Clarifying Data Citation and Sharing Some Practical Information for Journals

(Drawing on Slides by)
George Alter
ICPSR





Data-PASS is a partnership of eight social science data repositories

Data-PASS goals:

- Archive social science data collections at-risk of being lost.
- Catalog and promote access to archived collections in the Data-PASS shared catalog.
- Replicated preservation of archived collections.
- Advocate best practices in digital preservation.

DataPASS Partners

- Institute for Quantitative Social Science at Harvard University
- Howard W. Odum Institute for Research in Social Science at the University of North Carolina-Chapel Hill
- Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan
- Electronic and Special Media Records Service Division, National Archives and Records Administration
- Roper Center for Public Opinion Research
- Social Science Data Archive at the University of California, Los Angeles (UCLA)
- Qualitative Data Repository at Syracuse University
- Cornell Institute for Social and Economic Research (CISER) at Cornell University

Context

- Science under attack
- Defending the legitimacy of science
- Data access and research transparency (DA-RT)
- Why data citation is important

Science is Under Attack



NEW REPUBLIC

Political Science in the Crosshairs

Republicans defund academic studies whose lessons they don't want to learn

BY TIMOTHY NOAH | March 22, 2013

On March 20 the Senate <u>de-funded</u> political science grants from the National Science Foundation "except for research projects that the Director of the National Science Foundation certifies as promoting national security or the economic interests of the United States." Since political science research, like most scientific research, is seldom undertaken to promote national security or the economic interests of the U.S., it seems doubtful there will be many such exceptions.

And it's not just Political Science

- Threats to the legitimacy of science
 - Fraud and Mistakes
 - Publication bias
 - Political pressures



"When Contact C

Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals," Quarterly Journal of Political the canvassers ex Science, 2008, 3: 313-326.

How do we defend the legitimacy of science?

What distinguishes the voice of science from every other voice on TV,

radio, Internet, ...?

• Science has norms and ethics:

Transparency

Reproducibility

• Who promotes these norms?

- Professional associations
- Journals

MEMORANDUM

TO: APSA Council

FROM: Arthur Lupia and Colin Elman

DATE: September 2010

RE: INCREASING DATA ACCESS AND RESEARCH TRANSPARENCY (DA-RT)

1. BACKGROUND/MOTIVATION

Political Science advances through scholars making, critiquing, and defending evidence-based claims. These claims take many forms. Common forms include interpretations, descriptions, descriptive generalizations, and causal inferences.

This memorandum seeks to increase the quantity and credibility of all such claims. Our proposed means for achieving these ends are higher standards and changed expectations in matters of data access and research transparency. These changes will increase the availability

Political Science has been a leader in DA-RT

- 2010 DA-RT memo
- 2012 APSA Ethics Guide revised
- 2014 Journal Editors
 Transparency Statement

MEMORANDUM

TO: APSA Council

FROM: Arthur Lupia and Colin Elman

DATE: September 2010

RE: INCREASING DATA ACCESS AND RESEARCH TRANSPARENCY (DA-RT)

1. BACKGROUND/MOTIVATION

Political Science advances through scholars making, critiquing, and defending evidence-based claims. These claims take many forms. Common forms include interpretations, descriptions, descriptive generalizations, and causal inferences.

This memorandum seeks to increase the quantity and credibility of all such claims. Our proposed means for achieving these ends are higher standards and changed expectations in matters of data access and research transparency. These changes will increase the availability

Political Science has been a leader in DA-RT

- 2010 DA-RT memo
- 2012 APSA Ethics Guide revised
- 2014 Journal Editors
 Transparency Statement



- 6. Researchers have an ethical obligation to facilitate the evaluation of their evidence based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated.
 - 6.1 **Data access**: Researchers making evidence-based knowledge claims should reference the data they used to make those claims. If these are data they themselves generated or collected, researchers should provide access to those data or explain why they cannot.
 - 6.2 **Production transparency**: Researchers providing access to data they themselves generated or collected, should offer a full account of the procedures used to collect or generate the data.
 - 6.3 **Analytic Transparency**: Researchers making evidence-based knowledge claims should provide a full account of how they draw their analytic conclusions from the data, i.e., clearly explicate the links connecting data to conclusions.

