

Hyperlink persistence in astrophysics papers

P. Wesley Ryan (ASCL, wes@ascl.net)

Why do we care whether hyperlinks in papers work?

Astrophysics papers frequently contain hyperlinks to external resources. These hyperlinks may refer to code, data, or other information crucial to the paper.

Failure of these links over time may decrease research transparency, understanding, and reproducibility.

What we did

We extracted embedded hyperlinks from all papers published in *Astronomy and Astrophysics* (A&A) and the *Astrophysical Journal* (ApJ) in 2015, and all papers published in A&A in 2010, to measure the persistence of these links.

How we did it

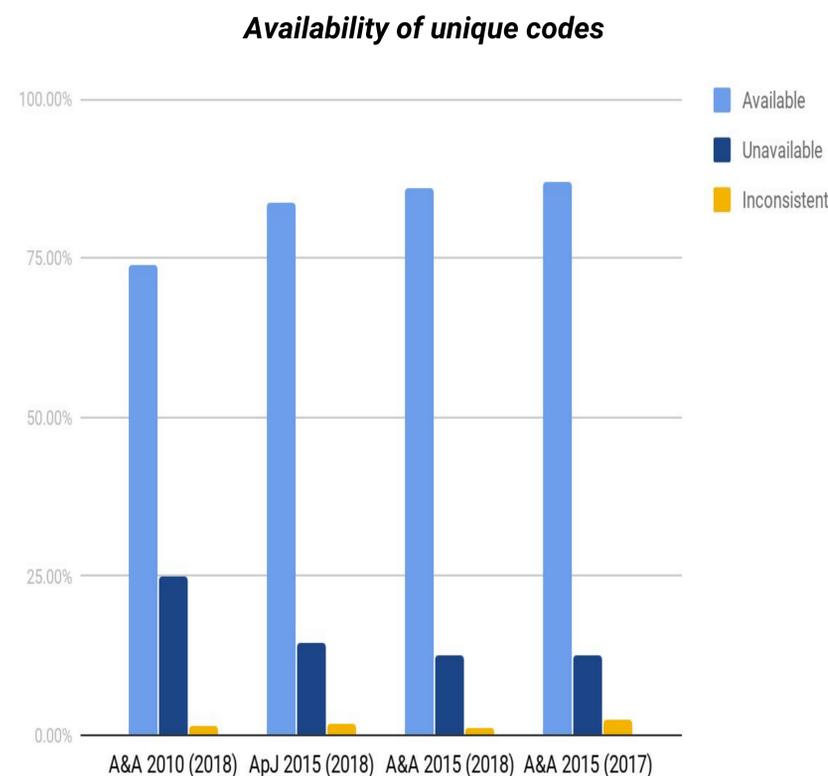
In September–October 2017, we extracted the clickable HTTP(S) hyperlinks from all 2015 A&A papers into a database, filtering out hyperlinks to email addresses, DOIs, and frequently-cited ‘infrastructure’ Web sites that we knew to be available, to build files of hyperlinks.

One year later, we tested the same set of links again in order to measure link persistence over time. We also extracted, filtered, and tested links in the 3,050 papers published in ApJ in 2015, using the same methods.

In December 2018, we repeated this process for the papers published in A&A in 2010.

What we found out

Links start going bad in less than 2 years, and we found 25% loss in only 8 years.



Links (time of testing)	Up	Down	Inconsistent
A&A 2010 (2018)	73.8%	24.9%	1.3%
ApJ 2015 (2018)	83.6%	14.6%	1.8%
A&A 2015 (2018)	86.1%	12.7%	1.2%
A&A 2015 (2017)	86.9%	10.6%	2.5%

Conclusions

A quarter of the links in our 2010 sample were inaccessible after only eight years, and similar results have been found in computational biology, medicine, and other fields, recently and even more than a decade ago.

Future directions

We plan to monitor the links in our current dataset, and collect new datasets from papers published in different journals and different years, in order to expand our understanding of link decay.

Similar research in other disciplines

Mangul *et al.* (2018) found that 26% of computational biology software tools were not accessible through the URLs referenced in published papers, and 57% of the tools sampled for a usability test could not be installed by strictly following the instructions provided in the manual.

Mangul *et al.* also found that 32% of software linked in computational biology papers published before 2012 was inaccessible by the provided URL, and 14% of software linked in papers published after 2012 was.

The use of URLs in published research was shown to be an issue well over a decade ago, as Crishlow *et al.* (2004) clearly demonstrated.

Citations

Allen A., Teuben, P. J. & Ryan, P. W. 2018 *ApJS*, 236, 10
Crichlow R, Davies S, Winbush N. *JAMA*. 2004;292(22):2720–2726. doi:10.1001/jama.292.22.2723-c
Howison, J., & Bullard, J. 2016, *Journal of the Association for Information Science and Technology*, 67, 2137. <http://dx.doi.org/10.1002/asi.23538>
Mangul, S., *et al*, 2018, bioRxiv pre-print. <http://dx.doi.org/10.1101/452532>
Ryan, P.W., 2017, *Astrophysics Source Code Library*, ascl:1711.018