Working with Data at its Source: Empowering Social Science Researchers to Share and Document Their Data for Archiving and Discovery in an Institutional Repository

Matt Marostica, Economics & Political Science Librarian
Ron Nakao, Data & Computational Social Science Librarian
Stanford University Libraries

40th Annual IASSIST Conference (June 6, 2014)
Storyline

• A Tale of Two Platforms – Stanford Digital Repository & data.stanford.edu
• Adding Data to Collection Development
• Case Study – DIME Study Data
• SDR Self-Deposit
A Tale of Two Platforms

• Stanford Digital Repository
  – Long-term Preservation
  – Hydra (Fedora, Solr, Blacklight, Moab)

• data.stanford.edu
  – Metadata creation
  – Drupal
Stanford Digital Timeline

- (2000) Data Extraction Web Interface (DEWI)
- (2009) Electronic Theses & Dissertations (ETDs)
- (2010) Google Books ingest began
- (2011) DMP more forceful requirement by NSF
- (2011) data.stanford.edu goes live
Incentives to Data Sharing

- Faculty are busy and their “culture of research” is difficult to change.
- Incentive for change: saving time
- Incentive for change: easy & helps their research
- Incentive for change: quick gratification
- Incentive for change: saving time
Pilot with Professor Michael Rosenfeld (Sociology)

- “Painless” Researcher creation of metadata
- Quick turnaround to make data usable
- Public & free download of data Data page
  “ownership” for updating information
- Collect data download statistics
# Data Metadata

<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Citation</strong></td>
<td>Abstract, Principal Investigator, Funding Agency, Bibliographic Citation, Contact Email</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Introduction, Acknowledgements</td>
</tr>
<tr>
<td><strong>Methodology</strong></td>
<td>Universe, Unit of Analysis, Type of data collection, Time span, Time of data collection, Geographic coverage, Smallest geographic unit, Sample description, Sample response rate, Weights</td>
</tr>
<tr>
<td><strong>Documentation</strong></td>
<td>Document file(s), Web site or document download link(s)</td>
</tr>
<tr>
<td><strong>Data Download Link(s)</strong></td>
<td>Data file(s)</td>
</tr>
<tr>
<td><strong>Notes</strong></td>
<td>Errata, Data Notes</td>
</tr>
<tr>
<td><strong>News</strong></td>
<td>News Coverage</td>
</tr>
</tbody>
</table>
Is There a Role for the Subject Librarian/Department Liaison in Data Curation?

• Knowledge of the Field: what types of outcomes are researchers attempting to explain?

• Personal Relationships

• Knowledge of institutional and non-institutional resources for data curation
Database on Ideology, Money in Politics, and Elections (DIME)

Abstract:

The Database on Ideology, Money in Politics, and Elections (DIME) is intended as a general resource for the study of campaign finance and ideology in American politics. The database was developed as part of the project on Ideology in the Political Marketplace, which is an on-going effort to conduct a comprehensive ideological mapping of political elites, interest groups, and donors using the common-space CFscore scaling methodology (Bonica 2013). Constructing the database required a large-scale effort to compile, clean, and process data on contribution records, candidate characteristics, and election outcomes from various sources. The resulting database contains over 100 million political contributions made by individuals and organizations to local, state, and federal elections spanning a period from 1979 to 2012. A corresponding database of candidates and committees provides additional information on state and federal elections.

Principal Investigator:

Adam Bonica

How to Cite this Dataset:


For CFscore measures of ideology, please cite:

Contact Email:
DIME Results

• DIME in first nine months
• 7,413 hits
• 1,506 documentation file downloads
• 3,178 data file downloads
SDR Self Deposit Program

- Team Work Still Necessary
- Self-Help Guide Needed
- Primarily Undergraduate Data to Date
- Stumbling block: Customization of PI Data
Thank you for staying to the bitter end!!

Questions?

Matt Marostica (mmarost@stanford.edu)

Ron Nakao (ronbo@stanford.edu)