



Software Citation

IRIS-HEP Blueprint Workshop

November 23, 2022
Adriaan Klinkenberg



Several experimental workflows in Elsevier

1. Author submits article to any journal, with code linked or attached
 2. Author submits to dedicated software journal (e.g. CPC, SoftwareX)
 3. For select titles: Author is asked to submit to **Code Ocean** to support peer-review and reproducibility (approx 30% uptake when asked)
- 2a: CPC supports code in **Mendeley Data - CPC Program Library**
- 2b: SoftwareX supports code on **GitHub** repository and **Software Heritage**

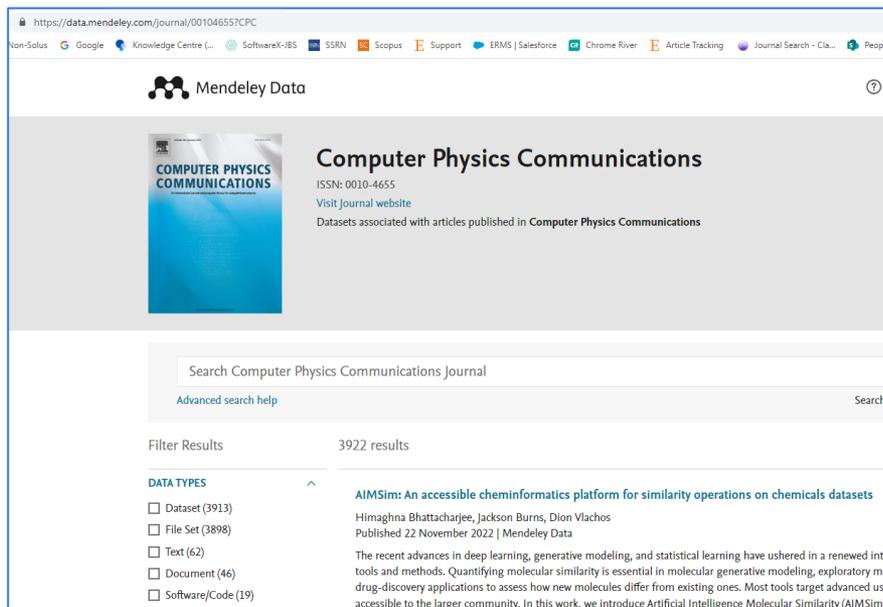
In all events, authors may have posted (versions of) their programs already on Zenodo, GitHub, figshare, arXiv etc.

All of these can be cited! And we are not consistent.



Computer Physics Communications – Code Library

- Editorial office ensures code is stored on Mendeley Data in CPC Code library



The screenshot shows the Mendeley Data page for the journal 'Computer Physics Communications'. The page includes the journal's cover image, ISSN (0010-4655), and a search bar. Below the search bar, there are filter results for 'DATA TYPES' and a list of datasets. The first dataset listed is 'AIMSim: An accessible cheminformatics platform for similarity operations on chemicals datasets' by Himaghna Bhattacharjee, Jackson Burns, and Dion Vlachos, published on 22 November 2022. The description of the dataset mentions the use of deep learning, generative modeling, and statistical learning for molecular similarity quantification.

https://data.mendeley.com/journal/00104655CPC

Mendeley Data

Computer Physics Communications
ISSN: 0010-4655
[Visit Journal website](#)
Datasets associated with articles published in **Computer Physics Communications**

Search Computer Physics Communications Journal
Advanced search help Search

Filter Results 3922 results

DATA TYPES

- Dataset (3913)
- File Set (3898)
- Text (62)
- Document (46)
- Software/Code (19)

