Citation of data and software in astronomy

A journal editor's perspective

Dr Keith T. Smith

Associate Editor,
Science magazine

4 April 2018
EWASS/NAM, Liverpool, UK

A study that puts human nature under the microscope p. 510
Sparse ice stresses polar bears pp. 524 & 568
HLA influences cancer therapy response pp. 526 & 582
Science
$15
2 FEBRUARY 2018
sciencemag.org
AAAS
A COSMOLOGICAL
CONUNDRUM
Motions of Centaurus A's satellite galaxies challenge structure formation pp. 520 & 534

2 Feb 2018 cover

Science | AAAS

Reproducibility

> *"If I have seen further it is by standing on the shoulders of Giants"*

Letter from Isaac Newton to Robert Hooke, 1675

- Most scientific advances build upon previous work
- This relies on reproducibility – the ability to check and extend previous work
- There are many current challenges to preserving this cycle

From *Blind Orion Searching for the Rising Sun*, Nicolas Poussin, 1658

Citations enhance reproducibility

- Verify claims
- Direct readers to full descriptions of the resources used
- Help readers obtain their own copies
- Should be *as specific as possible* e.g. to the relevant version
- Should be human legible and easy to locate

Citations give credit

- A citation assigns credit to previous work that influenced or enabled this one
- Can be intellectual (e.g. a concept) or practical (e.g. a tool)
- Citations are often used as a measure of research impact. This has numerous problems and is not advisable, but is nevertheless widespread.
- Should be automatically parsable

Citing data

- This is already common in observational astronomy, though not uniformly applied
- Previously published data:
  - Bad: We used SDSS
  - Poor: We used SDSS (York+ 2000)
  - Good: We used data release 12 of SDSS (Alam+ 2015)
- Cite the specific version used
- Give sufficient details to fully identify the exact dataset used

Citing data

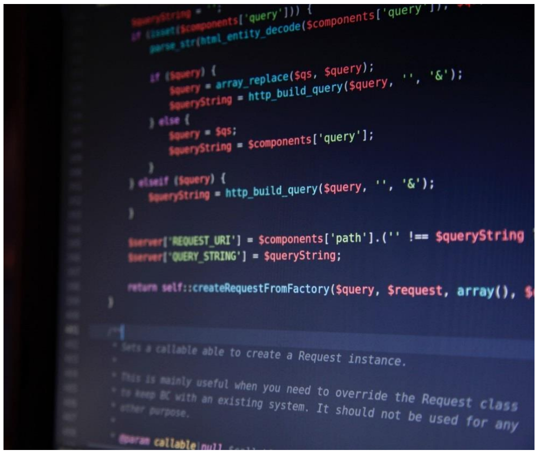
- Raw data:
  - Bad: from the ESO archive
  - Good: available in the ESO archive (archive.eso.org) under programme 67.B-0026
- Reduced data or simulation output:
  - Often simply not mentioned
  - Bad: available on request
  - Good: available at DOI 10.6084/m9.figshare.2075356

Citing software

- The same principles apply to software
- Previously published code:
  - Bad: we used GADGET
  - Poor: we used GADGET (Springel 2005)
  - Good: we used version 2.0.7 of GADGET (Springel 2005, mpa-garching.mpg.de/gadget/)
- New code:
  - Often simply not mentioned
  - Bad: available on request
  - Good: available at github.com/jobovy/gaia_tools

Citing software

- Problem: not all software has a suitable citation, especially for older code
- Check if/how the authors asked to be cited
- My order of preference:
  - Peer-reviewed paper
  - Non-peer-reviewed preprint or conference proceeding
  - Persistent unique identifier e.g. DOI or ASCL ID
  - Informal publication e.g. a user manual
  - Website URL
- Use a formal entry in the reference list



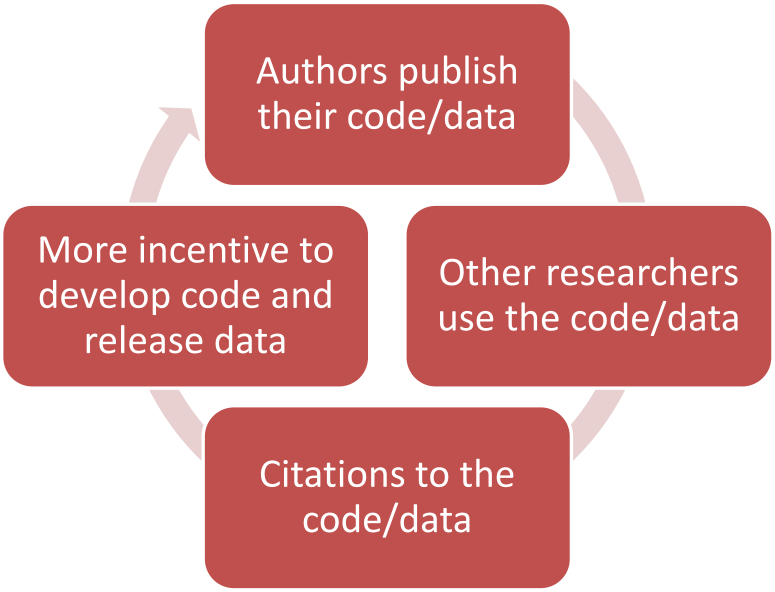
```
    $queryString = '';
    if (isset($components['query'])) {
        parse_str(html_entity_decode($components['query']),

        if ($query) {
            $query = array_replace($qs, $query);
            $queryString = http_build_query($query, '', '&');
        } else {
            $query = $qs;
            $queryString = $components['query'];
        }
    } elseif ($query) {
        $queryString = http_build_query($query, '', '&');
    }

    $server['REQUEST_URI'] = $components['path'].('' !== $queryString
    $server['QUERY_STRING'] = $queryString;

    return self::createRequestFromFactory($query, $request, array(),
}

/**
 * Sets a callable able to create a Request instance.
 *
 * This is mainly useful when you need to override the Request class
 * to keep BC with an existing system. It should not be used for any
 * other purpose.
 *
 * @param callable|null
```

Science | AAAS

Citing software

- It's not necessary or beneficial to cite generic, commercial or non-scientific software
  - e.g. Matlab, LaTeX, IDL
- Scientific software that was vital to the work, or required to reproduce it, should be cited
- There are some grey areas

A virtuous cycle



Authors publish their code/data

Other researchers use the code/data

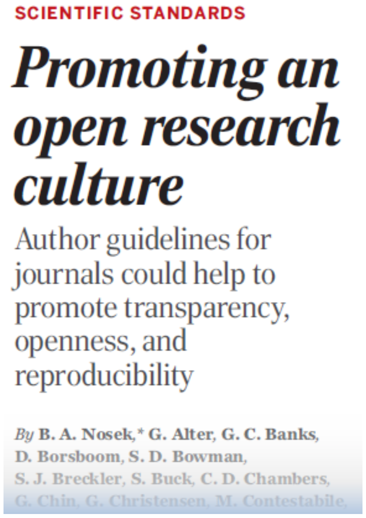
Citations to the code/data

More incentive to develop code and release data

TOP guidelines

- Transparency and Openness Promotion (TOP) is a policy framework for journals
- Developed by the Center for Open Science from 2013-15, with community & journal input
- Published in *Science*, one of the founding signatories
- Now has over 5,000 journals signed up - but none of the major astronomy journals

CENTER FOR OPEN SCIENCE



SCIENTIFIC STANDARDS

Promoting an open research culture

Author guidelines for journals could help to promote transparency, openness, and reproducibility

