The Dataverse Network – Combining Preservation with Scholarly Recognition

Mercè Crosas, Ph.D.
Director of Product Development
Institute for Quantitative Social Science (IQSS)
Harvard University
Acknowledgements

This work is done with contribution from:

• The Dataverse Network software development team at IQSS: Gustavo Durand, Ellen Kraffmiller, Kevin Condon, Leonid Andrev, Bob Treacy, Michael Heppler, Steve Kraffmiller

• Gary King, Albert J. Weatherhead III University Professor, Harvard University

• Micah Altman, Senior Research Scientist, IQSS, Harvard University
The Problem with Data

Professional archives focus on long term access by the wider community

- Persistent identifiers
- Fixity
- Backups and recovery
- Metadata standards
- Conversion standards
- Preservation standards

... but do not attract maximum contribution from data owners

Data owners focus on recognition and control

- Branding and visibility
- Data discovery
- Ease of use
- Scholarly citation
- Control over updates
- Terms of access and use

... but do not assure long-term use as would a professional archive
The Dataverse Network

Enables people to solve a political problem through technology – Combining centralized archiving with distributed ownership
Use Case 1: Dataverse for Researchers – Web Visibility and Branding

Sidney Verba
Carl H. Pforzheimer University Professor
Department of Government, Harvard University (email)

Sidney Verba’s Bio
Download CV

Professor Verba is the Carl H. Pforzheimer University Professor Emeritus and Research Professor of Government at Harvard University. He served as Director of the Harvard University Library for twenty-four years. At Harvard, he has also been chair of the Department of Government, Associate Dean of the Faculty for Undergraduate Education, and Associate Provost, among several other senior administrative posts. In addition, during these years he has also served as chair of the Board of Directors of the Harvard University Press and has been the author of University-wide reports on many complex subjects. Professor Verba received his B.A. from Harvard and his Ph.D. from Princeton. He has taught at Princeton, Stanford, the University of Chicago, and at Harvard for over thirty years.


Recent Publications
1. Fairness, Equality, and Democracy: Three Big Ideas
2. Family Ties: Understanding the Intergenerational Transmission of Participation
3. Equal at the Starting Line: Creating Parliaments across Generations and Among Groups
4. Who’s In? Political Equations, the Record of the Past Decade

What if the Dream of Participation Turned Out to be a Nightmare

Sidney Verba

Home
Publications
Data

Data

Sidney Verba
Carl H. Pforzheimer University Professor
Department of Government, Harvard University (email)

Search

Sidney Verba Dataverse

All IDSS Datasets

Sidney Verba

Search

Sidney Verba Dataverse

Sort by:

Studies: 11

Web Site Research-Not Released
by Sidney Verba

Abstract: Web site is a new data set for my research not yet released.

New Data for "BigMo"
by Sidney Verba

Abstract: Data are a file of which content has not been released. The original is called "BigMo" and is on the CV listing.

General Social Survey (GSS)

Abstract: The General Social Survey (GSS) conducts basic scientific research on the structure and development of American society with a data-collection program designed to both monitor social change within the...
• The researcher is the administrator of his/her dataverse.
• Manages data curation workflow.
• Sets permissions and terms of use to datasets.

**A Researcher ...**
Use Case 2: Dataverse for Journals – Replication of Published Work
Use Case 2: Dataverse for Journals – From Publication to Data

Journal, Book

Datasets, text, images

Formal Data Citation

Persistent identifier and url that never changes

Universal Numerical Fingerprint (UNF) to verify dataset
• A Journal for replication data can be set as an open dataverse.
• Authors upload data after registering.
• Journal reviews data before releasing it.
The Dataverse provides:

- **Versioning** – Keeps all changes to data and metadata to allow citation of old versions.
- **Deaccession** – Doesn’t delete studies permanently so citation is valid forever.
- **Universal Numeric Fingerprints** – Fixity that survives changes of format.
Use Case 3: Dataverse for Archives – Re-formatting for Preservation

Preservation and verifiable format by:
1. Extracting variable and format metadata.
2. Converting dataset to a format independent of software package.
3. Applying cryptographic algorithm to canonical format.
4. Obtaining an alphanumeric string based on semantic content:
   UNF5:EKgHvTNfkkS86dNzABhNw==
The Dataverse facilitates:
- Multiple copies of data
- Metadata exchange with other systems.
**New:** A ‘wiki’ dataverse is an archive to share and improve data through contributions (with a moderator to review them before their release).
Multiple contributors can be updating the same study.
Changes are only visible after being reviewed.
A new way to gather more data for a study and improve its cataloging.
Current Usage

• The IQSS Dataverse Network at Harvard holds:
  • 257 public dataverses from around the world
  • 36,000 studies (mostly in Social Science)
  • 640,000 files
• ... and there are over 10 more installations of the Dataverse Network at other Institutions.
Create a dataverse at: http://dvn.iq.harvard.edu
Or install the Dataverse Network software for your Institution.
For more information, go to: http://thedata.org
Contact me at: mcrosas@hmdc.harvard.edu

THANK YOU!