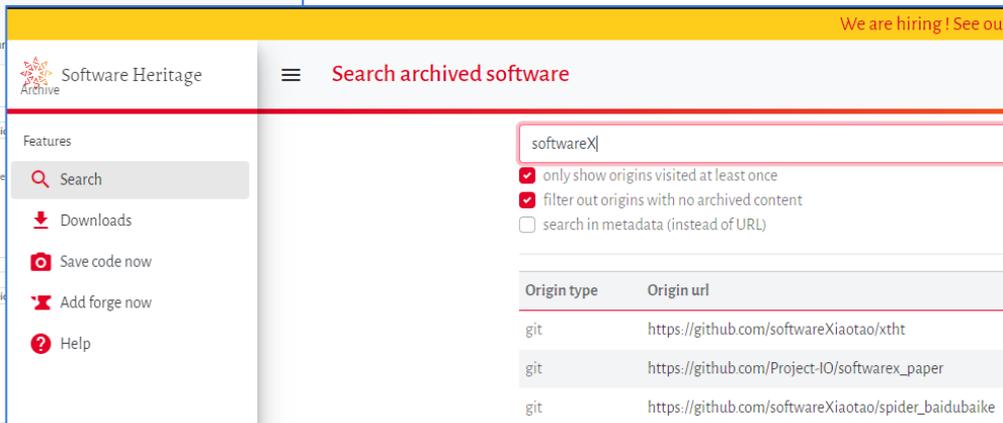
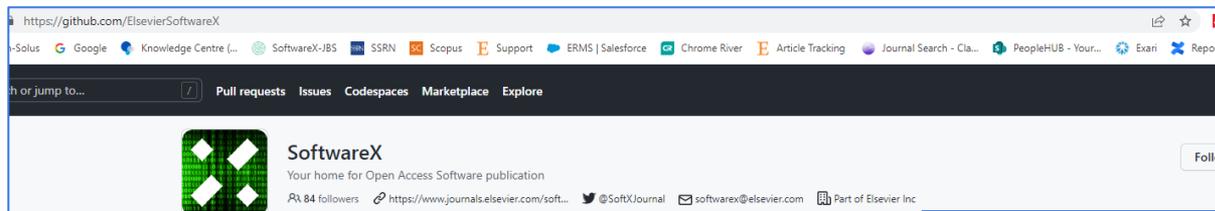
AIMSim: An accessible cheminformatics platform for similarity operations on chemicals datasets
Himaghna Bhattacharjee, Jackson Burns, Dion Vlachos
Published 22 November 2022 | Mendeley Data

The recent advances in deep learning, generative modeling, and statistical learning have ushered in a renewed interest in tools and methods. Quantifying molecular similarity is essential in molecular generative modeling, exploratory molecular discovery applications to assess how new molecules differ from existing ones. Most tools target advanced users and are not accessible to the larger community. In this work, we introduce Artificial Intelligence Molecular Similarity (AIMSim)



SoftwareX – re-publishing static “VoR” – useful?

- To facilitate citations and linking to the version described in the paper
- People should cite the version used. Versioning should be done consistently.



Code Ocean capsule available for 20-30% of SoftwareX or CPC papers

Creating additional citation option

The screenshot shows the Open Science Library interface. At the top, there's a search bar and navigation tabs for 'All' and 'SoftwareX'. Below that, a 'View capsule' button is visible. The main content area features a grid of capsules. One capsule is highlighted, titled 'CaTchDes: Matlab codes for Near-Optimal Regression Designs' by Mathematics, dated 22 Oct 2019. It includes a thumbnail image of a star-shaped grid with red dots. A 'Open capsule published in SoftwareX' link is at the bottom.

The screenshot shows the Code Ocean capsule interface for 'Essence Neural Networks' by Paul J. Blazek & Milo M. Lin. The interface is divided into several sections:

- Core Files:** A list of files including metadata, environment, and code files like adversarial_attacks.py, bdt_trees.py, etc.
- Metadata:** A section with a description: 'viable computational framework to understand the neural basis of cognition and pursue artificial general intelligence.' It includes tags like 'Deep Learning', 'explainable-ai', 'cognitive-psychology', 'Artificial Neural Network', 'computational-neuroscience', 'symbolic-computations', 'Artificial Intelligence', 'cognition', and 'Machine Learning'.
- Capsule:** A section with a DOI (10.24433/CO.7389497.v1) and a citation: 'Paul J. Blazek, Milo M. Lin (2021) Essence Neural Networks [Source Code]. https://doi.org/10.24433/CO.7389497.v1'. It also shows license information (Code Custom License, Data No Rights Reserved (CC0)).
- Associated Publication:** A section with a DOI (10.1038/s43588-021-00132-w), title 'Explainable neural networks that simulate reasoning', publication date 'September 2021', journal 'Nature Computational Science', and funding information.
- Reproducible Run:** A section on the right showing a timeline of runs, including 'Published Version 1.1' and 'Published Version 1.0'.



23.11.2022

Code / Software summary

in article header of *Original Software Publications (OSP)*

For less than 0.1% of our articles



23.11.2022

ELSEVIER

Program summary

Program Title: tauola-bbb

CPC Library link to program files: <https://doi.org/10.17632/3hyw3f7v32.1>

Licensing provisions: GPLv2

Programming language: FORTRAN/C++

Journal reference of previous version: *Comput. Phys. Comm.*, 232:220–236, 2018

Does the new version supersede the previous version?: No

Reasons for the new version: Meeting the needs of present day experiments.

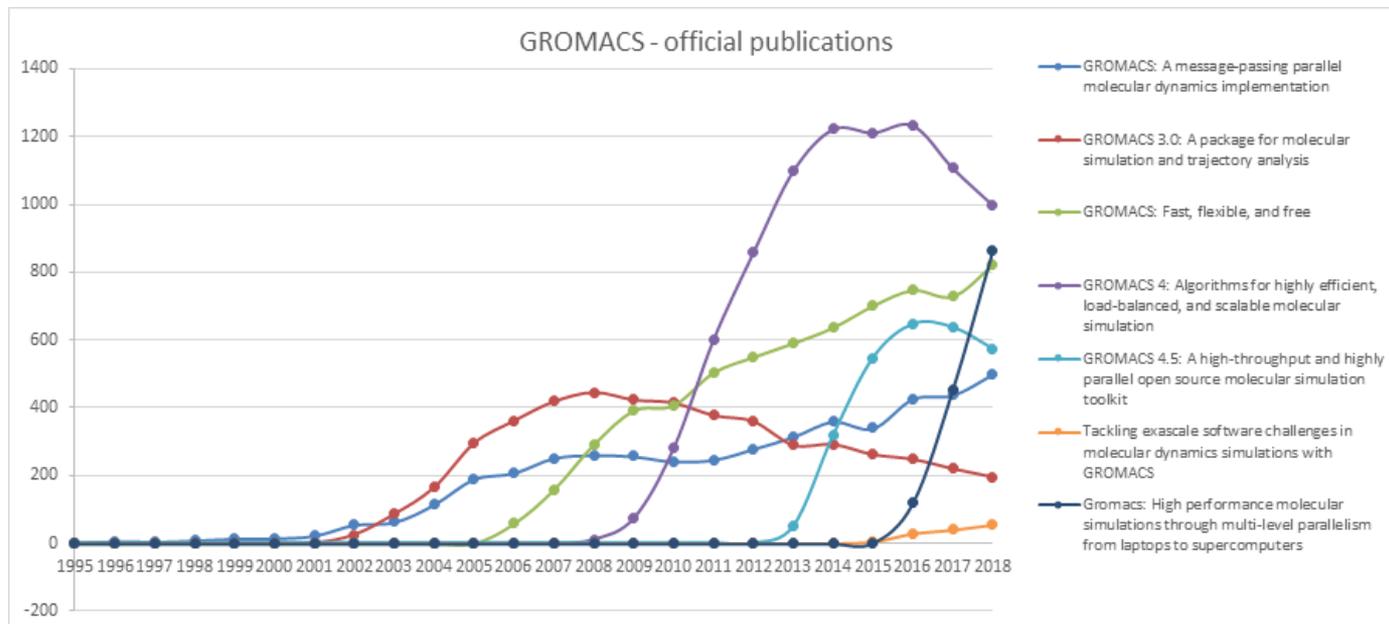
Summary of revisions: Phase space presamplers enriched with the possibility to generate narrow peaks of e^+e^- pairs, if present in matrix elements of 4- and 5-body decay channels and due to electromagnetic or New Physics interactions. The examples of physical models are implemented for: $\tau^- \rightarrow \pi^- e^- e^+ \nu_\tau$, $\tau^- \rightarrow \bar{\nu}_\mu \mu^- e^- e^+ \nu_\tau$ and $\tau^- \rightarrow \bar{\nu}_e e^- e^- e^+ \nu_\tau$ channels, which are used for the collected new algorithm validations.

