About Me

- I am a librarian at ICPSR
- I search for publications that analyze ICPSR data (in over 8,000 studies)
- One day scripts will take advantage of consistent data citing practices
- Making my job obsolete
Overview

• ICPSR’s experiment with linking data to published analyses

• Finding cited data in the social science literature

• How good citing practice affects stakeholders

• How ICPSR promotes good practice

• What’s next for IASSIST-SIGDC?
ICPSR’s Experiment with Linking Data to Literature

• Saw the value of clustering the literature with the data for instructors, students, researchers, funders

• Led to the creation of the Bibliography of Data-related Literature

• Studies link to publications, Bibliography links back to studies and out to full text
Bibliography of Data-Related Literature

The 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book chapters, government and agency reports, working papers, dissertations, conference papers, meeting presentations, unpublished manuscripts, magazine and newspaper articles, and audiovisual materials.

Submit Citations

ICPSR encourages its users to submit bibliographic citations to data we disseminate. If you would like to add a new citation to our bibliography, please submit it via our citation form, or email us at bibliography@icpsr.umich.edu.

Benefits of the Bibliography

The bibliography facilitates literature searches by social scientists, students, journalists, policymakers, and funding agencies. They use it to

- Identify much of the research that has already been undertaken in an area
- Replicate analyses of data
- Avoid duplicating analysis that has already been done
- Identify cross-disciplinary implications and uses of the data

The bibliography is also valuable for studying data as intellectual output. It allows investigators to

- Study how data resources are used
- Conduct citations analyses
- Investigate the life cycle of data
- Learn more about methodological issues, some of which are covered solely in the published literature
Chinese Household Income Project, 2002

Principal Investigator(s): Shi, Li

Summary: The purpose of this project was to measure and estimate the distribution of personal income and related economic factors 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. Data were collected through a series of questionnaire-based interviews conducted in rural and urban areas at the end of 2002. There are ten separate datasets. The first four datasets w... (view details)

Persistent URL: http://dx.doi.org/10.3886/ICPSR21741

Access Notes

- This study was originally processed, archived, and disseminated by DSIR, a project funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD).

What Can I Do With This Collection?

- Explore
  - View study description
  - Browse documentation files (login not required)
  - Variables:
    - List all variables in this study
    - Search the variables in this study
  - View related literature (52)
  - View related literature for the entire series

- Download
  - Download select files (select available analysis formats)
  - Download all files (232,656 KB)

- Analyze online
  - Analyze data using SDA (7)
Citation Search Results

If you weren't looking for citations, you can also perform your query against our data holdings or database of variables/questions. You may also wish to look over our searching tips.

You can also return to study page.

Query:

52 results found

Sort by: Pub Date (newest)

1 - 25 ▼

Citation  Pub. Year


Full Text Options: DOI, Worldcat, Google Scholar
Export Options: RIS/EndNote

Chen, Zhao, Lu, Ming, Wan, Guanghua. Inter-Industry Wage Differentials: An Increasingly Important Contributor to Urban China Income Inequality. Global COE Hi-Stat Discussion Paper Series. No. 130, Tokyo, Japan: Hitotsubashi University. 2010

Full Text Options: PDF
Export Options: RIS/EndNote


Full Text Options: DOI, Worldcat, Google Scholar
Export Options: RIS/EndNote
**View Record Details -- Journal Article**

<table>
<thead>
<tr>
<th>Title</th>
<th>Growing out of poverty: Trends and patterns of urban poverty in China 1988-2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Appleton, Simon&lt;br&gt;Song, Lina&lt;br&gt;Xie, Qingjie</td>
</tr>
<tr>
<td>Journal</td>
<td>World Development</td>
</tr>
<tr>
<td>Pub. Date</td>
<td>May 2010</td>
</tr>
<tr>
<td>Volume</td>
<td>38</td>
</tr>
<tr>
<td>Issue</td>
<td>5</td>
</tr>
<tr>
<td>Pages</td>
<td>665 - 678</td>
</tr>
<tr>
<td>ISSN</td>
<td>0305-750X</td>
</tr>
</tbody>
</table>

**Related Studies**

This publication is related to the following ICPSR dataset(s):

- [Chinese Household Income Project, 1988](https://www.icpsr.umich.edu/ICPSR/studies/8836) [ICPSR 8836]
- [Chinese Household Income Project, 1995](https://www.icpsr.umich.edu/ICPSR/studies/3012) [ICPSR 3012]
- [Chinese Household Income Project, 2002](https://www.icpsr.umich.edu/ICPSR/studies/21741) [ICPSR 21741]
Finding Cited Data in the Social Science Research Literature
Data "Sighting" (implicit) vs. Data Citing (explicit)
Issues with Attribution and Access

• 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
What We Think of as Citable Data

• Data are central to the argument

• Data are used to generate one table (or 20)

• Data are used in comparison to the central data used

• Data collection/methodology are described or critiqued
How Good Citing Practice Affects Stakeholders

- **Researchers and data producers** – Raises the visibility and legitimacy of their data as intellectual output and may have an impact on tenure and promotion through citation-related metrics.

