siam 2021 Conference on Computational Science and Engineering

Schrödinger's Code: A Preliminary Study on Research Source Code Availability and Link Persistence in Astrophysics

Alice Allen, ASCL/University of Maryland, aallen@ascl.net
Peter Teuben, University of Maryland

Abstract: How available is the software that enables astrophysics research? How persistent are links that scientists include in their research articles? We set out to answer these questions by examining a sample of astrophysics research articles published in 2015, looking through the papers for software, and categorizing it as to its availability. We also extracted hyperlinks from 1669 research articles published in 2015 by one journal, discarded those generally added by the publishing process, and tested the remaining links to see how many worked two years after the papers were published. We were pleasantly surprised to learn that 58% of the 285 unique codes discovered in papers were available, and distressed to find that only two years after publication, at best 90% of the hyperlinks we tested still worked. This talk will discuss our findings and subsequent work and its relevance to scientific software availability.

aallen@ascl.net

Software is the most used instrument in astronomy

aallen@ascl.net

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

Ince, Hatton, & Graham-Cumming, *The case for open computer programs*, Nature, v. 482, Feb. 23, 2012

aallen@ascl.net

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…

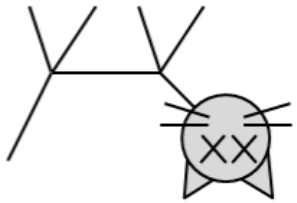


Image credits: tisKapot, Peter Kaptein, Olivia Cales

aallen@ascl.net

Looking for software

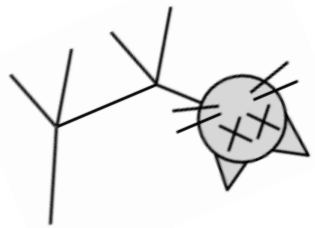
Papers examined for software use	166
Individual software “instances”	715
Identifiable software “instances”	418
Unique codes	285
Source code readily available	58%

aallen@ascl.net

Two problems!

Only ***418*** of the 715 were identifiable

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



aallen@ascl.net

Hyperlinks testing

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

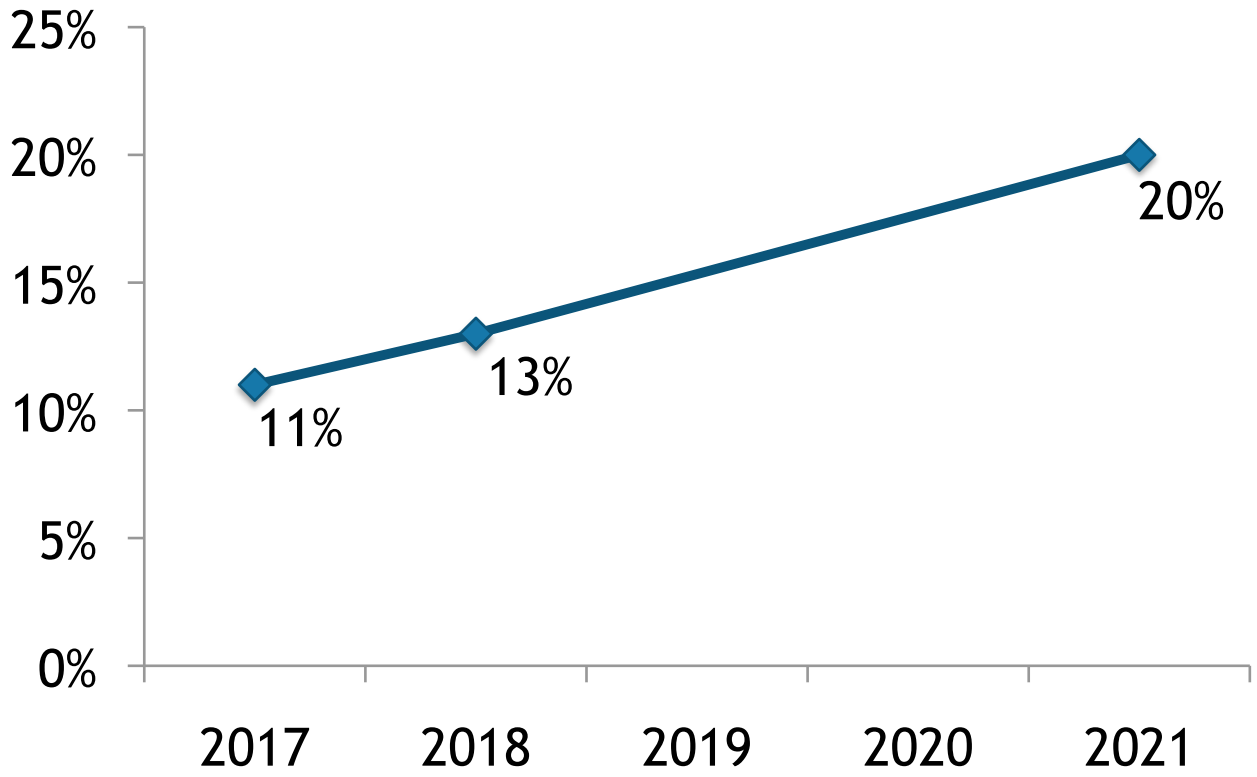
aallen@ascl.net

Another problem!

11% of all links were unreachable

aallen@ascl.net

Percentage of links failing



Data and code available at github.com/teuben/ascl-tools

aallen@ascl.net

Source code availability in other disciplines…

Collberg *et al* (2014)	***56%***
Howison & Bullard (2016)	***24-40%***

Hyperlink rot…

Mangul *et al* (2018)	***24%***

aallen@ascl.net

What will help/is helping?

Rewarding the effort involved

Training

Leadership from mentors, advisors, & referees

Changes in journal practices

Better support through technology

aallen@ascl.net

Resources

List o’ links

Check the ASCL blog for resources

https://ascl.net/wordpress/

Thank you!