



Discovering + Publishing Data with Dataverse

Elizabeth Quigley

User Experience Lead

The Dataverse Project (@dataverseorg)

equigley@iq.harvard.edu

What is Dataverse?

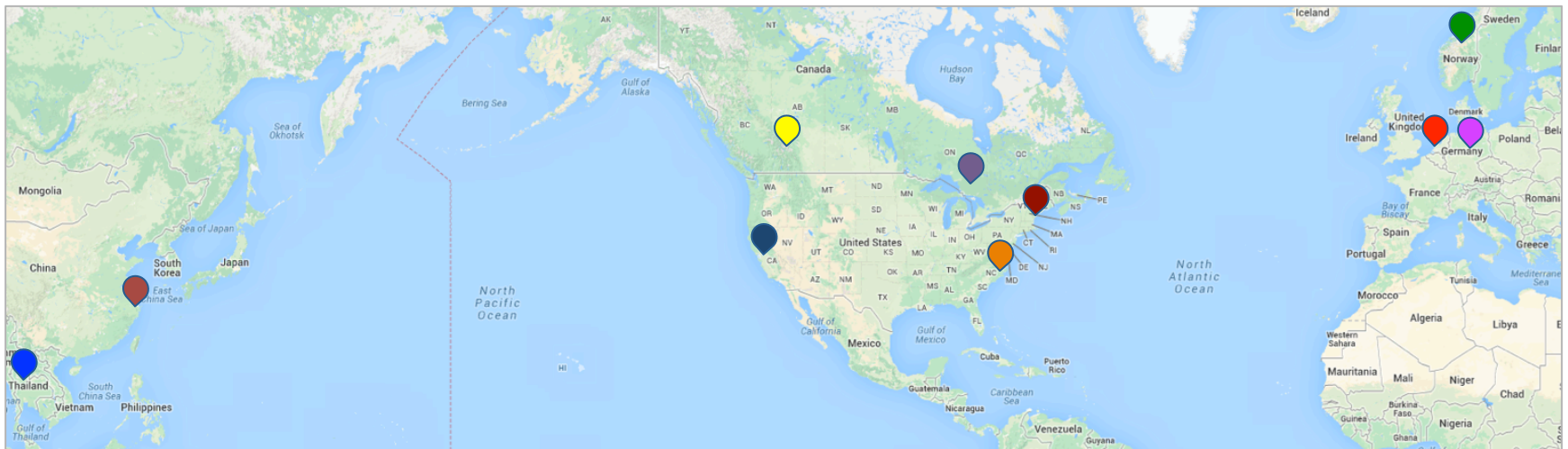
A software framework for publishing, citing and preserving research data (open source on [github](#) for others to install)

Developed by the Institute for Quantitative Social Science at Harvard University.

Provides incentives for researchers to share:

- Recognition & credit via data citations
- Control over data & branding
- Fulfill Data Management Plan requirements

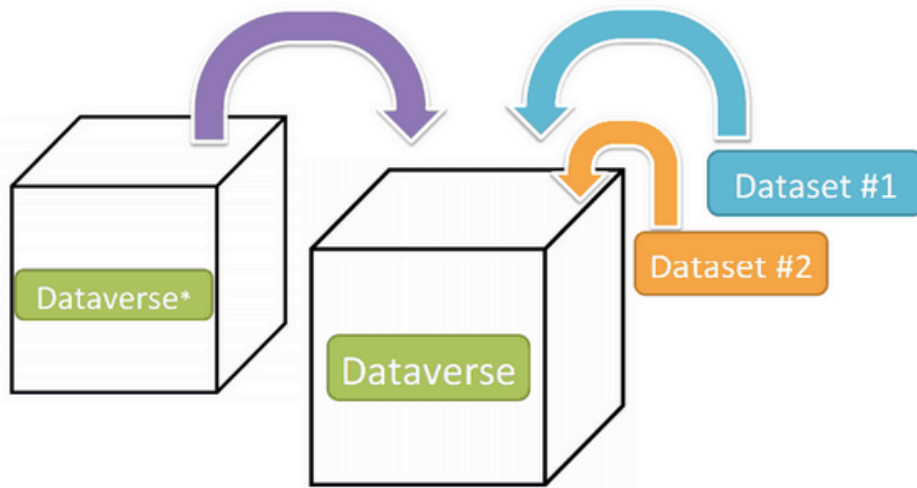
Dataverse Around the World



Institutions can setup/host their own Dataverse repository (UNC ODUM, Fudan Univ, Scholars Portal, DANS, etc) and within them can have dataverses for a variety of users (across all research domains): Researchers, Projects, Journals, etc.

What is a Dataverse or Dataset?

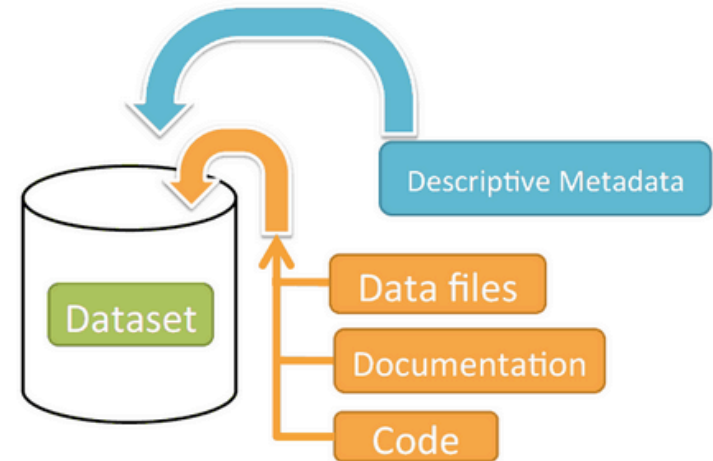
Schematic Diagram of a **Dataverse** in Dataverse 4.0



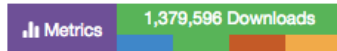
Container for your **Datasets** and/or **Dataverses***

* Dataverses can now contain other Dataverses (this replaces Collections & Subnetworks)

Schematic Diagram of a **Dataset** in Dataverse 4.0




Container for your data, documentation, and code.




Share, publish, and archive your data. Find and cite data across all research fields.


