



From Sighting Data to Citing Data

Alison Raab Labonte

Data Librarian, Research Library

Board of Governors of the Federal Reserve System

7 October 2016

Beyond the Numbers Conference, Federal Reserve Bank of St. Louis

Data Citing... or Data Sighting?

Bookmarks

- Shocks
 - C. Sensitivity Analysis
- III. Reasons for the Steepening of the Oil Demand Curve
- IV. Conclusions
- Appendix A: Data Description**
- Appendix B: Bayesian Estimation of a VAR with Time-Varying
 - A. Prior Distributions and Starting Values
 - B. Markov Chain Monte Carlo Algorithm for Simulating
- Appendix C: Generalized Impulse Responses and Sign
- REFERENCES

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.^[10] 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*.^[11] Since this series is only available from January 1974 onward, it was backcast until 1947:I 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^[12] 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:I as described in Mork (1989) and reconstructed backward to 1947:I 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
nea-
tion
asily
th to
stent

even
nial

reasur-
period
head-
sures.
sed in
f their

195

Citations in Academic Research

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

- H-Index; Web of Science Citation Index; Google Scholar
- NISO, Journal Article Markup Suite (JATS) ANSI/NISO Z39.96-2015

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

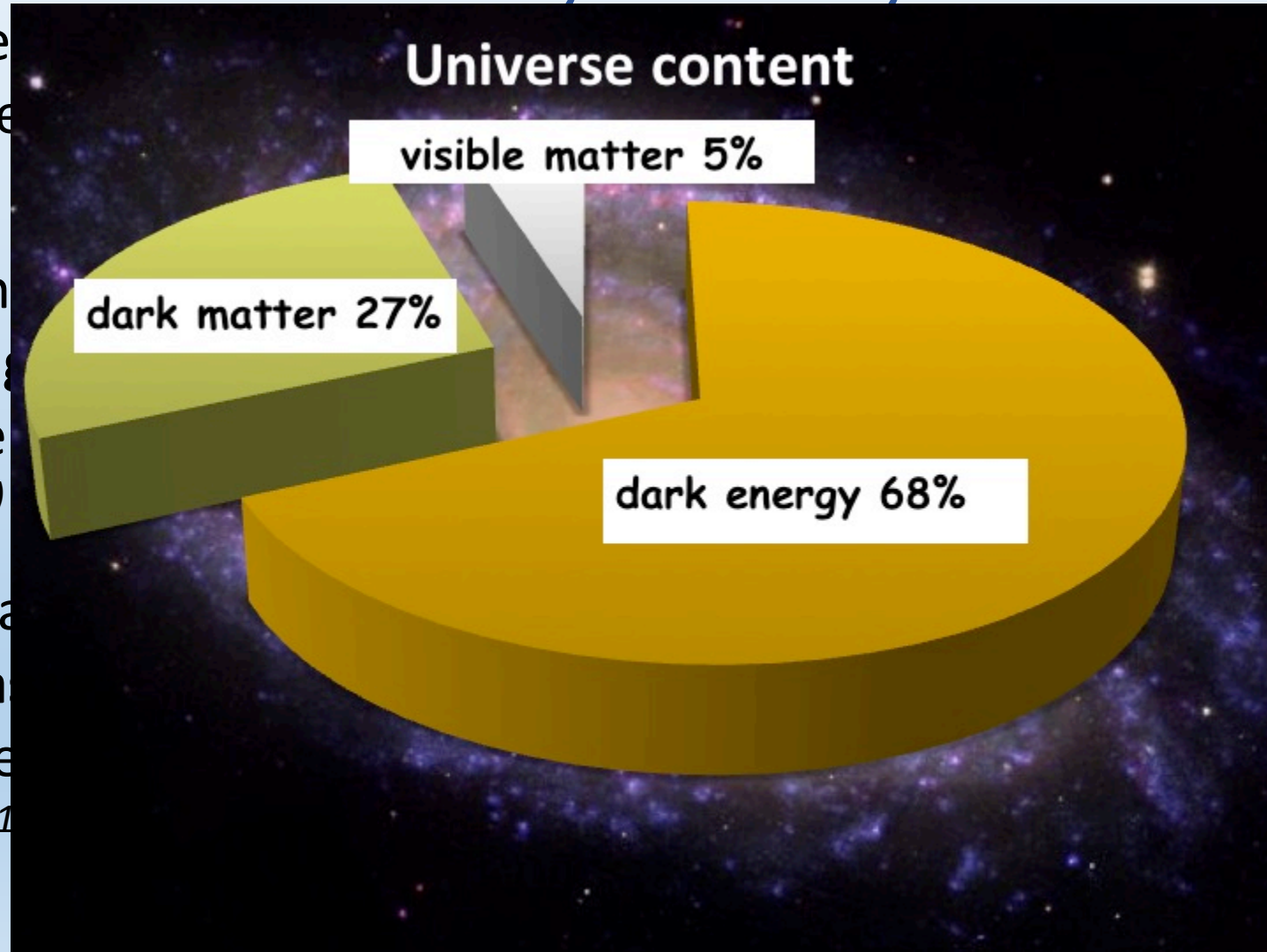


Data as the “Dark Matter” of the Scholarly Ecosystem

“Data has been the key to making research more accessible and aims to make research more accessible.”
Altman (2016)

Data citation is essential for ensuring long-term access and ensuring the integrity of research.
Rans, et al. (2013)

“On the surface, making data available seems like a good thing, but this and other factors have led to the development of a new process of research.”
Ball and Duke (2011)



data citation

has played in permanent

the process of consequence of to develop.”

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.bst 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](#).

