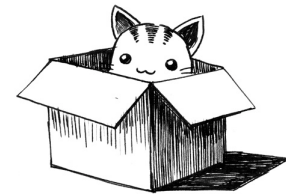


# Schrödinger's code: Opening the computational box



Alice Allen

Astrophysics Source Code Library (ASCL)/UMD

@asclnet aallen@ascl.net



ASCL.net

Astrophysics Source Code Library

*Making codes discoverable and Husky proud since 1999*



# Abstract

Though computational methods are widely used in many disciplines, many researchers do not share the source code they develop, making it difficult to replicate and reuse the work. This presentation will cover the changing landscape that includes funders' requirements, policy changes for existing journals, community resources, and more, that make it easy to release and archive codes to ensure they are available to support the research they enabled, improve the reproducibility of science, increase confidence in research, and meet new requirements made by funders and journals in many disciplines. It will also cover how the Astrophysics Source Code Library (ASCL), which has been working since 1999 to improve the transparency of research by registering open codes used in research, has made it possible for software to be cited as a first-order research object, and how researchers can garner credit for their codes by having them cited correctly and improve papers by including citations for the computational methods that enabled the research.

---

aallen@ascl.net

*"[Software] is the most prevalent of all the instruments used in modern science"*

*C. Goble, IEEE Internet Computing, Sept-Oct 2014*



*"... anything less than release of actual source code is an indefensible approach for any scientific results that depend on computation..."*

*Ince, Hatton, & Graham-Cumming, Nature, 2012*



# Schrödinger's Code

It's not until you open the box that you know whether the code is alive or dead.

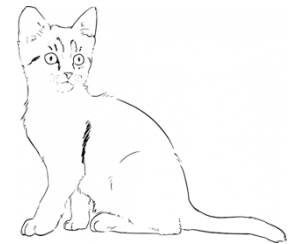
We opened the box...



aallen@ascl.net

# Schrödinger's Code

The good news: 58% were alive!



The rest...

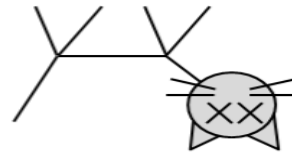


Image credits: tiskapot, Peter Kaptein, Olivia Cales

# Looking for software

Papers examined for software use	166
Individual software “instances”	715
Identifiable software “instances”	418
Unique codes	285
Source code readily available	58%

