

# Creating Data Citation Templates for Economics

---

Courtney Butler

July 2018

TB 18-04

<https://doi.org/10.18651/TB/TB1804>

FEDERAL RESERVE BANK *of* KANSAS CITY



# Creating Data Citation Templates for Economics

By Courtney Butler<sup>1</sup>

## **Abstract**

Copy and paste citation tools exist for traditional academic publication types from places like Citation Machine or Google Scholar, but similar plugins for datasets are scarce. In response, the Federal Reserve Bank of Kansas City built a data citation template for all acquired and select open data sources to allow economists to copy and paste data citations into their preferred word processing program.

---

<sup>1</sup> Data Curator, Courtney.Butler@kc.frb.org, Research Library, Center for the Advancement of Data and Research in Economics (CADRE) at the Federal Reserve Bank of Kansas City.

The scholarly community has put a number of mechanisms in place that support the attribution process. For example, APA, Chicago, and many other commonly used citation styles provide formatting guidance on a wide variety of source types, including books, periodicals, websites, reports, interviews, podcasts, and more (American Psychological Association, 2017; University of Chicago, 2017). This guidance dictates not only what information should be included in a citation but also the order information should be in and how it should be punctuated. This consistency allows readers familiar with the style to instantly recognize the information they are viewing and utilize it accordingly. However, familiarity is not equivalent to expertise. Lack of practice, intermittent use, and special circumstances can pose challenges to scholars attempting to apply standards in their works. Luckily, online tools have been developed that help simplify the formatting process for source types with accepted formatting standards. These tools often allow citation information to be quickly and easily copied and pasted directly into a document or reference library, as can be seen in Figure 1:

**Figure 1: Copy and Paste Citation Tools**

The figure illustrates three online tools for generating citations. The top-left tool, 'Cite', offers various citation styles (MLA, APA, Chicago, Harvard, Vancouver) and a 'BibTeX' option. The top-right tool, 'Citation Machine', provides a user interface for selecting a citation style (APA, MLA, Chicago) and displaying a formatted citation. The bottom-right tool, 'DOI Citation Formatter', allows users to input a DOI and generate a citation in a specific format.

**Cite Tool Output:**

MLA Basu, Susanto, and Brent Bundick. "Uncertainty shocks in a model of effective demand." *Econometrica* 85.3 (2017): 937-958.

APA Basu, S., & Bundick, B. (2017). Uncertainty shocks in a model of effective demand. *Econometrica*, 85(3), 937-958.

Chicago Basu, Susanto, and Brent Bundick. "Uncertainty shocks in a model of effective demand." *Econometrica* 85, no. 3 (2017): 937-958.

Harvard Basu, S. and Bundick, B., 2017. Uncertainty shocks in a model of effective demand. *Econometrica*, 85(3), pp.937-958.

Vancouver Basu S, Bundick B. Uncertainty shocks in a model of effective demand. *Econometrica*. 2017 May 1;85(3):937-58.

**BibTeX Citation Template:**

```
@article{basu2017uncertainty,
  title={Uncertainty shocks in a model of effective demand},
  author={Basu, Susanto and Bundick, Brent},
  journal={Econometrica},
  volume={85},
  number={3},
  pages={937--958},
  year={2017},
  publisher={Wiley Online Library}
}
```

**Citation Machine Output:**

Basu, Susanto, and Brent Bundick. "Uncertainty shocks in a model of effective demand." *Econometrica* 85, no. 3 (2017): 937-58.

**DOI Citation Formatter Output:**

DOI: 10.3886/ICPSR03547.v1

Format: chicago-author-date

Language and Country: en-US

Format Button: Format

Copy to clipboard Button: Copy to clipboard

Properly citing data offers the same benefits as properly citing traditional sources. Data citation allows readers to identify, access, and verify sources; provides appropriate attribution to data owners and providers; and facilitates collaboration and reuse, which ultimately leads to faster and more efficient research progress (DataCite, 2017; FORCE 11, 2014). Unfortunately, citing data is not an accepted or routine practice in the same way as it is for more traditional sources, and, at most, attribution is often included informally as a note below a graph or figure. While there are many reasons for this, including a broader discussion on data sharing in general, one major issue is the lack of universally accepted citation standards (Park & Wolfram, 2017). Other non-traditional source types such as musical recordings and YouTube videos receive acknowledgment in many styles, but formatting standards for data – which often serves as the basis for numerous scholarly outputs – are rare.