Nature of problem: Present day experiments, in particular Belle II, have the capability to measure extremely rare, not yet measured τ decays, as well as precise measurements of known decay channels. They require a Monte Carlo generator capable of generating all desired decays. They also need the possibility to modify and test new models for such decays. At the same time the default initialization of the event generator should contain decent modeling of known τ decays.

Solution method: The new version supplements previous publications on TAUOLA. Phase-space presamplers and new decay channels with additional light lepton pairs are introduced.

Additional comments including restrictions and unusual features: TAR BALL of the future

Versioning – authors must cite what they use



One version does not replace the other
An accumulated statistic also does not make sense

https://research-software-directory.org/software/citation-file-format

Cite this software

DOI: 10.5281/zenodo.1003150 COPY TO CLIPBOARD

Choose a version: <https://github.com/citation-file-format/citation-file-format/releases/tag/v0.9-RC1> ^

Choose a reference manager format: DOWNLOAD FILE

Wh

1. U
2. If
3. W
4. Pe

- <https://github.com/citation-file-format/citation-file-format/releases/tag/1.2.0>
- <https://github.com/citation-file-format/citation-file-format/releases/tag/1.1.0>
- <https://github.com/citation-file-format/citation-file-format.github.io/releases/tag/1.0.3-4>
- <https://github.com/citation-file-format/citation-file-format.github.io/releases/tag/1.0.3-3>
- <https://github.com/citation-file-format/citation-file-format.github.io/releases/tag/1.0.3-1>
- <https://github.com/citation-file-format/citation-file-format.github.io/releases/tag/1.0.3-1>
- <https://github.com/citation-file-format/citation-file-format.github.io/releases/tag/1.0.2>
- <https://github.com/citation-file-format/citation-file-format.github.io/releases/tag/1.0.3>
- <https://github.com/citation-file-format/citation-file-format.github.io/releases/tag/1.0.0>
- <https://github.com/citation-file-format/citation-file-format.github.io/releases/tag/1.0.1>
- <https://github.com/citation-file-format/citation-file-format/releases/tag/v0.9-RC1>

which makes it easy for visitors to cite your software or dataset

the metadata from your `CITATION.cff` file.

a [browser plugin](#).

Keywords
No keywords available

Programming language
Python 100%

License
CC-BY-4.0

Source code

DOI is changing when selecting different version; also supporting 5 reference manager formats



