AMERICAN ASTRONOMICAL SOCIETY

a data editor's view of the state of data and software citation in astronomy

topics

- who are the AAS Journal data editors?
- what are current models for software review and citation?
- where can open data linked to journal articles live (or die)?
- what would a review for “reproducibility” look like?

Data/Software Keywords:	Data/software review
This manuscript contained the following Data/Software keywords that may require review: github.com, sourceforge.net	1) We recommend that living code on github repositories (e.g. Be-synthesis-with-moog) place a "frozen" version on Zenodo (or other 3rd party repositories that issue DOIs) and then cite them in the article. A tutorial on how to do this is available here: https://github.com/AASJournals/Tutorials/tree/master/Repositories [Edit]

data review at submission:
a quick(ish) review of >90% of all manuscripts

➔ run scripts to identify linked code repositories;

➔ review links for remote, unarchived data/code/figures;

➔ request data be included or archived & linked to final article;

➔ review tables, figures & animations for size or accessibility;

➔ submit data/code recommendations to scientific editor for review.

example: animations (dominates our time)

Data/Software Review

Editor:

Please direct the author to the AAS Journal's Graphics Guide: animations are no longer supplemental material. Per the Guide, the author should modify the videos and improve the text/caption descriptions of the animations. More information and an example edit for Figure 5 can be found at these links:

https://journals.aas.org/graphics-guide/#animations
https://authortools.aas.org/AAS21849/

PA: I put a pretty good concat of the 4 Figure 5 animations at the above link.

[Edit]

example: data problem (published w/o fix)

Data/software review

The manuscript says that the catalog as described in Table 2 will be "released on-line". Can the authors be more specific in the revised manuscript? Do they mean with the published article in machine readable format? If on an external site we recommend a copy also be placed in a 3rd party repository that issues DOIs like Zenodo. A tutorial on how to set this up is available here:

https://github.com/AASJournals/Tutorials/blob/master/Repositories/UsingRepositories.md

example: software citations (fixes of var types)

Data/software review

It is great the authors are using \software to highlight the code used in the manuscript but citation should be given when known, e.g. Scipy (Jones et al. 2001), emcee (Foreman-Mackey et al. 2013), CASA (v5.1.1; McMullin et al. 2007), MIRIAD (Sault, Teuben & Wright M., 1995), & radiobear (de Pater et al. 2019).

Data/Software Review

Editor:
Even though the version they use is a modified GOTPM code, the original developers deserve attribution via citation: https://ui.adsabs.harvard.edu /abs/2004NewA....9..111D/abstract . For the purposes of reproducibility the text would best list more details of the modified version of GOTPM, e.g., https://astro.kias.re.kr/~kjhan/GOTPM/index.html.

example: (no cost) data request

Data/software review

The authors should consider providing the spectra shown in Figure 1 as the Data behind the Figure (DbF). The readers would surely appreciate having this important data set available with the published article.

example: accessibility (mainly informative)

Data/Software Review

Editor: The authors should begin using tools such as Color Oracle to check their figures for color accessibility. In a number of Figures color is used as the sole discriminator of point data whiledifferent symbols are better. Red/Green points or solid lines are especially inaccessible.

PA: At submission the author's wrote (and we should followup on at PA):

After acceptance, I would like to provide most of the figures in machine-readable format (ascii or fits files), so that other scientists can easily retrieve this extensive data set and play with it. The data set is extremely rich, and allows for

```
full range of possible parameterizations for the systematics that we investigated are detailed in
Section~\ref{sec_parameterization}. We used the \textsc{batman} package to model the transit
\citep{2015PASP..127.1161K} using a linear limb-darkening prescription.
374-
375:We used the Python MCMC package \textsc{emcee} to fit the parameters describing the transit and
the systematics simultaneously \citep{2013PASP..125..306F}. We fitted each observation separately. We
used the routines of \citet{2010PASP..122..935E} to convert the calendar dates in the headers of each
frame of each observation to BJD$_{\mathrm{TDB}}$.
376-
377-Figure~\ref{whitelight_corrected} shows the white lightcurves for each observation after removing
```