<https://www.aeaweb.org/journals/policies/sample-references>

AMERICAN ECONOMIC ASSOCIATION

Membership About AEA Journals Annual Meeting Careers Resources EconLit

Home > Journals > AEA Journals: Policies > Sample References

Journals

American Economic Review
AEJ: Applied Economics
AEJ: Economic Policy
AEJ: Macroeconomics
AEJ: Microeconomics
Journal of Economic Literature
Journal of Economic Perspectives
Back Issues
Institutional Subscriptions
Get Journal Alerts
Policies, Copyright, and Permissions
Advertise in AEA Journals
JSTOR access for AEA members
Athens Subscriber Login

Sample References

Styles of the AEA

We use the *Chicago Manual of Style Author-Date system* for all common publication types.

The following examples are intended to provide information for less common sources.

Datasets

When referencing datasets, please include the author name or name of the provider hosting the data, the year the data were collected or posted, the name or title of the dataset, the name of the database if applicable, and any other information necessary for one to retrieve the data. Please include the date accessed in parentheses at the end.

For data references specifically associated with a published paper, please include the Author Name(s). Year. "Paper Title: Dataset." *Journal Name*. Location of the data.

Examples

- 1 **Bureau of Labor Statistics.** 2000–2010. "Current Employment Statistics: Colorado, Total Nonfarm, Seasonally adjusted - SMS08000000000000000001." United States Department of Labor. <http://data.bls.gov/cgi-bin/surveymost?sm+08> (accessed February 9, 2011).
- 2 **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).
- 3 **Romer, Christina D., and David H. Romer.** 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks: Dataset." *American Economic Review*. <http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.3.763> (accessed August 22, 2012).

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

ADDITIONAL	
Location/Availability	
Version/Edition	
Access date	Crucial for reproducing analysis of continuously updated dynamic data Also useful for vendor/commercial data, as vendor names change.

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.