<




World Agroforestry Centre
ICRAF Dataverse



Population Services International
(PSI) Dataverse



INTERNATIONAL
FOOD POLICY
RESEARCH
INSTITUTE
IFPRI
International Food Policy
Research Institute (IFPRI)



Henry A. Murray
Research Archive
at Harvard University
Murray Research Archive
Dataverse

>

dataverse.harvard.edu

- Dataverses (1,220)**
 - Datasets (59,086)**
 - Files (281,149)**
- Dataverse Category**
- Researcher (359)
 - Research Project (134)
 - Organization or Institution (116)
 - Journal (47)
 - Teaching Course (9)
- Publication Date**
- 2015 (14,718)
 - 2011 (10,076)
 - 2007 (9,586)
 - 2012 (8,646)
 - 2009 (6,251)

1 to 10 of 60,306 Results

Sort <> < Previous 1 2 3 4 5 Next >

Replication Data for: Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects

Sep 10, 2015

Gonzalez-Navarro, Marco; Quintana-Domeque, Climent, 2015, "Replication Data for: Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects", <http://dx.doi.org/10.7910/DVN/6TC8RO>, Harvard Dataverse, V2

Review of Economics and Statistics: Forthcoming

Simulational Physics @ ETHZ Dataverse (ETH Zurich)

Sep 10, 2015

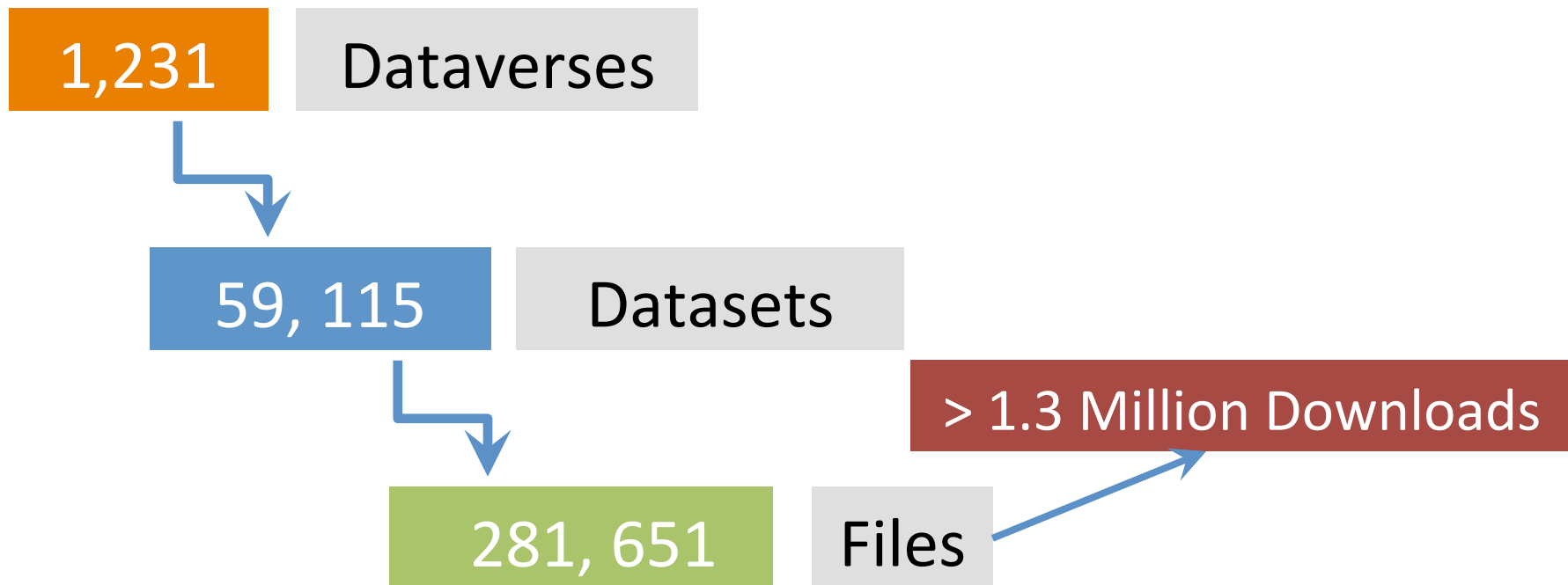
Administrative Regions

Sep 10, 2015 - A Tale of Cities: A Multi-source Dataset of Urban Life Dataverse

SpecialDat 2015 "Administrative Regions" <http://dx.doi.org/10.7910/DVN/6NM4V7>, Harvard Dataverse, V1

Harvard Dataverse

Open to all repository instance at Harvard currently has:



*number from September 15, 2015




DATAVERSE BEST PRACTICES

Dataverse Best Practices (1)

- Standard Metadata Schemas
 - DDI & OAI DC
 - New in 4.0:
 - DataCite 3.1
 - ISA-Tab (biomedical)
 - VO Resource (astronomy)
 - DC Terms
 - Metadata can be exported in JSON & XML

Dataverse Best Practices (2)

- Metadata is always public once a dataset is published
- By default, datasets receive CC0 Waiver The logo consists of a circular icon with a lightning bolt inside, followed by the text "PUBLIC DOMAIN" in a rectangular box.
- Even though default is CC0 and we encourage open/public data, when needed, data files in a dataset can be made restricted, or terms of use can be added

Dataverse Best Practices (3)

- Formal Data Citation
 - Originally based off Altman + King 2007
 - Endorse + comply w/ 2014 Joint Declaration of Data Citation Principles (FORCE11)
 - Lead by Merce Crosas, Director of Data Science @ IQSS
 - Versioning and File Fixity
- Persistent IDs: DOI (DataCite/EZID)
 - Resolve to a dataset landing page, not directly to the data files

Data Citation Example

Principle 2: Credit and Attribution (e.g. authors, repositories or other distributors and contributors)

Principle 4: Unique Identifier (e.g. DOI, Handle.). **Principle 5, 6 Access, Persistence:** A persistent identifier that provides access and metadata

Author(s), Year, Dataset Title, Data Repository or Archive, Version, Global Persistent Identifier

Principle 7: Specificity and verification (e.g. the specific version used).
Versioning or timeslice information should be supplied with any updated or dynamic dataset.

Dataverse Best Practices (4)

- Preservation format conversion for tabular data (extract column/variable metadata)
- File Fixity:
 - UNF (Altman, 2008) for tabular data
 - MD5 checksums for other files

Dataverse Best Practices (5)

- Data-PASS: (ICPSR, ODUM, NARA, ROPER,...)
 - Member of Data-PASS
- OAI-PMH: Harvesting metadata (DC, DDI)
 - From other Dataverse installations
 - From other OAI-DC compliant repositories
- If necessary: Deaccession a Dataset

DISCOVERING DATA

Searching for Data

- Search uses [Solr](#), an open source search platform
- Solr is also used by:

NETFLIX



Smithsonian Institution



Search this dataverse...

Q Find

Advanced Search

+ Add Data

 **Dataverses (1,219)**

 **Datasets (59,077)**

 **Files (281,071)**

Dataverse Category

Researcher (359)

Research Project (134)

Organization or Institution (115)

Journal (47)

Teaching Course (9)

Publication Date

2015 (14,708)

2011 (10,076)

2007 (9,586)

2012 (8,646)

1 to 10 of 60,296 Results

Sort ▾

« < Previous **1** 2 3 4 5 Next > »

An Economic Assessment of the Impact of Malpractice Liability on the Practice of Defensive Medicine

Sep 9, 2015 - Colgate ECON Student Data Repository Dataverse



Kovar, Christina, 2015, "An Economic Assessment of the Impact of Malpractice Liability on the Practice of Defensive Medicine", <http://dx.doi.org/10.7910/DVN/XUEK1R>, Harvard Dataverse, V1

Many argue that incentive costs of the malpractice system, namely practices such as defensive medicine, account for a substantial portion of growth in healthcare costs. However, aside from increasing healthcare costs, defensive medicine may also have negative effects on patients'...

Colgate ECON Student Data Repository Dataverse

Sep 9, 2015 Colgate University Data Repository Dataverse



This is a repository for student-generated data from Colgate University's Department of Economics. All the data in this repository is currently restricted and is only available to Colgate faculty, students, and staff.

