



Sharing Your Data with Dataverse

Elizabeth Quigley

User Experience Lead

The Dataverse Project (@dataverseorg)

equigley@iq.harvard.edu

What is Dataverse?

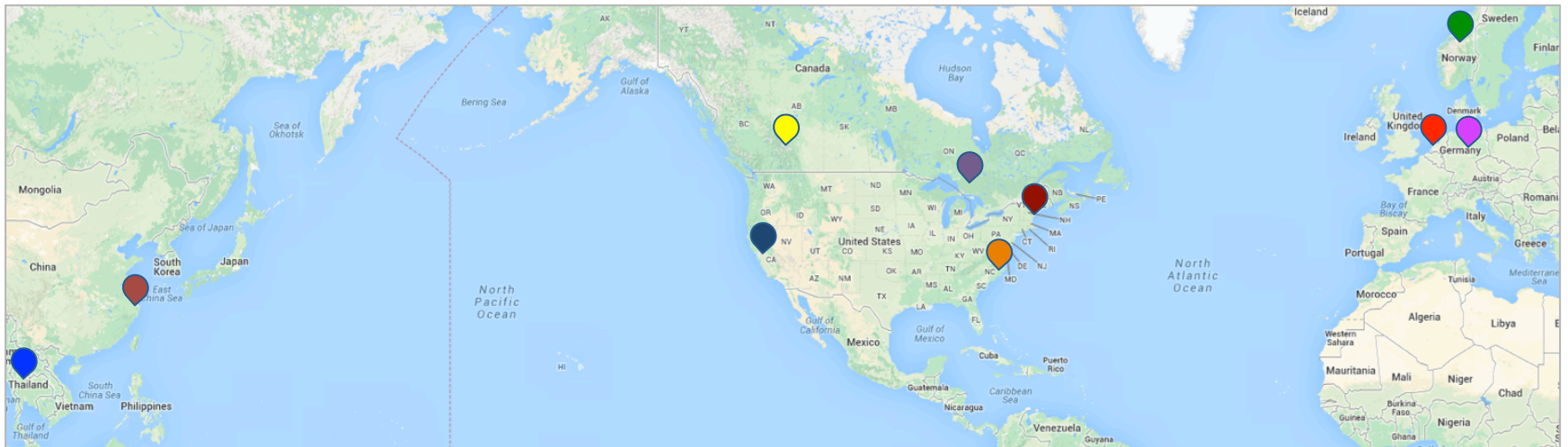
A software framework for publishing, citing and preserving research data (open source on [github](#) for others to install)

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

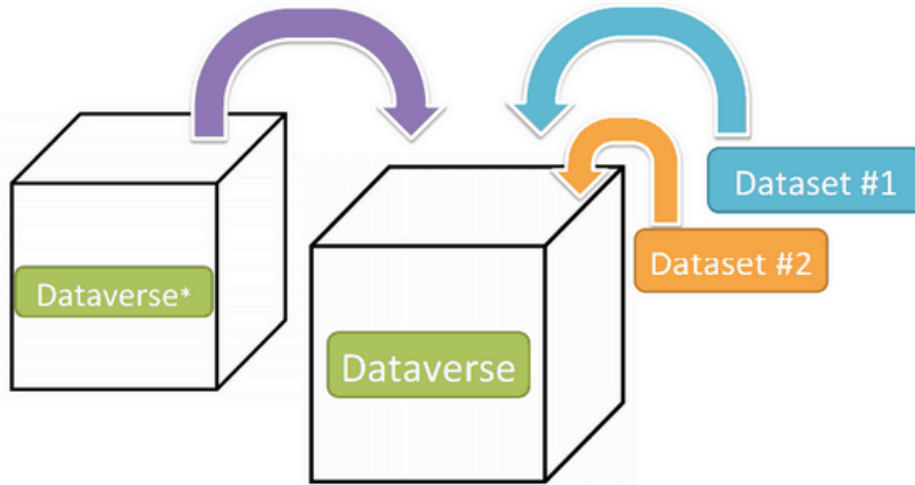
The Dataverse Community



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?

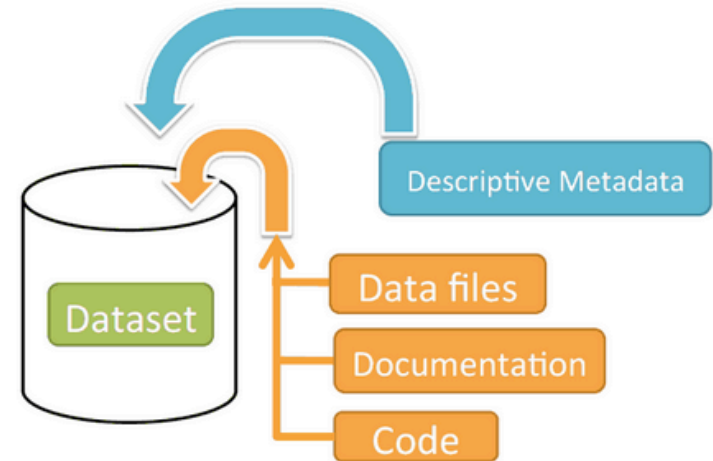
Schematic Diagram of a **Dataverse** in Dataverse 4.0