Citation for Bibliography/References		
Citation Style	https://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_v4.0.pdf	
		at Merged Database. 16, wrds-
2.2 Citation		
Creator [(Year published)]. <i>Title Data Set</i> . [Publisher], [accessed month/year] [, DOI or URL].	<p>Because many users of this schema are members of a variety of academic disciplines, DataCite remains discipline-agnostic concerning matters pertaining to academic style sheet requirements. Therefore, DataCite encourages rather than requires a particular citation format⁷. 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:</p> <p>Creator (PublicationYear): Title. Publisher. Identifier</p> <p>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:</p> <p>Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier</p> <p>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."</p> <p>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.</p> <p>Here are several examples:</p> <ul style="list-style-type: none">Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. V. 2.1. Geological Institute, University of Tokyo. http://doi.org/10.1594/PANGAEA.726855	<p>n Officer Opinion Survey ocs/snloansurvey. ov/public. Analytics, accessed May 3CD) (2014). "Contribution to loi.org/10.1787/0bb009ec-en. udy of Income Dynamics,</p>
Note: Informatic		

What can you do: Adapt a Data Citation Standard

Creator [(Year published)]. *Title of Data Set*. [Publisher] [, accessed month/year] [, DOI or URL].

For the confidential PSID data:

Panel Study of Income Dynamics (2014). Restricted use dataset. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI., accessed May 2015. <http://psidonline.isr.umich.edu/Guide/FAQ.aspx?ID=1040>.

For the CoreLogic data:

Corelogic, Inc. (2015). Real estate data, 2000-2015. Accessed May 2016. corelogic.com

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

The screenshot shows the DOI Citation Formatter beta website. The main heading is "DOI Citation Formatter beta". Below the heading are logos for DataCite, crossref, MEDRA, and Chinese DOI. The website has a search bar and navigation links (About, Guides, Support, Sign Up, Log In). The main content area features a DOI input field with the value "10.3886/ICPSR33802.v1", a style dropdown menu set to "apa", and a locale dropdown menu set to "en-US". A "Format" button is next to the locale dropdown. Below the input fields, a sample citation is displayed: "United States Department of Commerce. Bureau of the Census. (2013). American Community Survey (ACS): Public Use Microdata Sample (PUMS), 2009. ICPSR - Interuniversity Consortium for Political and Social Research. http://doi.org/10.3886/ICPSR33802.v1". At the bottom, there is a "Download" button and a "Text output for tables" button.

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

- NISO, Journal Article Markup Suite (JATS), version 1.1, ANSI/NISO Z39.96-2015) (Standardized Markup for Journal Articles) <http://jats.niso.org/1.1/>
 - Lapeyre, Deborah Aleyne (2015), “Citing Data in Journal Articles Using JATS,” FORCE11.
- 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

http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html Portal Ses



The Federal Reserve Board

Finance and Economics Discussion Series



The U.S. Treasury Yield Curve: 1961 to the present

Refet S. Gurkaynak, 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 makes public the Treasury yield curve estimates of the Federal Reserve Board from 1961 to the present. We use a well-known and simple method to compute the yield curve estimates. The resulting estimates can be used to compute forward rates for any horizon. We hope that the data, which are posted on the website <http://www.federalreserve.gov/pubs/feds/2006> and which will be updated periodically, 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 \(sdmx/zip\)](#)

[Home](#) | [Economic Research and Data](#) | [FEDS](#) | [List of 2006 FEDS papers](#) | [Accessibility](#) | [Contact us](#)

Last update: December 22, 2011

Increasing access to Board research and data

[Home](#) | [Browse All](#) | [Board Subject Taxonomy](#) | [Board Expert Finder](#) | [JEL codes](#) | [Board Research Library home](#) | [Ask-a-Librarian](#)

[Search](#) [Advanced Search](#)

[Home](#) > [FEDS Working Papers](#) > [The U.S. treasury yield curve : 1961 to the present / Refet S. Gurkaynak, Brian Sack, and Jonathan H. Wright](#)

[Reference URL](#) | [Add tags](#) | [Comment](#) | [Save to favorites](#)

The U.S. treasury yield curve : 1961 to the present / Refet S. Gurkaynak, Brian Sack, and Jonathan H. Wright.

