Formatting Data Citations - BibTeX Version

By LARS VILHUBER*

We illustrate how to create data citations with LATEXand BibTeX.

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. In consequence, there is a need to properly cite the digital inputs to our published output and to properly curate those inputs.

I. 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*)).

A. What to cite

In a nutshell, every dataset is to be cited. This is true for the main article as well as online appendices. In the past, use of data or code has been acknowledged in footnotes, and only rarely through bibliographic references. However, if the dataset is used, it should appear in the bibliography. The same is true for code reused from previous papers, or provided by authors.

B. How to cite

The AEA uses the Chicago style for citations and bibliographies (American Economic Association, 2018). However, the Chicago Style Manual (Chegg, 2018; Chicago Manual of Style Online, 2018) does not provide examples for data citations, and neither does the Citation Style Language[1] used by applications like Zotero[2] and Mendeley Desktop[3].

---

* Vilhuber: Cornell University, lars.vilhuber@cornell.edu.
[1]https://citationstyles.org/
[2]https://www.zotero.org/
[3]https://www.mendeley.com/download-desktop/

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

```
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,
1945--1968.''  Inter-University Consortium for Political and
Social Research, Ann Arbor, MI. ICPSR05404-v1.  doi:10.3886/ICPSR05404
(accessed February 8, 2011).
```

C.   Software

As part of our activities, the AEA prepress department has started the process of updating AEA templates available through such software.[4]

BibLaTeX. — Users of BibLaTeX should consult the companion document.

---

[4]For the technically inclined, this process involves updating an existing style or creating a new style on https://citationstyles.org/ and https://github.com/citation-style-language/styles, from where it propagates to a large number of software packages.

BIBTEX. — 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},
institution = {Murray Research Archive},
type = {},
number = {Version V2, UNF:5:obNHHq1gtV400a4T+Xrp9g==},
note = {\url{http://hdl.handle.net/1902.1/IOJHHXOOLZ}}
}
```

or

```
@techreport{leiss1999,
author = {Leiss, Amelia},
year = {1999},
title = {Arms Transfers to Developing Countries, 1945--1968},
institution = {Inter-University Consortium for Political and Social Research},
type = {},
number = {ICPSR05404-v1},
address = {Ann Arbor, MI},
note = {DOI:  10.3886/ICPSR05404 (accessed February 8, 2011).},
}
```

and thus generate "Duflo and Pande (2006*a*)" and "Leiss (1999*a*)" and the bibliographic entry in the References when using the pre-2018 `aea.bst` provided by the American Economic Association (AEA). Note the use of the note field to encapsulate the information. The `number` fields contains the key identifying information: version and (in this case), the UNF number generated by the Dataverse software.

A modification to the `aea.bst` file, tentatively named aea-mod.bst, might allow for a cleaner implementation:

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

and

```
@techreport{leiss1999-new,
        author = {Leiss, Amelia},
        year = {1999},
        title = {Arms Transfers to Developing Countries, 1945--1968},
        institution = {Inter-University Consortium for Political and Social Research},
        type = {},
        urldate = {February 8, 2011},
        number = {ICPSR05404-v1},
        address = {Ann Arbor, MI},
        doi = {10.3886/ICPSR05404},
}
```

which will generate "Duflo and Pande (2006*b*)" and "Leiss (1999*b*)". Note that we have not used the access date for either dataset, since both use persistent identifiers (handle or DOI).

REFERENCES

American Economic Association. 2018. "Sample References." (accessed on 2018-08-10).

Chegg. 2018. "Citation Machine: Chicago Manual Of Style 17th Edition (Author Date) format citation generator for journal article." Citation Machine. (accessed on 2018-10-02).

Chicago Manual of Style Online. 2018. "Author-Date: Sample Citations." <https://www.chicagomanualofstyle.org> (accessed 2018-10-02). (accessed on 2018-10-02).

DataONE. 2011*a*. "Data Citation and Attribution." (accessed on 2018-08-10).

DataONE. 2011*b*. "DataONE Tutorial on Data Citation." DataONE. http://www.dataone.org/sites/all/documents/L09_DataCitation.pptx (accessed 2018-05-22). (accessed on 2018-08-10).

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

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

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

Leiss, Amelia. 1999*a*. "Arms Transfers to Developing Countries, 1945–1968." Inter-University Consortium for Political and Social Research ICPSR05404-v1. DOI: 10.3886/ICPSR05404 (accessed February 8, 2011).

Leiss, Amelia. 1999*b*. "Arms Transfers to Developing Countries, 1945–1968." Inter-University Consortium for Political and Social Research ICPSR05404-v1, https://doi.org/10.3886/ICPSR05404 .