

# SOFTWARE DISCOVERY, PUBLICATION AND CITATION

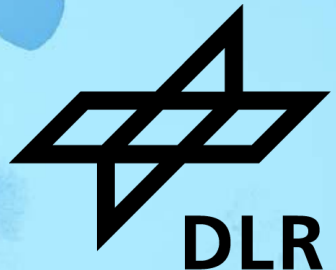
**Stephan Druskat**

German Aerospace Center (DLR), Institute for Software Technology, Sustainable Software Engineering

*Research Software Skills @ RSE Network (HUB, KCL), online, 2025-06-19*

License [CC-BY-4.0 International](#)

DOI [10.5281/zenodo.15696706](https://doi.org/10.5281/zenodo.15696706)



# About me

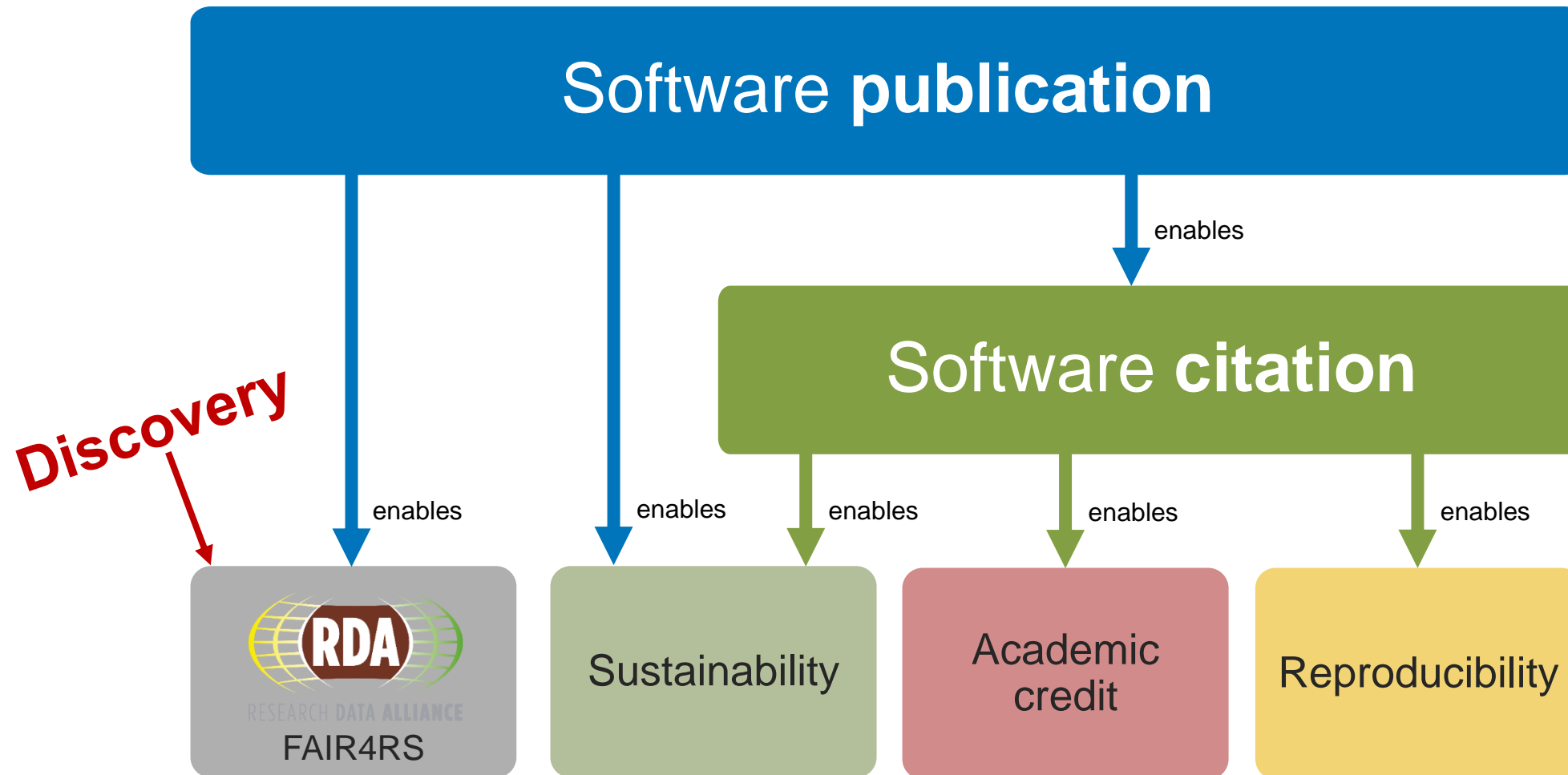


- M.A. *English Philology, Modern German Literature, Linguistics* (Free University Berlin)
- ~13 years *RSE* in several linguistic research projects
- Fellow of the [Software Sustainability Institute](#)
- Lead [Citation File Format](#) project
- Since 2019: Software engineering researcher at [DLR](#)
- PI [HERMES](#) project (*automating software publication with rich metadata*)
- Chair ReSA Task Force "[Software Authorship & Contribution](#)"

