location Keywords and categories



zenodo



5

## Why Research Software Metadata Matters

#### Software is fundamental to research!

But without good metadata, research software becomes very difficult to:

- Find and discover (Findable)
- Access and install correctly (Accessible)
- Use reproducibly across environments (Interoperable)
- Give proper credit for and understand usage rights (Reusable)



# The FAIR Principles and the community

Findable: Others can discover your software

Accessible: They can get to it and understand how to use it

Interoperable: It works with other tools and systems

Reusable: Others can build upon your work

FAIR Principles for Research Software (FAIR4RS Principles), https://doi.org/10.15497/RDA00068



#### COEOSC EVERSE

## The NumPy case

#### Before:

Inconsistent citations across papers, websites, or none After: 20,000+ verified citations revealing true impact Pattern repeats with pandas, scikit-learn, and other tools <u>Metadata Benefits:</u> Research Integrity: Precise version citation Knowledge Transfer: Version-specific documentation Recognition: Proper attribution for contributors Security: Targeted vulnerability patches Sustainability: Evidence-based funding decisions Innovation: Clear dependencies enable building on existing work

NumPy



deRSE25 Metadata for Research Software | Giacomo Peru



## **Core Concepts**

## **Essential Elements**

## **Common Formats**

Location

keywords contributor contIntegration codeRepository isPartOf copyrightHolder description funder review identifier name provider readme license @context version fileSize issueTracker @type runtimePlatform



The CodeMeta Project





deRSE25 Metadata for Research Software | Giacomo Peru

## Essential Elements of Research Software Metadata

Essential information for discoverability, attribution, and reuse



WHAT: Basic Identification



WHO: People and Affiliations



#### HOW: Usage Information

developmentStatus

softwareRequirements contIntegratio keywords contributor codeRepository isPartOf pyrightHolder description funder review identifier name provider eadme license @context version fileSize issueTracker @type runtimePlatform author relatedLink maintainer programmingLanguage softwareSuggestions releaseNotes

10



### **Basic Identification (WHAT)**

Purpose

Unique identification

Version tracking

Discovery

Classification

title: version: date-released: abstract: keywords:

# People and Organizations (Who)

Purpose

Attribution

Contact information

Institutional credit

Funding acknowledgment

#### authors:

- name:
- orcid:

affiliation

- name
- website

maintainer:

- email:

funder:

- name:
- grant:

ORCID

## Usage Information (How?)

Purpose

Enable reuse

Helps reproducibility

Support implementation

Enable citation

licence:

dependencies:

system requirements:

installation instructions:

documentation:

DOI:





#### (some) Metadata Standards

We will look at

CITATION.cff

codemeta

package specific

domain specific



#### The CodeMeta Project







#### 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

## CFF integrations: GitHub





- Detects CITATION.cff files
- Renders citation metadata (APA, BibTeX)
- Points to **CITATION.cff** instance
- <u>Documents</u> and supports creation