Guide for Authors for CPC

Reference to software

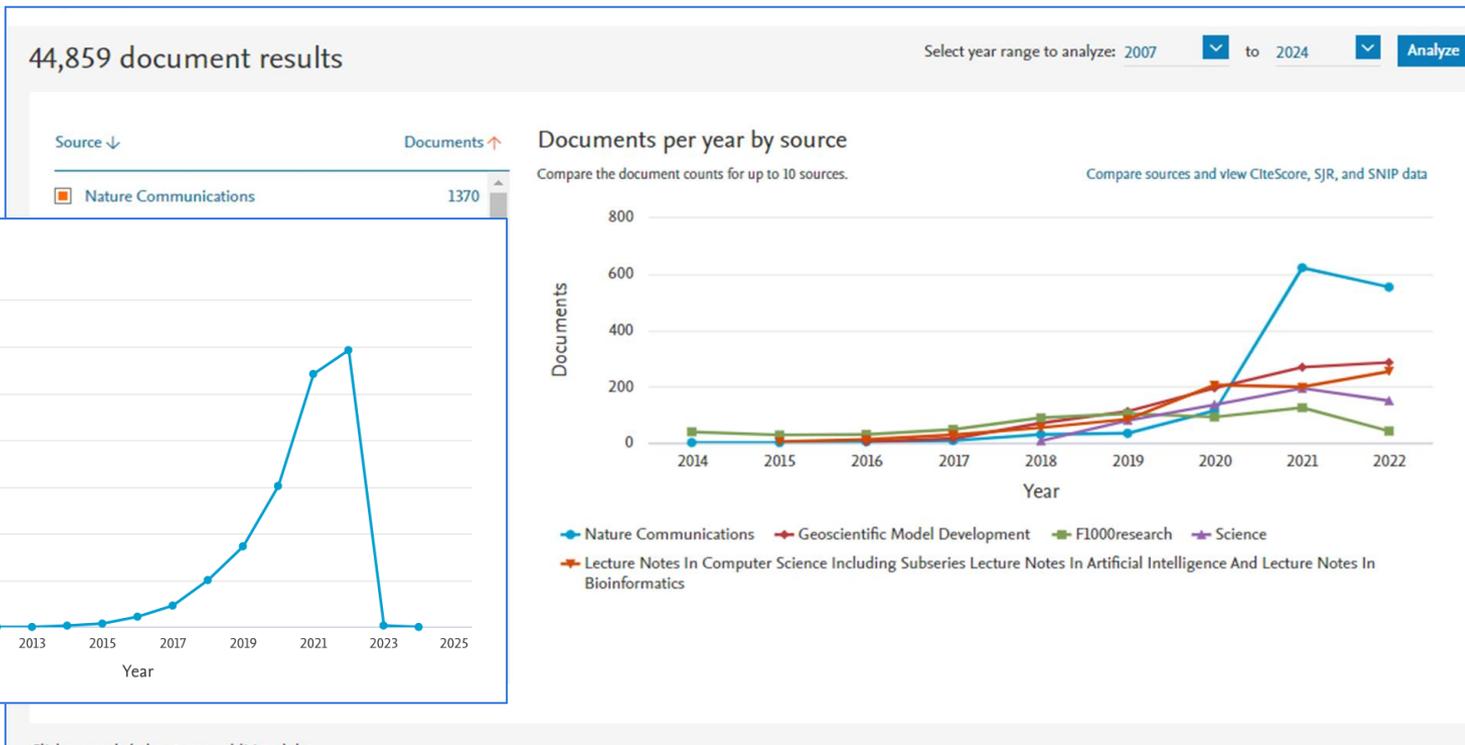
We recommend that software (including computational code, scripts, models, notebooks and libraries) should be cited in the same way as other sources of information to support proper attribution and credit, reproducibility, collaboration and reuse, and encourage building on the work of others to further research. To facilitate this, useful information is provided in this [article on the essentials of software citation](#) ↗ by FORCE 11, of which Elsevier is a member. A reference to software should always include the following elements: creator(s) e.g. the authors or project that developed the software, software title, software repository, version (where available), year, and global persistent identifier.

Data references

This journal encourages you to cite underlying or relevant datasets in your manuscript by citing them in your text and including a **data reference in your Reference List**. Data references should include the following elements: author name(s), dataset title, data repository, version (where available), year, and global persistent identifier. Add [dataset] immediately before the reference so we can properly identify it as a data reference. The [dataset] identifier will not appear in your published article.