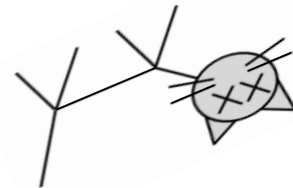




# Two problems!

Only **418** of the 715 were identifiable

Source code was not available for **40%**  
of the software used



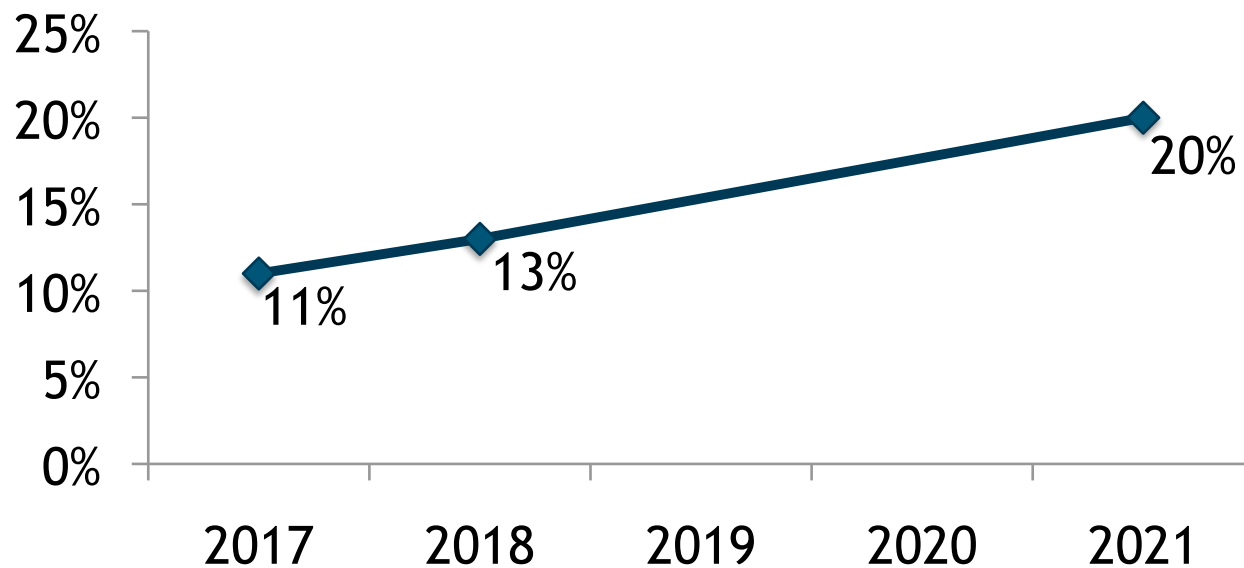
# Hyperlinks testing

Tested **2558** http(s) and ftp: protocol  
links extracted from 2015 papers

# Another problem!

**11%** of all links were unreachable

# Percentage of links failing



---

Ryan, Allen, & Teuben, ADASS XXVIII, 2019  
Data and code available at [github.com/teuben/ascl-tools](https://github.com/teuben/ascl-tools)



# Source code availability in other disciplines...

Collberg, Proebsting, & Warren (2014)	<b>56%</b>
Howison & Bullard (2016)	<b>24-40%</b>

## Hyperlink rot...

Mangul et al (2018)	<b>24%</b>
---------------------	------------



*...a change in a code researchers had not noticed led to incorrect results*

*M. Zorotovic, M. R. Schreiber and S. G. Parsons, Astronomy & Astrophysics, Aug 2014*

*...“a hidden coding error fueled a seven-year dispute between two of condensed matter’s top theorists.”*

*A. G. Smart, Physics Today, 22 Aug 2018*



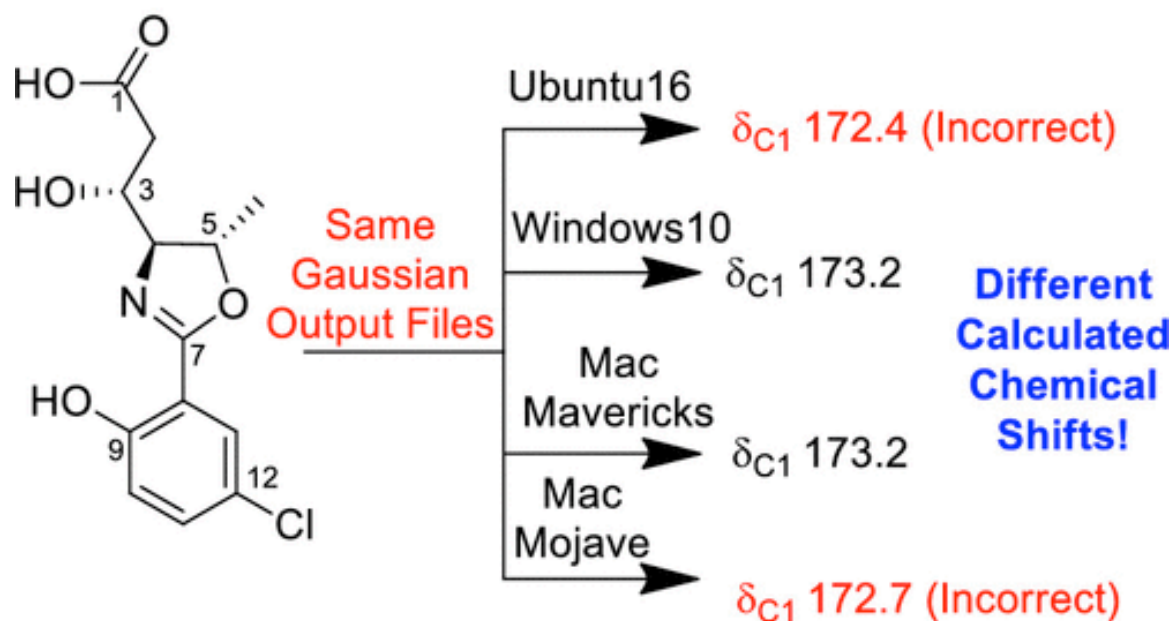
*"... calculations established the relative configurations of 1 and 2 and revealed that **the calculated shifts depended on the operating system when using the "Willoughby-Hoye" Python scripts ... a previously unrecognized flaw that could lead to incorrect conclusions.**"*