Political Science has been a leader in DA-RT

- 2010 DA-RT memo
- 2012 APSA Ethics Guide revised
- 2014 Journal Editors
 Transparency Statement
 - 27 journals



DATA ACCESS & RESEARCH TRANSPARENCY

HOME POLICIES RESOURCES EVENTS SISTER ORGS ABOUT

The Journal Editors' Transparency Statement (JETS)

For information on the petition, scroll to the bottom of the page. For those not familiar with JETS, here is the Journal Editors Transparency Statement (JETS) in its entirety:

Data Access and Research Transparency (DA-RT): A Joint Statement by Political Science Journal Editors

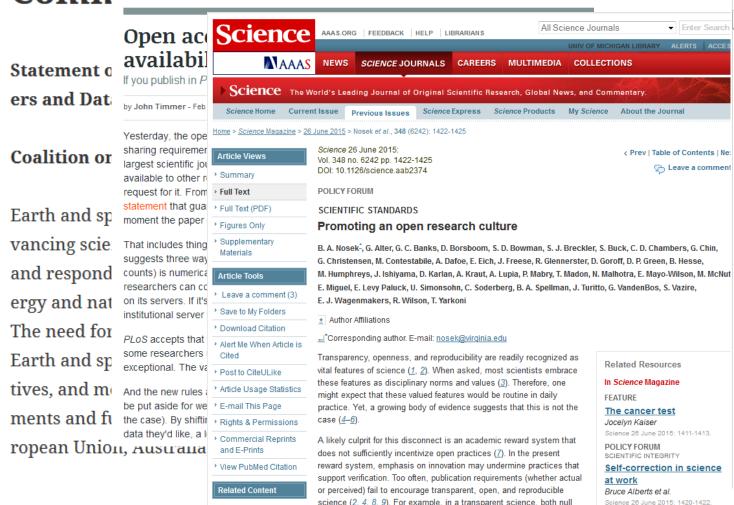
In this joint statement, leading journals commit to greater data access and research transparency, and to implementing policies requiring authors to make as accessible as possible the empirical foundation and logic of inquiry of evidence-based research. Please visit dartstatement.org for more information

Transparency requires making visible both the empirical foundation and the logic of inquiry of research. We agree that by January 15, 2016 we will:

 Require authors to ensure that cited data are available at the time of publication through a trusted digital repository. Journals may specify which trusted digital repository shall be used (for example if they have

Political Science is not alone

COPDESS Statement of Comn SCIENTIFIC METHOD / SCIENCE & EXPLORATION



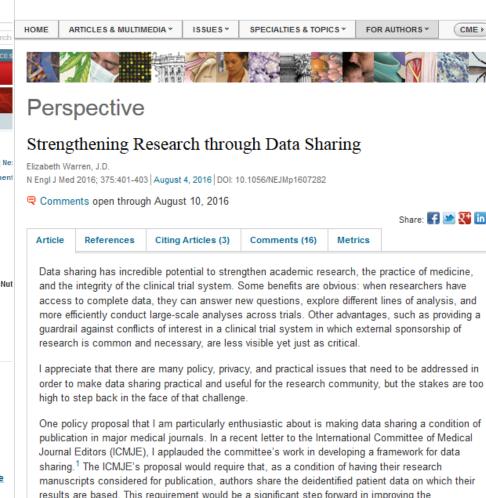


FOR AUTHORS *

Metrics

CME >

Share: F 🐸 🌠 🛅 🔠



TOP Guidelines are a template for journal policies.

nate

The Transparency and Openness Promotion Guidelines

TOP

Summary Table

Become a Signatory

Implementing the Guidelines

List of Signatories

Transparency, open sharing, and reproducibility are core values of science, but not always part of daily practice. Journals, funders, and scholarly societies can increase the transparency and reproducibility of research by adopting the TOP Guidelines.

Published in Science in 2015, TOP includes eight modular standards, each with three levels of increasing stringency. Journals select which of the eight transparency standards they wish to adopt for their journal, and select a level of implementation for each standard. These features provide flexibility for adoption depending on disciplinary variation, but simultaneously establish community standards.