Organizations across multiple disciplines have released guidance on the topic, but the guidance primarily focuses on the elements that should be included in a data citation rather than specific formatting (CODATA, 2017; DataCite, 2017; FORCE 11, 2014; ICPSR, 2017; National Research Council, 2012; Rauber, Asmi, van Uytvanck, & Proll, 2015; Strasser, 2015). Though this is a start, the lack of translation into common styles and thus consistency in practice undermines the ability of readers to easily recognize information elements as they pertain to data. It also presents challenges in integrating with online tools that automate formatting.

For the most part, online citation services do not support data citation at all, and those that do support data citation have significant limitations. For instance, the tool offered by a collaboration between DataCite, Crossref, mEDRA, and ISTIC requires the data to have a DOI. Datasets with DOIs are becoming more common, but persistent identifiers are still not entirely ubiquitous and DOI is not the only persistent identifier system (Klump & Huber, 2017). Citation Machine has a ‘raw data’ category, but it only offers Description, Contributor(s), and Publication Date/Location

fields (Citation Machine, 2018a; Citation Machine, 2018b; Citation Machine, 2018c). Users can add annotations at the end of the citation to provide more information, which may include a third-party distributor, access point, or unique identifier, but the lack of a dedicated field does not communicate to the user that this is important information to include.

In addition to citation-generating websites, many authors also use reference management resources, such as EndNote, RefWorks, and Zotero, to format and manage citations. Some of these tools allow authors to select ‘data’ as a type, but if the preferred citation style does not offer data formatting guidance then the information is often inserted into a similar template, such as ‘electronic resource’, instead. For example, BibTeX is a reference management tool that provides formatting assistance and allows authors to insert citations into documents written specifically in LaTeX, a word processing mark-up language widely used in economics and other academic disciplines. Authors simply provide information for specific elements according to the item type template, as can be seen for the @article type in Figure 1 and Figure 2, and the elements are formatted into the author’s selected style within the document using a simple command (Wikibooks, 2017). Though data citation can be accomplished by repurposing existing templates (e.g., @misc) or using a non-standard version (e.g., @electronic), there is currently no official @data template. These types of limitations diminish consistency and ease of use, which ultimately hinders the uptake of data citation practice.

**Figure 2: BibTeX Entry**

```
@article{blackholes,  
  author = "Rabbert Klein",  
  title = "Black Holes and Their Relation to Hiding Eggs",  
  journal = "Theoretical Easter Physics",  
  publisher = "Eggs Ltd.",  
  year = "2010",  
  note = "(to appear)"  
}
```

Despite these hurdles, data citation is becoming more common. Academic journals from several publishers now recommend that submitted papers include data citations in their references section, and publications from the Federal Reserve Bank of Kansas City's Research Division are also beginning to have formal data citation in its publications to align with industry practice (Elsevier, 2017; Springer Nature, 2017). Economists are accustomed to having a wide variety of tools and resources to assist with citation requirements, and thus the Federal Reserve Bank of Kansas City's Center for the Advancement of Data and Research in Economics developed citation templates for all acquired proprietary and select open data sources for authors to copy and paste into libraries and documents.

We began by drafting a general template based on modified Chicago in order to be consistent with in-house publication style. We then reviewed all active data contracts and the online terms of use for popular freely available resources. Specific data citation guidelines were identified when available, and data providers were contacted when supplied guidance was unclear. Difficulties arose when provider requirements did not fit well in the local template. For instance, some providers required that a specific and sometimes lengthy disclaimer appear as part of the official attribution. It was not immediately clear where these types of citations should appear – whether in the references section, beneath a chart, or as a footnote within the main text – in order to fit consistently with house style while also still meeting the obligations of the agreement. All such decisions were made in close collaboration with the Research Division Senior Editor, and the completed template is included in Figure 3. Examples can be found in the Appendix.

**Figure 3: Data Citation Template**

Author or Owner. Publication Year\*. “Supplement or Subsection.” *Dataset*, Date range. Accessed through Third-Party Publisher or Distributor. Available at URL or DOI.