Replication Data for Who's Careful: Regime Type and Target Selection

Searching for Data

(178)

Other (170)

Computer and Information Science

(69)

More...

Author Name

Digital Archive of Massachusetts

Anti-Slavery and Anti-Segregation

Petitions, Massachusetts Archives,

Boston MA (3,697)

U.S. Department of Commerce,

Bureau of the Census, Geography

Division (3,518)

Gallup Organization (1,529)

Harvard Map Collection, Harvard

College Library (1,315)

United States Information Agency

[USIA] (1,144)

More...



Thailand (2014): Rapid Assessment Findings. MSM in Bangkok: Drug Use and Sexual Risk Practices

Sep 9, 2015 - PSI/Thailand Dataverse



Dora Huang; Phakphoom Decha-anawong; Rapeepun Jommaroeng; Marc Theuss, 2015, "Thailand (2014): Rapid Assessment Findings. MSM in Bangkok: Drug Use and Sexual Risk Practices", <http://dx.doi.org/10.7910/DVN/VRIC0K>, Harvard Dataverse, V1

This exploratory study aimed to generate a more detailed understanding of the "scene" and context within which recreational drug use and risky sex occur among MSM in Bangkok, considering their behaviors and preferences, as well as socio-economic status and decision-making thought...

Replication Data for: Trends in Socioeconomic Inequalities in Motor Vehicle Accident Deaths in the United States, 1995–2010

Sep 9, 2015 - Sam Harper Dataverse



Harper, Sam, 2015, "Replication Data for: Trends in Socioeconomic Inequalities in Motor Vehicle Accident Deaths in the United States, 1995–2010", <http://dx.doi.org/10.7910/DVN/OW7M4S>, Harvard Dataverse, V2

The files in this dataset contain the publicly available data and code to reproduce the results of Harper S, Charters T.J, Strumpf E. Trends in Socioeconomic Inequalities in Motor Vehicle Accident Deaths in the United States, 1995–2010. Am J Epid 2015

Open Educational Resources (OER) analysis of NARS Institutions in India

Sep 9, 2015 - NAARM-ICM Dataverse



SK Soam; Bhavya Ogety, 2015, "Open Educational Resources (OER) analysis of NARS Institutions in India", <http://dx.doi.org/10.7910/DVN/74T77N>, Harvard Dataverse, V2

king

Find

Advanced Search

+ Add Data

Dataverses (4)

Datasets (467)

Files (186)

Dataverse Category

Research Project (1)

Researcher (1)

Publication Date

2011 (148)

2015 (133)

2007 (119)

2010 (100)

2008 (71)

1 to 10 of 657 Results

Sort

< Previous 1 2 3 4 5 Next >

Gary King Dataverse (Harvard University)

Jan 12, 2007

NEW KING

Anonymous RE Author Dataverse (ReproducibilityInRealEstate)

Apr 30, 2015

Data for Reproducing King County, WA Hedonic Pricing Study

King County, Washington : streams



Search term metadata field match displayed for each result

Adams, Greg (50)
Altman, Micah (55)
Benoit, Kenneth (55)

More...

File Type

tabulardata (162)

data (8)

document (8)

text (5)

application (2)

More...

and items have been removed from the AAT ...

Producer Name: King County (Wash.)

Sampling Procedure: King County, Washington data was obtained via a CD-ROM. The data were in the format of ArcINFO coverages and were converted to shapefiles using the "coverage to shapefile" wizard in ArcToolbox, version 8.3. The shapefiles were then imported to ArcSDE Layers using the ArcSDE command line importing tool, shp2sde

Author Name: King County (Wash.)

This dataset is harvested from our partners at Harvard Geospatial Library. Clicking the link will take you directly to the archival source of the data.

King County, Washington : landmarks

Dec 12, 2011 - Harvard Geospatial Library Dataverse



King County (Wash.); King County (Wash.). Transit Division., 2011, "King County, Washington : landmarks"

... Key landmarks within King County, Washington categorized into several major classes including but not limited to Government Sites; Major Commercial Sites; Transportation Facilities; Schools & Colleges; Sports & Leisure Facilities; Medical Facilities; Transit Center; Park and Ride Lots & Metro Tunnel Station; etc. ...

Producer Name: King County (Wash.)

Sampling Procedure: King County, Washington data was obtained via a CD-ROM. The data were in the format of ArcINFO coverages and were converted to shapefiles using the "coverage to shapefile" wizard in ArcToolbox, version 8.3. The shapefiles were then imported to ArcSDE Layers using

Search this dataverse...

 Find

[Advanced Search](#)

[+ Add Data](#)

 **Dataverses (1,219)**

 **Datasets (59,077)**

 **Files (281,071)**

Dataverse Category

Researcher (359)

Research Project (134)

Organization or Institution (115)

Journal (47)

Teaching Course (9)

Publication Date

2015 (14,708)

2011 (10,076)

2007 (9,586)

2012 (8,646)

2009 (6,251)

1 to 10 of 60,296 Results

 Sort ▾

« < Previous **1** 2 3 4 5 Next > »

An Economic Assessment of the Impact of Malpractice Liability on the Practice of Defensive Medicine

Sep 9, 2015 - Colgate ECON Student Data Repository Dataverse



Kovar, Christina, 2015, "An Economic Assessment of the Impact of Malpractice Liability on the Practice of Defensive Medicine", <http://dx.doi.org/10.7910/DVN/XUEK1R>, Harvard Dataverse, V1

Many argue that incentive costs of the malpractice system, namely practices such as defensive medicine, account for a substantial portion of growth in healthcare costs. However, aside from increasing healthcare costs, defensive medicine may also have negative effects on patients'...

Colgate ECON Student Data Repository Dataverse

Sep 9, 2015 Colgate University Data Repository Dataverse



This is a repository for student-generated data from Colgate University's Department of Economics. All the data in this repository is currently restricted and is only available to Colgate faculty, students, and staff.

Replication Data for Who's Careful: Regime Type and Target Selection

Sep 9, 2015 - Michael R. Kenwick Dataverse Dataverse



Bak, Dahee; Kenwick, Michael; Palmer, Glenn, 2015, "Replication Data for Who's Careful: Regime Type and Target Selection"

Thailand (2014): Rapid Assessment Findings. MSM in Bangkok: Drug Use and Sexual Risk Practices

Sep 9, 2015 - PSI/Thailand Dataverse



Dora Huang; Phakphoom Decha-anawong; Rapeepun Jommaroeng; Marc Theuss, 2015, "Thailand (2014): Rapid Assessment Findings. MSM in Bangkok: Drug Use and Sexual Risk Practices", <http://dx.doi.org/10.7910/DVN/VRIC0K>, Harvard Dataverse, V1

This exploratory study aimed to generate a more detailed understanding of the "scene" and context within which recreational drug use and risky sex occur among MSM in Bangkok, considering their behaviors and preferences, as well as socio-economic status and decision-making thought...

