Report for 2018 by the AEA Data Editor

By Lars Vilhuber*

Your abstract here.

The purpose of scientific publishing is the dissemination of robust research findings, exposing them to the scrutiny of peers. Key to this endeavor is documenting the provenance of those findings. For empirical articles, the foundations on which they reside are external to the article, and often to the journal, in which they are published. Our scientific community faces increasingly complex issues of privacy and confidentiality that prevent open access to those same sources. In consequence, there is a need to properly cite the digital inputs to our published output and to properly curate those inputs.

Many scientists, journals, learned societies, and funding agencies have called for greater transparency of research practices, and more assurance that published research is reproducible (Stodden et al., 2016; Fuentes, 2016; Moffitt, 2016; Camerer et al., 2016; Bollen et al., 2015; Joskow, 2015). This has lead to a focus on transparent access to research data and code (Coffman, Niederle and Wilson, 2017; Hoeffler, 2017; Duvendack, Palmer-Jones and Robert Reed, 2017; Hamermesh, 2017).

The AEA's data and code posting policy (American Economic Association, 2008), as well as that of other societies and journals, are intended to create a minimal framework from which to replicate empirical findings. Many partial solutions have been implemented. Journals have implemented verification of submitted code and data during the editorial process,[1] highlight the verification on data archives (Open Science Framework, 2017), maintain lists of acceptable third-party repositories,[2] and interlink with collaborating repositories to highlight authors' (and repositories') contributions to the data component of a scholarly work.[3] Outside of journals, several projects are working to educate the community to incorporate principles of replicability and traceability into their workflow.[4]Open Science Framework, Project TIER, BITSS are just a few of those active in that field (Gentzkow and Shapiro, 2014; Wilson et al., 2016). No journal currently does, in my opinion, an adequate job of providing information about restricted-access data, in part because most restricted-access data centers cannot provide

* Vilhuber: Cornell University, lars.vilhuber@cornell.edu.

[1]The American Journal of Political Science outsources this activity to the Odum Institute for Research in Social Science (CITE). The Journal of the American Statistical Association performs a "broad evaluation of quality and potential for usability of the code and data" since 2016 (Stodden et al., 2016).

[2]Nature Scientific Data maintains a list for its journals (Nature Scientific Data, 2016), and other institutions (CoreTrustSeal, FAIRsharing) have as their primary purpose to perform this kind of vetting.

[3]Elsevier interlinks, for instance, with ICPSR, highlighting the use of a repository on the article's web page.

[4]P

structured information about existence, modalities of access, or even data landing pages.[5] None of these solutions are widespread, and standards are only now being developed.

I. First Section in Body

Sample figure:

Figure here.

Figure 1. Caption for figure below.

Note: Figure notes without optional leadin.

Source: Figure notes with optional leadin (Source, in this case).

Sample table:

Table 1—Caption for table above.

	Heading 1	Heading 2
Row 1	1	2
Row 2	3	4

Note: Table notes environment without optional leadin.

Source: Table notes environment with optional leadin (Source, in this case).

II. Data Citations

Properly referencing data goes beyond just reproducibility - it is also proper scientific writing style. In the same way that we use bibliographic references to "printed" resources, we should also be using such references for data resources, to give and receive credit where credit is due. Not referencing an article or book is at best an oversight, and at worst plagiarism - and the same should apply to data objects. Numerous guides and tutorials exist (ICPSR, Force11, DataONE (2011*b*)).

The AEA uses the Chicago style for citations and bibliographies (American Economic Association, 2018). However, the Chicago Style Manual (REF) does not provide examples for data citations, and neither does the Citeproc database used by applications like Zotero (REF) and Mendeley (REF).

DataONE (DataONE, 2011*a*) suggests content and style that resemble the generic working paper or article citation style (adapted to Chicago style):

[5]Restricted-access data hosted on ICPSR and possibly Harvard Dataverse are notable exceptions. On the journal side, Elsevier journals have experimented with "Data Descriptions", but while the form is machine-readable, it is essentially free-form text, and checking the box "confidential data" essentially stops the process of filling in any information.

> Westbrook JW, Kitajima K, Burleigh JG, Kress WJ, Erickson DL, Wright SJ (2011) Data from: What makes a leaf tough? Patterns of correlated evolution between leaf toughness traits and demographic rates among 197 shade-tolerant woody species in a neotropical forest. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.8525

ICPSR (ICPSR, 2018) notes that a citation should include the following items:

- Title
- Author
- Date
- Version
- Persistent identifier (such as the Digital Object Identifier, Uniform Resource Name URN, or Handle System)

and provides a few examples, with some additional modifiers:

> Esther Duflo; Rohini Pande, 2006, "Dams, Poverty, Public Goods and Malaria Incidence in India", http://hdl.handle.net/1902.1/IOJHHXOOLZ UNF:5:obNHHq1gtV400a4T+Xrp9g== Murray Research Archive [Distributor] V2 [Version]

Finally, the AEA style guide (American Economic Association, 2018) suggests

> Leiss, Amelia. 1999. "Arms Transfers to Developing Countries, 19451968." Inter-University Consortium for Political and Social Research, Ann Arbor, MI. ICPSR05404-v1. doi:10.3886/ICPSR05404 (accessed February 8, 2011).