- Why are software publication and software citation important (for discovery)?
- What is software publication?
- What is software citation?
- How to make software citable?
- How to publish software?

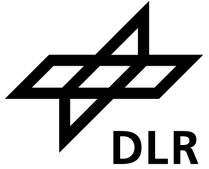
Q: Why are software publication and citation important?

A: *Because software is important for research!*



Adapted from Druskat et al. (2023)

# FAIR for Research Software (FAIR4RS)



- **F:** Software, and its associated metadata, is **easy for both humans and machines to find**.
- **A:** Software, and its metadata, **is retrievable via standardized protocols**.
- **I:** Software **interoperates with other software** by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.
- **R:** Software is both usable (**can be executed**) and reusable (**can be understood**, **modified, built upon, or incorporated** into other software).

Chue Hong et al. (2022)

## **Publication of software with metadata** **(Research) Software Engineering**

# FAIR for Research Software (FAIR4RS) supports discovery



- **F:** Software, and its associated metadata, is **easy for both humans and machines to find**.
- **A:** Software, and its metadata, **is retrievable via standardized protocols**.
- **I:** Software **interoperates with other software** by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.
- **R:** Software is both usable (**can be executed**) and reusable (**can be understood, modified, built upon, or incorporated** into other software).

Chue Hong et al. (2022)

**Publication of software with metadata**  
(Research) Software Engineering → **Discovery of research software**



# What is software publication?

Publication of metadata and artifacts\* for software **versions** in publication repositories

Persistent identifier for each version

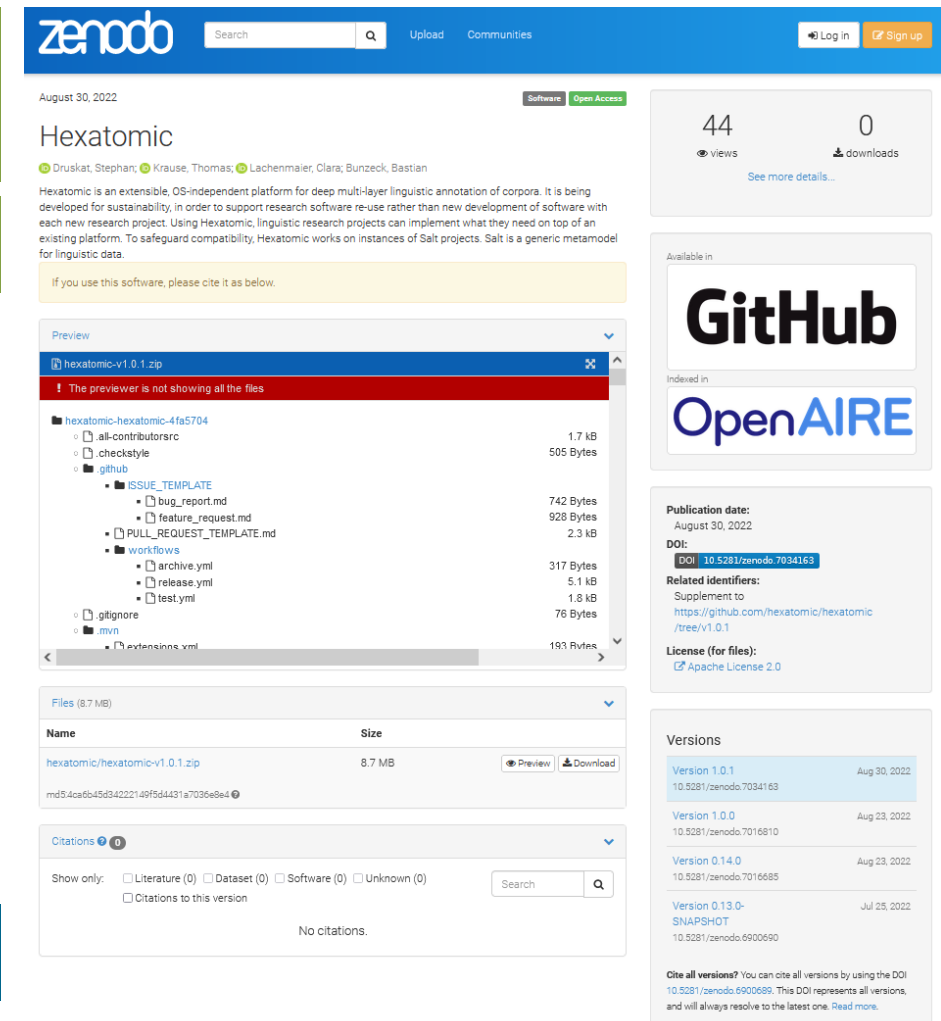
**Not** software publication:

- Software available on a source code platform (*GitHub, GitLab, etc.*)
- A paper *about* the software

Interim solution:

- software journals ([JOSS](#), [JORS](#), etc.)

Challenge: quality assurance



The screenshot shows the Zenodo page for the 'Hexatomic' software repository. The page includes a search bar, navigation links for 'Upload' and 'Communities', and a 'Log in' / 'Sign up' button. The main content area displays the software's title, authors (Druskat, Stephan; Krause, Thomas; Lachenmaier, Clara; Bunzeck, Bastian), and a description of the software as an extensible, OS-independent platform for deep multi-layer linguistic annotation. It also shows the software's version history, with the latest version being 1.0.1, published on August 30, 2022. The page includes a 'Preview' section showing the software's file structure, a 'Files' section with a table of files, and a 'Citations' section. The 'Files' table lists the following files and their sizes:

Name	Size
hexatomic/hexatomic-v1.0.1.zip	8.7 MB

The 'Citations' section shows a search bar and a list of citations, with the first citation being 'No citations'.

# What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic* version 1.4.2, Mar. 2023. Zenodo. DOI: [10.5281/zenodo.7778709](https://doi.org/10.5281/zenodo.7778709).

Software must be cited like a paper!

Smith et al. (2016)



# What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic* version 1.4.2, Mar. 2023. Zenodo. DOI: [10.5281/zenodo.7778709](https://doi.org/10.5281/zenodo.7778709).

A software **version** must be cited like a paper!

A software **project** is cited like a paper!

# What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic* ~~version 1.4.2~~, Mar. 2023. Zenodo. DOI: [10.5281/zenodo.6900689](https://doi.org/10.5281/zenodo.6900689).

A software **version** must be cited like a paper!

A software **project** must be cited like a paper!

# What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck,  
*Hexatomic* version 1.4.2, Mar. 2023. Software Heritage.  
**SWHID:** [swh:1:rel:2bc06dfdd14699b00c5acecc37213734988a15d6](https://swh.riken.jp/swh/id/swh:1:rel:2bc06dfdd14699b00c5acecc37213734988a15d6).

A software **version** must be cited like a paper, even if it isn't published!

A software **project** must be cited like a paper, even if it isn't published!

# What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck,  
*Hexatomic* version 1.4.2, Mar. 2023. GitHub. URL:  
<https://github.com/hexatomic/hexatomic/releases/tag/v1.4.2>.

A software **version** must be cited like a paper, even if it isn't published!

A software **project** must be cited like a paper, even if it isn't published!

# What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic* version a0c59ed, Mar. 2023. GitHub. URL: <https://github.com/hexatomic/hexatomic/commit/a0c59edf8be8c1d7c045745e4bd043f7369c7d1b>.

A software **version** must be cited like a paper, even if it isn't published!

A software **project** must be cited like a paper, even if it isn't published!

# What is software citation?



The Hexatomic authors, *Hexatomic*~~—version a0c59ed~~, n.d., GitHub. URL: <https://github.com/hexatomic/hexatomic/commit/a0c59edf8be8c1d7c045745e4bd043f7369c7d1b>.

A software **version** must be cited like a paper, even if it isn't published!

A software **project** must be cited like a paper, even if it isn't published!

# What is software citation?



S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic* version 1.4.2, Mar. 2023. Zenodo. DOI: [10.5281/zenodo.7778709](https://doi.org/10.5281/zenodo.7778709).

Academic credit



# What is software citation?



Discovery

S. Druskat, T. Krause, C. Lachenmaier, and B. Bunzeck, *Hexatomic* version 1.4.2, Mar. 2023. Zenodo. DOI: [10.5281/zenodo.7778709](https://doi.org/10.5281/zenodo.7778709).

Reproducibility

+

Findability, Accessibility, Reusability

# What is software citation?

## It is often still a best effort adventure!



???

