

Melissa HARRISON

EMBL-EBI

Open Science, Publications and Code

Introduction

- Until Nov 2021



- Since Nov 2021



- Europe PMC



- Software citation advocacy



Data: [10.6084/m9.figshare.19006721](https://figshare.com/figures/19006721) (M Parkin)

Problems

- Yet another thing I have to do to get published!
- How do I cite code? Here's the citation to the original research publication
- Software doesn't fit usual citation styles and publishers don't know how to tag them
- Software is not static

Solution?

- Persistent identifier for code
 - Archive code from GitHub to Zenodo to get a DOI
 - DOI is easy for publisher to recognise in reference list
 - Zenodo links back to the GitHub site



Code has a DOI...now what?

7000 random XML documents:

- **300,000** citations found
- **10** Zenodo citations identified, ie with Zenodo DOI



DOI	Publisher category	Datacite category
10.5281/zenodo.3899574	book	Software
10.5281/zenodo.3406690	book	Software
10.5281/zenodo.3551936	miscellaneous	Software
10.5281/zenodo.2531640	other	Software
10.5281/zenodo.1308157	other	Software
10.5281/zenodo.11640	software	Software
10.5281/zenodo.3724830	miscellaneous	Dataset
10.5281/zenodo.5092028	journal	Dataset
10.5281/zenodo.4033353	journal	Dataset
10.5281/zenodo.4407706	journal	Text*

- Zenodo category: Journal article

Issues



=

<code>



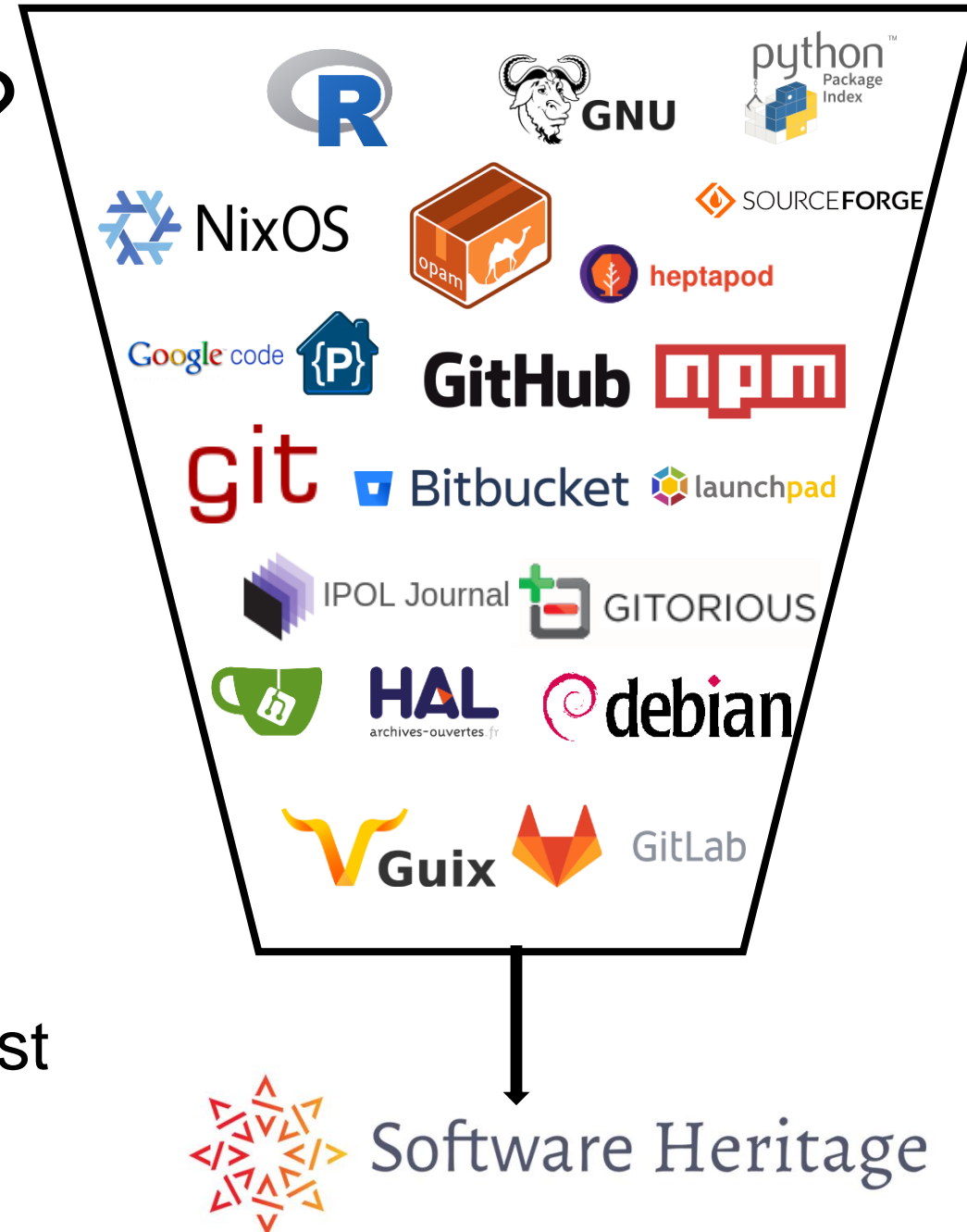
Publishers often validate references
against Crossref