[View Description](#)

[PDF](#) | [Text](#)

160.pdf
Link will provide options to open or save document.
File Format: Adobe Reader

XLS
Link will provide options to open or save document.
File Format: Microsoft Excel - xls

3.xls [97.83 KB]
Link will provide options to open or save document.
File Format: Microsoft Excel - xls

Data from Gurkaynak, Sack and Wright 2006-28 FEDS paper (Excel file)

[View Description](#)

Object Description

Author	Gurkaynak, Refet S. Sack, Brian Wright, Jonathan H.
Title	The U.S. treasury yield curve : 1961 to the present
Date	2006
Series	Finance and Economics Discussion Series
Board Taxonomy	Yield curve U.S. Treasuries
Author's keywords	Yield curve ; Forward rates ; On-the-run premium
Pagination	31 p.
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 makes public the Treasury yield curve estimates of the Federal Reserve Board from 1961 to the present. We use a well-known and simple method to compute the yield curve estimates. The resulting estimates can be used to compute forward rates for any horizon. We hope that the data, which are posted on the website http://www.federalreserve.gov/pubs/feds/2006 and which will be updated periodically, will provide a benchmark yield curve that will be useful to applied economists.
Publisher	Board of Governors of the Federal Reserve System (U.S.)
Cite as	Gurkaynak, Refet S., Brian Sack, and Jonathan H. Wright (2006). "The U.S. Treasury Yield Curve: 1961 to the Present," Finance and Economics Discussion Series 2006-28. Board of Governors of the Federal Reserve System, Washington, D.C.
Supporting data	(Excel): http://cdm16798.contentdm.oclc.org/cdm/ref/collection/p16798coll6/id/2 (XML): http://cdm16798.contentdm.oclc.org/cdm/ref/collection/p16798coll6/id/3
Later versions	[2007, Journal of Monetary Economics] http://cdm16798.contentdm.oclc.org/cdm/ref/collection/p16798coll7/id/2019

Description

Title	Data from Gurkaynak, Sack and Wright 2006-28 FEDS paper (Excel file)
Author	Gurkaynak, Refet S. Sack, Brian Wright, Jonathan H.
Date	2006
Board Taxonomy	Yield curve U.S. Treasuries
Series	Finance and economics discussion series ; 2006-28
Publisher	Board of Governors of the Federal Reserve System (U.S.)
Relation	The U.S. treasury yield curve : 1961 to the present / Refet S. Gurkaynak, Brian Sack, and Jonathan H. Wright: http://cdm16798.contentdm.oclc.org/cdm/ref/collection/p16798coll4/id/201

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.

**METADATA IS A
LOVE NOTE
TO THE FUTURE**

Thanks!

Any questions or comments?

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

Bibliography:

- Altman, Micah, and Gary King (2007), "A Proposed Standard for the Scholarly Citation of Quantitative Data," *D-Lib Magazine*, (April 2007) <http://www.dlib.org/dlib/march07/altman/03altman.html>
- Altman, Micah (2016), "The Role of Research Funding and Policy Community in Data Citation — Rewards, Incentives, and Infrastructure." (August 24, 2016), <http://informatics.mit.edu/blog/role-research-funding-and-policy-community-data-citation-%E2%80%94-rewards-incentives-and>
- 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>
- Buneman, Peter; Susan Davidson, and James Frew (2016), "Why Data Citation is a Computational Problem," *Communications of the ACM*, vol. 59, No. 9, pp. 50-57. <http://cacm.acm.org/magazines/2016/9/206243-why-data-citation-is-a-computational-problem/fulltext>
- Chavan, Vishwas, and Lyubomir Penev (2011), "The data paper: a mechanism to incentivize data publishing in biodiversity science," *BMC Bioinformatics*, 12 (supl 15): S2 (15 December 2011), <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-S15-S2>
- DataCite. Data Citation Schema, last updated September 2016), <http://schema.datacite.org>
- Executive Office of the President (2013), Office of Science and Technology Policy, (2013), "Memorandum of the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research," (2/22/2013), https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- Force11 (2014), "Data Citation Synthesis Group: Joint Declaration of Data Citation Principles," Martone M. (ed.) San Diego CA: FORCE11; 2014 (<https://www.force11.org/group/joint-declaration-data-citation-principles-final>)
- Hole, Brian (2013), "Incentivizing Data Sharing: The Journal of Open Public Health Data," Presentation at the Workshop for Open Access for Health Scientists and Professionals, Kampala (30 August 2013), <http://www.slideshare.net/brianhole/ubiquity-press-eifl130830>
- IASSIST, Data Citation Quick Guide. <http://www.icpsr.umich.edu/files/ICPSR/enewsletters/iassist.html>
- Kratz, John E. and Carly Strasser (2015), "Making data count," *Sci. Data* 2:150039. <http://www.nature.com/articles/sdata201539>
- 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
- Lapeyre, Deborah Aleyne (2015), "Citing Data in Journal Articles Using JATS," FORCE11, <https://www.force11.org/sites/default/files/d7/project/882/citing-data-in-jats-2015-06.pdf>