Zenodo references in reference list (in Scopus)



My personal take on this

- The academic community needs to reach consensus on how to consistently cite software, and share their outcome with Publishers. Publishers will not take the lead.
- Recognition can also come from ORCID and CRediT; not just citations
- Publishers can support the implementation process
- We can better instruct editors and referees what we expect from them
- Consolidated usage logging framework was never finished for article **usage** (not tried for citations)
- We currently support:
 - Data linking (through framework developed by Scholix, DataCite and CrossRef)
 - Software references in ref. list
 - Original Software Publication (OSP article type) with Code Summary (sharing URL to Code Ocean and/or GitHub)
 - Limited search capabilities





ELSEVIER

Thank you



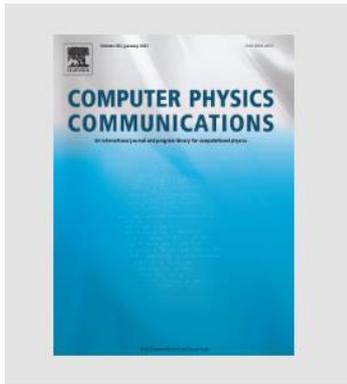
Several examples of software citations

URL (or DOI) in ref list

Jones et al., 2001 Jones, E., Oliphant, T., Peterson, P., et al., 2001. SciPy: Open source scientific tools for Python. URL <http://www.scipy.org/> [Online; accessed June 4, 2019].
[Google Scholar](#)

Reference to ASCL or SoftwareHeritage or CodeOcean (all by proxy)

van Diepen and Dijkema, 2018 van Diepen G., Dijkema T.J.
DPPP: Default Pre-Processing Pipeline
 Astrophysics Source Code Library (2018)
<https://ascl.net/1804.003>
[Google Scholar](#)



Footnote plus paper ref plus incomplete web reference

ASTRON (Netherlands Foundation for Research in Astronomy) with support from the Netherlands Foundation for Scientific Research (NWO). This research made use of Astropy,³⁴ a community-developed core Python package for Astronomy (Astropy Collaboration et al., 2013, Astropy Collaboration et al., 2018). This research made use of the Python Kapteyn Package (Terlouw and Vogelaar, 2015).

Footnote 34

<http://www.astropy.org>

[View in article](#)

Search Computer Phy

[Advanced search help](#)

Note:

Ref to Kapteyn Package is not linked or easily retrievable through this reference.

Google Scholar reports 64 citations to this particular reference.

5. Additional mosaicking software

To allow users a convenient workflow with the provided Apertif data products, especially continuum mosaicking and polarisation and line cubes, we developed the Aperlac independent mosaicking package *amosaic*.²⁹ This package is purely based on up-to-date Python3 routines and allows command line or Jupyter notebook

Footnote 29

<https://github.com/apertif/amosaic>

[View in article](#)

Filter Results

DATA TYPES

- Dataset (3912)
- File Set (3897)
- Text (62)
- Document (46)
- Software/Code (19)
- Other (8)
- Video (4)
- Image (2)

Even when cited, a sub-optimal version may be cited:

Janssen et al., 2019 Janssen M., Goddi C., Falcke H., van Rossum D., van Bemmel I., Kettenis M., Small D., Marti-Vidal I.
rPICARD: A CASA-based calibration pipeline for VLBI data
 (2019)
 arXiv e-prints, [arXiv:1905.01905](https://arxiv.org/abs/1905.01905)
[Google Scholar](#)

