# Preprints and Data Repositories
## Best Practices for Citation
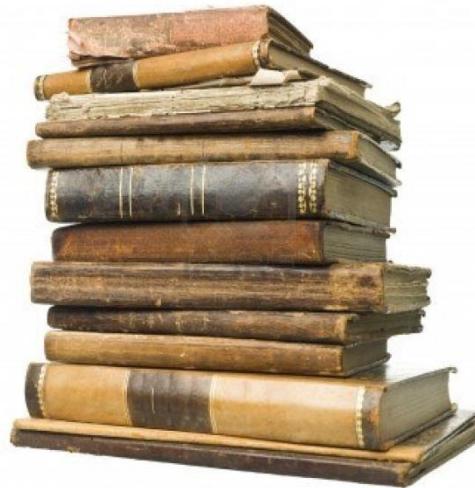
**Peter J. Olson, ELS**

Sheridan Journal Services

Waterbury, VT

# Preprints

## *What is a preprint?*

A preprint is an unpublished report of research that is posted online to encourage public discussion, feedback, and data sharing. Preprints are hosted on an online server and have *not* been peer reviewed.

# Preprint citations

**JCB** Journal of Cell Biology

### ARTICLE

## Kinesin-binding–triggered conformation switching of microtubules contributes to polarized transport

a third-rank tensor, which reflects the structure of the molecule (Psilodimitrakopoulos et al., 2013; Kaneshiro et al., 2018 *Preprint*). For a cylindrically symmetric polymer such as a microtubule, most elements of the SHG tensor vanish, and

Identifying a preprint as such in text makes it clear to the reader that a reference being used to support a particular statement has not been peer-reviewed.

Kaneshiro, J., Y. Okada, T. Shima, M. Tsujii, K. Imada, T. Ichimura, and T.M. Watanabe. 2018. Second harmonic generation polarization microscopy as a tool for protein structure analysis. *bioRxiv*. (Preprint posted June 4, 2018). https://doi.org/10.1101/338137

Identifying a preprint as such in the reference list distinguishes it from an accepted article and makes it clear that it is not a published-ahead-of-print article.

## *Updating preprint references*

Publish-ahead-of print (PAP) articles that are cited as references can often be updated before publication of the article in which they are cited:

Proc Natl Acad Sci U S A. 2018 Oct 15. pii: 201803505. doi: 10.1073/pnas.1803505115. [Epub ahead of print]

**Retrospective attribution of false beliefs in 3-year-old children.**

Király I[1,2], Oláh K[3], Csibra G[2,4], Kovács ÁM[2].

Retrospective attribution of false beliefs in 3-year-old children

Ildikó Király, Katalin Oláh, Gergely Csibra, and Ágnes Melinda Kovács

PNAS November 6, 2018 115 (45) 11477-11482; published ahead of print October 15, 2018 https://doi.org/10.1073/pnas.1803505115

However, the amount of time that elapses between the initial posting of a preprint and its publication as a peer-reviewed article may not allow for an update:

Once the paper is published, with page numbers and everything (which can often be years -- my paper above was accepted in April 2010, published January 2013), there is a form you can use to add the journal reference and DOI to the arXiv page. Conveniently, this does *not* generate a new version, so there's no reason not to (you can see this on my paper above -- the reference and DOI were added last week, but no new version of the paper was generated).

*From* Stack Exchange. Is it necessary to include and update the status of a paper on arXiv once it has been accepted for publication in a journal? https://academia.stackexchange.com/questions/8621/is-it-necessary-to-include-and-update-the-status-of-a-paper-on-arxiv-once-it-has. Accessed Nov 6, 2018.

# Data repositories

## *What is a data repository?*

A data repository (a.k.a "data warehouse" or "data mart") is an online, publically accessible library of a specific data population that is used to support current research and aid in the development and execution of future research.
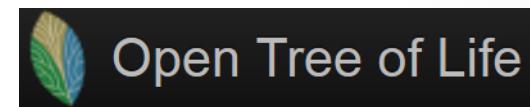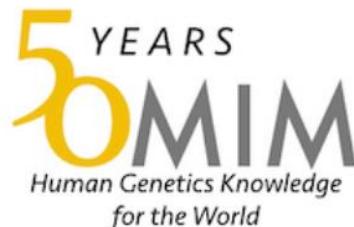
# Data repository citations: title page footnote



**PNAS**

## Embryophyte stress signaling evolved in the algal progenitors of land plants

Data deposition: The Transcriptome Shotgun Assembly projects have been deposited at DNA Data Bank of Japan/European Molecular Biology Laboratory (European Nucleotide Archive)/GenBank under accession nos. GFXX00000000, GFZF00000000, GFXY00000000, GFZG00000000, GFXZ00000000, and GFYA00000000. Single-read data have been deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive under accession nos. SRR5948987–SRR5949030 and SRR5990072–SRR5990080. The NCBI Bioproject ID is PRJNA399177.