Replication Data for: Trends in Socioeconomic Inequalities in Motor Vehicle Accident Deaths in the United States, 1995–2010

Sep 9, 2015 - Sam Harper Dataverse



Harper, Sam, 2015, "Replication Data for: Trends in Socioeconomic Inequalities in Motor Vehicle Accident Deaths in the United States, 1995–2010", <http://dx.doi.org/10.7910/DVN/OW7M4S>, Harvard Dataverse, V2

The files in this dataset contain the publicly available data and code to reproduce the results of Harper S, Charters TJ, Strumpf E. Trends in Socioeconomic Inequalities in Motor Vehicle Accident Deaths in the United States, 1995–2010. Am J Epid 2015

Open Educational Resources (OER) analysis of NARS Institutions in India

Sep 9, 2015 - NAARM-ICM Dataverse



SK Soam; Bhavya Ogety, 2015, "Open Educational Resources (OER) analysis of NARS Institutions in India", <http://dx.doi.org/10.7910/DVN/71377M>, Harvard Dataverse, V1

Browsing for Data

Computer and information science (69)

[More...](#)

Author Name

- Digital Archive of Massachusetts
- Anti-Slavery and Anti-Segregation Petitions, Massachusetts Archives, Boston MA (3,697)
- U.S. Department of Commerce, Bureau of the Census, Geography Division (3,518)
- Gallup Organization (1,529)
- Harvard Map Collection, Harvard College Library (1,315)
- United States Information Agency [USIA] (1,144)

[More...](#)

Browsing for Data

- All dataverses are able to select facets by going to the General Information option under the Edit button
- Facets available for all metadata domains supported in Dataverse

Browse/Search Facets

i Choose the metadata fields to use as facets for browsing datasets and dataverses in this dataverse.

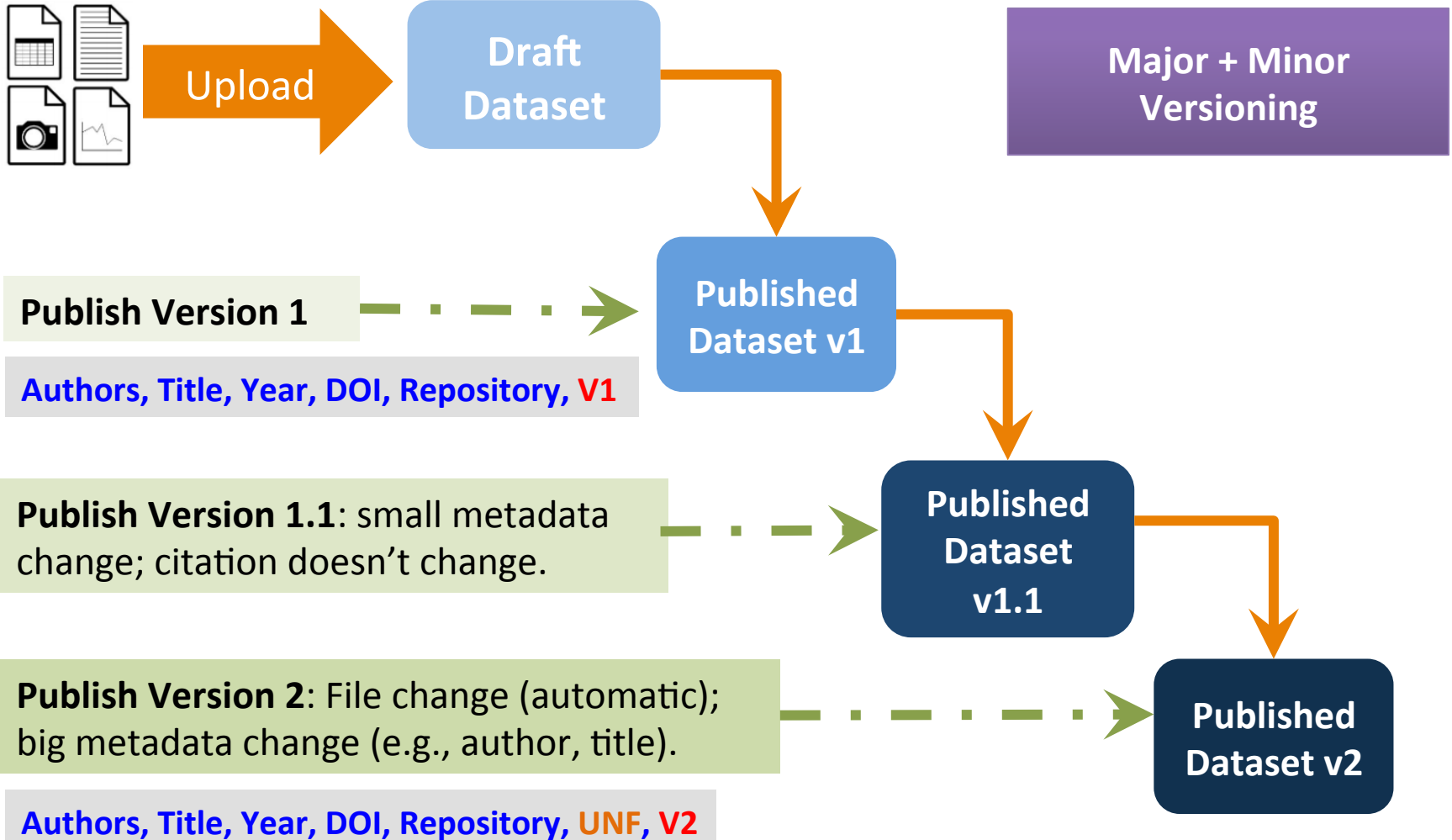
The interface shows a list of metadata fields on the left and a 'Selected' list on the right. The 'All Metadata Fields' dropdown is currently set to 'All Metadata Fields'. The list of fields includes: Author Affiliation, Keyword Term, Topic Classification Term, Language, Producer Name, Production Date, Contributor Type, and Contributor Name. The 'Selected' list contains: Subject and Author Name. Navigation buttons (→, →|, ←, |←) are located between the two lists.

All Metadata Fields	Selected
Author Affiliation	Subject
Keyword Term	Author Name
Topic Classification Term	
Language	
Producer Name	
Production Date	
Contributor Type	
Contributor Name	



PUBLISHING WITH DATAVERSE

Rigorous Data Publishing Workflows



Publishing a Dataset


Root Dataverse

Root Dataverse > Georgia Election Years Data from 1900-2000



Georgia Election Years Data from 1900-2000 Draft Unpublished

Quigley, Elizabeth, 2015, "Georgia Election Years Data from 1900-2000", <http://dx.doi.org/10.5072/FK2/J7EHC4>, Root Dataverse, DRAFT VERSION [UNF:6:keV+t8Adra7+wPM+G4PByw==] 


 Download Citation ▾

If you use these data, please add this citation to your scholarly resources. [Learn about Data Citation Standards.](#)

Description	Data on elections in Georgia.
Subject	Social Sciences

- Files
- Metadata
- Terms
- Versions


3 Files


 Upload + Edit Files



[dataversecm2015attendees.tab](#)

Tabular Data - 27.3 KB - Sep 10, 2015 - 0 Downloads
Original File MD5: 806db14e92ac205e1d585e7da57eb63b;
14 Variables, 170 Observations - UNF:6:keV+t8Adra7+wPM+G4PByw==


 Explore

 Download ▾



[dataversecm2015attendees.xlsx](#)


Adobe PDF - 38.4 KB - Sep 10, 2015 - 0 Downloads
MD5: e9a0470edeac4fa08b6da72dbb6e7222;

