Data Citing... or Data Sighting?

APPENDIX A: DATA DESCRIPTION

Monthly world oil production data measured in thousands of barrels of oil per day were obtained from the US Energy Information Administration’s (EIA) *Monthly Energy Review* starting in January 1973. Monthly data for global production of crude oil for the period 1953:4 to 1972:12 were taken from the weekly *Oil & Gas Journal* (issue of the first week of each month). For the period 1947:1 to 1953:3, monthly data were constructed by interpolation of yearly world oil production data by means of the Litterman (1983) methodology using US monthly oil production data from the EIA as an indicator variable. Annual oil production data were obtained from *World Petroleum* (1947–1954), the *Oil & Gas Journal* (end-of-year issues, 1954–1960), and the EIA’s *Annual Energy Review* (1960–2010). Consistency between these different data sources was checked at each of the overlapping periods. Quarterly data are averages of monthly observations.

The nominal US refiners’ acquisition cost of imported crude oil was taken from the *Monthly Energy Review*. Since this series is only available from January 1974 onward, it was backcast until 1947:1 with the quarterly growth rate of the producer price index (PPI) for crude oil retrieved from the Bureau of Labor Statistics (BLS) database (WPU0561). Data were converted to quarterly frequency before backcasting by averaging over months. For the robustness checks with regard to the choice of the oil price variable, we use the quarterly average of the West Texas Intermediate (WTI) spot oil price obtained from the Federal Reserve Economic Data (FRED) database maintained by the St. Louis FED (OILPRICE) and of the nominal US refiners acquisition cost of composite crude oil from the *Monthly Energy Review*. The latter was adjusted for price controls on domestic oil production for the period 1971:III to 1974:1 as described in Mork (1989) and reconstructed backward to 1947:1 in the same way as the imported refiners’ acquisition cost series.

Quarterly seasonally adjusted series for US real GDP (GDPC96: real gross domestic product, billions of chained 2005 dollars) and for the US GDP deflator

lags (DP), data on is mation asly th to stent even nominal even nominal
Citations in Academic Research

Research Citations: “Coin of the academic realm” (Kratz (2015))

- H-Index; Web of Science Citation Index; Google Scholar

Does data count in citation analysis?

- The good news:
  - Research that cites data, gets cited more (Piwowar, Day & Fridsma (2007))
- The bad news: Low compliance with recommendations
  - Social science papers cite data 17% of the time (Mooney (2011))
  - My unscientific survey of AEA journals – 23% list data in references
“Data has been the ‘dark matter’ of the scholarly ecosystem — data citation aims to make the role of data visible.”

Altman (2016)

Data citation “must echo the role that traditional journal citation has played in ensuring longevity of the scholarly record, acting as a bridge to permanent access and enabling reward systems.”

Rans, et al. (2013)

“On the surface, citing datasets is a trivially easy thing to do.... The process of making datasets citable, however, is rather more difficult in consequence of this and other factors. A culture of citing datasets has been slow to develop.”

Ball and Duke (2011)
AEA recommendations for data citation

- American Economics Association (AEA)
  - https://www.aeaweb.org/journals/aer/submissions/accepted-articles/styleguide
  - Sample references: https://www.aeaweb.org/journals/policies/sample-references

---

J. Reference list:

Use full names of authors and/or editors. List all authors/editors up to/including ten names. Authors of articles and books and materials without specific authors or editors, such as government documents or bulletins, are to be listed alphabetically. Please follow the Chicago Manual of Style’s “Author-Date” style, LaTeX and Scientific Word users, please use the aea.lst file.

- We encourage you to use bibliographic software when preparing your reference list. If you are using software please select “Chicago Author-Date” when using bibliographic software.
- Not all document types are covered in reference list software. We include helpful guidelines for several document types at our sample references page.
- Data References: References to datasets should be included in your reference list. You can find examples of how to cite datasets here: View Sample References.
Key players in data citation standards

• DataCite – https://www.datacite.org/
  • Metadata schema (updated Sept 2016): http://schema.datacite.org/
  • Useful for: a data citation schema you can implement.
  • Its schema is used by ICPSR, Dataverse, etc.
  • Handy data citation formatter! https://www.datacite.org/citation.html

• CoDATA http://www.codata.org/ and Force11 https://www.force11.org/about
  • Joint Declaration of Data Citation Principles: https://www.force11.org/group/joint-declaration-data-citation-principles-final
  • Data Citation Primer: http://force11.github.io/data-citation-primer/authors/
  • Useful for: theory and major principles behind data citation.

• IASSIST: http://www.iassistdata.org/topic/other-topics/data-citation
  • Handy quick guide to data citation: http://iassistdata.org/sites/default/files/quick_guide_to_data_citation_high-res_printer-ready.pdf

• ICPSR: https://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp

• California Digital Library: https://datapub.cdlib.org/datacitation/
# Elements of Data Citation

## CORE

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author/creator</td>
<td>the person(s), organization, agency responsible for creating the dataset</td>
</tr>
<tr>
<td>Year of Publication</td>
<td>the year the dataset was made available (not the date coverage of the data)</td>
</tr>
<tr>
<td>Title or description</td>
<td>complete title (if no title exists, create a brief description of the data)</td>
</tr>
<tr>
<td>Publisher</td>
<td>entity (organization, database, archive, journal) that hosts/provides/licenses the data</td>
</tr>
<tr>
<td>DOI/URI</td>
<td>the unique identifier if the data set is online (Not so easy for some data sources, such as vendor data)</td>
</tr>
</tbody>
</table>

## ADDITIONAL

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location/Availability</td>
<td></td>
</tr>
<tr>
<td>Version/Edition</td>
<td></td>
</tr>
<tr>
<td>Access date</td>
<td>Crucial for reproducing analysis of continuously updated dynamic data Also useful for vendor/commercial data, as vendor names change.</td>
</tr>
</tbody>
</table>
### What can you do:
Adapt a Data Citation Standard

- **Example at the Board** – modified DataCite, Chicago Manual of Style (16th Ed.), and local editing practices.