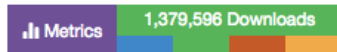
Container for your **Datasets** and/or **Dataverses***

* 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)**

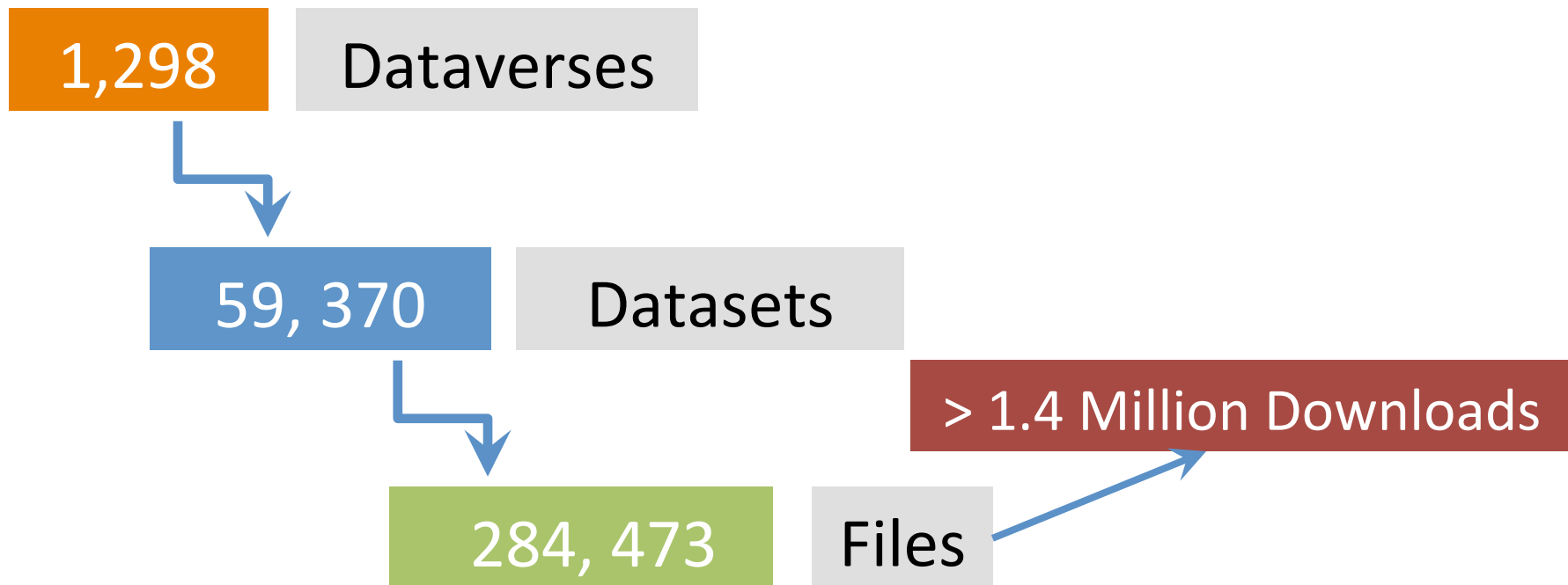
Sep 10, 2015
- Administrative Regions**

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 October 25, 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 symbol, 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 + PUBLISHING DATA WITH DATAVERSE

Benefits for Researchers Using Dataverse

- Add your own custom branding and increase your datasets' visibility by embedding them on your website with our widgets.
- Safe and long-term data storage in preservation format.
- Share your data with everyone, or only specific individuals you approve.
- No need to translate data when statistical software formats change.
- Domain specific metadata: making it easy for others to find your data and associated scholarship.
- Allow users to download your data in any format and run many advanced statistical methods online.

Welcome to Harvard University's Data Sharing Portal

The institutional data available on this site are being shared to encourage the creativity and innovation of the Harvard University community. In this collection you will find both datasets that are open to the world, and some that are only accessible to members of the University. Please check in often to see what new sources of data we add to the collection.

HARVARD SHARED INSTITUTIONAL DATA

1 to 5 of 5 Results

Sort ▾

Harvard Faculty Finder



Sep 11, 2015

Waldo, Jim, 2015, "Harvard Faculty Finder", <http://dx.doi.org/10.7910/DVN/PLMNRW>, Harvard Dataverse, V1

The Harvard Faculty Finder creates an institution-wide view of the breadth and depth of Harvard faculty and scholarship, and it helps students, faculty, administrators, and the general public locate Harvard faculty according to research and teaching expertise. More inform...

Library Cloud



Sep 11, 2015

Weinberger, David, 2015, "Library Cloud", <http://dx.doi.org/10.7910/DVN/WXAKPE>, Harvard Dataverse, V1

Catalog metadata++ about Harvard Library's collection + VIA images + Archival items. Total of about 17M items. Provides an API, which is documented at <https://wiki.harvard.edu/confluence/display/LibraryStaffDoc/Library+Cloud>. A overview is attached as the data file. Entry point...

