DATA SHARING ACROSS THE DISCIPLINES, REVISITED: ACADEMIC JOURNALS AND REPLICATION POLICIES

An empirical study
Replication, Replication

Gary King, Harvard University
An Introduction/Overview

- Some Background
- Measuring Journal Policies on Sharing Data – Methods and Findings
- Closing Comments
- [Context – Data Associated with Individual Researchers in the Social Sciences and Their Publications]
Some Background

- “The Age of Big Data” and “The Data Deluge”
- Public Expectations that Data Should be Open
- Academic Debates On Sharing Research Data:
  - Scholars in the social sciences should be more willing to share data that they collect
  - Journals should make sharing of data part of the publication process, or even as a condition for publication
Benefits of Sharing Data

- Reduce Duplication of Effort
- Externalities
- Visibility of Research

- Replication:
  - Research is not replicable if data are not available
  - Research that is not replicable is not scientific
  - Transparency and Methodological Rigor

IASSIST: June 6, 2012
Benefits of Sharing Data

- “The confirmation of research findings through replication by other researchers is an essential part of scientific methodology” (Dewald, Thursby, and Anderson, (1986))

- “[If] the empirical basis for an article or book cannot be reproduced, of what use to the discipline are the conclusions?” (King (1995))

- “The credibility of quantitative social science benefits from policies that increase confidence that results reported by one research can be verified by others” (Freese (2007))
Implications for Data Sharing

- What Should Actually be Shared?
  - The Data Themselves:
  - The Documentation/Metadata (alas, not always forthcoming!)
  - Code/Procedures Used to Assemble the Data, Transform Them, and Perform the Published Analysis

- Data Alone Are Not Sufficient
Data Sharing and Replication

- “Authors of quantitative empirical articles must make their data available for replication purposes … Required material would include all data, specialized computer programs, program recodes, and an explanatory file describing what is included and how to reproduce the published results.”

Our Question – How Have Journals Reacted to This Debate?:

- Are journals expecting or requiring authors to make data available?
- Do policies differ across fields in the social sciences?
- Have policies become more prevalent and/or “strict” over time?

Our Approach: Classify Journal Policies Using an Index of Data Sharing to Compare Across Time/Disciplines
Methods: Overview

- Template - Gleditsch and Metelits (2003):
  - Identified Top 15 Journals in Political Science and in International Relations
  - Journals Selected Based on JCR Citation Rankings
  - Reviewed Data-Sharing Policies of Those Journals

- Our Approach:
  - Expand Range of Academic Fields
  - Select Larger Number of Journals Per Field
Methods: Fields Chosen

- Business/Financial
- Economics
- International Relations
- Political Science
- Sociology
Methods: Selection and Coding

- Selected Top 20 Journals in Each Field
- 96 Unique Titles
- Examine Data-Sharing Policies in 2003 and in 2012
- Create Data-Sharing Index:
  - Range of 0-10
  - Higher Score – More “Strict” Policy
- Use Index to Code/Score Journal Policies
Methods: Data-Sharing Index

- Does Journal Have an Explicit Policy?
- Terms of Policy – “Encouraged” v. “Required”
- Materials Required
  - Description of Data
  - Data Themselves
  - Code for Analysis and for Transformation of Data
- Timeframe for Submission – Pre- or Post-Publication
- Location for Sharing
Findings: Journals With Data Policies

% of Journals with Data-Sharing Policies, by Field

- No Stated Policy
- Stated Policy

<table>
<thead>
<tr>
<th>Field</th>
<th>2003</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus./Fin.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECON</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int. Rel.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POLS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IASSIST: June 6, 2012
Findings: Mean Index Scores

Mean Scores for Journal Data-Sharing Policies, by Field

2003 2012

Bus./Fin.  ECON  Int. Rel.  POLS  SOC

Index Value

IASSIST: June 6, 2012
## Findings: Summary Statistics

<table>
<thead>
<tr>
<th>Field</th>
<th>Year</th>
<th>Number</th>
<th>Mean</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Business/Finance</strong></td>
<td>2003</td>
<td>2</td>
<td>0.55</td>
<td>7.0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>8</td>
<td>1.85</td>
<td>7.0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Economics</strong></td>
<td>2003</td>
<td>4</td>
<td>0.90</td>
<td>6.0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>15</td>
<td>4.35</td>
<td>9.0</td>
<td>0</td>
</tr>
<tr>
<td><strong>International Relations</strong></td>
<td>2003</td>
<td>4</td>
<td>1.00</td>
<td>6.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>6</td>
<td>2.00</td>
<td>8.0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Political Science</strong></td>
<td>2003</td>
<td>4</td>
<td>0.93</td>
<td>6.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>8</td>
<td>2.20</td>
<td>8.0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Sociology</strong></td>
<td>2003</td>
<td>0</td>
<td>0.00</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>4</td>
<td>0.95</td>
<td>7.0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2003</td>
<td>14</td>
<td>0.68</td>
<td>7.0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>41</td>
<td>2.27</td>
<td>9.0</td>
<td>0</td>
</tr>
</tbody>
</table>
Examples of Journal Policies

- Gender & Society – requires “[a] methods section systematically describing the methods used in collecting the data for the paper” but does not require sharing of the data themselves.

- Public Opinion Quarterly – “To permit competent professionals to confirm the results and analyses, authors are expected to retain raw data for a minimum of 5 years after publication of the research.”

- Review of Economic Studies – “Authors of accepted papers must provide to the Review, prior to publication, the data, programs, and other details of the computations sufficient to permit replication (…) We reserve the right to refuse publication of papers whose authors do not comply with these requirements.”

IASSIST: June 6, 2012
Barriers/Obstacles to Sharing

- Data Sharing in the Social Sciences is Not Necessarily the Norm

- Reasons:
  - Confidential Data
  - Proprietary Data
  - Philosophical/pedagogical concerns
  - Incentives – Time/Effort
    - “Documentation, for instance, is often thought of as a waste of time” (Gleditsch and Metelits (2003), p. 73)
  - Incentives – Professional Considerations
Special Thanks to the Following People:

- Joel Herndon
- Various Journal Editors in Business/Finance
- Various Journal Editors in Economics
- Various Journal Editors in Political Science
- Various Journal Editors in International Relations
- Various Journal Editors in Sociology
Questions From the Audience?

I have no idea what you’re talking about...

...so here’s a bunny with a pancake on its head.
Findings: Social Science Journals

Graphs by fieldlab

IASSIST: June 6, 2012