at post-acceptance: data editing

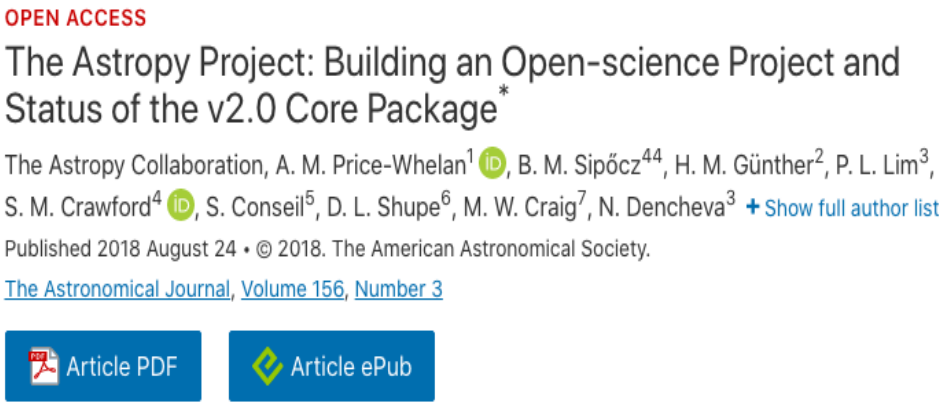
- ➔ 15-20% of accepted manuscripts enter post-acceptance data editing;
  - ➔ tables are standardized; interactive/animated materials edited;
- ➔ data repositories or PID requested for some raw or submitted materials;
  - ➔ run scripts to pick out code mentions, check for references;
- ➔ request authors archive and cite their codes & find preferred citations.

models for software citation

➔ Journal article or Journal like citation;

➔ Indexed in a Code Library;

➔ Direct Citation of Software (by version/release).



OPEN ACCESS

The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package*

The Astropy Collaboration, A. M. Price-Whelan[1], B. M. Sipőcz[44], H. M. Günther[2], P. L. Lim[3], S. M. Crawford[4], S. Conseil[5], D. L. Shupe[6], M. W. Craig[7], N. Dencheva[3] + Show full author list

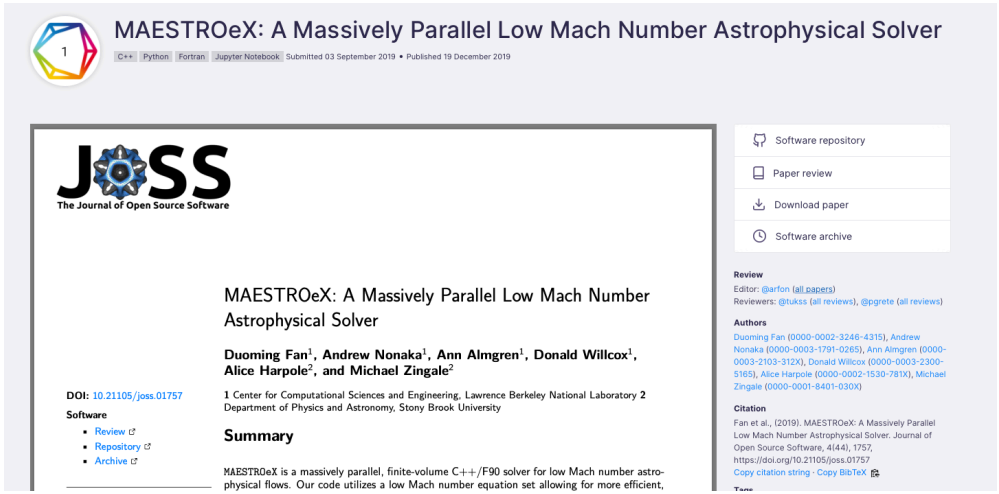
Published 2018 August 24 • © 2018. The American Astronomical Society.