 Download



[Screen Shot 2015-09-09 at 3.27.07 PM.png](#)

PNG Image - 448.0 KB - Sep 10, 2015 - 0 Downloads
MD5: 69b6ff658cd2143dfdd1c84271964aa5;

 Download

Publishing Dataverse + Dataset

Election Data Dataverse Unpublished

Root Dataverse > Election Data Dataverse > New Mexico Election Data 1900-2000

Metrics 0 Downloads ✉ 🔄 📄 Publish ✎ Edit

New Mexico Election Data 1900-2000 Draft Unpublished

Quigley, Elizabeth, 2015, "New Mexico Election Data 1900-2000", <http://dx.doi.org/10.5072/FK2/TIZXA9>, Root Dataverse, DRAFT VERSION [UNF:6:keV+t8Adra7+wPM+G4PByw==] ? ☰ Download Citation

If you use these data, please add this citation to your scholarly resources. [Learn about Data Citation Standards.](#)




Publish Dataset ✕

⚠ This dataset cannot be published until [Election Data Dataverse](#) is published. Would you like to publish both right now?

ℹ Once you publish this dataset it must remain published.

Yes, Publish Both Cancel

3 Files

-  **Screen Shot 2015-08-24 at 5.25.39 PM.png**
Adobe PDF - 38.4 KB - Sep 10, 2015 - 0 Downloads
MD5: e9a0470edeac4fa08b6da72dbb6e7222; 📄 Download
-  **Screen Shot 2015-08-24 at 5.25.39 PM.png**
PNG Image - 48.1 KB - Sep 10, 2015 - 0 Downloads
MD5: 3fd362d548fa2e1d052dda10c0112aef; 📄 Download
-  **Screen Shot 2015-08-24 at 5.25.39 PM.tab**
Tabular Data - 27.3 KB - Sep 10, 2015 - 0 Downloads
Original File MD5: 806db14e92ac205e1d585e7da57eb63b; 🔍 Explore 📄 Download

Publishing a Dataset

Root Dataverse

Root Dataverse > Georgia Election Years Data from 1900-2000

Metrics 0 Downloads

✉️ 🔄 🌐 Publish ✎ Edit

Georgia Election Years Data from 1900-2000

Draft Unpublished

Quigley, Elizabeth, 2015, "Georgia Election Years Data from 1900-2000", <http://dx.doi.org/10.5072/FK2/J7EHC4>, Root Dataverse, DRAFT VERSION [UNF:6:keV+t8Adra7+wPM+G4PByw==]

If you use these data, please add this citation to your scholarly resources. [Learn about Data Citation Standards.](#)

Description

Subject

Files Metadata Terms Ver




3 Files

+ Upload + Edit Files

Publish Dataset

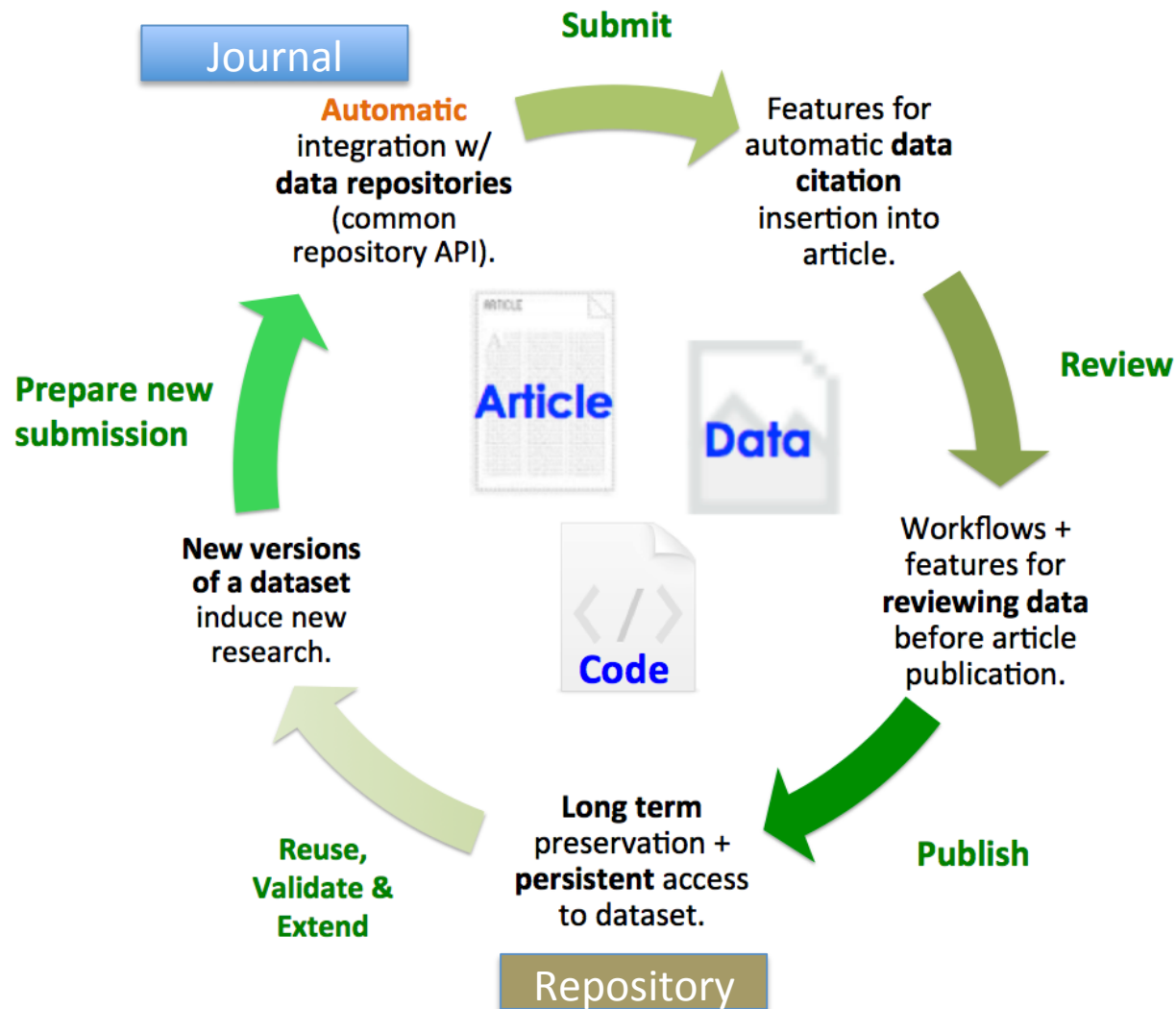
⚠️ Are you sure you want to publish this dataset? Once you do so it must remain published.