Read the complete TOP Guidelines PDF or wiki

Implementing TOP

Follow these steps to implementing the TOP Guidelines, or read how Science and Ecology Letters will be implementing the TOP Guidelines.

- Science: Taking up TOP
- Ecology Letters: Ecology Letters, and Transparency and Openness Promotion (TOP) guidelines

Become a signatory

716 Journals and 62 organizations have already become signatories of the TOP Guidelines.

Guidelines for Transparency and Openness Promotion in Journal Policies and Practices

Citation Standards	Article is not published until providing appropriate citation for data and materials following journal's author guidelines.
Data Transparency	Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Analytic Methods (Code) Transparency	Code must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Research Materials Transparency	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Design and Analysis Transparency	Journal requires and enforces adherence to design transparency standards for review and publication
Preregistration of studies	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
•	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal uses Registered Reports as a submission option for replication studies with peer review prior to observing the study outcomes.

TOP Guidelines are a template for journal policies.

The TOP Guidelines

1. Citation Standards

Summary: Citation of articles is routine and well-formulated. Similar standards can be applied to citation of data, code, and materials to recognize and credit these as original intellectual contributions. Level 1 recommends citation standards, Level 2 requires adherence to citation standards, and Level 3 requires and enforces adherence to citation standards.

Level 1

All data, program code and other methods should be appropriately cited. Such materials should be recognized as original intellectual contributions and afforded recognition through citation.

- a. All data sets and program code used in a publication should be cited in the text and listed in the reference section.
- b. References for data sets and program code should include a persistent identifier, such as a Digital Object Identifier (DOI). Persistent identifiers ensure future access to unique published digital objects, such as a text or data set. Persistent identifiers are assigned to data sets by digital archives, such as institutional repositories and partners in the Data Preservation Alliance for the Social Sciences (Data-PASS).
- c. Data set citation example:
 - Campbell, Angus, and Robert L. Kahn. American National Election Study, 1948. ICPSR07218-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1999. http://doi.org/10.3886/ICPSR07218.v3

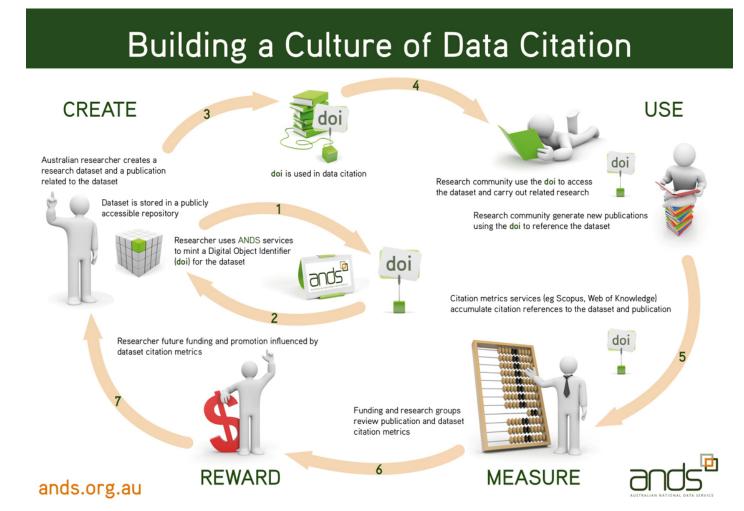
Level 2

Same text as Level 1 with additions in **bold** (full text <u>here</u>):

All data, program code and other methods **must** be appropriately cited. Such materials **are** recognized