Zenodo and many other DOIs are
sourced from DataCite

Solution?

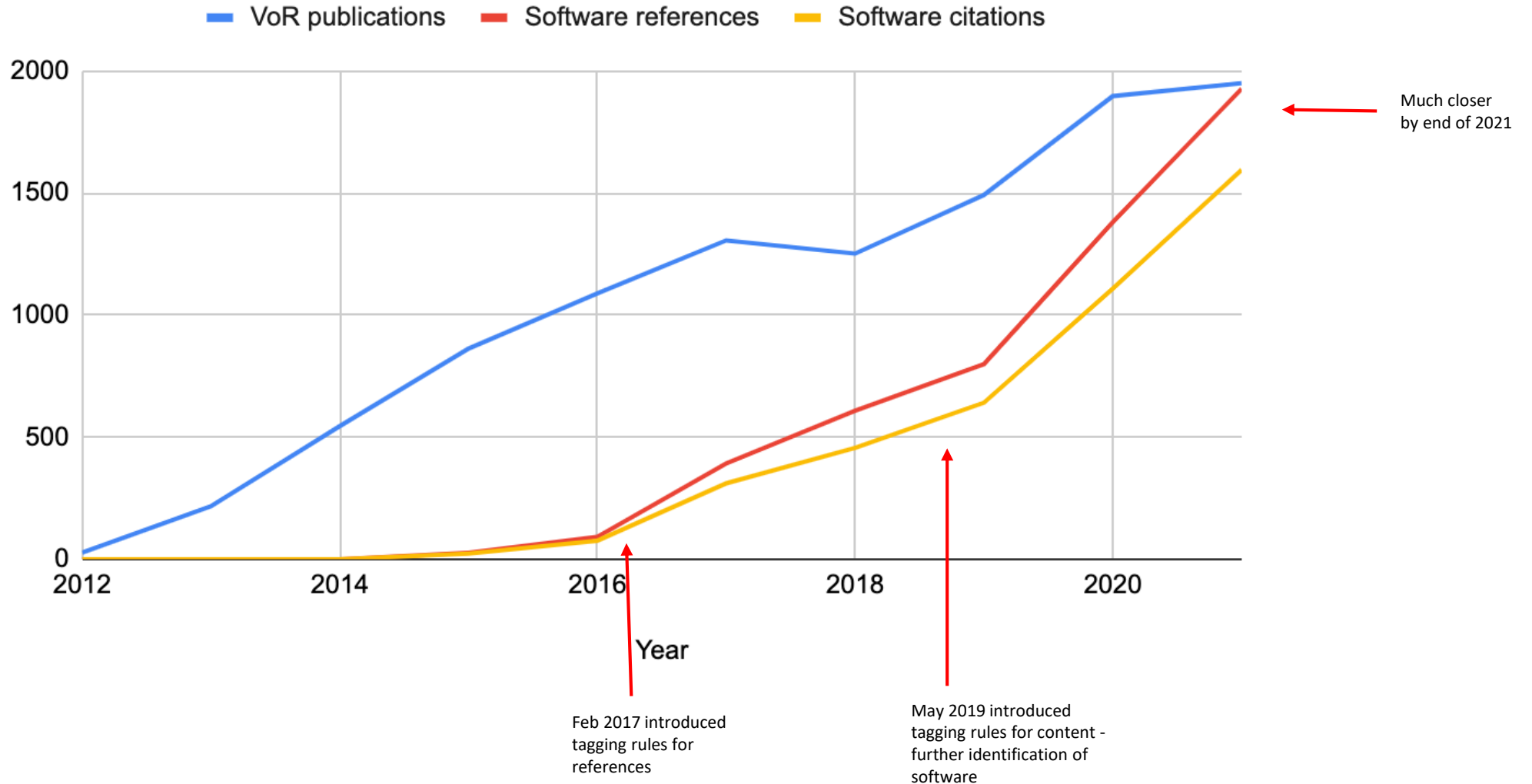
- Persistent identifier for code
- Archive code to Software Heritage
- Use SWH to link to specific parts of code
- SWH re-ingests and updates
- SWH indexes more sources than just GitHub



