

Dataverse

Sharing Research Data; Building Social Science

Gary King

Institute for Quantitative Social Science
Harvard University

Talk at the American Political Science Association meetings, August 2014

A Data Publishing Solution: Key Requirements

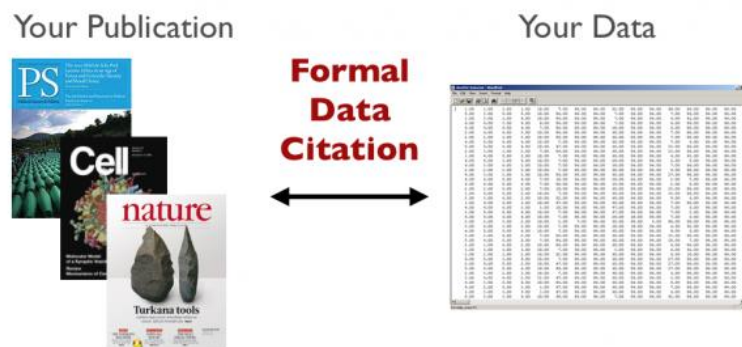
1. Long term preservation:

- Following **international standards** of preservation formatting, documentation, attribution, code, metadata
- Care by **professional archivists** (hint: that's probably not you!)
- A **trusted data repository**, backed by endowment or institution
- An active, supportive **community**
 - Assist, not replace, existing archives
 - Open source development, user groups, installations worldwide

2. Credit to the author (not only the archive):

- **Formal data citation**: including persistent identifier (DOI), universal numeric fingerprint (UNF), etc.
- **Web visibility and Control**: a virtual archive with ownership by author (everyone gets 100% of the credit)
- **Ever improving software**: sophisticated on-line analytics (Zelig), download in every format, high visibility for distribution, etc.

Options for Archiving Replication Data at Journals



Option A: Do it yourself

1. Publish data to authors' own Dataverse; get citation
2. Cite data in paper; journal only needs to check for citation

Option B: With journal verification

1. Submit draft of replication data to Journal Dataverse
2. Journal reviews it, and approves it for release
3. Dataset is published with a formal data citation & back to the article.

Option C: Full Automation

1. Seamless Integration between journal submission system and Dataverse
2. Automatic Link created between article and data
3. Easy for journal and author, and less error-prone.

Journals can have their own Dataverses too

Political Science Research and Methods

Harvard Dataverse Network >

Political Science Research and Methods (PSRM) Data

Data and replication code repository for Political Science Research and Methods (PSRM) journal. Instructions for contributors are available here.

11 datasets
100 downloads

by Dawn Brancati

Description: copy directly from abstract in PSRM publication

Replication data for: Explaining Fixed Effects: Random Effects modelling of Time-Series Correlation by Bell, Andrew; Jones, Kelvin

Description: This article challenges Fixed Effects (FE) modelling as the 'default' for time-series-correlation data. Understanding differences between within- and between-effects is crucial when choosing modelling strategies.

Replication data for: The Distributive Effects of Corruption

by Mircea Popa

Description: Abstract This paper puts forward an explanation for the resilience of corruption, by arguing that corrupt actors gain to a much larger share of the population than just to corrupt government officials. Corruption is a result of the distribution of power and resources.

Replication data for: Pivotality and Turnout: Evidence from a Field Experiment in the Aftermath of the 2008 US Presidential Election by Fowler, Anthony

Description: Many democratic citizens habitually abstain from the political process, and the reasons for this are complex. This paper examines the role of social norms and social capital in explaining turnout, and the role of social capital in explaining the relationship between turnout and political participation.

Replication data for: Does Survey Participation Increase Voter Turnout? Re-examining the National Election Studies by Persson, Mikael

Description: A Hawthorne effect found in election studies is that pre-election survey participation increases turnout. Swedish National Election Studies, Granberg and Holmberg (1992) showed evidence in support of this claim.

Replication data for: Candidate Confidence and Electoral Appeal: An Experimental Study on Voter Evaluations

by Dumitrescu, Delia; Gidengil, Elisabeth; Stol, Dietlind

Description: The nonverbal display of confidence is strongly associated with leadership and power. This paper examines the role of confidence in explaining voter turnout, and the role of confidence in explaining the relationship between confidence and political participation.