- **Publishers/publishing platforms** – Makes the linkage between publications and underlying data explicit and enhances the rich web of associations linking content creators and their scholarly output.
How Good Practice Affects Stakeholders

- **Data users** – Enables authors to cite data used in an article and allows secondary analysts to find data resources of interest and to conduct literature reviews.

- **Librarians and information professionals who help data users** – Supports instruction in and assistance with data discovery and reuse.
How Good Practice Affects Stakeholders

- **Funding agencies** – Permits funders to measure the impact of their investments

- **Solutions and standards developers** – Facilitates standardization and machine-actionability

- **All of us** – Our research investment is not lost
Data Sharing Increases Research Productivity . . .

• The majority of social science data are not archived publicly
• Younger researchers are more protective of their research than established ones
• When data are archived (formally), 2-3 times more publications are created
• Need to ensure all productivity is measured

-- Pienta, Amy M.; Alter, George C.; Lyle, Jared A. (2010) The enduring value of social science research: The use and reuse of primary research data. [http://hdl.handle.net/2027.42/78307](http://hdl.handle.net/2027.42/78307)
How ICPSR Promotes Good Practice

• Makes it easy for users to cite our data

• Aids the cultural shift among data producers, authors, and publishers

• Encourages development of technology infrastructure
Provide Citations for Studies

- ICPSR has been providing citations to its data since 1990

- Based on “Cataloging Machine-Readable Data Files” by Sue Dodd, American Library Association, 1982
Include Standard Elements and Persistent, Unique Identifier

- Content Creator/Principal Investigator
- Title
- Distributor [ICPSR]
- Distribution place and date
- ICPSR study number
- Version number
- Materials designation [Computer file]
- DOI
Register DOIs for Studies

- DOIs are persistent URLs that ensure resources are always uniquely identified and findable (machine-readable)

- Could lead to automatic collection of citations into an index, for rapid creation of study bibliographies and other associated lists

- Since 2008 ICPSR has been assigning DOIs to data at the study level
Accommodate Journals’ Interest in Accessing Authors’ Data

- Journals more frequently requesting DOIs to ensure that data have been archived before articles are published

- Data depositors are coming to ICPSR for DOIs pre-publication, generally depositing into the Publication-Related Archive
Make Citations and DOIs More Prominent

- Provides FAQ: Why and How Should I Cite Data

- Provides RIS export for data citations into bibliographic citation software

- Highlights the data citation and DOI in several places

Principal Investigator(s): Schneider, Barbara; Waite, Linda J

Summary: The 500 Family Study was designed to obtain in-depth information on middle class, dual-career families living in the United States. To understand the complex dynamics of today's families and the strategies they use to balance the demands of work and family, over 500 families from 8 cities across the United States were interviewed. The study investigated different issues facing parents with older and younger children, families with adolescents and kindergarteners were included in the sample. (view details)

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

Access Notes:
- One or more files in this study are restricted; consult the detailed metadata for more information.
- This study is provided by Child Care & Early Education Research Connections.

What Can I Do With This Collection?
- Explore
  - View study description
  - Browse documentation files (login not required)
  - Variables:
    - List all variables in this study
    - Search the variables in this study
  - View related literature (183)

- Download
  - Download documentation files (login not required)
  - Download select files (select available analysis formats)
  - Download all files (159,864 KB)

- Analyze online
  - Analyze data using SDA (?)
**Description & Citation--Study No. 4549**

- Bibliographic Description
- Scope of Study
- Methodology
- Access and Availability

**Bibliographic Description**

**ICPSR Study No.:** 4549

**Title:** The 500 Family Study [1998-2000: United States]

**Principal Investigator(s):**
- Schneider, Barbara, University of Chicago. National Opinion Research Center (NORC). Alfred P. Sloan Center on Parents, Children and Work