eLife experience



Software citations in eLife over time



Time and effort

- Quality control and instructions
- Documentation
- Instructions and training
- Automated checks

What we found

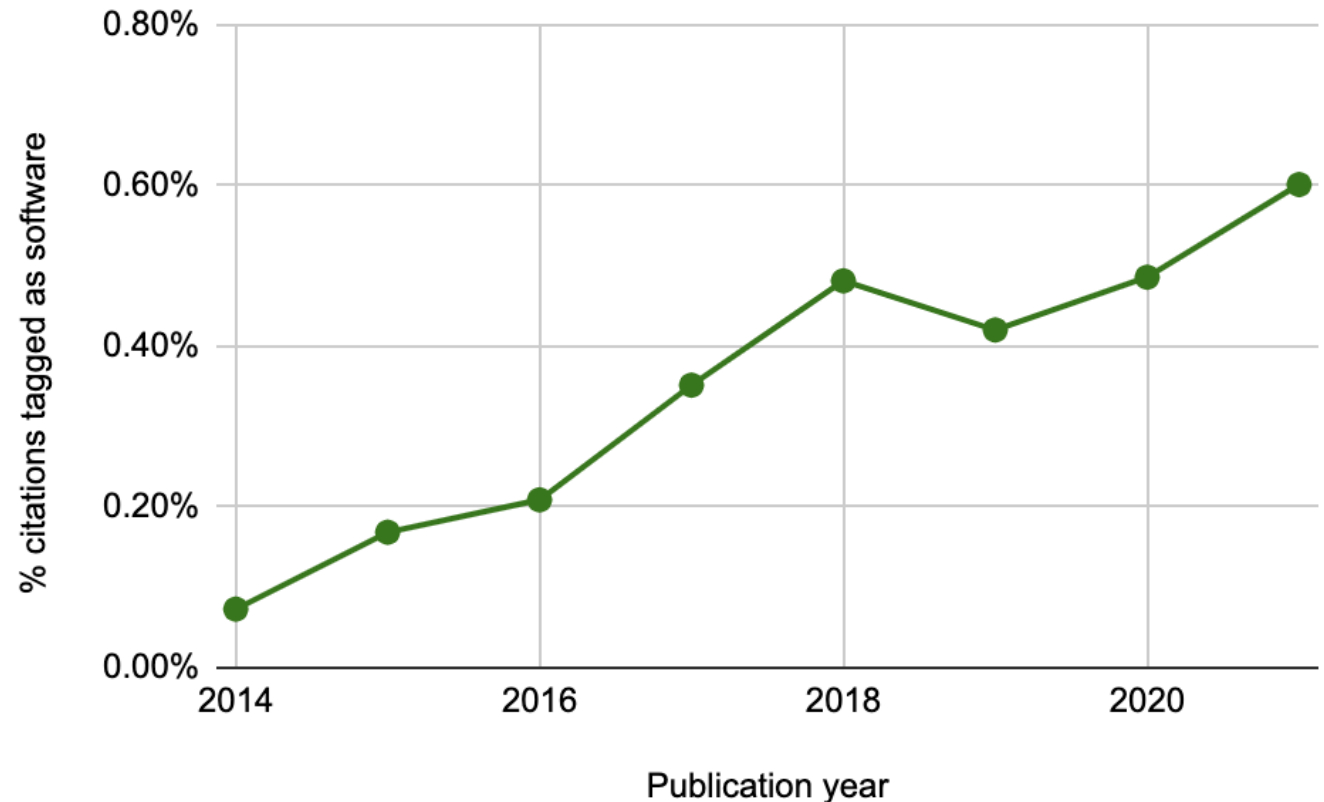


- Using the same 7000 XML documents:
 - 15 references with “software” @publication-type
- Looked at eLife and Peer J articles since 2014