PSRM

ajps American Journal of Political Science

Harvard Dataverse Network >

American Journal of Political Science (AJPS) Dataverse

POWERED BY THE Dataverse PROJECT v. 3.6.2

Create Account Log In

OXFORD JOURNALS CONTACT US MY BASKET MY ACCOUNT

POLITICAL ANALYSIS

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Oxford Journals > Social Sciences > Political Analysis

Harvard Dataverse Network >

Political Analysis Dataverse

The Society for Political Methodology, the Political Methodology Section of the American Political Science Association, and the central web site for the political methodology community. The primary purpose of this site is to serve as the gateway to the central web site for the political methodology community. There is also information on the page covering our annual summer meetings held over twenty years. This site also serves as the gateway to the publications, including The Political Methodology Journal and Political Analysis, the official journal of the society. Additionally, we house a collection of syllabi for undergraduate and graduate courses in political methodology. Please direct questions to politicalanalysis@hss.caltech.edu.

OPEN DATVERSE

220 datasets
13,500 downloads

Political Analysis

Linked study

Replication data for: A Case for Registering Studies of Political Outcomes: An Application in the 2010 House Elections

Volume 08, 1999 - 2000

Volume 09, 2001

Volume 10, 2002

Volume 11, 2003

Volume 12, 2004

Volume 13, 2005

Volume 14, 2006

Volume 15, 2007

Volume 16, 2008

Volume 17, 2009

Volume 18, 2010

Search Studies

Search

Advanced Search Tips

Studies: 220 | Downloads: 13501

Sort By: Global ID

Replication data for: Geographic Boundaries as Regression Discontinuities by Keele, Luke; Titiunik, Rocio

Description: Data to illustrate our geographic RD methodological framework. Illustration is a re-examination of the effects of political advertisements on voter turnout during a presidential campaign, exploiting the exogenous variation in the volume ...

Continue [-]

Replication data for: Fuzzy Sets on Shaky Ground: Parameter Sensitivity and Confirmation Bias in fsQCA by Krogslund, Chris; Choi, Donghyun Danny; Poertner, Mathias

Description: Scholars have increasingly turned to fuzzy set Qualitative Comparative Analysis (fsQCA) to conduct small- and

doi:10.7910/DVN/26453

2 downloads

Last Released: Aug 22, 2014

doi:10.7910/DVN/27100

144 downloads

Last Released: Aug 20, 2014

FOR MILITANT

< View Previous Study Listing

140 datasets
6,500 downloads

ion can be replicated using these data.

Neil Malhotra, and Jacob N. Shapiro, "Poverty and Support for Militant Politics: Evidence from Pakistan", American Journal of Political Science.

itation Details

Support for Militant Politics: Evidence from Pakistan

by; C. Christine Fair (Georgetown University); Neil Malhotra (Stanford University); Jacob N. Shapiro (University of California, Berkeley)

The Institute for Quantitative Social Science at Harvard University

gblair@princeton.edu

tion and Scope

and extremist violence frequently cite poverty as a root cause of support for the evidence to support this contention, particularly in the Pakistani case. We exposed to the negative externalities of militant violence, and may in fact be less than these hypotheses we conducted a 6000-person, nationally representative