**Funding:** Alfred P. Sloan Foundation (2006-14 and 2003-6-20 (2nd funding cycle))


**Scope of Study**

**Summary:** The 500 Family Study was designed to obtain in-depth information on middle class, dual-career families living in the United States. To understand the complex dynamics of today's families and the strategies they use to balance the demands of work and family, over 500 families from 8 cities across the United States were studied. To address different issues facing parents with older and younger children, families with adolescents and families with kindergartners were included in the study.
ICPSR Members Promote Data Citing to Students

• “How to Cite a Data Set”, University of North Carolina, University Libraries

• Anticipate evolving user needs

• Fill in the holes left by style guides
Work with Vendors to Promote Data Linking from the Literature

- Thomson Reuters to display data linkages in Web of Science
- Full and summary records in Web of Science will link to related data when appropriate
- ICPSR is providing a periodic data feed of datasets and related publications to TR
- TR is integrating data feeds from others including UK Data Archive
- The service will be ready in early 2012
Coordinate with Partners to Influence Journals

- On behalf of the Data-PASS partners, ICPSR wrote to ASA and APSA urging them to raise the standards for data citations in their journals.

- Professional associations are in a position to set standards for their members and for journal editors (including copy editors).
Build on Successful Efforts

• The letter stressed the importance of citing data the same way that publications are cited and the value of persistent identifiers.

• Both organizations discussed the letters at recent national meetings.

• *American Sociological Review’s* new submission guidelines.
Update Citation Software

- ICPSR worked with EndNote (part of Thomson Reuters), and “Dataset” is now a Reference Type in EndNote
- Metadata fields in this reference type overlap substantially with ICPSR metadata fields
- ICPSR is now working with RefWorks to add a new “dataset” Reference Type
- Zotero is building a Dataset Type, and could use input from the social sciences
Data Citations

Professional associations in the social sciences are increasingly recognizing the importance of properly citing data in their publications to encourage the replication of scientific results, to improve research standards, and to give proper credit to data producers.

The Data-PASS partners are committed to promoting standards and improving practices for the citation of data. This site offers guidelines and best practices for citing social science research data in order to promote vigorous and consistent attribution of datasets.

The American Sociological Review has already adopted a set of standards for citing data after an appeal from the Data-PASS partners. As other peer-reviewed journals and data stakeholders follow suit, consistently applied data citation standards will ensure that research data can be: discovered; reused; replicated for verification; credited for recognition; and tracked to measure usage and impact.

In short, accurate citation of data promotes more and better science, and we believe all data stakeholders can do more to improve data citation. Below are guidelines on how to cite data and what you can do to help.

How to Cite Data

Citing data is straightforward. Each citation must include the basic elements that allow a unique dataset to be identified over time:

- Title
- Author
- Date
- Version
- Persistent identifier (such as the Digital Object Identifier, Uniform Resource Name URN, or Handle System)

Here are some examples:


Esther Duflo; Rohini Pande, 2006, 'Dams, Poverty, Public Goods and Malaria Incidence in India', http://hdl.handle.net/1902.1/10JH1K9OCLZ UNF:5:obNHHq1gtV400aT-Xx9p9g— Murray Research Archive [Distributor] V2 [Version]


In addition to the above basic elements, we also recommend the addition of fidelity information, such as a checksum or Universal Numeric Fingerprint, which enables verification that data used later matches data originally cited.

What You Can Do

As the below diagram from the Australian National Data Service shows, your actions can make a difference in building a culture of data citation - whether you're a data producer, author, or journal editor.
Open to Help with Efforts to Standardize Data Citation

• DataCite, and their Metadata Kernel for the Publication of Research Data

• IQSS Data Citation Principles Workshop

• CODATA Data Citation Standards and Practices (International Council for Science: Committee on Data for Science and Technology)
What’s next for IASSIST-SIGDC?

• Represent the interests of the social science community

• Identify cross-over with efforts to create standards for the cultural and technological shift
IASSIST Special Interest Group on Data Citation: Stated Goals

- Recommend data citation standards and best practices for their use
- Work with vendors and projects to ensure that data are cited and linked to publications
- Advance the use of persistent identifiers
IASSIST Special Interest Group on Data Citation: Stated Goals

• Support data citation-related initiatives and programs

• Collaborate and liaison with similar groups in other organizations

• Promote awareness and develop pedagogy for data literacy and the data lifecycle as it relates to citation
Today’s Meeting at 5 PM, Harbour Centre Room 2050

• Members include representatives of most of the stakeholder groups

• Brainstorm areas of focus and identify gaps

• Divide into small action teams
Thank you...

Questions?