Neupane et al, Organic Letters 2019 21 (20), 8449-8453





## "Willoughby-Hoye" Scripts from 2014 Nature Protocols



Neupane et al, Organic Letters 2019 21 (20), 8449-8453

aallen@ascl.net



**Lorena Barba**  
@LorenaABarba



The lesson here is not merely that a piece of research code had a bug that led to operating-system-dependent results, but that PUBLISHING THE CODE led to discovery of the problem  
[vice.com/en\\_us/article/...](https://www.vice.com/en_us/article/...) #Reproducibility



# Policy changes



- All proposals submitted to the Office of Science for research funding must include a Data Management Plan (DMP).
- DMPs should provide a plan for making all research data displayed in publications resulting from the proposed research open, machine-readable, and digitally accessible to the public at the time of publication.
- Research data is ... the recorded factual material commonly accepted in the scientific community as necessary to validate research findings. The term *digital data* encompasses ... experimental, observational, and simulation data; **codes, software and algorithms** ...

aallen@ascl.net

## Frequently Asked Questions

### Do I need to submit a Data Management Plan (DMP)?

**1. Will a proposal be funded if it does not include a Data Management Plan?**

No. The Office of Science reserves the right to reject, without merit review, any proposal that does not include a DMP.

**2. Will a proposal be funded if the DMP does not follow the [Requirements](#) of the Office of Science Statement on Digital Data Management?**

No.

**3. I am submitting a research proposal in response to the Office of Science "[Annual Funding Opportunity Announcement](#)", do I need to include a data management plan with my proposal?**

Yes.

<https://science.osti.gov/Funding-Opportunities/Digital-Data-Management/FAQs>

---



# Policy changes



- Software ... created as part of a ROSES award, should be made publicly available when it is practical and feasible to do so, and when there is scientific utility in doing so.
- SMD expects that the source code, with associated documentation sufficient to enable use of the code, will be made publicly available as Open Source Software (OSS) under an appropriately permissive license.
- This includes all software developed with SMD funding used in the production of data products, as well as software developed to discover, access, visualize, and transform NASA data.

# Policy changes



- Fastlane will not permit submission of a proposal that is missing the DMP.
- Significant software or code developed as part of the project should be distributed open-source, and include a description of how users can access the code, how to obtain documentation on how to use the code, and the conditions under which they can use and modify the code. A software license should be explicitly specified.
- Results from Prior NSF Support: A description of data and other products created or generated during the research supported by an NSF award must be included in the section 'Results from Prior NSF Support'.



aallen@ascl.net

## Changes in existing journals

- Encourage or require code release
- Encourage or require software citations
- Allow software articles without research results





# Policy changes

## nature portfolio

---

nature > nature portfolio > editorial policies

- Authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications.
- Upon publication, Nature Portfolio journals consider it best practice to release custom computer code in a way that allows readers to repeat the published results.
- A subset of Nature journals undertake peer review of custom code or mathematical algorithm, and software... we require release of code/algorithm during the peer review process, verification by peer reviewers and release of code/algorithm at publication.



