Session 11

Data collection with Internet

Srdjan Mrkić
United Nations Statistics Division
Data collection using the Internet

- Using Internet self-response as a collection mode
  - Advantages
  - Challenges
  - Requirements
- Implementation process
- Contact and communication strategy
- Identification and authentication of respondents
- Design of questionnaire and application features
- Support for respondents
- Managing and monitoring response
- Security
Using Internet self-response as a collection mode

- **Avantages**
  - Improved coverage and data quality
  - Potential cost saving
  - Greater convenience and lower response burden
  - Improved timeliness
  - Protecting privacy
  - Other benefits (environmentally friendly, generating a large amount of metadata)
Using Internet self-response as a collection mode

- **Challenges**
  - Providing reliable linkage between household and location
  - Coordination of multi-mode collection
  - Developing systems infrastructure
  - Protecting data security – portal may be crashed/respondent data may be corrupted or lost
  - Extended collection period (hard to capture picture as at census day)
  - Mode effect and non-response bias
  - High initial cost
Using Internet self-response as a collection mode

- **Requirements**
  - High literacy rate
  - High level of access to computers and the Internet, and computer literacy
  - Making reasonable assumptions about take-up rate
  - Availability of address/building/dwelling list
  - Legal authority
  - Questionnaire design for mobile devices
  - Public trust and acceptance of Internet for official business
Implementation process

- A first step towards the use of Internet for future census

*Feasibility study to assess the success of internet collection to obtain a measure of:*

  - Public acceptance
  - Take-up rate
  - Impact on overall response
  - Distribution of returns over the collection period
  - Data quality
  - Impact on field management
  - IT infrastructure requirements
  - Impact on response burden
  - Impact on under-enumeration or over count
Implementation process

- Planning phase should include:
  - Assessment of local circumstances
  - Review of international experience
  - Establishment of viable options
  - Preferred strategy, on technology and the balance between in-house work versus outsourced contracts
  - Risk analysis
  - Implementation plan
  - Short-term and long-term vision
  - Evaluation plan
  - Setting achievable targets
Contact and communication strategy

- Multi-phase contact approach
- Offering a paper questionnaire
  - at initial contact or;
  - during the reminder phase or;
  - on request
- Choosing the best approach
- Communication strategy (invitation letter/reminder letters, etc.)
## Multi-phase contact strategies

<table>
<thead>
<tr>
<th>JAPAN (2015)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PHASE 1:</strong> Enumerators distributed internet IDs to all households</td>
<td></td>
</tr>
<tr>
<td><strong>PHASE 2:</strong> Enumerators distributed questionnaires to households which have not responded through the Internet. Questionnaires were submitted by handing over to enumerators or by postal mail</td>
<td></td>
</tr>
<tr>
<td><strong>PHASE 3:</strong> Non-response follow-up by field enumerators</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KOREA (2011)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PHASE 1:</strong> Internet option available for 10 days just before Census Day. Incentives included handing out gifts by drawings and giving students two-hours credits for volunteer work.</td>
<td></td>
</tr>
<tr>
<td><strong>PHASE 2:</strong> Field interviews during 15 days just after Census Day.</td>
<td></td>
</tr>
</tbody>
</table>
Identification and authentication of respondents

- **Benefits**
  - Better linkage of households to address of dwellings
  - Ability to deliver different forms to different households
  - Reduced risk of impersonation
  - Reduced risk of duplicate responses
  - Better security both actual and perceived

- **Challenges**
  - Confidential delivery of identifiers
  - Proper linkage of identifiers to households
  - Increased risk of non-response (if identifiers are misplaced)
Development of data collection application and portal

- Questionnaire design and application features
  - Log-in screen (user experience starts here)
  - Screen visibility
  - Intuitive and easy to complete
  - Format: matrix or sequential
  - Automated skip patterns
  - Response options and menus
  - Validation messages
  - Progress through/save/submit questionnaire
  - Mobile friendly, responsive design (multiple browsers)
  - Embedded support/help for respondents
Support for respondents

- Online help
- Frequently asked questions
- Embedded help text in the online questionnaire
- Questionnaire guide
- Census helpline (Call Centre)
Management and monitoring of Internet response

- Metadata -- *During enumeration*
  - Daily returns
  - Cumulative returns
  - Comparison with predicted returns
  - Concurrent users
  - Saved, abandoned or incomplete returns
  - Reports on IT infrastructure stability and capacity
Management and monitoring of Internet response

- Metadata -- *For later analysis*
  - Number of edits
  - Use of help pages
  - Item non-response rate
  - Average completion time
  - Use of foreign languages
  - Comparison of socio-demographic variables for households using different devices
Management and monitoring of Internet response

- Coordination with non-response follow-up
  - Critical in multi-mode collection
  - Central and integrated data collection operation control system,
  - Non-response follow-up procedures need to have flexibility
Security

- To maintain security of personal information:
  - Secure log-in
  - Internet application should ensure zero footprint on respondent computer
  - Timing-out after period of inactivity
  - Encryption
  - Powerful firewalls, intrusion detection
  - Strong access control procedures
  - Contingency plans for temporary service interruptions
  - Communication strategies to assure respondents
Testing

- Questionnaire -- cognitive/qualitative testing
- Planning -- experimental testing to estimate take-up rate, various metrics for planning
- IT systems -- infrastructure and system testing
User-friendly Forms - Left hand panel displays

- Allowed respondents to navigate to respective screens easily
- Allowed respondents to monitor their progress
## Internet participation rate, South Korea

<table>
<thead>
<tr>
<th>Year</th>
<th>Data collection</th>
<th>Key promotion</th>
<th>Internet Participation Rate (%)</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>Short &amp; Long form</td>
<td>none</td>
<td>0.9</td>
<td>Initial</td>
</tr>
<tr>
<td>2010</td>
<td>Short &amp; Long form</td>
<td>Volunteer time, Gift voucher</td>
<td>47.9</td>
<td>Development</td>
</tr>
<tr>
<td>2015</td>
<td>Long form</td>
<td>Gift voucher</td>
<td>48.6</td>
<td>Mature</td>
</tr>
</tbody>
</table>
Item Non-Response by Mode

Evaluation & Challenges

<table>
<thead>
<tr>
<th>Category</th>
<th>Internet (%)</th>
<th>Interview (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupancy status</td>
<td>0.01%</td>
<td>0.78%</td>
</tr>
<tr>
<td>Type of living quarters</td>
<td>0.00%</td>
<td>1.13%</td>
</tr>
<tr>
<td>Sex</td>
<td>0.05%</td>
<td>2.34%</td>
</tr>
<tr>
<td>Age</td>
<td>0.05%</td>
<td>2.37%</td>
</tr>
<tr>
<td>Relationship to head of household</td>
<td>0.03%</td>
<td>2.60%</td>
</tr>
<tr>
<td>Education attainment</td>
<td>0.06%</td>
<td>2.65%</td>
</tr>
<tr>
<td>Place of birth</td>
<td>0.07%</td>
<td>3.22%</td>
</tr>
<tr>
<td>Status in commuting</td>
<td>0.03%</td>
<td>0.95%</td>
</tr>
<tr>
<td>Marital status</td>
<td>0.08%</td>
<td>1.30%</td>
</tr>
</tbody>
</table>

Regional Workshop on the 2020 World Programme on Population and Housing Censuses: International standards and contemporary technologies
Ankara, Turkey, 12-15 March 2019