\*If the publication date is not clear or available, use “Accessed [Year]” instead

Once the provider issues were addressed, the template was translated into both LaTeX scripting and plain text for use with Microsoft Word. Citation information for select resources was filled in where possible, and the resulting resources were placed on an intranet site from which authors could copy and paste. Though components such as Author/Owner were pre-filled, there were still elements that had to be filled in by the researchers based on the specific cut of data being used for analysis. This led to numerous questions on points we originally thought clear and straightforward. For example, does Publication Date refer to the date the entire dataset first came into existence, the date the specific subset of data used for analysis was originally made available, the date the data was made available by a third-party aggregator, or something different altogether? And should the Publication Date include months or days to match the units in the data or just years? Also, should the Dataset field use official, legal names, or should they instead use colloquial titles more familiar to other researchers in the field? These and other questions made what originally seemed like a simple plug-and-play task turn into something more complicated than anticipated, and the template will continue to be iterative as questions are raised and addressed.

Though its development presented a number of unexpected challenges, having copy and paste data citations compiled in a centralized location along with a (relatively) easy-to-follow template has significantly streamlined the data citation process for Federal Reserve Bank of Kansas City publications. The Data Services Librarian currently works very closely with authors to develop and approve data citations, and this process will continue to improve as we refine citation guidance and increase efforts to raise awareness of these resources. Future directions include integrating the

citations into a BibTeX template or similar reference management tool and creating copy and paste citations for other non-traditional items such as software and code.

Our locally developed templates do not address the lack of standardization across the industry, but they can help our researchers become more used to citing data and related resources as they would any other scholarly source. We hope the proliferation of data citation as a practice will encourage industry movement in this direction, and all template components are subject to change as standards become more normalized.

## References

- American Psychological Association. (2017). *Table of Contents: Reference Examples*. Retrieved from Publication Manual of the American Psychological Association, Sixth Edition: <http://www.apastyle.org/manual/table-of-contents.pdf>
- Citation Machine. (2018a). *Generate American Psychological Association 6th edition citations for Raw Data*. Retrieved from Citation Machine: A Chegg Service: <http://www.citationmachine.net/apa/cite-a-rawdata/manual>
- Citation Machine. (2018b). *Generate Chicago Manual of Style 16th edition (full note) citations for Raw Data*. Retrieved from Citation Machine: A Chegg Service: <http://www.citationmachine.net/chicago/cite-a-rawdata/manual>
- Citation Machine. (2018c). *Generate Modern Language Association 8th edition citations for Raw Data*. Retrieved from Citation Machine: A Chegg Service: <http://www.citationmachine.net/mla/cite-a-rawdata/manual>
- CODATA. (2017). *International Series of Data Citation Workshops*. Retrieved from CODATA-ICSTI Data Citation Standards and Practices: <http://www.codata.org/task-groups/data-citation-standards-and-practices/international-series-of-data-citation-workshops>
- DataCite. (2017). *Cite Your Data*. Retrieved from <https://www.datacite.org/cite-your-data.html>
- DataCite, Crossref, mEDRA, & ISTIC. (n.d.). *DOI Citation Formatter*. Retrieved from <https://citation.crosscite.org/>
- Elsevier. (2017). *Data Citation*. Retrieved from <https://www.elsevier.com/about/open-science/research-data/data-citation>
- FORCE 11. (2014). *Joint Declaration of Data Citation Principles*. Data Citation Synthesis Group, Marone M. (ed.), San Diego, CA. Retrieved from <https://doi.org/10.25490/a97f-egykh>

ICPSR. (2017). *Citing Data*. Retrieved from

<https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/citations.html>

Klump, J., & Huber, R. (2017). 20 Years of Persistent Identifiers – Which Systems are Here to Stay?