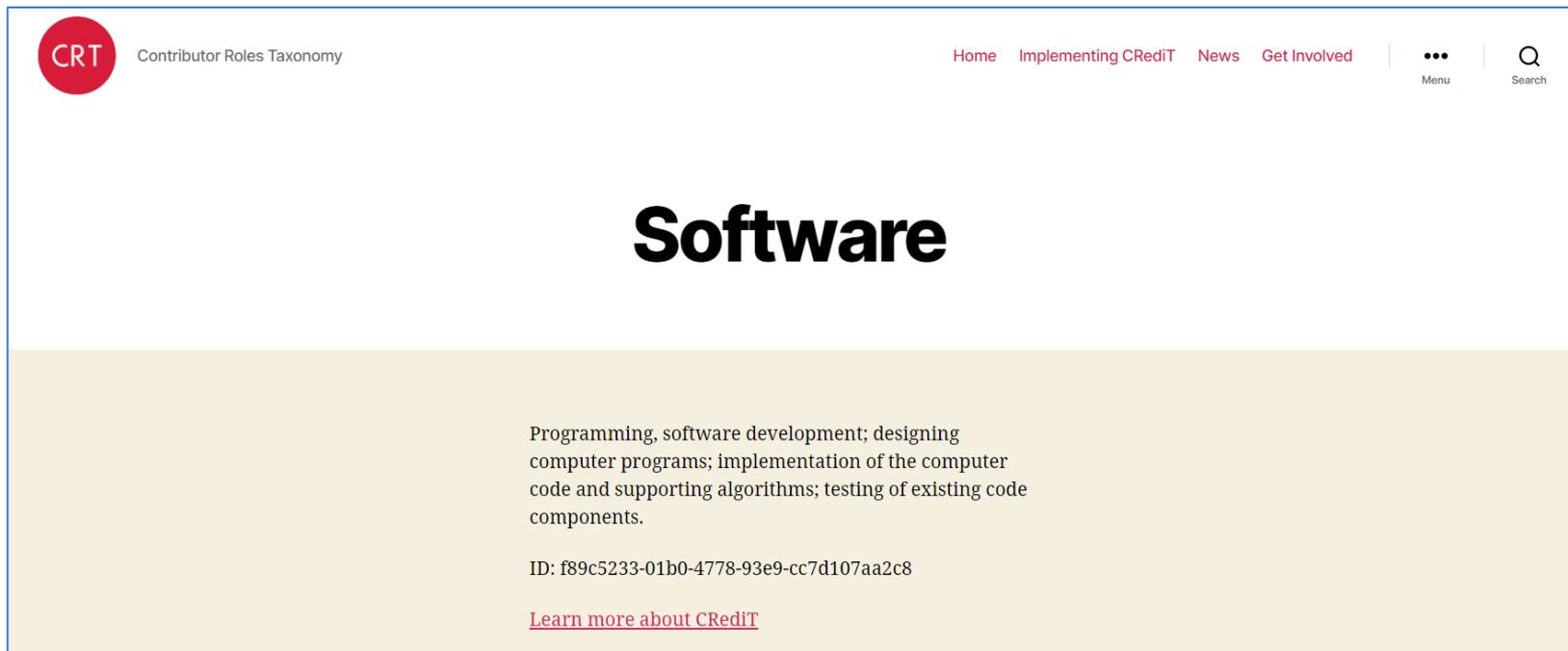
Published VorR: A&A (2019)
<https://doi.org/10.1051/0004-6361/201935181>



23.11.2022

Preprint; not using article

CReditT – one of 14 contributor roles



The screenshot shows the CReditT website interface. In the top left corner, there is a red circular logo with the letters 'CRT' and the text 'Contributor Roles Taxonomy' next to it. In the top right corner, there is a navigation menu with links for 'Home', 'Implementing CReditT', 'News', and 'Get Involved', followed by a 'Menu' icon (three dots) and a 'Search' icon (magnifying glass). The main content area features the word 'Software' in a large, bold, black font. Below this, there is a light beige background containing the following text: 'Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.' Below this description is the ID: 'f89c5233-01b0-4778-93e9-cc7d107aa2c8'. At the bottom of the beige area, there is a red link that says 'Learn more about CReditT'.



Recommendations

- In designing a way forward:
Let's focus on what we can already measure – i.e. collect automatically, using articles and existing workflows. You can only expect so much from a publisher/indexation service, not a new metric.
- *“Both journal editors and reviewers can help with this through more active enforcement of **mandatory** data and code sharing policies.”*

(source: Hamilton, D.G., Page, M.J., Finch, S. *et al.* How often do cancer researchers make their data and code available and what factors are associated with sharing?. *BMC Med* **20**, 438 (2022). Published November 9, 2022. <https://doi.org/10.1186/s12916-022-02644-2>)