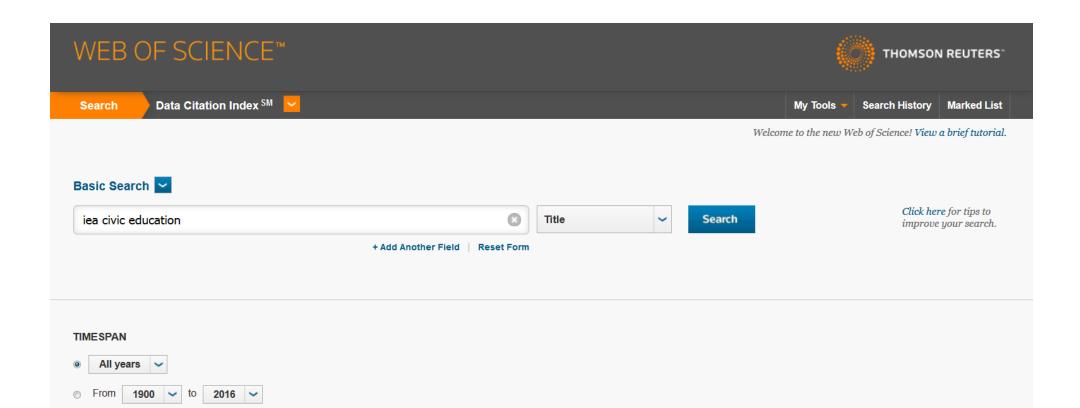
Why focus on data citation?

- Citation has been essential to scholarship for centuries
 - Data producers deserve recognition
- Data citation encourages data sharing



Why focus on data citation?

- Links between publications and data expand the visibility of both
 - Citation counts (and impact factors) are the currency of scholarship today
 - An example: Web of Science Data Citation Index



Web of Science Data Citation Index





WEB OF SCIENCE™

IEA Civic Education Study, 1999: Civic Knowledge an European Countries, 2 Latin American Countries, Ho

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

Group Author(s): Humboldt University of Berlin; University of Maryland-College Park