Bibliography (continued)

- Michigan State University (MSU), Data Citation Guide, <http://libguides.lib.msu.edu/citedata>
- Mooney, H. (2011), "Citing data sources in the social sciences: do authors do it?" *Learned Publishing* 24, (2011), pp. 99–108, <https://pdfs.semanticscholar.org/c52e/db2eb663e1e43f5ad03738d73923307566f6.pdf>
- Mooney, H, MP Newton (2012), "The Anatomy of a Data Citation: Discovery, Reuse, and Credit," *Journal of Librarianship and Scholarly Communication*, 1(1):eP1035. <http://dx.doi.org/10.7710/2162-3309.1035>
- Moss, Elizabeth, Christin Cave, Jared Lyle (2015), "Sharing and citing research data: A repository's perspective," <https://deepblue.lib.umich.edu/handle/2027.42/115490>
- NISO, Journal Article Markup Suite (JATS), version 1.1, ANSI/NISO Z39.96-2015) (Standardized Markup for Journal Articles) <http://jats.niso.org/1.1/>
- 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>
- Piwowar, Heather A., Roger S. Day, Douglas B. Fridsman (2007), "Sharing Detailed Research Data Is Associated with Increased Citation Rate," *PLOS* (March 21, 2007), <http://dx.doi.org/10.1371/journal.pone.0000308>
- Rans, Jonathan, Michael Day, Monica Duke, and Alex Ball (2013), "Enabling the citation of datasets generated through public health research," The Wellcome Trust, <https://wellcome.ac.uk/sites/default/files/enabling-citation-of-datasets-through-public-health-research-phrdf-2013.PDF>.
- Robinson-Garcia, Nicholas, Evaristo Jiménez-Contreras, & Daniel Torres-Salinas (2015), "Analyzing data citation practices according to the Data Citation Index," *arXiv:150106285*, <https://arxiv.org/ftp/arxiv/papers/1501/1501.06285.pdf>
- Sieber, P.JE, BE Trumbo (1995), "(Not) Giving credit where credit is due: citation of data sets," *Science and Engineering Ethics*, 1, pp. 11–20.
- Simons, Natasha (2012), "Implementing DOIs for Research Data," *D-Lib Magazine*, Vol 18, Number 5/6 (May/June 2012), <http://www.dlib.org/dlib/may12/simons/05simons.html>
- US OMB Circular A-110 https://www.whitehouse.gov/omb/circulars_a110

Image credits:

Slide 3, "I [Love] Footnotes" image from: HistoryGypsy (2011), "Haunted and Taunted by My Own Wretched Foot," *While we're paused* (blog) <https://lanternhollowpress.com/2011/07/10/haunted-and-taunted-by-my-own-wretched-foot/> (accessed 9/28/2016).

Slide 4: "Dark Matter" slide: <http://www.quantumdiaries.org/wp-content/uploads/2013/06/disk-dark-matter.jpg>

Slide 14: "Metadata is a Love Note to the Future" Quote from Jason Scott, NYPL Labs. Image from Rebekah Cummings (2015 Research data management and sharing for social and behavioral sciences and humanities Rebekah Cummings, Research Data Management Librarian J. Willard Marriott Library September 15, 2015 <http://www.slideshare.net/RebekahCummings/research-data-management-and-sharing-for-the-social-sciences-and-humanities>