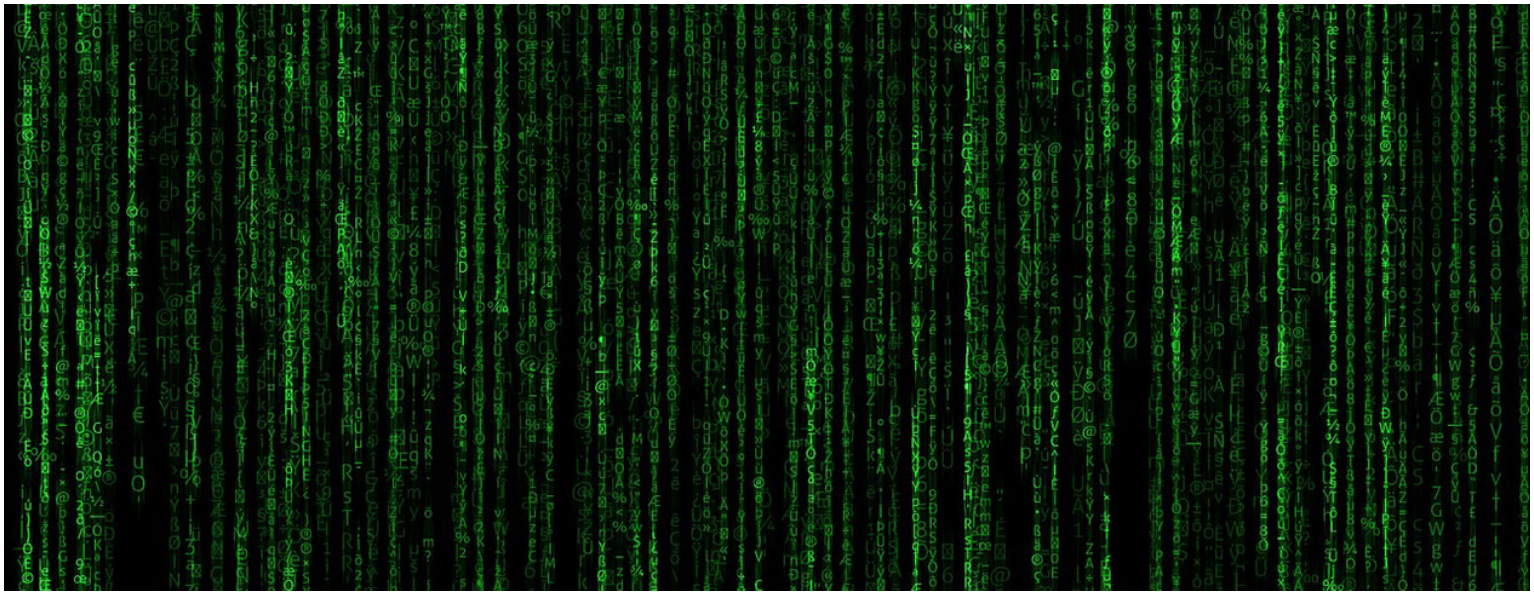


Gary King: an update on Dataverse

ISSN 1047-1987

POLITICAL ANALYSIS

www.pan.oxfordjournals.org

VOLUME 22 NUMBER 1 WINTER 2014

Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments
Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto

A New Multinomial Accuracy Measure for Polling Bias
Kai Arzheimer and Jocelyn Evans

Evaluating Sensitivity of Parameters of Interest to Measurement Invariance in Latent Variable Models
Daniel L. Oberski

Defining and Modeling State-Dependent Dynamic Systems
Justin Esarey and Jacqueline H. R. DeMeritt

Scoring from Contests
Keith E. Schnakenberg and Elizabeth Maggie Penn

Estimating the Political Center from Aggregate Data: An Item Response Theory Alternative to the Stimson Dyad Ratios Algorithm
Anthony J. McGann

Experiments to Reduce the Over-Reporting of Voting: A Pipeline to the Truth
Michael J. Hanmer, Antoine J. Banks, and Ismail K. White

The official journal of The Society for Political Methodology and the Political Methodology Section of the American Political Science Association

OXFORD
UNIVERSITY PRESS

Political Analysis

Political Analysis (http://pan.oxfordjournals.org/) chronicles the exciting developments in the field of political methodology, with contributions to empirical and methodological scholarship outside the diffuse borders of political science. It is published on behalf of The Society for Political Methodology (http://polmeth.wustl.edu/). *Political Analysis* (http://pan.oxfordjournals.org/) is ranked #5 out of 157 journals in Political Science by 5-year impact factor, according to the 2012 ISI Journal Citation Reports. Like *Political Analysis* on Facebook and follow @PolAnalysis (https://twitter.com/polanalysis) on Twitter.