Inter-university Consortium for Political and Social Research

DOI: http://dx.doi.org/10.3886/ICPSR21661

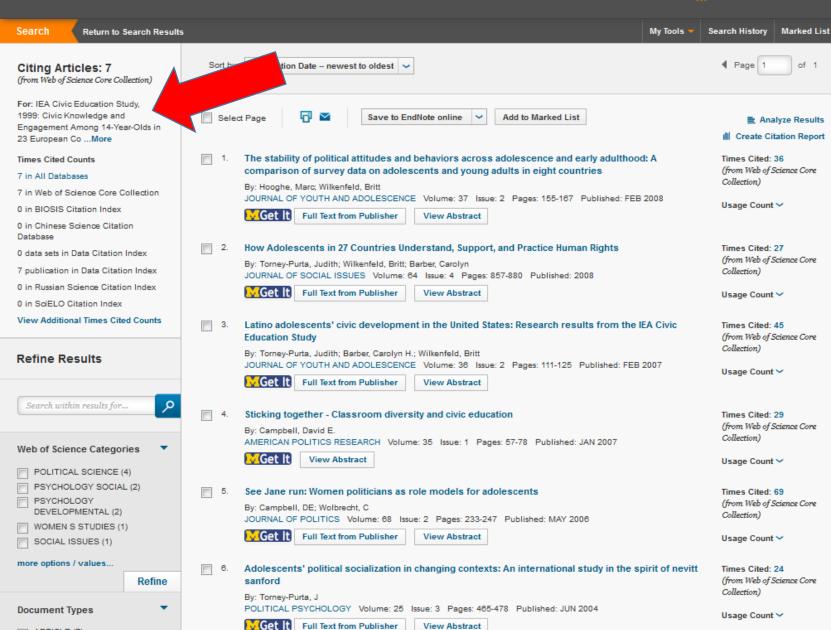
Viewed Date: 06 Jan 2016

Published: 2008

Abstract

The IEA Civic Education Study (known as CIVED) is an international assessment of the countries. Data were collected in 1999 under the auspices of the International Associa headquartered in Amsterdam. Begun in 1958, IEA is a consortium of educational and International Steering Committee (University of Maryland, College Park) and an intern International Coordinating Center (Humboldt University of Berlin) coordinated the day-Germany) processed the data. The Study's National Research Coordinators developed phase a more qualitative set of case studies in 24 countries and the second phase a te attention to macrosystem elements such as the country's history of democracy, transnat civic education goals are carried into practice, especially the role of the formal commi-Publications from the study have addressed each of these theories and approaches an specialists, comparative scholars, political scientists, and psychologists. The following Chile, Colombia, Cyprus, Czech Republic, Denmark, England, Estonia, Finland, Germ Norway, Poland, Portugal, Romania, Russian Federation, Slovak Republic, Slovenia, 5 nationally-representative samples ranged from 2,076 students tested in Belgium (Fren measured by a total civic knowledge scale composed of two subscales: civic content ar characteristics of democracies, and civic skills refer to the interpretative skills needed to political cartoon or an election leaflet). In addition, survey items measured students' α civic issues such as immigrants' or women's rights, trust in governmental institutions, an Students also answered questions pertaining to their background characteristics, their c classroom. Item responses gathered from approximately 95,000 students in total and 2 response theory are included on the file. Questionnaires also solicited background info

ARTICLE (7)



in eight countries . JOURINAL OF

Linking publications to data is difficult

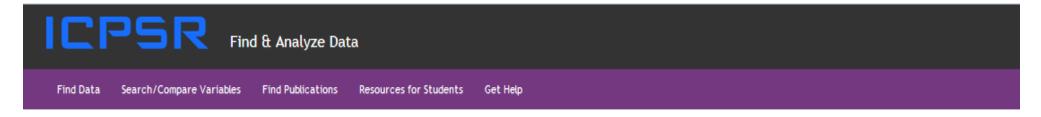
Data citation has been haphazard at best

Excerpt from DataPASS letter to APSA leadership, August 25, 2010

3 articles4 ways of citing data

It is easy to illustrate this problem. A quick look at 3 articles published in volume 104 of the American Political Science Review showed no consistency in the way publicly-available data are cited. One article (Singer) used two data sources, but provided a citation in the references section for only one with no unique identifier or Web address. The other data source was mentioned in both the figures/tables and the footnotes with a Web address. A second article (Hansford and Gomez) cited data within the text, with an online source named. Other data in the article were cited in the footnotes, giving an online module as the source. Neither data source appeared in the references list. The third article (Layman, Carsey, Green, Herrera, and Cooperman) named the data sources in the text, footnotes, and figures/tables, with no information about the data's location (beyond a reference to another article describing the data). None of the data sources appeared in the references list.

Linking publications and data should be easy...



Quick Download -

(28 datasets; 726,109 KB)

Table of Contents

- Access Notes
- Dataset(s)
- Study Description
- Citation
- Funding
- Scope of Study
- Methodology
- Version(s)
- Related Publications

IEA Civic Education Study, 1999: Civic Knowledge and Engagement Among 14-Year-Olds in 23 European Countries, 2 Latin American Countries, Hong Kong, Australia, and the United States (ICPSR 21661)

Principal Investigator(s): Humboldt University of Berlin; University of Maryland-College Park

Summary:

The IEA Civic Education Study (known as CIVED) is an international assessment of the civic knowledge and skills of 14-year-olds (8th and 9th graders) in 28 countries. Data were collected in 1999 under the auspices of the International Association for the Evaluation of Educational

Achievement (I

Data citation is a simple as cut-and-paste

titutes in more

This study is maintained and distributed by Civic Learning, Engagement, and Action Data Sharing (CivicLEADS). CivicLEADS provides infrastructure for researchers to share and access high-quality datasets to study civic education, civic action, and the relationships between the two.