What is the **most reliable** metadata I can find that **best** supports:

Academic credit ?

+ ?

Reproducibility ?

+ ?

Findability, Accessibility, Reusability? ?

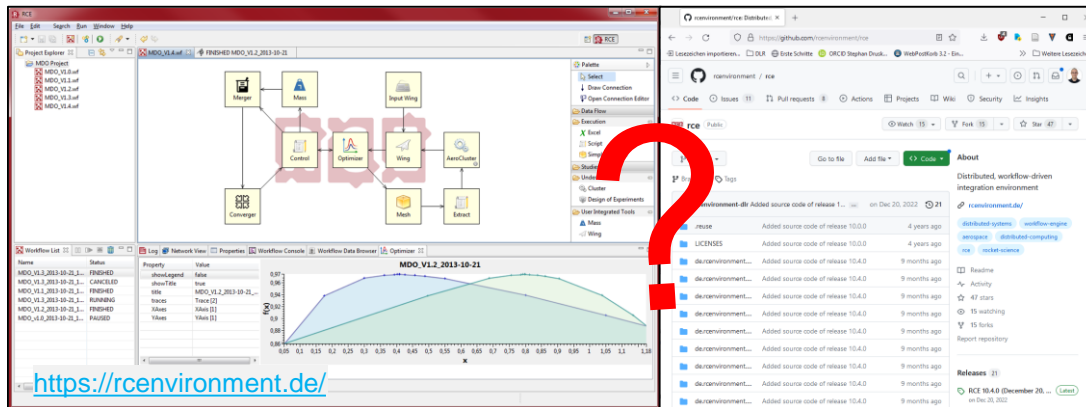
# How to make software citable?

## Metadata

[citation-file-format.github.io](https://citation-file-format.github.io)

Authors? Name? Version? Publication date?

Provide metadata in `CITATION.cff` in the repository!



[Cite this repository](#)

```
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
    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"
```

**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](#)

Druskat et al. (2021)

- YAML: citation-relevant metadata, including references (e.g., to primary dependencies)
- Easy creation/update/validation (web tool [cffinit](#), IDE support), compatibility, tools
- Open community format (~ 25,000 files on GitHub; integrations: Zenodo, Zotero, JabRef, ...)

# How to publish software?

## Publication repositories

DOIs for versions & projects, metadata, persistence

Simple automation with CFF, GitHub & Zenodo!

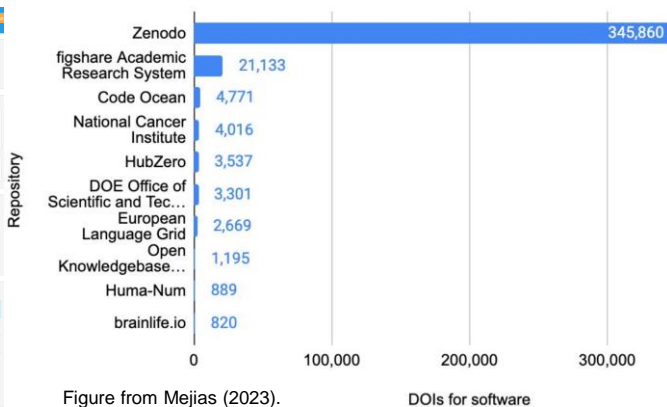
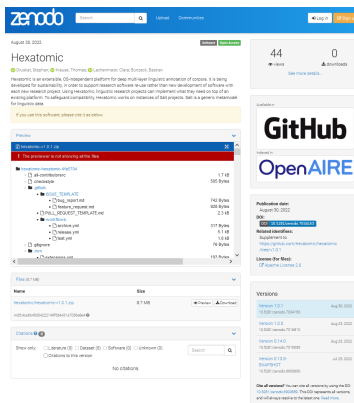
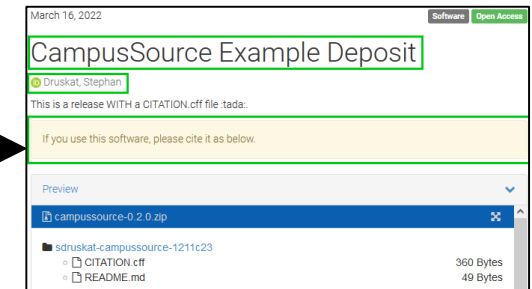
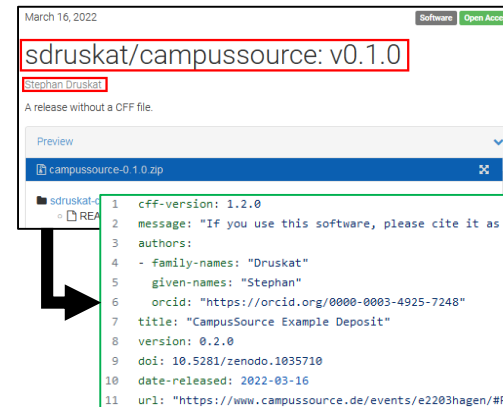


Figure from Mejias (2023).

