Incentives for Data Producers to Create Archive/Ready Data: Implications for Archives and Records Management

Margaret Hedstrom
Jinfang Niu
School of Information
University of Michigan
Why are Incentives for Data Producers Important?

- Archivists expect/require data producers to provide
  - Data
  - Metadata
  - Documentation

sufficient for

- Validation of accessions
- Preparing data for long-term preservation
- Appropriate release or dissemination
‘Archive-Ready’ Data

- ‘Archive-Ready’ = Content and metadata sufficient for archivists to prepare the data for long-term preservation and dissemination
- In OAIS terminology: The Submission Information Package is complete and accurate enough to produce an Archival Information Package without recourse to the data producer
- ‘Dissemination-Ready’ requires additional effort
Compliance and Incentive Issues

- Data producers are unaware of recordkeeping, archival and data deposit requirements
- Data producers do not consider archival requirements until projects are completed
- Permanent records and data are not transferred to archives
- Submissions are incomplete

Data Producer Study

“Best Case” Scenario

- Social Science researchers required to deposit data for federal funding
- Designated Repository: National Archive of Criminal Justice Data at ICPSR
- Guidelines and training available to assist grantees
- Grantees are expected to deposit data at the completion of their funded projects
## Deposit and Processing Delays

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deposit Delay</strong> (days)</td>
<td>767</td>
<td>664</td>
<td>-27</td>
<td>2630</td>
</tr>
<tr>
<td><strong>Processing Delay</strong> (days)</td>
<td>355</td>
<td>376</td>
<td>20</td>
<td>1187</td>
</tr>
<tr>
<td><strong>Total Delay</strong> (days)</td>
<td>1160</td>
<td>1122</td>
<td>263</td>
<td>2846</td>
</tr>
</tbody>
</table>
Reasons for Delays

- Grantees do not submit data at end of the grant
- Archivists require additional information/data from data producers
- Data producers review and approve revisions

Processing Time (in hours)

<table>
<thead>
<tr>
<th>Processing time (hours)</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>79</td>
<td>60</td>
<td>8.5</td>
<td>359</td>
</tr>
</tbody>
</table>
Data Producers’ Attitudes toward Depositing Data (n=54)

☐ 51 grantees (94.4%) were aware of the deposit requirement

☐ Attitude toward deposit requirement:
  - Very favorable (34%)
  - Favorable (31%)
  - Neutral (24%)
  - Unfavorable (11%)

☐ 57% would deposit their data even if they were not required to do so
Obstacles to Deposit

- Would like to publish more papers before releasing data (44%)
- Confidentiality (35%)
- Loss of control over the data (31%)
- Loss of exclusive use of the data (28%)
- Costs of preparing data (20%)
Incentives: I would be more likely to deposit my data if:

<table>
<thead>
<tr>
<th>Incentive</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>I thought the data would really benefit others</td>
<td>34</td>
<td>65.4%</td>
</tr>
<tr>
<td>If deposit were mandatory to receive new funding from NIJ</td>
<td>26</td>
<td>50.0%</td>
</tr>
<tr>
<td>If the data and documentation counted as a publication</td>
<td>19</td>
<td>36.5%</td>
</tr>
<tr>
<td>If citations to the data counted like citations to publications</td>
<td>17</td>
<td>32.7%</td>
</tr>
<tr>
<td>If depositing data was a requirement for publishing a paper</td>
<td>14</td>
<td>26.9%</td>
</tr>
<tr>
<td>If I received monetary compensation</td>
<td>8</td>
<td>15.4%</td>
</tr>
</tbody>
</table>
Incentives

- Extrinsic Incentives
  - Material rewards (payments, promotions)
  - Moral rewards (praise, public commendation, reputation)
  - Coercive laws and policies (avoidance of punishment)

- Intrinsic Incentives
  - Enjoyment
  - Contributing to a common good
  - Sense of fulfilling an obligation or responsibility to a group or society
Incentives for Social Science Researchers

- Emphasize intrinsic incentives
- Inform data producers of when and how their data are used.
- Make depositing the norm, not the exception
- Permit limited embargo of data between deposit and release
Implications for Archives and Records Management

- Large investments are being made in repositories with less attention to producer/archive relationships and processes.
- Heavy reliance on extrinsic incentives (laws, policies, guidelines).
- Explore intrinsic incentives:
  - Resonate with social and cultural values in the public sector, academia, and not-for-profits.
  - Effective intrinsic rewards do not require additional intervention.
Thank you

☐ This research was funded by the National Science Foundation NSF Award # IIS 0456 022