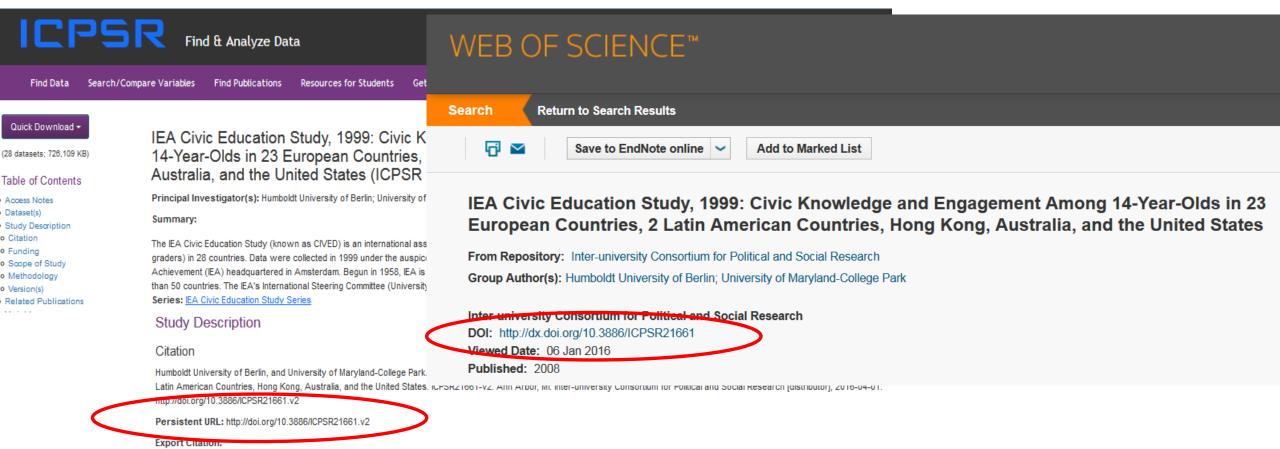
Citation

Humboldt University of Berlin, and University of Maryland-College Park. IEA Civic Education Study, 1999: Civic Knowledge and Engagement Among 14-Year-Olds in 23 European Countries, 2 Latin American Countries, Hong Kong, Australia, and the United States. ICPSR21661-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016-04-04. http://doi.org/10.3886/ICPSR21661.v2

Persistent URL: http://doi.org/10.3886/ICPSR21661.v2

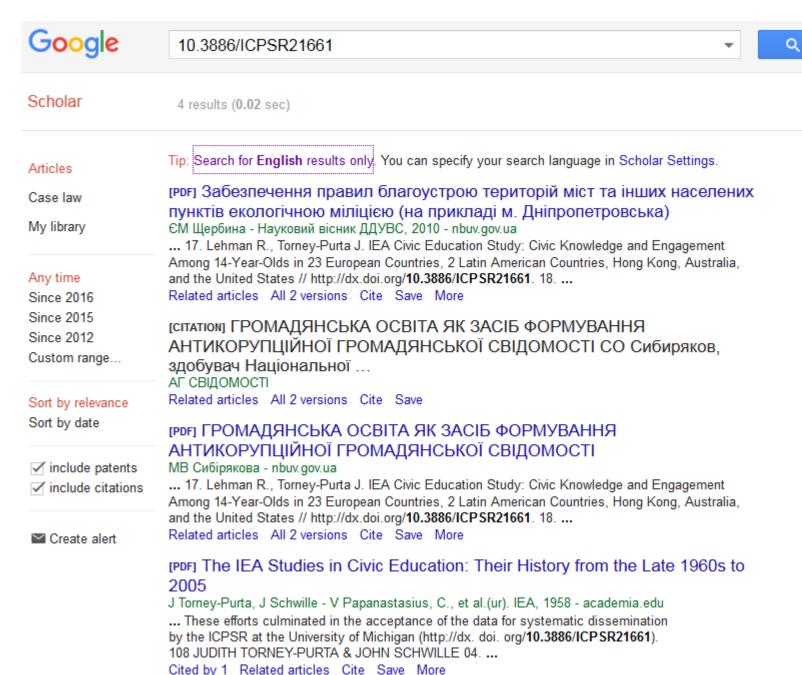
EndNote XML (EndNote X4.0.1 or higher)

Persistent Identifiers can be linked by machines



RIS (generic format for RefWorks, EndNote, etc.)
 EndNote XML (EndNote X4.0.1 or higher)

If all publications put persistent identifiers in their citations, we could use Google to link data and publications.



Sessions today

- Data Citation Sebastian
- Data Management Tom and Bill
- Journal Workflow Tom and Gustavo
- Journal Policies Colin

Outcomes

We want to convince you that

- Data citation is important
- Data citation is easy for authors and for editors
- Persistent identifiers are an essential part of citations
- Depositing data in a DataPASS repository is a simple process
- Help is available