# Policy changes



- All data, program code, and other methods must be appropriately cited.
- In general, all computer code central to the findings being reported should be available to readers to ensure reproducibility.
- Author-written source code that is not yet publicly available should be archived in a permanent public repository prior to publication and likewise cited.

aallen@ascl.net

# Policy changes



- AAS Journals welcome articles which describe the design and function of software of relevance to research in astronomy and astrophysics.
- Such articles need not include research results produced using the software, although including examples of applications can be helpful.
- We highly recommend that authors release code described in an article under an appropriate open source license and archive the published version of their code.



aallen@ascl.net

## New journals

*Journal of Open Research  
Software*

*Astronomy and Computing*

*Computational Astrophysics and  
Cosmology*

*Software X*

*Journal of Open Source  
Software*

*Computing and Software for  
Big Science*



aallen@ascl.net

# Citation methods and services

Better ways to cite software explicitly

New tools to help with citation



## Community resources

More places to put software and information about software

Indexers capture/track software citations

Broader efforts cross disciplines and influence others



## Browsing Codes

Results 126–250 of 2456 (2415 ASCL, 41 submitted)

[Previous](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#)  
[18](#) [19](#) [20](#) [Next](#)

Order

[Title](#)



[Date](#)



Mode

[Abstract](#)

[Compact](#)

Per Page

[50](#)

[100](#)

[250](#)

[All](#)

### [[ascl:2010.007](#)] [stella: Stellar flares identifier](#)

[Feinstein, Adina D.](#); [Montet, Benjamin T.](#); [Ansdell, Megan](#)

stella creates and trains a neural network to identify stellar flares. Within stella, users can simulate flares as a training set, run a neural network, and feed in their own data to the neural network model. The software returns a probability at each data point as to whether that data point is part of a flare; the code can also characterize the flares identified.

### [[ascl:2010.006](#)] [LaSSI: Large-Scale Structure Information](#)

[Schaan, Emmanuel](#)

LaSSI produces forecasts for the LSST 3x2 point functions analysis, or the LSSTxCMB S4 and LSSTxSO 6x2 point functions analyses using a Fisher matrix. It computes the auto and cross correlations of galaxy number density, galaxy shear and CMB lensing convergence. The software includes the effect of Gaussian and outlier photo-z errors, shear multiplicative bias, linear galaxy bias, and extensions to  $\Lambda$ CDM.

### [[ascl:2010.005](#)] [GRAPUS: GRAvitational instability PopUlation Synthesis](#)

[Forgan, Duncan](#); [Rice, Ken](#); [Hall, C.](#); [Meru, F.](#)



aallen@ascl.net

# Astrophysics Source Code Library

Registry of software used in refereed astrophysics research

Can archive software upon request, but mostly point to source code download site

Started in 1999 by Dr. Nemiroff

Housed at MTU



# Benefits of registering your code

Unique identifier

Listing in indexing services, including Web of Science  
and Google Scholar

Trackable citation method

Increased discoverability



aallen@ascl.net

# How to register your code

Use handy [online submissions form](#)

Email info to [editor@ascl.net](mailto:editor@ascl.net)

Submit via [codemeta.json file](#)





**GREATER RECOGNITION**

aallen@ascl.net

1. Release your software with a license
2. Specify citation method



## Citation methods

Software itself via ASCL, JOSS, DOI from archiving service

Article using or describing the code

~~GitHub, SourceForge, BitBucket repo URL~~ **NO!**

~~URL to personal institutional page~~ **NO!**

~~URLs in general~~ **NO!**

1. Release your software with a license
2. Specify citation method
3. Register your code
4. Archive your code
5. Cite other people's codes well
6. Include a software section in articles





aallen@ascl.net

RAWR



# Save science: Release your codes!

Links to the resources mentioned are available at:

<https://tinyurl.com/MTUtalk>