Continue Cancel

	dataversecm2015attendees.tab Tabular Data - 27.3 KB - Sep 10, 2015 - 0 Downloads Original File MD5: 806db14e92ac205e1d585e7da57eb63b; 14 Variables, 170 Observations - UNF:6:keV+t8Adra7+wPM+G4PByw==	Explore Download
	dataversecm2015attendees.xlsx Adobe PDF - 38.4 KB - Sep 10, 2015 - 0 Downloads MD5: e9a0470edeac4fa08b6da72dbb6e7222;	Download
	Screen Shot 2015-09-09 at 3.27.07 PM.png PNG Image - 448.0 KB - Sep 10, 2015 - 0 Downloads MD5: 69b6ff658cd2143dfdd1c64271964aa5;	Download

DATAVERSE + JOURNALS

An Integrated & Automated Journal / Data Publishing Workflow

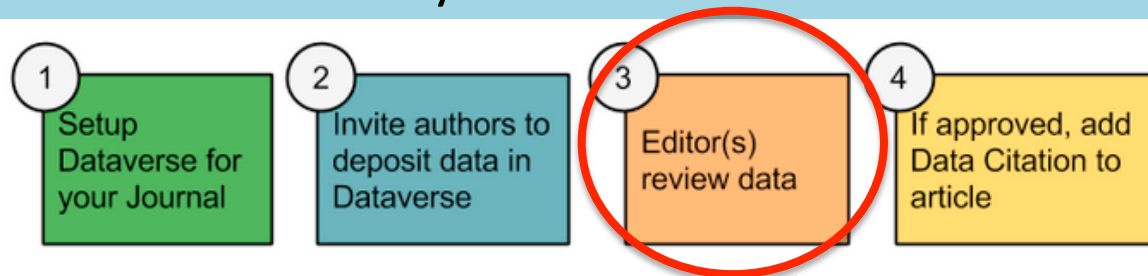


Current Workflows in Dataverse: To Connect Data to Journals

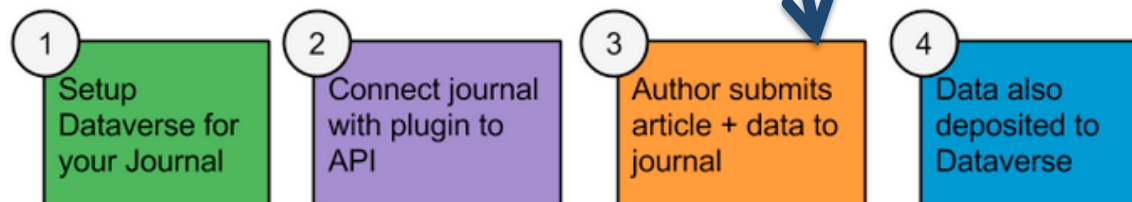
A. Journals include Dataverse as a Recommended Repository




B. Authors Contribute Directly to a Journal's Dataverse




C. Automated Integration of Journal + Dataverse (e.g., OJS)



Metadata for Journals

 **Datasets (203)**

 **Files (1,562)**

Publication Date

- 2014 (67)
- 2013 (64)
- 2015 (45)
- 2012 (27)

Author Name

- Hainmueller, Jens (5)
- Broockman, David (3)
- Huber, Gregory A. (3)
- Shor, Boris (3)
- Bechtel, Michael M. (2)

[More...](#)

Keyword Term

- Public opinion (18)
- Elections (16)
- Field experiments (11)
- Representation (10)
- Surveys (9)

[More...](#)


Journal Volume

- 58 (52)
- 57 (38)
- 59 (38)
- Forthcoming (30)
- 56 (2)


Journal Issue

- 3 (39)
- 2 (32)
- 1 (31)
- 4 (28)


Journal Publication Date

Replication Data for: Segregation and Inequality in Public Goods 


Sep 2, 2015

 Trounstine, Jessica, 2015, "Replication Data for: Segregation and Inequality in Public Goods", <http://dx.doi.org/10.7910/DVN/4LZXTY>, Harvard Dataverse, V1 [UNF:6:YJcDrrl2KgHaxc576EUQzw==]


Many United States cities function without regular problems. They have well-kept roads, sewers that never overflow, and public parks with swing sets and restrooms. Others struggle to maintain balanced budgets, fail to adequately equip or staff their police forces, and offer littl...

Replication Data for: Experiential Learning and Presidential Management of the U.S. Federal Bureaucracy: Logic and Evidence from Agency Leadership Appointments 


Sep 2, 2015

 Krause, George; Anne Joseph O'Connell, 2015, "Replication Data for: Experiential Learning and Presidential Management of the U.S. Federal Bureaucracy: Logic and Evidence from Agency Leadership Appointments", <http://dx.doi.org/10.7910/DVN/E9UQ0S>, Harvard Dataverse, V1 [UNF:6:HziRrSquxXKHSXfAIM1a1w==]


Presidents become increasingly effective at managing the bureaucracy because of the information and expertise that they acquire from on-the-job experience. In their appointment choices, this theory predicts that presidents become better at reducing information asymmetries incurre...

Replication Data for: Are Voters Equal under Proportional Representation? 


Aug 28, 2015

 Kedar, Orit; Harsgor, Liran; Sheinerman, Raz A., 2015, "Replication Data for: Are Voters Equal under Proportional Representation?", <http://dx.doi.org/10.7910/DVN/OJQCQ5O>, Harvard Dataverse, V1 [UNF:6:Udvi1wt/+pekhWW+r+KIFw==]


We develop and apply a new conceptual framework and measure for evaluating electoral systems, focusing on (in)equality in parliamentary representation. Our main arena of interest is proportional representation with districts, an electoral system employed by more than half of demo...

Replication Data for: Does Compulsory Voting Increase Support for Leftist Policy? 


Aug 26, 2015

 Bechtel, Michael M.; Hangartner, Dominik; Schmid, Lukas, 2015, "Replication Data for: Does Compulsory Voting Increase Support for Leftist Policy?", <http://dx.doi.org/10.7910/DVN/29591>, Harvard Dataverse, V1 [UNF:6:YBaBag8Cm89+fqhdRxtcw==]

Citizens unequally participate in referendums and this may systematically bias policy in favor of those who vote. Some view compulsory voting as an important tool to alleviate this problem while others worry about its detrimental effects on the legitimacy and quality of democrati...

Replication Data for: When Governments Regulate Governments 

Aug 18, 2015

 Konisky, David, 2015, "Replication Data for: When Governments Regulate Governments", <http://dx.doi.org/10.7910/DVN/INXGMT>, Harvard Dataverse, V1 [UNF:6:3QWE6iWLG9cdZT5oDvR5pQ==]

This article advances a political theory of regulation that accounts for the choices of regulators and regulated entities when both are governments



ORGANIZING AN INSTITUTION'S DATA

Researcher/Faculty Member

GARY KING

Gary King Dataverse (Harvard University) <http://gking.harvard.edu/>

Harvard Dataverse > Gary King Dataverse



Search this dataverse... Find Advanced Search

Datasets (0)

Datasets (47)

Files (1,307)

Publication Date

- 2007 (23)
- 2009 (6)
- 2014 (4)
- 2015 (4)
- 2008 (3)

More...

Subject

Social Sciences (4)

Author Name

- King, Gary (44)
- Zeng, Langche (8)
- Gelman, Andrew (4)
- Soneji, Samir (4)
- Honaker, James (3)

More...

Keyword Term

- Causal inference; EPBR; Matching (1)
- causal inference, community intervention trials, field experiments, group-randomized trials, health policy, matched-pair design, noncompliance (1)
- congressional districts (1)
- election districts (1)
- election returns (1)

More...