fftracker / CITATIO	N.cff	in main			Cancel char	nges
ding a <b>CITATION.cff</b> file erview. Learn more.	e helps users to easil	ly cite your software fr	om the repository		Insert exa	ample
> Edit new file	⊙ Preview	Spaces	\$ 2	\$	No wrap	\$
cff-version: 1.2.0						
message: "If you us	e this software, p	please cite it as b	elow."			
authors:						
- family-names: "YO	UR_NAME_HERE"					
given-names: "YOUR_NAME_HERE"						
orcid: "https://o	rcid.org/0000-0000	9-0000-0000"				
- family-names: "Li	sa"					
given-names: "Mon	a"					
orcid: "https://orcid.org/0000-0000-0000-"						
title: "cfftracker"						
version: 1.0.0						
doi: 10.5281/zenodo	.1234					
date-released: 2021	-08-19					
url: "https://githu	b.com/sdruskat/cf1	ftracker"				
	fftracker / CITATION Idding a CITATION.cff file erview. Learn more. > Edit new file cff-version: 1.2.0 message: "If you us authors: - family-names: "YOU orcid: "https://o the".cff.racker" version: 1.0.0 doi: 10.5281/zenodo date-released: 2021 url: "https://githu	fftracker / CITATION.cff Iding a CITATION.cff file helps users to easile erview. Learn more. > Edit new file Cff-version: 1.2.0 message: "If you use this software, p authors: - family-names: "YOUR_NAME_HERE" given-names: "YOUR_NAME_HERE" orcid: "https://orcid.org/0000-0000 - family-names: "Nona" orcid: "https://orcid.org/0000-0000 title: "cfftracker" version: 1.0.0 doi: 10.5281/zenodo.1234 date-released: 2021-08-19 url: "https://github.com/sdruskat/cff	fftracker / CITATION.cff in main kding a CITATION.cff file helps users to easily cite your software fr erview. Learn more. > Edit new file Preview Spaces cff-version: 1.2.0 message: "If you use this software, please cite it as b authors: - family-names: "YOUR_NAME_HERE" given-names: "YOUR_NAME_HERE" orcid: "https://orcid.org/0000-0000-0000-0000" - family-names: "Nona" orcid: "https://orcid.org/0000-0000-0000-0000" tils: "cffracker" version: 1.0.0 doi: 10.5281/zenodo.1234 date-released: 2021-08-10 url: "https://github.com/sdruskat/cfftracker"	fftracker / CITATION.cff in main kding a CITATION.cff file helps users to easily cite your software from the repository erview. Learn more. > Edit new file Preview Spaces \$ 2 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" igiven-names: "YOUR_NAME_HERE" given-names: "YOUR_NAME_HERE" 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"	fftracker / CITATION.cff in main kding a CITATION.cff ie helps users to easily cite your software from the repository erview. Learn more.  Edit new file Preview Previ	fftracker / CITATION.cff in main Cancel char kding a CITATION.cff file helps users to easily cite your software from the repository Insert exa review. Learn more. Insert exa Section 2 Concerns and the software from the repository Insert exa Edit new file Preview Spaces \$ 2 \$ No wrap cff-version: 1.2.0 message: "If you use this software, please cite it as below." authors: - family-names: "YOUR_NAME_HERE" orcid: "https://orcid.org/0000-0000-0000-0000" - family-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"



Software 🔓 Open

Hide affiliations

Sandbox). 1. The first release from the GitHub API. 2. The

- GitHub-Zenodo bridge reads metadata from CITATION.cff
- Previously: via API (<u>GitHub release</u>, <u>Zenodo Sandbox deposit</u>)
- Now: author-supplied metadata (GitHub release, Zenodo Sandbox deposit)

_				
	1	+ cff-version: 1.2.0	Published March 18, 2024   Version with-CITATION.cff	Software
	5	+ title: pseudoSoftware		
	5	+ message: >-		
	7	+ If you use this software, please cite it using the	nseudoSoftware	
	3	+ metadata from this file.		
	9	+ type: software	-Druskat, Stephan <sup>1</sup> 💿	Hide af
1	3	+ authors:		
1	1	+ - given-names: Stephan	1. German Aerospace Center (DLR)	
1	2	+ family-names: Druskat		
1	3	+ email: stephan.druskat@dlr.de		
1	1	+ affiliation: German Aerospace Center (DLR)	This "software" is a demonstration-only repository with two releases. For both releases, the Zenodo GitHub integr	ration
1	5	+ orcid: 'https://orcid.org/0000-0003-4925-7248'	contains no CITATION.cff file. Hence, the Zenodo Sandbox record is populated with metadata retrieved from the	GitHub API. 2
1	5	+ repository-code: 'https://github.com/citable-s0ftware/cff-github-zenodo'	second release contains a CITATION off file. Hence, the Zenodo Sandbox record is populated with metadata retri	eved from the
1	7	+ url: 'https://citation-file-format.github.io'	CITATION.cff file.	
1	3	+ abstract: >-		
1	Э	+ This "software" is a demonstration-only repository with	Notes	
2	3	+ two releases.		
2	1	+	If you use this software, please cite it using the metadata from this file.	
2	2	. For both polosos, the Zenode GitHub integration		

## CFF integrations: reference managers



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

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

	💃 untitled" – JabRef – 🗆 🗙						
Info Notes Tags Related	File Edit Library Quality Lookup Tools View Options Help						
Citation Key: druskatHexatomic2021	Q Search		へ ※ 「自   赤 % ぐ   »				
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	● ↑ ↓ ③ untitled*× ilter groups	Title Citation File Format	Year Journal/Booktitle 2021				
Series Title	<ul> <li>Required fields</li> <li>Optional fields</li> <li>Deprecated</li> </ul>	fields 🗅 Other fields 🗆 Allgemein 🗅 Zusammenfassung 🗅 Komr	mentare Related articles {} biblatex source 🞄 Li 🤿				
Version 0.6.0-SNAPSHOT		Software (Druskat2021)					
Date 2021-03 ym	Author Stephan Druskat and Jurriaan H. Spaaks and	Neil Chue Hong and Robe Druskat, Stephan / Spaaks, Jurriaan	H. / Hong, Neil Chue / Haines, Robert / Baker,				
System		Alexander	n, Egon / Pelez-Suarez, David / Konovalov,				
Place	9	2021-08-09					
Company	Editor	Abstract: CITATION off files are plai	n text files with human- and machine-readable				
Prog. Language Java	ÿ	citation information for software. Cod	e developers can include them in their				
ISBN	Title Citation File Format	specification for the Citation File Form	nat.				
Short Title		Comment:					
URL https://github.com/hexatomic/hexatomic		If you use CFF in your research, plea	se cite it using these metadata.				
Rights Apache-2.0	Add group Date 2021-08-09	•					
Archive	Web ↑ ↓ 🔕 🔦						
Loc. in Archive	M Portal	<ul> <li>Generate</li> </ul>					
Library Catalog GitHub	Search · ·						