By **B. A. Nosek,* G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile,**

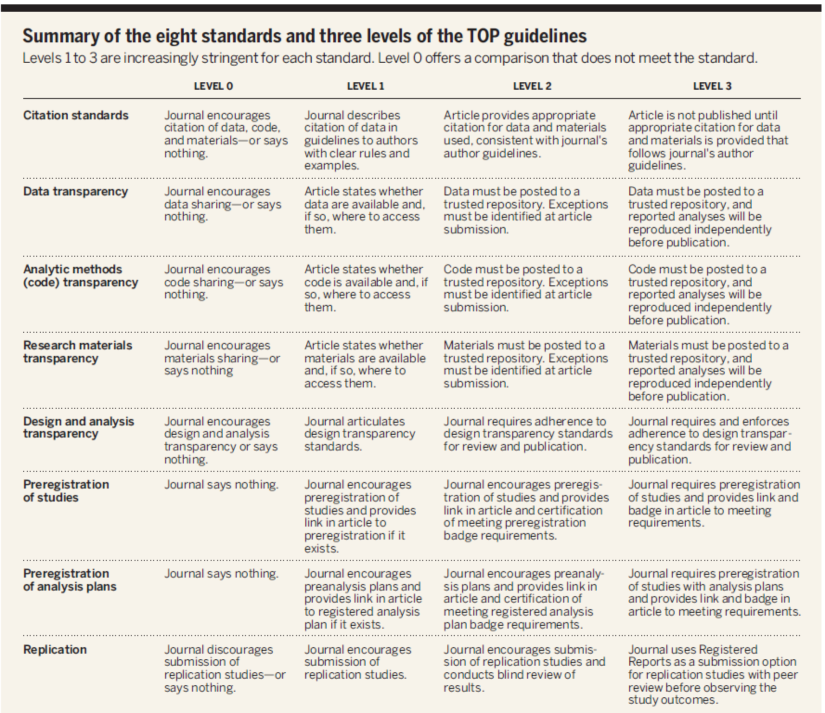
Nosek+, *Science*, 2015

Science | AAAS

TOP guidelines

Nosek+, *Science*, 2015

- Three levels of compliance in eight areas
- Each journal can choose which levels to adopt
  - – Not all signatories are equal
- Provides *standardisation* of policies across many journals
- *Science* is mostly at level 2



Summary of the eight standards and three levels of the TOP guidelines

Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Citation standards	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Analytic methods (code) transparency	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Research materials transparency	Journal encourages materials sharing—or says nothing	Article states whether materials are available and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Design and analysis transparency	Journal encourages design and analysis transparency or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
Preregistration of studies	Journal says nothing.	Journal encourages preregistration of studies and provides link in article to preregistration if it exists.	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Preregistration of analysis plans	Journal says nothing.	Journal encourages preanalysis plans and provides link in article to registered analysis plan if it exists.	Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies—or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts blind review of results.	Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes.

Current *Science* policies

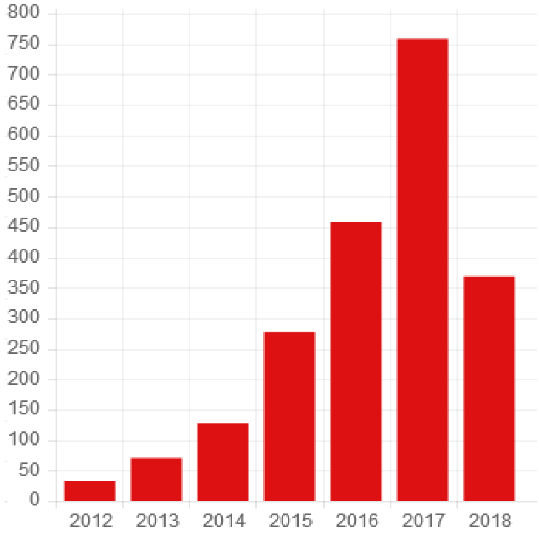
- "All data, program code, and other methods must be appropriately cited using DOIs, journal citations, or other persistent identifiers."
- "All data used in the analysis must be available to any researcher for purposes of reproducing or extending the analysis."
- "We require that all computer code used for modelling and/or data analysis that is not commercially available be deposited in a publicly accessible repository upon publication."

Summary excepts, full policies at sciencemag.org/authors/science-journals-editorial-policies

Outlook

- Citation of data & software is becoming more routine, but not yet universal
- Journal policies make a real difference; many are modernising in this area
- Editors and referees should push authors and be aware of developing community standards
- Raising author awareness helps

Citations to Astrophysics Source Code Library entries by year



ascl.net/dashboard on 28 Mar 2018

Science | AAAS

Summary

- Appropriate citations enhance reproducibility and give credit
- Use existing bibliographic reference systems wherever possible
- Give sufficient details e.g. version numbers, persistent IDs, author names. Not just an URL.
- TOP guidelines provide a standardised framework for journal policies
- Increasing community uptake

Any questions?

CO_2 turbines combat CO_2 emissions pp. 796 & 805
Mapping proteins in human cells pp. 806 & 820
Neurons that drive binge-like eating in mice p. 853
Science
$15
26 MAY 2017
sciencemag.org
AAAS
UNCOVERING
JUPITER
Juno sees the gas giant from new angles pp. 821 & 826

26 May 2017 cover