For users of BibTex, a generic database entry might look like

```
@techreport{duflopande2006,
author = {Esther Duflo and Rohini Pande},
year = 2006,
title = {Dams, Poverty, Public Goods and Malaria Incidence in India},
howpublished = {},
institution = {Murray Research Archive [Distributor]},
note = {\url{http://hdl.handle.net/1902.1/IOJHHXOOLZ UNF:5:obNHHq1gtV400a4T+Xrp9g==} V2
}
```

or

```
@techreport{leiss1999,
author = {Leiss, Amelia},
year = {1999},
```

```
title = {Arms Transfers to Developing Countries, 19451968},
institution = {Inter-University Consortium for Political and Social Research},
address = {Ann Arbor, MI},
note = {ICPSR05404-v1. DOI:  10.3886/ICPSR05404 (accessed February 8, 2011).},
doi = {10.3886/ICPSR05404},
}
```

and thus generate "Duflo and Pande (2006)" and "Leiss (1999)" and the bibliographic entry in the References.

REFERENCES

American Economic Association. 2008. "Data Availability Policy." *https://www.aeaweb.org/journals/policies/data-availability-policy*, Accessed: 2017-4-6.

American Economic Association. 2018. "Sample References." *https://www.aeaweb.org/journals/policies/sample-references*, (accessed on 2018-08-10).

Bollen, Kenneth, John T Cacioppo, Robert M Kaplan, Jon A Korsnick, and James L Olds. 2015. "Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science." Subcommittee on Replicability in Science, National Science Foundation Directorate for Social, Behavioral, and Economic Sciences.

Camerer, Colin F, Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu. 2016. "Evaluating replicability of laboratory experiments in economics." *Science*, 351(6280): 1433–1436.

Coffman, Lucas, Muriel Niederle, and Alistair J Wilson. 2017. "Replications: A Proposal to Increase their Visibility and Promote them."

DataONE. 2011*a*. "Data Citation and Attribution." *https://www.dataone.org/citing-dataone*, accessed on 2018-08-10.

DataONE. 2011*b*. "DataONE Tutorial on Data Citation." *http://www.dataone.org/sites/all/documents/L09_DataCitation.pptx*, accessed on 2018-08-10.

Duflo, Esther, and Rohini Pande. 2006. "Dams, Poverty, Public Goods and Malaria Incidence in India." Murray Research Archive [Distributor]. `http://hdl.handle.net/1902.1/IOJHHXOOLZUNF:5:obNHHq1gtV400a4T+Xrp9g==` V2 [Version].

Duvendack, Maren, Richard Palmer-Jones, and W Robert Reed. 2017. "What is Meant by 'Replication' and Why Does It Encounter Resistance in Economics?"

Fuentes, Montse. 2016. "Reproducible Research in JASA." *http://magazine.amstat.org/blog/2016/07/01/jasa-reproducible16/*, Accessed: 2017-4-4.

Gentzkow, M, and Jesse Shapiro. 2014. "Code and data for the social sciences: A practitioner's guide." University of Chicago.

Hamermesh, Daniel. 2017. "What is Replication? The Possibly Exemplary Example of Labor Economics."

Hoeffler, Jan H. 2017. "Replication and Economics Journal Policies."

ICPSR. 2018. "Data Citations." *https://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp*, accessed on 2018-08-10.

Joskow, Paul L. 2015. "President's Letter, Alfred P. Sloan Foundation Annual Report 2014." Alfred P. Sloan Foundation.

Leiss, Amelia. 1999. "Arms Transfers to Developing Countries, 19451968." Inter-University Consortium for Political and Social Research, Ann Arbor, MI. ICPSR05404-v1. DOI: 10.3886/ICPSR05404 (accessed February 8, 2011).

Moffitt, Robert. 2016. "Report: American Economic Association Committee on Statistics (AEAStat)." *American Economic Review*, 106(5): 788–793.

Nature Scientific Data. 2016. "Nature Scientific Data recommended repositories." *figshare*.

Open Science Framework. 2017. "Badges to Acknowledge Open Practices Wiki." *https://osf.io/tvyxz/wiki/home/*, Accessed: 2017-10-18.

Stodden, Victoria, Marcia McNutt, David H Bailey, Ewa Deelman, Yolanda Gil, Brooks Hanson, Michael A Heroux, John P A Ioannidis, and Michela Taufer. 2016. "Enhancing reproducibility for computational methods." *Science*, 354(6317): 1240–1241.

Wilson, Greg, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, and Tracy K Teal. 2016. "Good Enough Practices in Scientific Computing."

Reviewers