Data Citation at ICPSR

Jared Lyle, Elizabeth Moss, Christin Cave
Data Citation Workshop: Developing Policy & Practice
Washington, D.C.
12 July 2016
[1] What are we doing with data citation?

[2] How could better data citation impact us?
What are we doing with data citation?
ICPSR has provided data citations since 1990, and assigned digital object identifiers (DOIs) since 2008.

Study Description

Citation

Persistent URL: http://doi.org/10.3886/ICPSR04549.v1

Export Citation:
- **RIS** (generic format for RefWorks, EndNote, etc.)
- **EndNote XML** (EndNote X4.0.1 or higher)
Machine-readable Data Citation

<h1 itemprop="name">The 500 Family Study [1998-2000: United States] (ICPSR 4549)</h1>


<url>http://doi.org/10.3886/ICPSR04549.v1</url>
<datePublished>2008-05-30</datePublished>
Attribution Requirement in Terms of Use

Citing Data

You agree to reference the recommended bibliographic citation in any publication that employs resources provided by ICPSR. Authors of publications based on ICPSR data are required to send citations of their published works to ICPSR for inclusion in a database of related publications (bibliography@icpsr.umich.edu).
References


U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

doi:10.3886/ICPSR21240
Our users value data citations
Researchers (n=247) were asked: “How interested would you be to know each of the following about the impact of your data?

*White dots show the mean on a scale of one-to-four. All error bars depict 95% confidence intervals calculated by basic bootstrap with 10,000 resamplings.

• **Downloads:** Download counts, on the other hand, are both highly valuable and practical to collect. Downloads were a resounding second-choice metric for researchers and 85% of repositories already track them.

• **Citations:** Citations are the coin of the academic realm. They were by far the most interesting metric to both researchers and data managers. Unfortunately, citations are much more difficult than download counts to work with, and relatively few repositories track them. Beyond technical complexity, the biggest challenge is cultural: data citation practices are inconsistent at best, and formal data citation is rare. Despite the difficulty, the value of citations is too high to ignore, even in the short term.

https://datapub.cdlib.org/2015/08/04/2334/
Funders also value data citations
Data citation allows us to answer:

- Who uses the data?
- How are they used?
- With what impact?
Track and Link Data and Publications

Find Publications

The ICPSR Bibliography of Data-related Literature is a continuously-updated database of thousands of citations of works using data held in the ICPSR archive. The works include journal articles, books, book chapters, government and agency reports, working papers, dissertations, conference papers, meeting presentations, unpublished manuscripts, magazine and newspaper articles, and audiovisual materials.

Find Publications

Find Publications

- View all citations
- Browse by author
- Browse by journal
- Browse by study

Search Tips

- Our citation search indexes only the literal citations, not the full text of the publications.
- Instead of entering a research question, use only one or two search words, and use the filters to narrow
Your query returned 69384 citations.

<table>
<thead>
<tr>
<th>Pub. Year</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Author(s)</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
</tr>
<tr>
<td>2013</td>
<td>Erol, R.Y., Ortín, U.</td>
</tr>
<tr>
<td>2013</td>
<td>Offer, Shira</td>
</tr>
<tr>
<td>2011</td>
<td>Offer, Shira, Schneider, Barbara</td>
</tr>
<tr>
<td>2011</td>
<td>Weston, William</td>
</tr>
<tr>
<td>2009</td>
<td>Inter-university Consortium for Political and Social Research</td>
</tr>
</tbody>
</table>
Seclusion, decision-making power, and gender disparities in adult health: Examining hypertension in India

Samuel Stroope

doi: 10.1016/j.ssresearch.2015.05.013

Highlights

- Women’s seclusion is associated with higher odds of women’s hypertension.
- Women’s seclusion is associated with lower odds of men’s hypertension.
- Women’s low decision-making power is associated with higher odds of women’s hypertension.

Data for this Article

ICPSR Interuniversity Consortium for Political and Social Research Social and behavioral science research data

From Repository: Inter-university Consortium for Political and Social Research.

Source: Inter-university Consortium for Political and Social Research

DOI: http://dx.doi.org/10.3886/ICPSR06836

Abstract: The purpose of this project was to measure and estimate the distribution of income in both rural and urban areas of the People's Republic of China. The principal investigators based their definition of income on cash payments and on a broad range of additional components: payments in kind valued at market prices, agricultural output produced for self-consumption valued at market prices, the value of rent coupons and other direct subsidies, and the imputed value of housing. The rural component of the collection consists of two data files, one in which the individual is the unit of analysis and a second in which the household is the unit of analysis. Individual rural respondents reported on their employment status, level of education, Communist Party membership, type of employer (e.g., public, private, or foreign), type of economic sector in which employed, occupation, whether they held a second job, retirement status, monthly pension, monthly wage, and other sources of income. Demographic variables indicate relationship to household, gender, age, and student status. Rural households reported extensively on the character of the household and residence. Information was elicited on type of terrain surrounding the house, geographic position, type of house, and availability of electricity. Also reported were sources of household income (e.g., farming, industry, government, rents, and interest), taxes paid, valued form, total amount and type of cultivated land, financial assets and debts, quantity and value of various crops (e.g., grains, cotton, flax, sugar, tobacco, fruits and vegetables, tea, seeds, nuts, lumber, livestock and poultry, eggs, fish, shrimp, wool, honey, and silkworm cocoons), amount of grain purchased or provided by a collective, use of chemical fertilizers, gasoline, and oil, quantity and value of agricultural machinery, and all household expenditures (e.g., food, fuel, medicine, education, transportation, and electricity). The urban component of this collection also consists of two data files, one in which the individual is the unit of analysis and a second in which the household is the unit of analysis. Individual urban respondents reported on their economic status within the household, Communist Party membership, sex, age, nature of employment, and relationship to the household head. Information was collected on all types and sources of income from each member of the household whether working, working, or retired, all revenue received by owners of private or individual enterprises, and all in-kind payments (e.g., food and durable and nondurable goods). Urban households reported total income (including salaries, interest on savings and bonds, dividends, rent, leases, allowance, gifts, and boarding fees), all types and values of foodstuffs received, and total debt. Information was also gathered on household accommodations and living conditions, including number of rooms, total living area in square meters, availability and cost of running water, sanitary facilities, heating and air-conditioning equipment, kitchen availability, location of residences, ownership of homes, and availability of electricity and telephones. Households reported on all of their expenditures including amounts spent on food items such as wheat, rice, edible oils, pork, beef and mutton, poultry, fish, and seafood, sugar, and vegetables by means of both coupons from state-owned stores and at free market prices. Information was also collected on rents paid by the households, fuel availability, type of transportation used, and availability and use of medical and childcare.

Document Type: Data study

Accession Number: DRCI:DATA2012176001845894

Language: English

Funding:

<table>
<thead>
<tr>
<th>Funding Agency</th>
<th>Grant Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ford Foundation</td>
<td></td>
</tr>
</tbody>
</table>

Times Cited: 9

This article has been cited 9 times in Web of Knowledge.


Liu, ZG. The economic impact and determinants of investment in human and political capital in China, ECONOMIC DEVELOPMENT AND CULTURAL CHANGE, JUL 2003.

How could better data citation impact us?
Bibliography

My own mind was my source due to experience.

http://www.flickr.com/photos/papertrix/38028138/
(CC BY-NC 2.0)
Challenges of Data Citation

- Poor and inconsistent citing practices
- Emerging data citation standard
- Ambiguous descriptions of data used in abstract, methodology, acknowledgments
- Requires inefficient (human) searching and browsing to track data and keep up with the demand
- Without standard practice, it is very difficult to quantify the impact of data sharing
Data “Sighting” (implicit) vs. Data Citing (explicit)

Sample?
Methods?
Abstract?
Acknowledgements?
Charts and Tables?
Discussion?
Footnotes?
Appendices?
References!
Examples of poor citation practice

• Sample described, not named, no author information, no access information, only a publication cited
• Data named in text, with some attribution, but no access information
• Cited in reference section, but with no permanent, unique identifier, so difficult for indexing scripts to find to automate tracking
Examples of a poor data citation

Poorly described and cited data

differently. Using survey data from 272 staff who worked at a Midwestern maximum-security state prison, this study examined the association between the occupational stressors of perceived dangerousness of the job, role conflict, role ambiguity, repetitiveness, and work-on-family conflict with the three forms of organizational commitment. The effects of the occupational stressors varied for each form of organizational commitment. Specifically, work-on-family conflict had a significant positive association with continuance commitment. Role conflict and repetitiveness had negative associations with moral commitment. Finally, all five stressors had significant negative associations with affective commitment.

Excessive human search effort, extensive collection knowledge

= Too costly, too questionable for confident measure of impact
Examples of a good data citation

Citing data with a DOI


Minimal human search effort

= High hit accuracy for the cost, and better confidence of impact measures
Ordinary or Peculiar Men? Comparing the Customers of Prostitutes With a Nationally Representative Sample of Men

MA Monto, C Milrod - International Journal of Offender Therapy and ..., 2013

Page 1. International Journal of Offender Therapy and Comparative Criminology XX(X) 1 –19 ©

The Author(s) 2013 Reprints and permissions: sagepub.com/journalsPermissions.nav DOI:

10.1177/0306624X13480487 ijo.sagepub.com Article Ordinary or Peculiar Men? ...
Make Your Data Count!

• If it’s not cited, it can’t be counted
• Without counting data use, there is no accurate way to measure the impact of your shared data
• Without a well-formed citation, your data cannot take advantage of the potential of linked scholarly publishing
• Store your data where citations are unique and persistent
• Cite your own data and others’ in your publications
Thank you!

Jared Lyle
lyle@umich.edu
Audience

Sep 24, 2012 - Sep 23, 2013

Unique Visitors

- Studies: 308,967
- Bibliography: 47,379
- SSVD: 37,753

Pageviews

- Studies: 4,563,388
- Bibliography: 898,195
- SSVD: 1,442,353

Source: Google Analytics
The Scholix initiative is a high level interoperability framework for exchanging information about the links between scholarly literature and data. It aims to build an open information ecosystem to understand systematically what data underpins literature and what literature references data.

http://www.scholix.org/
Placeholder to request persistent IDs for government reports
No Common Practice of Formal Data Citation

- Abstract?
- Acknowledgements?
- Charts and Tables?
- Appendices?
- Discussion?
- Footnotes?
- Sample?
- Methods?

**References!**

- Without an explicit citation, reader must infer or be out of luck
- No attribution—no credit
- No access—no reuse
- No discernible impact!
Example

Abstract A huge literature shows that teen mothers face a variety of detriments across the life course, including truncated educational attainment. To what extent is this association causal? The estimated effects of teen motherhood on schooling vary widely, ranging from no discernible difference to 2.6 fewer years among teen mothers. The magnitude of educational consequences is therefore uncertain, despite voluminous policy and prevention efforts that rest on the assumption of a negative and presumably causal effect. This study adjudicates between two potential sources of inconsistency in the literature—methodological differences or cohort differences—by using a single, high-quality data source: namely, The National Longitudinal Study of Adolescent Health. We replicate analyses across four different statistical strategies: ordinary least squares regression.

Data

We use data from Add Health (the National Longitudinal Study of Adolescent Health), a school-based, nationally representative sample of 20,745 seventh through twelfth graders in 1994–1995. Respondents were reinterviewed in 1996 (Wave II), 2001–2002 (Wave III),

Acknowledgments Kane is the corresponding author; she carried out most of the statistical analysis and wrote the first draft of this article. She was also responsible for coordinating input from coauthors on subsequent drafts. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwistle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). This research received support