READ MORE (HTTP

BY GARY KING (HTTPS://BLOG.OUP.COM/AUTHORS/GARY-KING/) AND R. MICHAEL ALVAREZ (HTTPS://BLOG.OUP.COM/AUTHORS/R-MICHAEL-ALVAREZ/)

DECEMBER 7TH 2014

At the American Political Science Association (http://www.apsanet.org/) meetings earlier this year, Gary King (http://gking.harvard.edu/), Albert J. Weatherhead III University Professor at Harvard University, gave a presentation on Dataverse (http://gking.harvard.edu/presentations/dataverse-sharing-research-data-building-social-science). Dataverse is an important tool that many researchers use to archive and share their research materials. As many readers of this blog may already know, the journal that I co-edit, *Political Analysis (http://pan.oxfordjournals.org/)*, uses Dataverse to archive and disseminate the replication materials (https://blog.oup.com/2014/11/replication-data-access-transparency-social-science/) for the articles we publish in our journal. I asked Gary to write some remarks about Dataverse, based on his APSA presentation. His remarks are below.

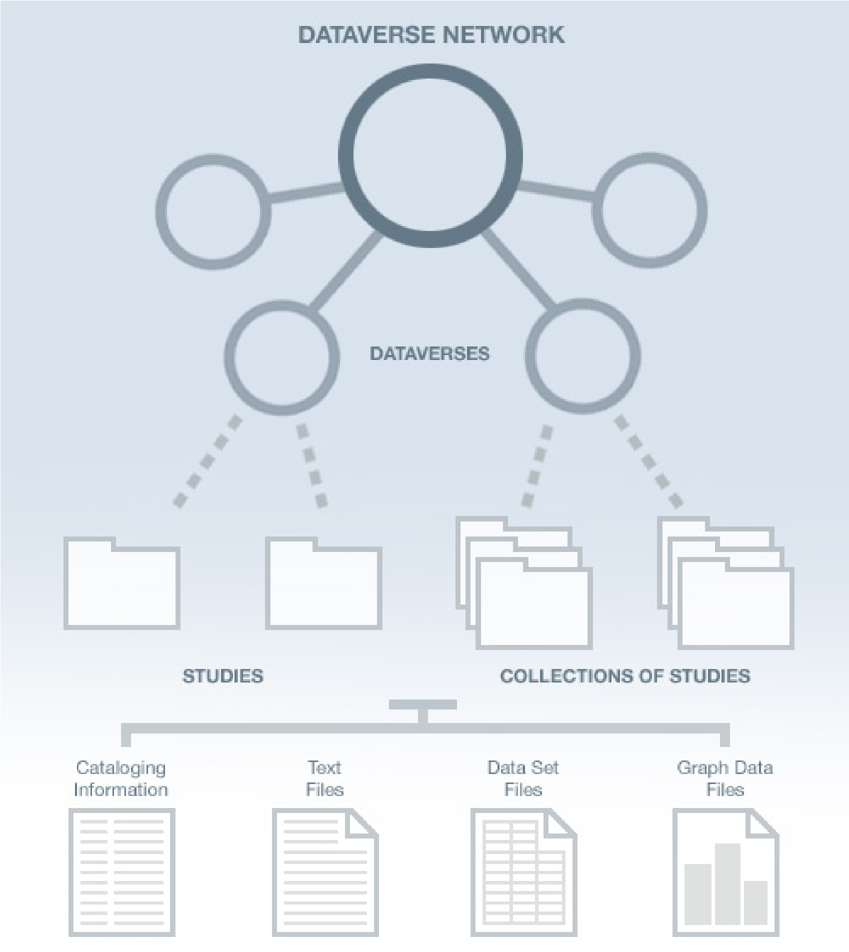
* * * * *

An update on Dataverse

By Gary King

If you're an academic researcher, odds are you're not a professional archivist and so you probably have more interesting things to do when making data available than following the detailed protocols and procedures established over many years by the archiving community. That of course might be OK for any one of us but it is a terrible loss for all of us. The Dataverse Network Project (http://thedata.org) offers a solution to this problem by eliminating transaction costs and changing the incentives to make data available by giving you substantial web visibility and academic citation credit for your data and scholarship (King, 2007 (http://gking.harvard.edu/files/abs/dvn-abs.shtml)). Dataverse Networks are installed at universities and other institutions around the world (e.g., here (http://www.oxfordjournals.org/our_journals/polana/for_authors/general.html) is the Dataverse network at Harvard's IQSS (http://www.iq.harvard.edu/)), and represent the world's largest collection of social science research data. In recent years, Dataverse has also been adopted by an increasingly diverse array of other fields and protocols and procedures are being built out to enable numerous fields of science, social science, and the humanities to work together.