<table>
<thead>
<tr>
<th>Citation Style</th>
<th>Citation for Bibliography/References</th>
</tr>
</thead>
<tbody>
<tr>
<td>DataCite</td>
<td><img src="https://example.com/datacite_image.png" alt="DataCite Image" /></td>
</tr>
</tbody>
</table>

**2.2 Citation**

Because many users of this schema are members of a variety of academic disciplines, DataCite remains discipline-agnostic concerning matters pertaining to academic discipline sheet requirements. Therefore, DataCite encourages rather than requires a particular citation format. In keeping with this approach, the following is the preferred format for rendering a DataCite citation for human readers using the first five properties of the schema:

Creator (PublicationYear): Title. Publisher. Identifier

It may also be desirable to include information from two optional properties, Version and ResourceType (as appropriate). If so, the preferred form is as follows:

Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier

For citation purposes, DataCite prefers that DOI names are displayed as linkable, permanent URLs. The Identifier may appear in its original format. If the original format is chosen, be sure to include the characters “doi:” pre-pended to the Identifier as in “doi:10.1234/abc.”

For resources that do not have a standard publication year value, DataCite suggests that PublicationYear should include the date that is preferred for use in a citation.

Here are several examples:

- **Irino, T; Tada, R (2009):** Chemical and mineral compositions of sediments from ODP Site 1277-797, V. 2.1. Geological Institute, University of Tokyo. [http://doi.org/10.1594/PANGAEA.726855](http://doi.org/10.1594/PANGAEA.726855)

**Note:** Information extracted from a merged database.
16. wrds-

What can you do: Adapt a Data Citation Standard

For the confidential PSID data:

For the CoreLogic data:
What can you do:
Make Data Citation Easy.

Make data citation as easy as possible
- Provide recommended citations
- Export functions to EndNote, RIS, etc. (ICPSR, publishers)
- Citation formatter (DataCite/CrossRef)
- LaTeX code for easy integration into research and work
What can you do: Normalize data citation (and data deposit!)

At your organization
- Educate about the value of citation
- Provide tools to make data citation happen
- Example at the Fed: Data review team at the Board

In scholarly communication and publishing
- If possible, get involved in the publishing of work products at your workplace
  - Example at the Fed: proposed transition of FEDS Working Papers from html to a repository with a linked data collection
The U.S. Treasury Yield Curve: 1961 to the Present

Refet S. Gürkaynak, Brian Sack, and Jonathan H. Wright
2006-28

Abstract: The discount function, which determines the value of all is the most basic building block of finance and is usually inferred from the yield curve. It is therefore surprising that researchers and practitioners do not have a long history of high-frequency yield curve estimates. This paper provides the Treasury yield curve estimates of the Federal Reserve Board from 1961 to the present. We use a well-known and simple model to fit the data very well. The resulting estimates can be used to provide yields for any horizon. We hope that the data, which are now available at http://www.federalreserve.gov/pubs/feds/2006 and which will be updated, will provide a benchmark yield curve that will be useful to applied economists.

Keywords: Yield curve, forward rates, on-the-run premium, treasury

Full paper (2532 KB PDF) | Full paper (Screen Reader Version) | Data - Excel file (30 MB XLS) | Data - Screen reader | Data - XML (sdmx/zip)
Incentivize Data Citation.

“Academic researchers as a class are drawn to research and scholarship through an interest in puzzle-solving, but they are also substantially incented by recognition and money.” (Altman 2016)

• Make data citation valuable to your users and organization
• Value may differ for: economists; research assistants; data managers; librarians; budget team.
Thanks!

Any questions or comments?

Alison Raab Labonte, alison.r.labonte@frb.gov
Bibliography:

- American Economic Association, “AEA recommendations on Citation,) https://www.aeaweb.org/journals/aer/submissions/accepted-articles/styleguide
- Ball, Alex and Monica Duke (2011), “How to Cite Datasets and Link to Publications,” (July 19), http://opus.bath.ac.uk/28006/1/idccws6-ab-slides.pdf
- DataCite. Data Citation Schema, last updated September 2016), http://schema.datacite.org
- IASSIST, Data Citation Quick Guide. http://www.icpsr.umich.edu/files/ICPSR/eneletters/iassist.html
- Kunze, John, Robert Cook (2009), “Defining the Data Citation Problem in the DataNet Context.” DataONE Presentation (December 2009), openwetware.org/images/3/37/Datanet_citation_agu_jak.ppt
Bibliography (continued)

- Michigan State University (MSU), Data Citation Guide, http://libguides.lib.msu.edu/citedata
- Pienta AM, GC Alter GC, and JA Lyle (2010), “The enduring value of social science research: the use and reuse of primary research data,” https://deepblue.lib.umich.edu/handle/2027.42/78307

Image credits: