SUPPORTING DATA ACCESS and MANAGEMENT with THE DATaverse

Mercè Crosas, Ph.D.
Director of Product Development
Institute for Quantitative Social Science, Harvard University
mcrosas@iq.harvard.edu
WHY did we build the Dataverse Network?
A PROBLEM
Different needs for Archives, Data Libraries, Researchers, Journals, ...

- We need to preserve the data forever
- We need persistent links
- I want recognition for my data
- I want to curate the data
- I want to review submissions and control the access
A SOLUTION

Centralized Professional Archiving with
Distributed Control and Recognition

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

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

Mercè Crosas, LIBER June 2011
HOW does it work?
DATAVERSE NETWORK

A centralized software installation and data repository

Individual distributed data archive with its own branding
RICH DATA SUPPORT

Data Files

Tabular Data:
STATA
SPSS
CSV + control card
Tab delimited + DDI

Social Network Data:
GraphML
+ complementary files in any format

- Convert to preservation format (primary data + variable metadata)
- Calculate Universal Numerical Fingerprint (UNF)
- Define Terms of Use and Permissions
- Download in multiple formats
- Download a subset of the data
- Generate summary statistics
- Apply Zelig (R) statistical methods
- Visualize time series
# FORMAL, PERSISTENT DATA CITATIONS

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sidney Verba; Joseph DiNunzio; Christina Spaulding, 1985, &quot;Harvard Sexual Harassment Survey, 1983&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>1985</td>
</tr>
<tr>
<td>Title</td>
<td>&quot;Harvard Sexual Harassment Survey, 1983&quot;</td>
</tr>
<tr>
<td>Persistent URL and ID</td>
<td><a href="http://hdl.handle.net/1902.1/00686">Persistent URL</a> UNF:3:eG6QJsi2VsgIMV49FtGcDA== Murray Research Archive [Distributor] V1 [Version]</td>
</tr>
</tbody>
</table>
DATA VERSIONING AND MANAGEMENT

Contributors, curators, admins view

End user view

Christopher Casillas; Peter K. Enns; Patrick C. Wohlfarth, "How Public Opinion Constrains the Supreme Court", http://hdl.handle.net/1902.1/14568 UNF: 5:YBYuXzOp6STpakyTEIoScQ== V1 [Version]

Christopher Casillas; Peter K. Enns; Patrick C. Wohlfarth, "How Public Opinion Constrains the Supreme Court", http://hdl.handle.net/1902.1/14568 UNF: 5:dI8qi49P0uIB9pLfXA3RCw== V2 [Version]
A study is never permanently deleted, after is released. Curators or Admins can **deaccession** the study.

Christopher Casillas; Peter K. Enns; Patrick C. Wohlfarth, "How Public Opinion Constrains the Supreme Court", [http://hdl.handle.net/1902.1/14568](http://hdl.handle.net/1902.1/14568) UNF: 5:dI8qi49P0uIB9pLfXA3RCw== VN [Version]

[Go to this other study]
WORKFLOWS FOR ALL

Regular Dataverse

Open Dataverse

Wiki Dataverse

Option A - The admin or curator adds Mike as a contributor. Mike can now add studies and edit his own studies in this Dataverse.

or

Option B - The admin or curator adds Mike as a contributor. Mike can add studies and edit any study in this Dataverse.

Ann creates an account and starts adding and editing her own studies in an open Dataverse. The admin or curator reviews the studies before releasing them.

Josep creates an account and starts adding new studies and editing any study in a wiki Dataverse. The admin or curator reviews the studies before releasing them.
REGULAR DATAVERSE
OPEN DATaverse
This page contains replication data and code for the articles published in the Review of Economics and Statistics. Please notice that most articles in volumes previous to volume 92 were not subject to the journal's data availability policy.

Review of Economics and Statistics

Sort By: ▼

Replication data for: New Eurocoin: Tracking Economic Growth in Real Time
by Filippo Altissimo; Riccardo Cristadoro; Mario Forni; Marco Lippi; Giovanni Veronese

Replication data for: Inequality and Redistribution: Evidence from U.S. Counties and States, 1890-1930
by Rodney Ramcharan

Replication data for: Corruption and Bilateral Trade Flows: Extortion or Evasion?
by Pushan Dutt; Daniel Traca

Replication data for: Universal Vouchers and Racial and Ethnic Segregation
by Eric Brunner; Jennifer Imazeki; Stephen Ross
WIKI DATaverse
Data

Election Data Archive Dataverse

The Harvard Election Data Archive is a space to share and improve election data. You can create an account and either correct the cataloging information for the studies in this dataverse or upload new data files. Thank you to our colleagues Michael McDonald and David Bradlee for their help with data collection.

The Harvard Election Data Archive by Stephen Ansolabehere and Jonathan Rodden is licensed under a Creative Commons Attribution 3.0 Unported License.

Election Data Archive

Sort By:  

Studies: 20

New Jersey Data Files
by Stephen Ansolabehere; Jonathan Rodden
Abstract: New Jersey Shapefile and Election Results

New Hampshire Data Files
by Stephen Ansolabehere; Jonathan Rodden
Abstract: New Hampshire Shapefile and Election Results
DATAVERSE
BEYOND
SOCIAL SCIENCE
The whole-Galaxy CO survey presented in Dame et al. (2001) is a composite of 37 separate surveys that are described and numbered in Table 1 of the paper. The data from most of these surveys can be accessed by clicking on the survey number in the map below, which is Figure 1 from the paper. Larger composites of these individual surveys are available from the link below.
- Study and variable metadata are exported into **XML** (Dublin Core, Data Descriptive Initiative - DDI, ...) and MARC.
- **OAI-PMH** for harvesting metadata.
- **LOCKSS** for data duplication in multiple locations.
- **Z39.50** for distributed search.
- **E-Z Proxy** to authenticate for data access.
THANK YOU
GRÀCIES
http://thedata.org
Should Libraries play a role promoting good practices and standards for data sharing and management?

Should research institutions and funding agencies require data sharing? and should their Libraries be the place for data deposit?