Seamless Integration

OJS Journal



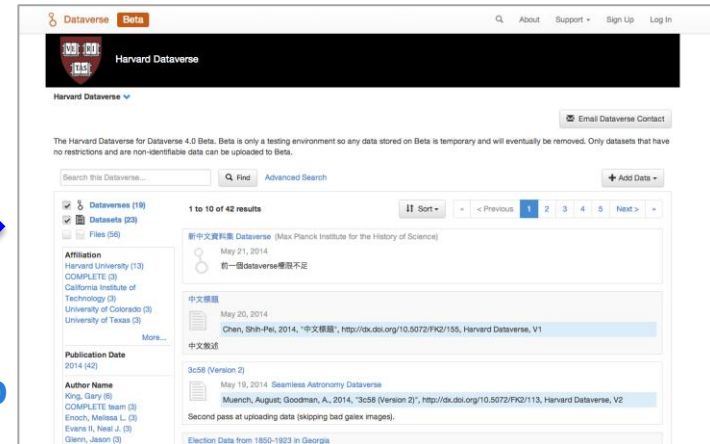
Citation
to Data



Citation
to
Article

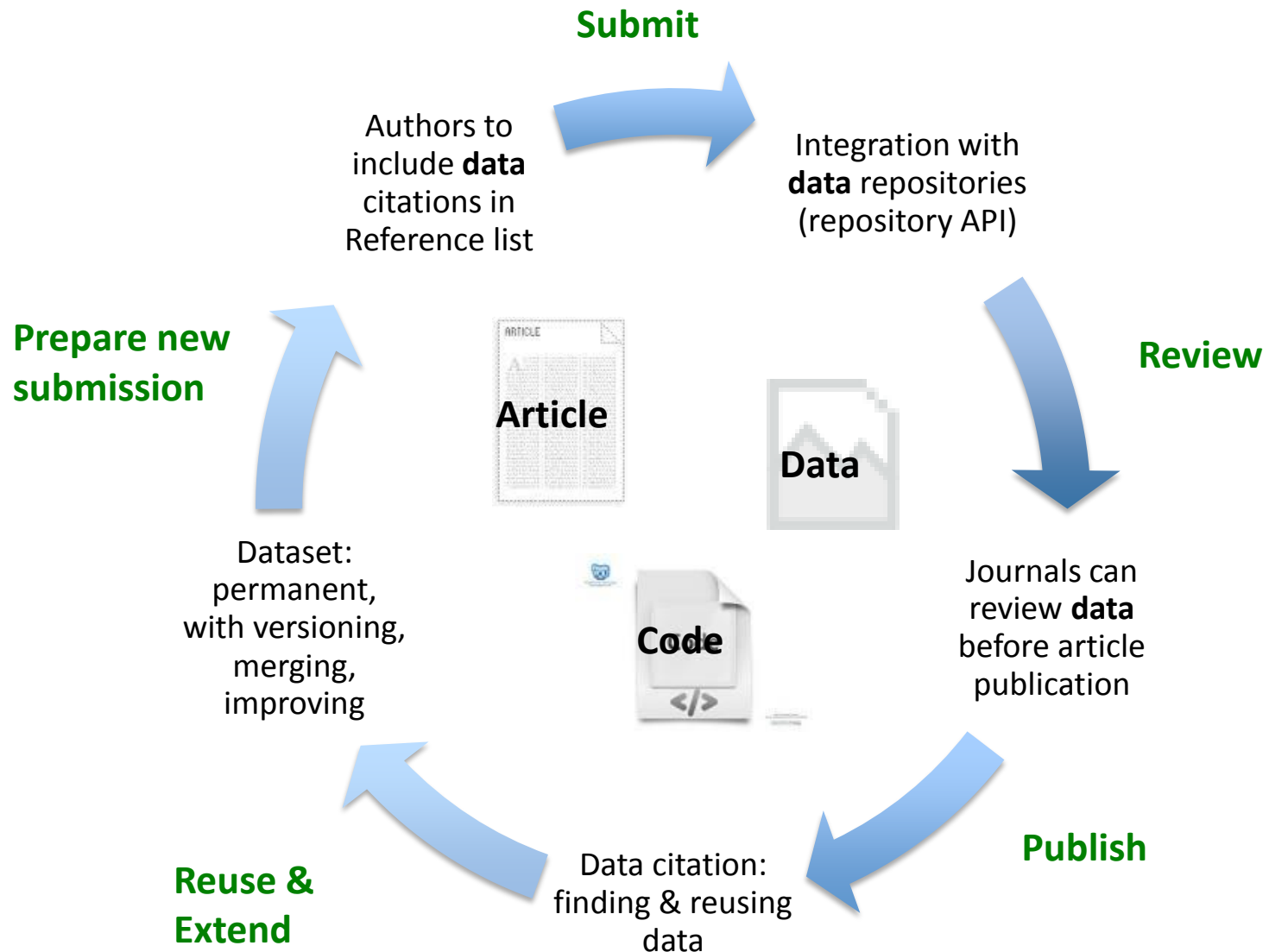


Journal Dataverse



- Sloan funded project to integrate PKP's Open Journal System (OJS) with Dataverse software via API
- Full integration thus far: ~**400 journals** (out of 5000 using OJS)
- Many more integrations in progress

An Integrated, Living Workflow



In Progress

- **Confidentiality, Privacy:**
 - DataTags: self-customizable data risk and sensitivity labels
 - Highly secure storage, with multiple security levels
- **New Data Challenges:**
 - Very large scale data sets
 - Streaming data (continuous versioning)
 - Tools for online fast data access, queries, and visualizations
 - Analysis methods for when data cannot be moved
 - Use (huge volume of) web analytics to improve dataverse and Zelig

For more information

TheData.org

GaryKing.org