Software citations in eLife & PeerJ

Based on "publication-type" attribute in PMC JATS XML



What are they “referencing”?

source	n	percentage
Not available	2868	46.22
GitHub	1151	18.55
software heritage	387	6.24
zenodo	142	2.29
r package	70	1.13
cran	56	0.90
r package version	54	0.87
gitlab	35	0.56

- 6205 “software” references identified
- Top 8 listed, then long tail of 1055 other sites – 932 are unique “source”

Software citations



Validator tool Who's involved Contact us Privacy policy

Home > Recommendations > Software citations

Search recommendations... e.g. 'affiliatio

Contents

Provenance

Context

Description

Additional reading

Recommendation

Examples

History

Software citations

Status: Published

Version: 1

License: this recommendation document is licensed under [CC BY-ND 4.0](#)

DOI: <https://doi.org/10.3789/niso-rp-40-2021>

ISBN (13): 978-1-950980-16-1

Provenance

JATS4R software citation subgroup members:

Mohamed Abdelhakeem, *Hindawi*, Jeffrey Beck, *NCBI/NLM/NIH* (<https://orcid.org/0000-0002-1798-9797>), Daina Bouquin, *Harvard-Smithsonian Center for Astrophysics* (<https://orcid.org/0000-0003-2626-3688>), Shani Chachamu, *Oxford University Press*, Neil Chue Hong, *University of Edinburgh* (<https://orcid.org/0000-0002-8876-7606>), Andreas Czerniak, *Bielefeld University* (<https://orcid.org/0000-0003-3883-4169>), Chris Erdmann, *AGU* (<https://orcid.org/0000-0003-2554-180X>), Michael Evans, *F1000*, Kara Handren, *University of Toronto* (<https://orcid.org/0000-0003-3877-8371>), Brooks Hanson, Melissa Harrison (Chair), *eLife* (<https://orcid.org/0000-0003-3523-4408>), Trevor J Hiblar, *Silverchair*, Daniel S. Katz, *University of Illinois* (<https://orcid.org/0000-0001-5934-7525>), Lauren Kmec, *AAAS* (<https://orcid.org/0000-0003-2301-5208>), Thomas Lemberger, *EMBO*, Alain Monteil, *INRIA*, August Muench, *American Astronomical Society* (<https://orcid.org/0000-0003-0666-6367>); Luke Prescott, *Hindawi*, Shelley Stall, *AGU* (<https://orcid.org/0000-0003-2926-8353>), Jake Yeston, *AAAS* (<https://orcid.org/0000-0002-8252-0848>); Viviana Letizia, *Elsevier* (<https://orcid.org/0000-0003-1088-5255>)

History

Emerging references subgroup*

Working: September 27, 2017 – Jan 3, 2020

Software citation subgroup

Working: March 1, 2021 – Sept 1, 2021

JATS4R Committee review: September 7, 2021 – October 1, 2021

Public review: October 1, 2021 – November 1, 2021

JATS4R Steering Committee review: November 2 – November 5, 2021

NISO Topic Committee review: November 5 – 17 November 2021

Published: 19 November 2021

Software citations

Recommendation

1. `@publication-type="software"` on `<mixed-citation>` or `<element-citation>`. Use "software" as the value of `@publication-type` to indicate that the citation is to software. Do not attempt dual tagging of references using the `@publication-type` attribute (e.g. data | software), one or the other is required.
9. `<pub-id>`. If there is a well-defined identifier for software this element should be used, for example doi, accession number, or SWHID. As per existing JATS4R recommendations on [data citations](#) (recommendation 6), this element should be used to hold both the repository ID for the software, in the element content (see [Examples 1A-C](#)), and, if applicable, the full URI to the software, in the `@xlink:href` attribute (see [Examples 5A-B](#)).
13. `<ext-link>`. Must be used to provide the URI for the software available online if there is not a pub-id containing a resolvable link. Use the standard xlink attributes to provide a live link. Do not use for DOIs (see [Examples 7A-B](#))

- 6205 “software” references identified
- 202 had pub-id – all DOIs
- 2524 had external link
- 3483 had no pub-id and no external link

Conclusions

Long road ahead

- Which path to take?
- Many contributors
- Education
- Update tooling