1 to 10 of 47 Results

Sort

« < Previous 1 2 3 4 5 Next > »

Replication data for: Reverse Engineering Chinese Censorship: Randomized Experimentation and Participant Observation



May 22, 2015

King, Gary; Pan, Jennifer; Roberts, Margaret, E., 2014, "Replication data for: Reverse Engineering Chinese Censorship: Randomized Experimentation and Participant Observation", <http://dx.doi.org/10.7910/DVN/26212>, Harvard Dataverse, V5

Chinese government censorship of social media constitutes the largest coordinated selective suppression of human communication in recorded history. Although existing research on the subject has revealed a great deal, it is based on passive, observational methods, with well known...

Replication data for: Systematic Bias and Nontransparency in US Social Security Administration Forecasts



May 8, 2015

Kashin, Konstantin; King, Gary; Soneji, Samir, 2015, "Replication data for: Systematic Bias and Nontransparency in US Social Security Administration Forecasts", <http://dx.doi.org/10.7910/DVN/28122>, Harvard Dataverse, V1

We offer an evaluation of the Social Security Administration demographic and financial forecasts used to assess the long-term solvency of the Social Security Trust Funds. This same forecasting methodology is also used in evaluating policy proposals put forward by Congress to modi...

Replication data for: Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts



May 8, 2015

Kashin, Konstantin; King, Gary; Soneji, Samir, 2015, "Replication data for: Explaining Systematic Bias and Nontransparency in US Social Security Administration Forecasts", <http://dx.doi.org/10.7910/DVN/28323>, Harvard Dataverse, V1

The accuracy of U.S. Social Security Administration (SSA) demographic and financial forecasts is crucial for the solvency of its Trust Funds, other government programs, industry decision making, and the evidence base of many scholarly articles. Because SSA makes public little rep...

Replication data for: A Unified Approach To Measurement Error And Missing Data: Overview



Mar 23, 2015

Blackwell, Matthew; Honaker, James; King, Gary, 2015, "Replication data for: A Unified Approach To Measurement Error And Missing Data: Overview", <http://dx.doi.org/10.7910/DVN/29606>, Harvard Dataverse, V1

Although social scientists devote considerable effort to mitigating measurement error during data collection, they often ignore the issue during data analysis. And although many statistical methods have been proposed for reducing measurement error-induced biases, few have been wi...

Replication data for: A Unified Approach To Measurement Error And Missing Data: Details And Extensions.



Mar 23, 2015


Blackwell, Matthew; Honaker, James; King, Gary, 2015, "Replication data for: A Unified Approach To Measurement Error And Missing Data: Details And Extensions." <http://dx.doi.org/10.7910/DVN/29610>, Harvard Dataverse, V1

Research Project/Archive




The Henry A. Murray Research Archive is the endowed, permanent repository for quantitative and qualitative research data at the Institute for Quantitative Social Science. Our collection comprises over 125 terabytes of data, audio, and video. More information about the Murray Archive can be found at our [website](#).


<




Diversity Datasets: Race, Ethnicity, Sexual Orientation, Religion Dataverse



Early Head Start Research and Evaluation Project Dataverse



Economic Theory and Demography Dataverse



Education Dataverse

>

[Q Find](#) [Advanced Search](#)

[Dataverses \(7\)](#)

[Datasets \(444\)](#)

[Files \(6,757\)](#)

Dataverse Category
[Research Project \(7\)](#)

Publication Date
[2007 \(303\)](#)
[2010 \(88\)](#)
[2009 \(30\)](#)
[2008 \(15\)](#)
[2013 \(8\)](#)

[More...](#)


Topic Classification Term
[1 \(231\)](#)
[mra \(177\)](#)
[yes \(159\)](#)
[mixed \(155\)](#)
[female, male \(153\)](#)

1 to 10 of 451 Results Sort ▾

« < Previous **1** 2 3 4 5 Next > »

Oral History of the Tenured Women in the Faculty of Arts and Sciences at Harvard University, 1981 📄


Jul 20, 2015

 Judith B. Walzer, 2009, "Oral History of the Tenured Women in the Faculty of Arts and Sciences at Harvard University, 1981", <http://hdl.handle.net/1902.1/00709>, Harvard Dataverse, V7

The purpose of this 1981 study was to document the history of women faculty at Harvard University. Of the then 13 women tenured in the Faculty of Arts and Sciences at Harvard in 1981, 11 professors and the husband of one deceased professor participated in interviews. Nine of these...

A Comparison of Cross-Generational Attitudes About Filial Obligations, 1982 📄


Jul 17, 2015

 Harold Feldman; Margaret Feldman, 2007, "A Comparison of Cross-Generational Attitudes About Filial Obligations, 1982", <http://hdl.handle.net/1902.1/00864>, Harvard Dataverse, V2

The purpose of this study was to assess the attitudes of students and their parents about their obligations toward each other and toward the elderly. A total of 483 high school and college students and one parent or older relative (mostly mothers) completed a six-page, closed-end...

Success and Failure Incidents From Self-Employed Women, 1979 📄

May 15, 2015

 Nancy Ann Flexman, 2007, "Success and Failure Incidents From Self-Employed Women, 1979", <http://hdl.handle.net/1902.1/00565>, Harvard Dataverse, V2

Organization or Institution

Avahan Dataverse (Bill & Melinda Gates Foundation)

[Harvard Dataverse](#) > **Avahan Dataverse**





In 2003, the Bill & Melinda Gates Foundation launched Avahan, the India AIDS Initiative, to reduce the spread of HIV in India. Avahan's ten-year charter had three distinct parts. The first was to build and operate a scaled HIV prevention program, with saturated coverage for key population most at risk, in the six states which account for the bulk of HIV infections in India. The second was to transfer the program to the Government of India and other implementers in the country; and the third encouraged the replication of best practices by fostering and disseminating learnings from the program.


Avahan provided funding and support to targeted HIV prevention programs in the six Indian states with the highest HIV prevalence, and along the nation's major trucking routes. Gathering and using data was critical for all of Avahan's goals to continuously refine the program and its many moving parts, to inform other HIV prevention efforts including the national prevention program and its direction, to measure impact and to capture best practices. Data in this Dataverse represent the full range of data collected and used by the lead implementing partners, and some evaluation, knowledge building and capacity building partners in Avahan. The range of data encompasses routine program monitoring data, survey data used for monitoring and for evaluation, and special studies to better understand the HIV epidemic in the program areas.


This dataverse has been divided into different sub-domains such as 'Migration Research', 'Media and Advocacy Studies', etc. Each sub-domain contains all the studies under that thematic area, which is further grouped by the institution that has conducted the study and by specific topics/key populations.

◀


**Avahan - Organizational
Development Monitoring
Dataverse**


**Avahan - Media and Advocacy
Dataverse**


**Avahan - Program Monitoring,
Evaluation, STI Monitoring and
STI Studies Dataverse**


**Avahan - Community Mobilization
Monitoring and Research
Dataverse**

▶

Search this dataverse...