## CodeMeta is a vocabulary

#### Codemeta (v2.0) in a nutshell:

Property	Туре	Description
softwareSuggestions	SoftwareSourceCode	Optional dependencies , e.g. for optional features, code development, etc.
maintainer	Person	Individual responsible for maintaining the software (usually includes an email contact address)
contIntegration	URL	link to continuous integration service
buildInstructions	URL	link to installation instructions/documentation
developmentStatus	Text	Description of development status, e.g. Active, inactive, suspended. See repostatus.org
embargoDate	Date	Software may be embargoed from public access until a specified date (e.g. pending publication, 1 year from publication)
funding	Text	Funding source (e.g. specific grant)
issueTracker	URL	link to software bug reporting or issue tracking system
referencePublication	ScholarlyArticle	An academic publication related to the software.
readme	URL	link to software Readme file

#### Schema.org extension (findable by search engines)

## CodeMeta is a machine-readable vocabulary

## CodeMeta access URL: <u>https://doi.org/10.5063/schema/codemeta-2.0</u>

"@context": {
 "type": "@type",
 "id": "@id",
 "schema":"http://schema.org/",
 "codemeta": "https://codemeta.github.io/terms/",
 "Organization": {"@id": "schema:Organization"},

"softwareSuggestions": { "@id": "codemeta:softwareSuggestions", "@type": "@id"},
"contIntegration": { "@id": "codemeta:contIntegration", "@type": "@id"},
"buildInstructions": { "@id": "codemeta:buildInstructions", "@type": "@id"},
"developmentStatus": { "@id": "codemeta:developmentStatus", "@type": "@id"},
"embargoDate": { "@id": "codemeta:embargoDate", "@type": "@id"},
"funding": { "@id": "codemeta:funding" },
"readme": { "@id": "codemeta:issueTracker", "@type": "@id" },
"issueTracker": { "@id": "codemeta:issueTracker", "@type": "@id" },
"referencePublication": { "@id": "codemeta:issueTracker", "@type": "@id" },
"maintainer": { "@id": "codemeta:maintainer" }

## JSON-LD vocabulary

## CodeMeta is a crosswalk

Crosswalk for WikiData Properties

- Crosswalk for DataCite metadata
- Crosswalk for Debian packages
- Crosswalk for DOAP Ontology
- Crosswalk for GitHub API
- Crosswalk for Java's Maven metadata
- Crosswalk for NodeJS package.json
- Crosswalk for Python distutils
- **R** Crosswalk for R Packages



Crosswalk for Ruby gems

Mapping between vocabularies

### Crosswalks have been curated by the community

#### E.g., for ASCL (<u>http://ascl.net/</u>):

https://github.com/codemeta/codemeta/blob/master/crosswalks/

<u>ASCL.csv</u>

44	description	abstract
45	identifier	ascl_id
46	name	title
47	sameAs	
48	url	
49	relatedLink	
50	givenName	
51	familyName	
52	email	email

### Propagating your metadata across regstries





CodeMeta.jsonId

Once your CodeMeta file is on your code repository, it can be propagated **upon release** between different platforms

#### Language-specific Standards

Ecosystem-Specific Standards

Python: setup.cfg, pyproject.toml

**R: DESCRIPTION** 

Node.js: package.json

Each solves: Making software installable and usable







#### **Domain-Specific Standards**

Field-Specific Metadata Needs

#### Examples

- BioSchemas (Life Sciences)
- Climate Science Standards
- Digital Humanities Requirements

#### Purpose

- Method documentation
- Domain relationships
- Specialized features







#### Metadata Locations - Overview

Where Does Metadata Live?