DOIs for software



- Most popular: [Zenodo](#) (developed & run at CERN), GitHub integration
- Persisted records, persistent identifiers ([DataCite](#) DOIs w/ metadata), landing pages
- Policies (institution/funder) may mandate specific repositories

# How to publish software (advanced)?

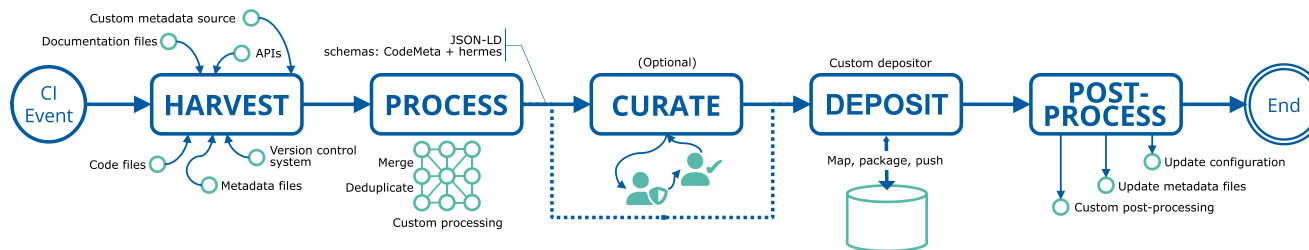


## Continuous publication with HERMES

[hermes.software-metadata.pub](https://hermes.software-metadata.pub)

Run publication workflow in CI

Extensible to support different infrastructure



Kernchen, Meinel, Druskat, et al. (2025)



- Full metadata control; includes metadata curation & updating repository metadata
- Different metadata sources (including CFF) & publication targets (including Zenodo)
- Developed as FLOSS at Helmholtz – documentation, training materials

# Conclusion



- Software citation and FAIR publication improve discoverability
- Software citation is part of good research practice (e.g., DFG, 2025)
  - Software metadata provided by authors
  - Cite software (versions) like papers
- Software publication is part of good research practice, and ...
  - ... highly recommended ([Helmholtz Open Research Software Indicator](#)),
  - ... possible (Zenodo),
  - ... automatable (HERMES),
  - but lacks an established step for external quality assurance!
- You can start today!

# Thanks!



Projects:           **Citation File Format:** <https://citation-file-format.github.io/>  
                      **HERMES:** <https://hermes.software-metadata.pub>

Me:                   [stephan.druskat@dlr.de](mailto:stephan.druskat@dlr.de) | [@sdruskat@fediscience.org](https://sdruskat@fediscience.org)

References:        S. Kernchen, M. Meinel, S. Druskat, et al., "Extending and applying automated HERMES software publication workflows," Electronic Communications of the EASST, vol. 83, 2025, doi: [10.14279/ECEASST.V83.2624](https://doi.org/10.14279/ECEASST.V83.2624).  
                      S. Druskat, O. Bertuch, and A. Struck, "Towards Research Software-ready Libraries," ABI Technik, vol. 43, no. 3, pp. 168–178, 2023, doi: [10.1515/abitech-2023-0031](https://doi.org/10.1515/abitech-2023-0031).  
                      G. Mejias, "DOIs for research software: Increasing Visibility, Connectivity, Citability," 2023. doi: [10.5281/zenodo.7985199](https://doi.org/10.5281/zenodo.7985199).  
                      N. P. Chue Hong, et al., "FAIR Principles for Research Software (FAIR4RS Principles)," 2022, doi: [10.15497/RDA00068](https://doi.org/10.15497/RDA00068).  
                      S. Druskat, J. H. Spaaks, et al., "Citation File Format (Version 1.2.0)," 2021, doi: [10.5281/zenodo.5171937](https://doi.org/10.5281/zenodo.5171937).  
                      A. M. Smith, et al., "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).  
                      Deutsche Forschungsgemeinschaft, "Guidelines for Safeguarding Good Research Practice. Code of Conduct," 2025, doi: [10.5281/zenodo.14281892](https://doi.org/10.5281/zenodo.14281892).