The Role PI & Institutional Characteristics Play in Shaping Data Sharing Behavior

Amy Pienta, ICPSR
University of Michigan
Goals

I. To identify the research data funded by NSF related to the social and behavioral sciences.

II. Quantify how much research data is shared either formally (through a data archive or repository) versus informally.

III. To describe the characteristics of PIs and institutional characteristics associated with data sharing behavior.
Awards in a database held locally at ICPSR

- Social science and/or behavioral science awards made by NSF
- Original or primary data collection proposed, including assembling a database from existing (archival) sources
NSF Grant Awards in LEADS Database

LEADS contains 17,194 awards made by NSF

LEADS spans 30 years of awards - 1976 to 2005
# Records Reviewed | # Social Science Data
--- | ---
Recent NSF (1976+) | 17,194 | 2,537
Historic NSF (Pre-1976) | 96,403 | 4,019
National Survey of Social Science PIs

- 1,499 NSF PIs emailed a survey
  - living, email address located
- 411 responded (27.4% response rate)
- 316 collected data (77% screened in correctly)
- 283 after dissertation awards dropped
Results: Award Characteristics

- 88.8% funded by SBE (CSE=15, OPP=16)
- Awards made between 1985-2001
- Awards were 1-8 years long (more than \( \frac{1}{2} \) were 2-3 years long)
- Mean award amount 312K
  - Archived - 591K
  - Informal - 231K
  - Not Shared - 225K
Results: Prevalence of Data Sharing among Social/Behavioral PIs

<table>
<thead>
<tr>
<th></th>
<th>Archived Data</th>
<th>Shared Formally</th>
<th>Not Shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSF (N=277)</td>
<td>23.1%</td>
<td>43.3%</td>
<td>33.6%</td>
</tr>
</tbody>
</table>
## Results: Prevalence of Data Sharing among Social/Behavioral PIs

<table>
<thead>
<tr>
<th></th>
<th>Archived Data</th>
<th>Shared Formally</th>
<th>Not Shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSF</td>
<td>23.1%</td>
<td>43.3%</td>
<td>33.6%</td>
</tr>
<tr>
<td>N=277</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIH</td>
<td>7.7%</td>
<td>46.2%</td>
<td>46.2%</td>
</tr>
<tr>
<td>N=732</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data sharing rates in greater depth...

• Archived includes:
  – self-reported archived (verified most)
  – placed in an Institutional Repository

• Informal includes:
  – Department and personal website
  – Data upon request (95%)
  – Other
Results: Characteristics of NSF PIs

- 35% Female
- 83.4% White
- 60.7% Tenured (v. non-tenured & non-faculty)
- Mean Age = 43 (range, 27-75)
- Mean # Lifetime Federal Awards 6.6 (range, 1-100)
Results: Race and Data Sharing Behavior among NSF PIs (n=277)
Results: Rank and Data Sharing Behavior among NSF PIs (n=277)
# Results: Discipline of NSF PIs

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociology</td>
<td>24.2</td>
</tr>
<tr>
<td>Economics</td>
<td>20.6</td>
</tr>
<tr>
<td>Political Science</td>
<td>9.4</td>
</tr>
<tr>
<td>Psychology</td>
<td>22.4</td>
</tr>
<tr>
<td>Other</td>
<td>23.5</td>
</tr>
</tbody>
</table>
Results: Disciplinary Differences in Data Sharing Behavior

- Sociology
  - Not Shared
  - Informal
  - Archived
- Economics
  - Not Shared
  - Informal
  - Archived
- Political Science
  - Not Shared
  - Informal
  - Archived
- Psychology
  - Not Shared
  - Informal
  - Archived
Contextual Factors

• Geographic region of the institution
  – Northeast, Mid-West, Southern, West
  – Similar data sharing practices across

• Carnegie classification of the institution
  – Private Research Organizations
  – Research Institutions
  – Non-research Institutions
Type of Institution

<table>
<thead>
<tr>
<th>Type of Institution</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>80%</td>
</tr>
<tr>
<td>Non-research</td>
<td>40%</td>
</tr>
<tr>
<td>Private Research Organization</td>
<td>20%</td>
</tr>
</tbody>
</table>

- **No Sharing**
- **Informal**
- **Archived**
# Results: Perceived Barriers to Sharing Data from NSF Award

<table>
<thead>
<tr>
<th>Perceived Barrier</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do not have adequate time to prepare data</td>
<td>49.4</td>
</tr>
<tr>
<td>Difficult to prepare documentation</td>
<td>45.3</td>
</tr>
<tr>
<td>Concerned about protecting confidentiality</td>
<td>31.4</td>
</tr>
<tr>
<td>Concerned about potential for others misinterpreting data</td>
<td>24.0</td>
</tr>
<tr>
<td>Concerned others will publish before me</td>
<td>21.6</td>
</tr>
<tr>
<td>Others not interested in topic</td>
<td>20.9</td>
</tr>
<tr>
<td>Informed consent language prevents it</td>
<td>14.9</td>
</tr>
<tr>
<td>IRB prevents me from sharing</td>
<td>10.2</td>
</tr>
</tbody>
</table>
Recapping the Highlights

• Disciplinary differences result in the greatest differences in data sharing behaviors.
  – Economists and Political Scientists share the most. Informal sharing is greater among the economists.
  – Psychologists share data less often than all others social/behavioral scientists awarded projects by NSF.
Recapping the Highlights

• White PIs are also more likely to share.
• Tenured faulty archive data more often than non-tenured faculty.
  – But, non-tenured faculty share more informally.
  – Other individual characteristics matter less.
Recapping the Highlights

• Institutional differences exist as well.
  – PIs located at private research organizations are most successful at sharing data – especially informally.

• The most important barriers to data sharing among NSF PIs is time to prepare the data and documentation.