The Astronomical Journal, Volume 156, Number 3

Article PDF Article ePub

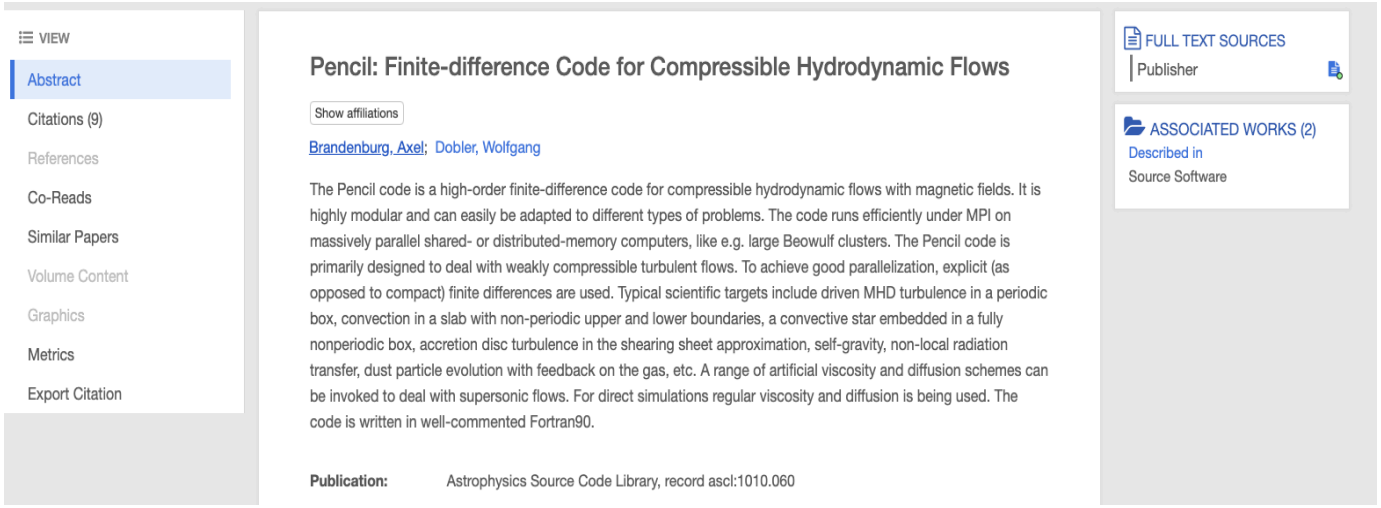
models for software citation: Journal article

- ➔ Publish a proxy for the software via a “Journal article.”
- ➔ Examples: Astropy Collaboration et al. (2013, 2018)
- ➔ Accumulates new authors/developers via new publication;
- ➔ “Astronomy” Journals do not directly review the software.
- ➔ https://www.astropy.org/acknowledging.html



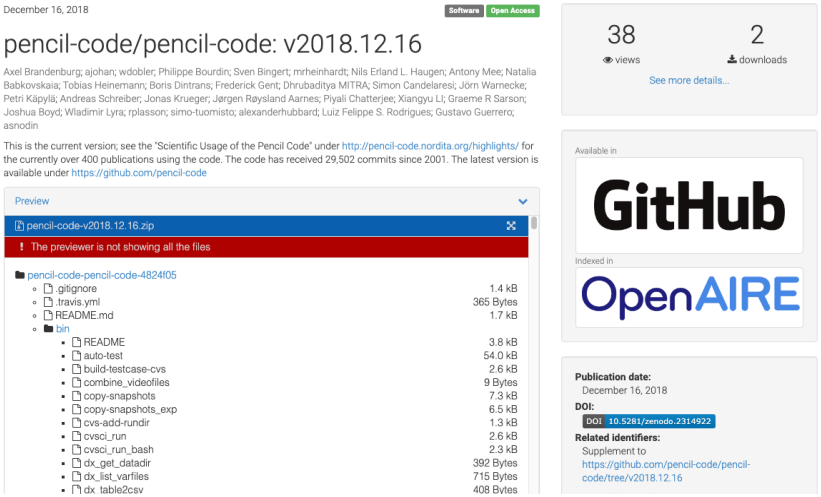
models for software citation: Journal “like” article

- ➔ The Journal of Open Source Software: https://joss.theoj.org/
- ➔ Scripted, detailed peer-review of code;
- ➔ Accumulates new authors/developers via new publication;
- ➔ A “science” submission to the AAS Journals can have a parallel “software” peer-review.
  - ➔ https://doi.org/10.3847/1538-4357/ab4f75
  - ➔ https://doi.org/10.21105/joss.01757



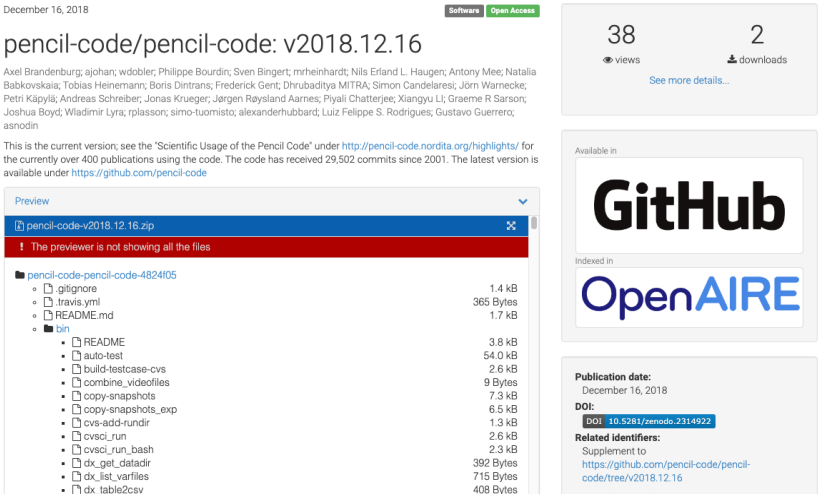
models for software citation: Code Index