*Data Science Journal*, 16, 9. Retrieved from <http://doi.org/10.5334/dsj-2017-009>

National Research Council. (2012). *For Attribution -- Developing Data Attribution and Citation*

*Practices and Standards: Summary of an International Workshop*. Washington, DC: The

National Academies Press. Retrieved from <https://doi.org/10.17226/13564>

Park, H., & Wolfram, D. (2017). An examination of research data sharing and re-use: implications

for data citation practice. *Scientometrics*, 111(1), 443-461. Retrieved from

<https://doi.org/10.1007/s11192-017-2240-2>

Princeton University. (2017). *Academic Integrity at Princeton*. Office of Communications. Office of the

Dean of the College. Retrieved from

<https://www.princeton.edu/pr/pub/integrity/pages/academic-integrity-2017.pdf>

Rauber, A., Asmi, A., van Uytvanck, D., & Proll, S. (2015). *Data Citation of Evolving Data:*

*Recommendations of the Working Group on Data Citation (WGDC)*. Research Data Alliance.

Retrieved from <http://dx.doi.org/10.15497/RDA00016>

Springer Nature. (2017). *Research Data Policies and Services*. Retrieved from

<http://www.springernature.com/gp/authors/research-data-policy/faqs/12327154>

Strasser, C. (2015). *Research Data Management: A Primer Publication of the National Information Standards*

*Organization*. Baltimore, MD: National Information Standards Organization. Retrieved from

<http://www.niso.org/publications/research-data-management-primer>

Terrebonne Parish Library. (n.d.). *Lesson Plan 8: Why and When to Cite Sources*. Retrieved from

<http://mytpl.org/wp-content/uploads/TFT-LP-8-When-and-Why-to-Cite-Sources1.pdf>

University of Chicago. (2017). *Chapter 14: Notes and Bibliography*. Retrieved from The Chicago Manual of Style Online: Contents:

<http://www.chicagomanualofstyle.org/book/ed17/part3/ch14/toc.html>

Wikibooks. (2017). *Bibliography Management*. Retrieved from LaTeX:

[https://en.wikibooks.org/wiki/LaTeX/Bibliography\\_Management](https://en.wikibooks.org/wiki/LaTeX/Bibliography_Management)

## Appendix

### Citation Style Guidance

#### Citing data in work products

Data are cited for the same reasons that books, articles, and reports are cited - to acknowledge the original creator or author and help other researchers find the data.

Data citations have many of the same components as other citations, including author or creator and title, along with other information when relevant or available, such as URL access information or a digital object identifier (DOI), publisher/access provider, year published, edition, and so on.

#### Tips for citing data

- Provide relevant information so that your reader will be able to find the data.
  - Include citations in a source note below tables and figures; for text, include a parenthetical citation.
  - Include data sources in your reference list, just as you would list books and articles.
    - The Research Library strongly recommends using a permanent URL or DOI to point readers to the data source.
- Check license agreements and terms of use for specific requirements.
  - For all licensed (and some public) data, contact the Data Services Librarian. In some cases, datasets are acquired from an access provider or publisher. Authors must acknowledge the provider. For example, CBOE Volatility Index data accessed via FRED must include the note "Accessed through the Federal Reserve Bank of St. Louis" in the citation.

If you use data from a public website, be sure to check the terms of use on the website.

<b>Citation Style Guidance*</b>		
<b>Type</b>	<b>Template</b>	<b>Example(s)</b>
In-Text	(Author/Owner) OR (Author/Owner : Distributor) If there's no author, the parenthetical reference uses the first few words of the citation on the References page.	(Butler) (U.S. Census Bureau) (International Monetary Fund : Federal Reserve Bank of St. Louis)
Figures, Tables, Charts, Graphs	Sources: Author/Owner1, Author/Owner2 (Distributor), Author/Owner3	Sources: Butler, International Monetary Fund (Federal Reserve Bank of St. Louis), and U.S. Census Bureau.
References	Author or Owner. Publication Year*. "Supplement or Subsection." <i>Dataset</i> , Date range. Accessed through Publisher or Distributor. Available at URL or DOI  *If the publication date is not clear or available, use "Accessed [Year]" instead	U.S. Census Bureau and U.S. Bureau of Labor Statistics. 2017. <i>Current Population Survey (Household Survey): Civilian Employment-Population Ratio</i> , January 2000 - December 2016. Accessed through the Federal Reserve Bank of St. Louis. Available at notarealurl.com  U.S. Census Bureau and U.S. Bureau of Labor Statistics. Accessed 2017. "Annual Social and Economic Supplement." <i>Current Population Survey</i> , March 2014. Available at doi.org/notarealDOI