Supporting Australia’s 2021 Census with Administrative Data

ESCAP Stats Café Series April 2022
Ross Watmuff, Census Futures, ABS
The 2016 Australian Census was the first to be supported by a national **Address Register**. This improved operations and lowered costs significantly.

The 2021 Australian Census is the first to be supported by national, integrated **person level administrative data**. This has significantly improved data quality outcomes.

Administrative data currently plays an indirect, supportive role for the Australian Census. Many more benefits of this data remain to be explored for future Censuses.
Administrative data used for the 2021 Census

- Dwelling admin data:
  - Address Register
  - Electricity data (used at area level)
  - Rentals data

- Person admin data:
  - Data from the ‘Multi Agency Data Integration Project’, which is:
    - Combined persons on Health, Benefits and Tax Registers
    - Analytical data linked from many different admin sources
Two admin data assets created to support the Census

Admin Population Asset
Based on MADIP, “scoped” to Census night population
Age, Sex, Location, Indigenous Status, Relationships, Country of Birth, Income, Previous locations...

Admin Dwelling Asset
Based on Address Register at Census time
Dwelling info, electricity usage, rental status, MADIP persons and characteristics...
How admin data supported the 2021 Census

Use in planning

Key benefit:
Help maximise Census response

Level of data used:
Area (mesh block)

Elements of use:
✓ Inform Aboriginal and Torres Strait Islander urban collection strategy
✓ Provide dwelling occupancy expectations to monitor response rates

Use in data processing

Key benefit:
Improve the Census count

Level of data used:
Dwelling and area level

Elements of use:
✓ Improve decision on whether dwellings are unoccupied
✓ Improve imputed counts of people for dwellings that don’t respond
✓ Quality assurance of Census counts

Preparing a low response contingency

Key benefit:
Repair of Census data for unexpected response impacts

Level of data used:
Dwelling and person level

Elements of use:
✓ Data repair for areas affected by a natural disaster, pandemic lockdown etc
✓ Data repair for reduced response across the board or large-scale loss of data
As part of 2021 Census development, conducted a privacy impact assessment on the use administrative data to support the Census.

To increase transparency and engender public trust, released a series of articles in the lead up to the Census.
Help to maximise the Census response

- Analyse 2016 response patterns by looking at admin characteristics of responding/non-responding dwellings
- Inform the Aboriginal and Torres Strait Islander urban collection strategy
- Provide dwelling occupancy expectations to help monitor response rates
Use in processing

- Improving Census counts
  - Informing occupancy determination
  - Informing donor selection during imputation

- Quality assurance
  - Comparing area-level counts for Census and admin data
Model to predict Census night dwelling occupancy – used a mix of dwelling and area level inputs

<table>
<thead>
<tr>
<th></th>
<th>Dwelling level inputs</th>
<th>Area level inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data sources</td>
<td>• Medicare, Tax, Centrelink</td>
<td>• Electricity usage (smart meter 30%, quarterly 70%)</td>
</tr>
<tr>
<td></td>
<td>• Address Register</td>
<td>• Previous Census</td>
</tr>
<tr>
<td>Influential Information</td>
<td>• Dwelling receives benefits, child present</td>
<td>• Area occupancy rates derived from electricity usage</td>
</tr>
<tr>
<td></td>
<td>• Listed rental vacancy</td>
<td>• Previous Census occupancy rate</td>
</tr>
<tr>
<td></td>
<td>• Dwelling structure</td>
<td></td>
</tr>
</tbody>
</table>
Use in processing – expected improvements to Census count from 2016 research

- Improved imputation guided by admin data and occupancy model
- 2016 imputation – too many older people, too many people in inner-city dwellings (false occupied)

Estimate of persons who should have been imputed (2016 Post enumeration survey)
Preparing a low response contingency - proposed decision pathway

- Census response is lower than expected
  - Has Census data quality been significantly affected?
    - Is a whole area missing, or are there big gaps for important populations?
    - Are important planning and policy decisions going to be affected?
  - Targeted communication
    - Additional field work
  - Census response is still too low
  - Move to decision process below

1. Has Census data quality been significantly affected?
   - Yes
     - Does administrative data of high enough quality to fix the impacts?
       - Does administrative data cover the gaps for the affected areas or populations?
       - Does it have the right information?
     - No
   - No

2. Is there admin data of high enough quality to fix the impacts?
   - Yes
     - Do benefits outweigh the costs, particularly delays to critical results?
       - What is the delay to critical results like population counts?
       - Are the additional costs to data processing feasible?
     - No
   - No

3. Do benefits outweigh the costs, particularly delays to critical results?
   - Yes
     - Do the data owners agree to us using their administrative data for this reason?
       - Are signed agreements in place?
       - Are data owners being kept informed of plans?
     - No
   - No

4. Do the data owners agree to us using their administrative data for this reason?
   - Yes
     - Have we been transparent and are we protecting privacy?
       - Has there been a privacy impact assessment of this approach?
     - No
   - No

5. Have we been transparent and are we protecting privacy?
   - Yes
     - Use administrative data to fill in gaps in the Census
     - No
   - No

Continue normal output processes for the Census
Preparing a low response contingency - proposed data repair approach

Admin data + 2021 Census = Repaired 2021 Census
2021 response rates were very good – the contingency was not required!

However, the preparations have begun demonstrating the potential of administrative data to support imputation and data repair in future Censuses.
Questions?