 Find [Advanced Search](#)

Dataverses (8)

Datasets (44)

Files (261)

Dataverse Category
[Organization or Institution \(3\)](#)
[Research Project \(2\)](#)

Publication Date
[2014 \(28\)](#)
[2015 \(17\)](#)
[2013 \(7\)](#)

1 to 10 of 52 Results

 Sort ▾

« < Previous **1** 2 3 4 5 Next > »

STI_MnE Strategic Areas Dataverse (Avahan)

May 21, 2015 Avahan - Program Monitoring, Evaluation, STI Monitoring and STI Studies Dataverse



This dataverse provides tools, methodology, data and information related to publications based on program monitoring, evaluation and research studies. The monitoring data includes: centralized management information system (CMIS), STI management information system (STI-MIS). The...

Avahan_STI_MnE _Institutions Dataverse (Avahan Program Monitoring, Evaluation, STI Monitoring and STI Studies)

May 21, 2015 Avahan - Program Monitoring, Evaluation, STI Monitoring and STI Studies Dataverse



This dataverse provides tools, methodology, data and information related to publications based on program monitoring, evaluation and research studies. The monitoring data includes: centralized management information system (CMIS), STI management information system (STI-MIS). The...

[Avahan Dataverse](#)

Department



Harvard Dataverse > CfA Dataverse



This is the Astronomy data repository at Harvard. It is currently open to all scientific data from astronomical institutions worldwide. Administration and support is provided by the Harvard-Smithsonian Center for Astrophysics (CfA) in collaboration with Harvard Library (HL) and the Institute for Quantitative Social Science (IQSS). Infrastructure is provided by Harvard University Information Technology Services.

The Astronomy Dataverse Network plays an important role in fulfilling your Data Management Plan requirements (e.g. as mandated by NSF), and for providing data re-use and citation opportunities. Find out more about our team by exploring the Seamless Astronomy and Wolbach Library teams at the CfA. We are also connecting the Astronomy Dataverse to the indexing services provided by the SAO/NASA Astrophysical Data Service (ADS).

Search this dataverse... Find Advanced Search

Datasets (35)

Files (1,852)

Dataverse Category

Researcher (5)

Research Project (2)

Publication Date

2011 (69)

2013 (38)

2014 (24)

2015 (11)

2012 (7)

Author Name

Thaddeus, P. (38)

Dame, T. M. (25)

COMPLETE team (20)

Hartmann, Dap (19)

Bronfman, L. (7)

More...

Subject

Astronomy and Astrophysics (13)

Social Sciences (2)

Keyword Term

Astrophysics Data System (5)

1 to 10 of 149 Results

Sort

« < Previous 1 2 3 4 5 Next > »

The International Political Economy Data Resource

Aug 30, 2015 - Benjamin A.T. Graham's Dataverse

Graham, Benjamin A.T., 2015, "The International Political Economy Data Resource", <http://dx.doi.org/10.7910/DVN/28003>, Harvard Dataverse, V5 [UNF:6:pVw4sA44AUMaaHhW3Qfv0Q==]

Quantitative scholars in international relations often draw repeatedly on the same sources of country-year data across a diverse range of projects. The IPE Data Resource seeks to provide a public good to the field by standardizing and merging together 55 of core IPE data sources...

APEX CMZ SHFI-1 survey

Aug 20, 2015 - APEX CMZ H2CO Dataverse

Ginsburg, Adam, 2015, "APEX CMZ SHFI-1 survey", <http://dx.doi.org/10.7910/DVN/27601>, Harvard Dataverse, V5

APEX SHFI-1 survey of the CMZ

Replication data for: Deep 3.8 Micron Observations of the Trapezium Cluster

Jun 23, 2015 - August Muench Dataverse

Muench, August; Alves, Joao; Lada, Charles; Lada, Elizabeth, 2015, "Replication data for: Deep 3.8 Micron Observations of the Trapezium Cluster", <http://dx.doi.org/10.7910/DVN/28977>, Harvard Dataverse, V2

This is the data behind the paper, "Deep 3.8 Micron Observations of the Trapezium Cluster," by Lada et al. (2004). It includes FITS image files and reference comparison files that would prove useful for interpreting the FITS image files. A note on the images: two images are given...

Replication Data for: YSOVAR: Mid-infrared Variability in the Star-forming Region Lynds 1688

May 15, 2015 - Hans Moritz Günther Dataverse

Guenther, Hans Moritz, 2015, "Replication Data for: YSOVAR: Mid-infrared Variability in the Star-forming Region Lynds 1688", <http://dx.doi.org/10.7910/DVN/LNRC6A>, Harvard Dataverse, V1

This dataset contains all YSOVAR lightcurves of L1688 as published in YSOVAR: Mid-infrared Variability in the Star-forming Region Lynds 1688. The data

Courses



[Project TIER: Teaching Integrity in Empirical Research Dataverse](#) [Visit the Project TIER website](#)

Harvard Dataverse > [Project TIER: Teaching Integrity in Empirical Research Dataverse](#)



This dataverse supports a protocol for teaching undergraduates to document the statistical analysis they do for empirical research projects in such a way that their results are completely reproducible and verifiable. The protocol is guided by the principle that the documentation prepared to accompany an empirical research project should be sufficient to allow an independent researcher to replicate easily and exactly every step of the data management and analysis that generated the results reported in the study. You will find in this dataverse examples of the protocol as applied in senior thesis and introductory statistics projects. We hope that requiring students to follow this protocol will not only teach them how to document their research appropriately, but also instill in them the belief that it is an important professional responsibility to do so. For more information, visit the [Project TIER website](#).



Search this dataverse...

Q Find [Advanced Search](#)

 **Dataverses (4)**

 **Datasets (22)**

 **Files (194)**

Dataverse Category

[Teaching Course \(3\)](#)
[Organization or Institution \(1\)](#)

Publication Date

[2015 \(16\)](#)
[2014 \(7\)](#)
[2013 \(3\)](#)

Author Name

[Adams, Gaines \(1\)](#)
[Auer, Alexis \(1\)](#)
[Bonsu, Nicole \(1\)](#)
[Bracker, Mason \(1\)](#)
[Brennan, Claire \(1\)](#)

1 to 10 of 26 Results

Sort ▾

« < Previous **1** 2 3 Next > »

[Colgate Dataverse](#) (Colgate University)
Jul 8, 2015



[Haverford Dataverse](#) (Haverford College)
Jun 17, 2015



[Simplicity Versus WAR: Examining Salary Determinations in Major League Baseball's Arbitration and Free Agent Markets](#)
May 29, 2015



Studnitzer, Joshua, 2015, "Simplicity Versus WAR: Examining Salary Determinations in Major League Baseball's Arbitration and Free Agent Markets", <http://dx.doi.org/10.7910/DVN/28782>, Harvard Dataverse, V2

This paper examines salaries given to arbitration eligible players in Major League Baseball from 2008-2013 and compares them to free agent contracts from the same period. Anecdotal evidence suggests that simpler statistics are more successful in Major League Baseball's final offer...

[More...](#)

Thank you!

Any questions?

Contact: equigley@iq.harvard.edu

Learn more: dataverse.org

Demo Dataverse: dataverse-demo.iq.harvard.edu



@dataverseorg

References

Altman M. A Fingerprint Method for Verification of Scientific Data. In A Fingerprint Method for Verification of Scientific Data. Springer-Verlag; 2008.

Altman M, King G. A Proposed Standard for the Scholarly Citation of Quantitative Data. D-Lib Magazine [Internet]. 2007;13(3/4).