- ➔ List your software in the Astrophysics Source Code Library (ASCL);
- ➔ Provides links between the source codebase and documentation.
- ➔ Stores the “preferred” citation for a developer team;
- ➔ No versioning but author list can be modified at any time for any reason.
- ➔ ASCL Record for Pencil: https://ascl.net/1010.060



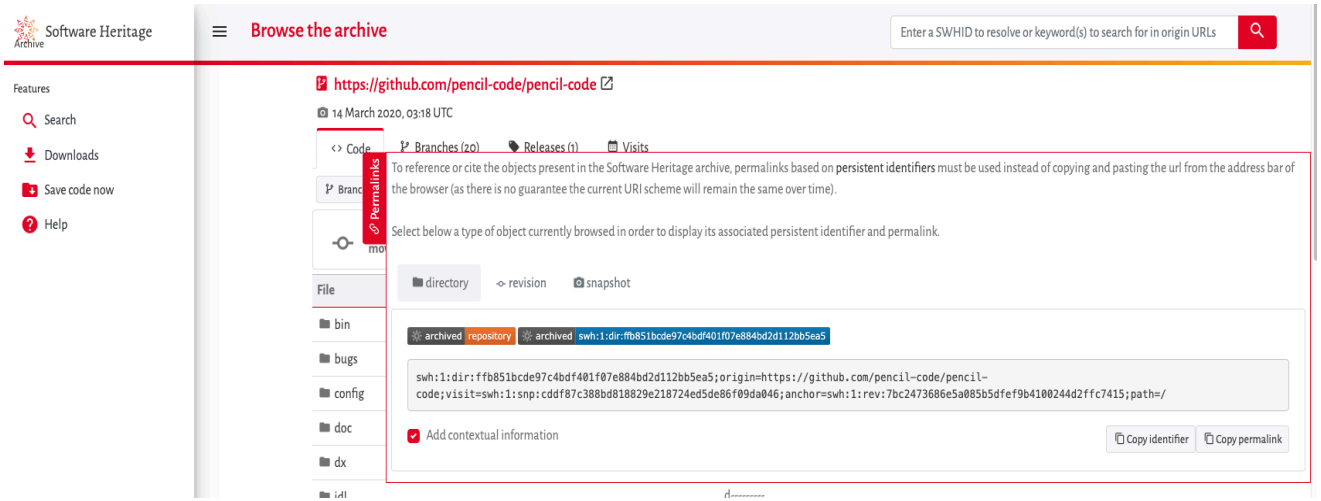
models for software citation: Direct Citation of Software

- ➔ Citation of a digital object representing a persistent version of software
- ➔ Github is not considered a preservation platform.
- ➔ Examples:
  - ➔ Developer export code to a separate platform (e.g., Zenodo)
  - ➔ Code is automatically archived (e.g., Software Heritage)



models for software citation: Direct Citation of Software

- ➔ Zenodo archiving requires developers to:
  - ➔ Use versioning as a proxy for citation;
  - ➔ Review and maintain all the Zenodo object metadata for each version;
- ➔ ADS Indexes any cited version of a Zenodo software object:
  - ➔ doctype:software and bibstem:"zndo"
- ➔ Latest Release (v2018.12.16): https://doi.org/10.5281/zenodo.2314922



models for software citation: Direct Citation of Software

- ➔ Software Heritage automatically (not quite) indexes all software repositories
- ➔ Citations are granular (down to the commit);
- ➔ Citations have no authorship model (yet) nor can they be indexed in ADS.
- ➔ https://archive.softwareheritage.org/browse/origin/directory/?origin_url=https://github.com/pencil-code/pencil-code

models for software citation: Pencil Code

- ➔ There is no preferred citation model on the Pencil landing pages
- ➔ A user who goes to ASCL finds the ASCL entry is the “preferred” citation;
- ➔ The Zenodo entry has never been cited and thus is not indexed in ADS;
- ➔ What does the Pencil Code community desire from these software citation models?

models for data citation and reproducibility: Pencil Code

- ➔ Examples of data citation: Zenodo AAS Community, MAST, MESA inlists
- ➔ What could the Pencil Code Community imagine from a reproducibility review?