Protecting Confidentiality in Canadian Research Data Centres

Cynthia Cook
Senior Research Data Centre Analyst,
Statistics Canada
Four relevant aspects of the Canadian experience

1. Think about multiple, complimentary modes of data access.
2. Choose between perturbing data or perturbing results.
3. Institutional costs from breeches of disclosure can influence data access.
4. Researcher Engagement can be your best line of defense against disclosure.
Trade-offs in Data Access

- Perturbed Data
  - Data Liberation Initiative
  - Loss of Access & Utility
- Public Access
  - Greater Risk of Disclosure
- Restricted Access
  - Research Data Centre Program
- Unperturbed Data
<table>
<thead>
<tr>
<th>Data Liberation Initiative</th>
<th>Research Data Centres</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Library</td>
<td>Secure computing lab</td>
</tr>
<tr>
<td>Statistics Canada household surveys</td>
<td>Statistics Canada household surveys</td>
</tr>
<tr>
<td>Perturbed, partially suppressed data files (without personal identifiers)</td>
<td>Unperturbed, full data files (without personal identifiers)</td>
</tr>
<tr>
<td>Mostly cross-sectional data</td>
<td>Mostly longitudinal data</td>
</tr>
<tr>
<td>No restrictions on the release of analytical output</td>
<td>Only analytical output leaves the centre after vetting for disclosure</td>
</tr>
<tr>
<td>No access to foreign researchers</td>
<td>Conditional access to foreign researchers</td>
</tr>
<tr>
<td>No contractual obligation to produce a product</td>
<td>A contractual obligation to produce a product</td>
</tr>
<tr>
<td>Rely on data to determine the risk of disclosure</td>
<td>Rely on researchers to determine risk of disclosure</td>
</tr>
</tbody>
</table>
Perturb Data or Perturb Results?

- The output from modeling data poses less of a risk to disclosure than the data itself.
- Some researchers want the output from modeling data.
- Invest in controlling the results, rather than controlling the data to satisfy data user needs.
- Eg. Gini Coefficients.

Research Data Centre Program
Disclosure Control without Data Perturbation

- Legally
- Physically
- Electronically
- Analytically → Disclosure Risk Management

Protect Response Rates Of Voluntary Household Surveys
Resources for Conducting Disclosure Analysis

- The Researcher
- Experience of the RDC Analyst
- Hard and fast rules
- Intuitive guidelines
- Statistics Canada Methodologists
Challenges

• Engaging researchers
• Keeping the volume down
• Keeping disclosure as a priority given time constraints and research goals
• Knowing when to get help

Solution: Open lines of communication
To learn more …

Research Data Centre Program:

www.statcan.ca/francais/rdc/
www.statcan.ca/english/rdc/
rdc-cdr@statcan.ca

All Data Access Mechanisms to Statistics Canada Data:

http://www.statcan.ca/english/studies/ixres.htm
http://www.statcan.ca/francais/studies/ixres_f.htm