Three Main Locations:

- Repository Root Files
- Package Configuration Files
- External Registries





#### COEOSC EVERSE

#### Repository and Package Files

#### Repository Root & Package Files

#### Repository Root Files:

- README.md First point of contact
- CITATION.cff Citation information
- LICENSE Usage terms
- CONTRIBUTORS.md Attribution list

#### Package Configuration Files:

- Python: setup.cfg, pyproject.toml
- R: DESCRIPTION
- Node.js: package.json

🗋 README.md	
	CITATION.cff
	CONTRIBUTORS.md
🗋 setup.py	pyproject.toml
DESCRIPTION	
	🗋 package.json

#### COEOSC EVERSE

## **External Archives**

Archive and preserve

Software Heritage - Code preservation

Zenodo - DOI assignment

**PyPI/CRAN/npm** - Package distribution

Institutional Repos - Academic archiving









Different Approaches to CITATION.cff

scikit-image and napari

Two scientific Python image analysis tools

- scikit-image
- napari

#### CITATION.cff with different approaches

- <u>scikit-image</u>: Team-focused, paper citation
- <u>napari</u>: Individual contributor focus, software DOI

Installation	Gallery	Documentation	Comr	munity	O Source	S	earch documenta	ition
Stable (release r 0.25.2rc0 - Febr 2025 Download Development pre-0.26	<mark>lotes</mark> ) uary	Image p scikit-image processing. restriction. reviewed co volunteers.	rocess is a colle It is avail We pride ode, writt	ction of al able free ourselves en by an a	<b>Python</b> gorithms for in of charge and f s on high-quali active commun	nage <mark>ree of</mark> ty, peer- ity of		
Ownload								
		If you find t	his proje	ct useful,	please cite:			[BiBT
Contribute a Get Help advice & Discuss cont	get involved community tributor talk	Stéfan van o Boulogne, J image contr https://doi.o	der Walt, oshua D. ibutors. org/10.77	Johannes Warner, N scikit-ima 717/peerj.	L. Schönberge leil Yager, Emn ge: Image proc 453	r, Juan Nunez-I nanuelle Gouilla s <b>essing in Pyth</b> o	glesias, Franço art, Tony Yu an on. PeerJ 2:e45	is d the sciki 53 (2014)
StackOverflow	coding help							
na	ıpari		Usage	Plugins	Community	Contributing	API – Reference	na hu
<b>Q</b> sear	ch	x K	🏫 > Cont	ributing				
<ul> <li>Contrib</li> <li>Core D</li> <li>napari</li> <li>napari</li> <li>(NAPs)</li> <li>Roadm</li> </ul>	uting guide eveloper guide architecture guide Advancement Proj aps	e posals	We we contrib napari partici	ntribu loome your bute to napa has a <u>Code</u> pates in the <u>untact</u> : how	ting contributions! H ari. <u>of Conduct</u> that napari comm to contact us.	ere you will find r should be honore unity.	esources to help d by everyone wl	you ho
			• <u>Ro</u> po	admaps: co ssible work	onvey the genera plans. w <b>ides</b> : resources	l direction for na	pari developmen	t and

Contributing guide

<u>Testing</u>

29

Development installation

scikit-imade



## scikit-image

Simple maintenance

Academic	paper
citation	

Collective recognition

type: software authors:
– name: The scikit–image team
preferred-citation:
type: article
title: "scikit-image: image processing in Python"
authors:
– family–names: van der Walt
given–names: Stéfan J.
[more authors]

doi: 10.7717/peerj.453

cff-version: 1.2.0
title: scikit-image

### napari

Software DOI citation

**Detailed authors** 

cff-version: 1.2.0 title: 'napari: a multi-dimensional image viewer for Python' identifiers: - type: doi value: 10.5281/zenodo.3555620 authors: - given-names: Nicholas family-names: Sofroniew affiliation: Chan Zuckerberg Initiative orcid: https://orcid.org/0000-0002-3426-0914 alias: sofroniewn

31



## Hands-on: Creating your CITATION.cff



Go to:

https://citation-file-format.github.io/cff-initializer-javascript/

Enter basic info:

Title, author(s), version, date, repository URL

Add optional details:

Abstract, keywords, license

Copy generated code  $\rightarrow$  save as 'CITATION.cff' in repository root

Commit and push to see GitHub citation button appear



deRSE25 Metadata for Research Software | Giacomo Peru



# Thank you

# Now onto FAIR assessment for research software with Elena Breitmoser and Faruk Diblen



Funded by the European Union

deRSE25 Metadata for Research Software Giacomo Peru