Harvard Library Bibliographic Dataset

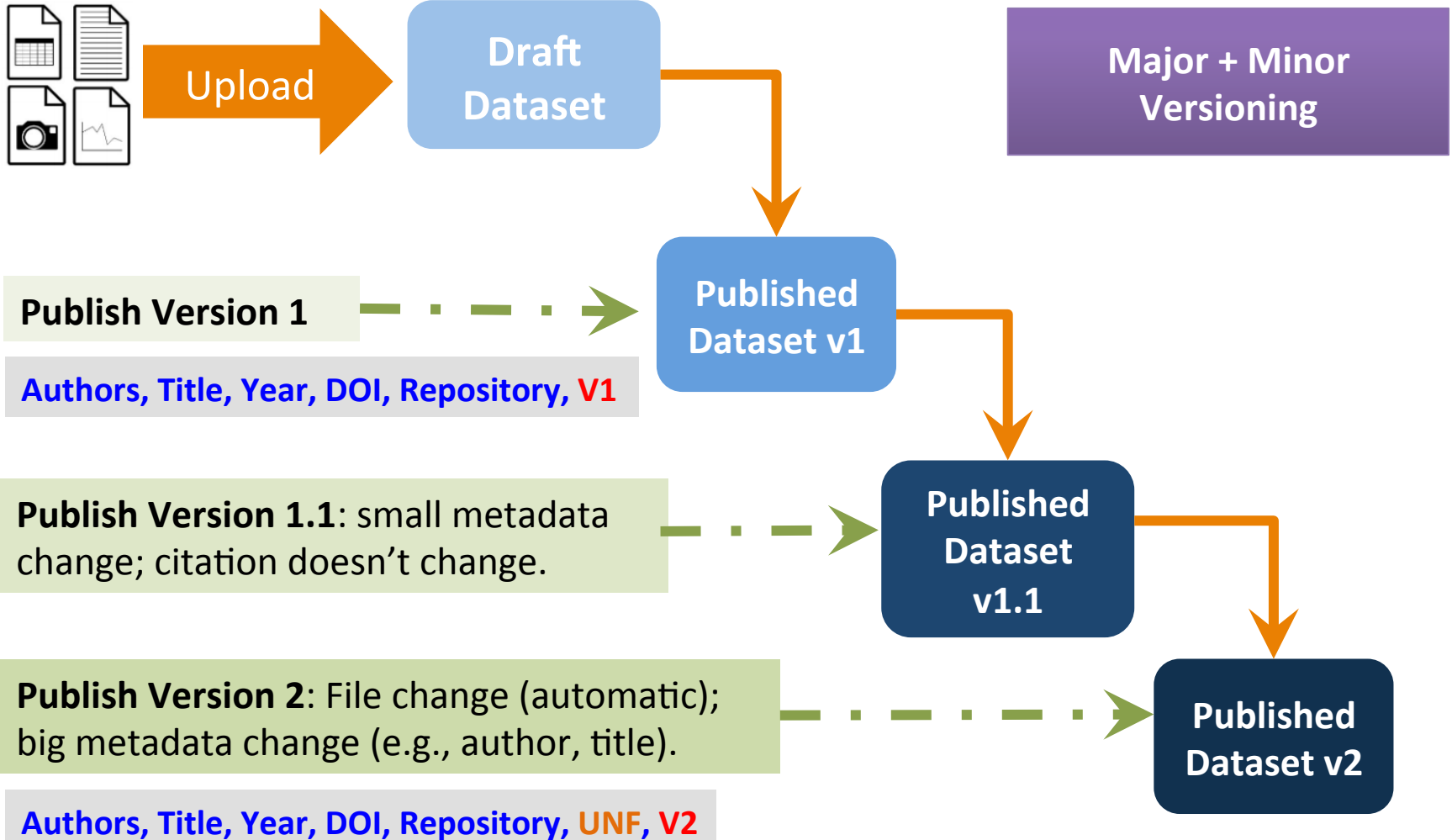


Sep 11, 2015

Waldo, Jim, 2015, "Harvard Library Bibliographic Dataset", <http://dx.doi.org/10.7910/DVN/LZDQYN>, Harvard Dataverse, V1

This dataset contains over 12 million bibliographic records for materials held by the Harvard Library, including books, journals, electronic resources, manuscripts, archival materials, scores, audio, video and other materials. The metadata has been created, acquired and modified...

Rigorous Data Publishing Workflows





LIVE DEMO TIME!

demo.dataverse.org

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... [Advanced Search](#)

Dataverses (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](#).

Navigation carousel with four categories:

- [Diversity Datasets: Race, Ethnicity, Sexual Orientation, Religion Dataverse](#)
- [Early Head Start Research and Evaluation Project Dataverse](#)
- [Economic Theory and Demography Dataverse](#)
- [Education Dataverse](#)

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

1 to 10 of 451 Results

Sort: [icon] | Page: 1 | < Previous | 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

- [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 \(153\)](#)

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

Try out Dataverse: demo.dataverse.org



@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).