With a few minutes of set-up time, you can add your own Dataverse to your homepage with a list of data sets or replication data sets you make available, with whatever levels of permission you want for the broader community, and a vast array of professional services (e.g., here's my Dataverse (http://gking.harvard.edu/data) on my homepage (http://gking.harvard.edu/)). People will be able to more easily find your data and homepage, explore your data and scholarship, find connections to other resources, download data in any format, and learn proper ways of citing your work. They will even be able to analyze your data while still on your web site with a vast array of statistical methods through the transparent and automated connection Dataverse has built to Zelig: Everyone's Statistical Software (http://projects.iq.harvard.edu/zelig), and through Zelig to R (http://www.r-project.org/). The result is that your data will be professionally preserved and easier to access — effectively automating the tasks of professional archiving, including citing, sharing, analyzing, archiving, preserving, distributing, cataloging, translating, disseminating, naming, verifying, and replicating data.



Dataverse Network Diagram, by Institute for Quantitative Social Science. CC-BY-2.0 via Wikimedia Commons (http://commons.wikimedia.org/wiki/File:Dataverse_Network_Diagram.jpg).

Dataverse is an active project with new developments in software, protocols, and community connections coming rapidly. A brand new version of the code, written from scratch, will be available in a few months. Through generous grants from the Sloan Foundation (http://www.sloan.org/), we have been working hard on eliminating other types of transaction costs for capturing data for the research community. These include deep integration with scholarly journals so that it can be trivially easy for an editor to encourage or require data associated with publications to be made available. We presently offer journals three options:

- Do it yourself. Authors publish data to their own dataverse, put the citation to their data in their final submitted paper. Journals verify compliance by having the copyeditor check for the existence of the citation.
- Journal verification. Authors submit draft of replication data to Journal Dataverse. Journal reviews it, and approves it for release. Finally, the dataset is published with a formal data citation and back to the article. (See, for example, the *Political Analysis* Dataverse (https://thedata.harvard.edu/dvn/dv/pan), with replication data back to 1999.)
- Full automation: Seamless integration between journal submission system and Dataverse; Automatic Link created between article and data. The result is that it is easy for the journal and author and many errors are eliminated.

Full automation in our third option is where we are heading. Already today, in 400 scholarly journals in the Open Journal System (https://pkp.sfu.ca/ojs/), the author enters their data as part of submission of the final draft of the accepted paper for publication, and the citation, permanent links between the data and the article, and formal preservation is taken care of, all automatically. We are working on expanding this as an option for all of OJS’s 5,000+ journals, and to a wide array of other scholarly journal publishers. The result will be that we capture data with the least effort on anyone’s part, at exactly the point where it is easiest and most important to capture.

We are also working on extending Dataverse to cover new higher levels of security that are more prevalent in big data collections and those in public health, medicine, and other areas with informative data on human subjects. Yes, you can preserve data and make it available under appropriate protections, even if you have highly confidential, proprietary, or otherwise sensitive data. We are working on other privacy tools (http://privacytools.seas.harvard.edu/) as well. We already have an extensive versioning system in Dataverse, but are planning to add support for continuously updated data such as streamed from sensors, tools for online fast data access, queries, visualization, analysis methods for when data cannot be moved because of size or privacy concerns, and ways to use the huge volume of web analytics to improve Dataverse and Zelig.

This post comes from the talk I gave at the American Political Association Meetings August 2014, using these slides (http://gking.harvard.edu/presentations/dataverse-sharing-research-data-building-social-science). Many thanks to Mike Alvarez for inviting this post.

Featured image: Matrix code computer by Comfreak (http://pixabay.com/en/users/Comfreak-51581/). CC0 via Pixabay (http://pixabay.com/en/matrix-code-computer-pc-data-356024/).

***Gary King** is the Albert J. Weatherhead III University Professor at Harvard University, based in the Department of Government in the Faculty of Arts and Sciences. He also serves as Director of the Institute for Quantitative Social Science. King and his research group develop and apply empirical methods in many areas of social science research, focusing on innovations that span the range from statistical theory to practical application. You can follow him at @kinggary (https://twitter.com/kinggary). He is the author of "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It" (http://oxford.ly/1rOJlv3) (available to read for free for a limited time) in Political Analysis.*

***R. Michael Alvarez** is a professor of Political Science at Caltech. His research and teaching focuses on elections, voting behavior, and election technologies. He is editor-in-chief of Political Analysis (http://pan.oxfordjournals.org/) with Jonathan N. Katz.*

POSTED IN: EDUCATION (HTTPS://BLOG.OUP.COM/CATEGORY/SOCIAL_SCIENCES/EDUCATION/) / JOURNALS (HTTPS://BLOG.OUP.COM/CATEGORY/SUBTOPICS/OXFORDJOURNALS/) / MEDIA (HTTPS://BLOG.OUP.COM/CATEGORY/ARTS_AND_HUMANITIES/MEDIA/) / POLITICS (HTTPS://BLOG.OUP.COM/CATEGORY/SOCIAL_SCIENCES/POLITICS-POLITICAL-SCIENCE/) / SOCIAL SCIENCES (HTTPS://BLOG.OUP.COM/CATEGORY/SOCIAL_SCIENCES/) / TECHNOLOGY (HTTPS://BLOG.OUP.COM/CATEGORY/SCIENCE_MEDICINE/TECHNOLOGY/)