**NOTE** Direct links to web pages for specific accession numbers can be generated via eXtyles® character styles.
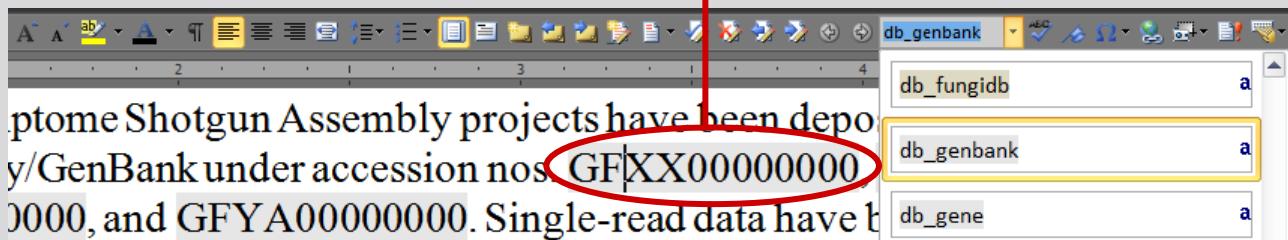
# *Data repository citations: data and software availability statement*

**Cell Reports** Resource

## Large-Scale Human Dendritic Cell Differentiation Revealing Notch-Dependent Lineage Bifurcation and Heterogeneity

**DATA AND SOFTWARE AVAILABILITY**

The accession numbers for the raw single cell RNA-sequencing data reported in this paper are GEO: GSE108546 and GSE115405.

**NOTE** The "data and software availability statement" can appear in the front matter, back matter, and/or full text HTML version of the article.

*From* Balan S et al. Large-scale human dendritic cell differentiation revealing notch-dependent lineage bifurcation and heterogeneity. Cell Rep 2018; 24:1902-1915.e6.

# *Data repository citations: reference list*

**ENDOCRINOLOGY**

## Association of *In Utero* Persistent Organic Pollutant Exposure With Placental Thyroid Hormones

31. Li ZM, Hernandez-Moreno D, Main KM, Skakkebaek NE, Kiviranta H, Toppari J, Feldt-Rasmussen U, Shen H, Schramm KW, De Angelis M. Data from: Association of in utero persistent organic pollutant exposure with placental thyroid hormones. Dryad 2018. Deposited 4 June 2018. https://doi.org/10.5061/dryad.9594945.

**NOTE** Certain data repositories will not provide working links for newly deposited data until the article in which the data are cited has been published in some form (namely, publish-ahead-of-print). As such, the author may need to request a functional link from the repository before final publication.

## *Querying the author*

### Sample query

> All cited data must be deposited in a publicly accessible database before publication. For any data cited in the text and/or the data deposition footnote, please provide complete reference information for the applicable database(s). If any data mentioned the data deposition footnote are not also cited in the text, please (a) insert a citation at an appropriate point in the text and (b) provide a complete reference for the applicable database.

### Critical components

Ask the author for:

- A description of the data

- The name of the database

- The accession number(s)

- Confirmation that the data have been deposited (for newly deposited data)

**NOTE**  Consider crafting a standard (or "standing") query to ensure that the same message is always delivered to your authors.

# Food for thought…

## *Preprints*

- Should preprint references be identified as such, or is a DOI sufficient?

- If a referenced preprint is eventually published as a peer-reviewed article, should the reference be updated to match the article?

- Is citing your own preprint a form of self-citation?

> **NOTE**  Authors often refer to preprints as "working papers"; make sure to distinguish these  from "manuscript in preparation" citations that do not have a DOI.

> **NOTE**  Preprint ≠ published ahead of print.

## *Data repositories*

- Do data repositories need to appear in the reference list, or is a footnote/link sufficient?

- Are specific accession numbers being cited? If so, is a separate reference required for each accession number, regardless of how many are cited?

> **NOTE**  Distinguish previously deposited data from newly deposited data by including an "accessed" or "deposited" date, respectively, in the reference list.

# Further Reading

**What are preprints, and how do they benefit authors?**
Mudrak B. AJE Scholar. https://www.aje.com/en/arc/benefits-of-preprints-for-researchers/.

**Preprints and citations: should non-peer reviewed material be included in article references?**
Crotty D. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2018/03/14/preprints-citations-non-peer-reviewed-material-included-article-references/.

**Plan for new medical preprint server receives a mixed response**
Enserink M. Science. https://www.sciencemag.org/news/2017/09/plan-new-medical-preprint-server-receives-mixed-response.

**What's up with data citations?**
Vines T. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2018/05/28/whats-up-with-data-citations/.

# Thank you!

**Peter J. Olson, ELS**
Sheridan Journal Services
Waterbury, VT
peter.olson@sheridan.com