Transforming the Census

David Martin (NCRM, UK Data Service)
NCRM Autumn School, 2017
Transforming the Census

- Censuses are changing!
- Why does it matter?
- International context
- ONS Census Transformation
- Administrative data
- Internet enumeration
- Big data
- Research implications
- [discussion]
Censuses are changing!

- Increasing demand for timely data
- Difficulty accessing addresses and people
- Falling response rates
  - More effort to achieve same response
- Increasing costs
- Dissolving definitions
  - Usual residence?
  - Household?
  - Main employment?...
- Development of new methods
Photos: David Martin
## Census types

<table>
<thead>
<tr>
<th>Traditional+Internet</th>
<th>Register+Linked Admin data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional Census</td>
<td>Traditional+Internet+Admin data</td>
</tr>
</tbody>
</table>
International context 2011

- Traditional+Internet
  - England and Wales, New Zealand, Canada, Australia, Ireland, Portugal, Malta, Northern Ireland
  - Traditional Census

- Register+Linked Admin data
  - Sweden, Finland, Netherlands, Belgium
  - Traditional+Internet+Admin data
International context 2011

- Traditional+Internet: USA, France
- Rolling Survey
- Traditional Census: England and Wales, New Zealand, Canada, Australia, Ireland, Portugal, Malta, Northern Ireland
- Register+Linked Admin data: Sweden, Finland, Netherlands, Belgium
- Traditional+Internet+Admin data
International context 2011

- Traditional+Internet
- Register+Linked Admin data
- Traditional Census
- Traditional+Internet+Admin data
Why does it matter?

(From perspective of England and Wales, 2017)

- Only source of high quality small area population distributions
- Unique combination of attribute detail and spatial resolution
- Small area denominator population for prevalence, standardized rates, deprivation indicators
- Census as driver of basic statistical geography – output areas, super output areas
- Range of integrated data products
- Multiple interactions linked via current address/address one year ago/workplace address
- Persons linked via household questions
Range of integrated outputs

- National big picture on small populations
- Small area aggregate data
- Integrated statistical boundary and georeferencing system
- Interaction data – flows between small areas
- Microdata samples
- Integration with Longitudinal Studies
Two big shifts

- Traditional + Internet
- Register + Linked Admin data
- Traditional Census
- Traditional + Internet + Admin data
2021 England and Wales Recommendations from the National Statistician:

- Increased use of administrative data and surveys in order to enhance the statistics from the 2021 Census and improve annual statistics between censuses.
- An online census of all households and communal establishments in England and Wales in 2021 as a modern successor to the traditional, paper-based decennial census. ONS recognises that special care would need to be taken to support those who are unable to complete the census online.
International context 2021

Traditional Census
- Traditional+Census
- Traditional+Internet
- Traditional+Internet+Admin data

Register+Linked Admin data
- England and Wales, New Zealand, Canada, Australia, Northern Ireland
- Sweden, Finland, Netherlands, Belgium

Rolling Survey
- USA, France
- Ireland, Portugal

Malta

https://statswiki.unece.org/display/censuses/2020+Population+Census+Round
Administrative data

• If we are collecting loads of administrative data and using it to check the census, couldn’t we just use it to replace the census?
• Why are we paying to do this twice?
• Other countries do it - usually in combination with a population register…
• How quickly can we get there?
• What are the obstacles?
  – Technical, legal, practical, public acceptability ++ issues
Administrative data census

- Use existing government administrative data, collected in course of health care, education, benefits, taxation etc.
- Link at the person level to a statistical population database
- Each administrative data source brings additional variables
- Could replace/enhance existing
- Potential to update data annually, hence outputs always more up to date
### 2011 Census QA Pack - Bristol, City of

#### Age-sex population distributions for All Persons, Males and Females with sex ratios

<table>
<thead>
<tr>
<th>Age</th>
<th>2011 Census Count</th>
<th>2011 Census Estimated</th>
<th>Lower 95% Confidence Interval</th>
<th>Upper 95% Confidence Interval</th>
<th>Rolled forward estimates</th>
<th>Patient Register 2011</th>
<th>School Census 2011</th>
<th>Social Security and Revenue Information 2011</th>
<th>Comparator Lower Bound</th>
<th>Comparator Upper Bound</th>
<th>Response Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>26,100</td>
<td>26,800</td>
<td>28,200</td>
<td>23,100</td>
<td>24,200</td>
<td>24,300</td>
<td>29,500</td>
<td>23,400</td>
<td>29,500</td>
<td>21,400</td>
<td>30,700</td>
</tr>
<tr>
<td>5-9</td>
<td>20,400</td>
<td>22,800</td>
<td>21,100</td>
<td>23,800</td>
<td>21,300</td>
<td>21,400</td>
<td>24,300</td>
<td>21,500</td>
<td>24,600</td>
<td>21,400</td>
<td>24,800</td>
</tr>
<tr>
<td>10-14</td>
<td>15,500</td>
<td>17,200</td>
<td>19,700</td>
<td>16,600</td>
<td>17,900</td>
<td>17,900</td>
<td>21,900</td>
<td>18,900</td>
<td>21,800</td>
<td>18,600</td>
<td>23,400</td>
</tr>
<tr>
<td>15-19</td>
<td>24,400</td>
<td>27,000</td>
<td>22,400</td>
<td>23,600</td>
<td>22,400</td>
<td>22,400</td>
<td>26,400</td>
<td>22,900</td>
<td>26,400</td>
<td>23,500</td>
<td>25,400</td>
</tr>
<tr>
<td>20-24</td>
<td>32,300</td>
<td>44,400</td>
<td>42,400</td>
<td>48,700</td>
<td>46,700</td>
<td>46,700</td>
<td>47,600</td>
<td>44,100</td>
<td>47,600</td>
<td>45,200</td>
<td>48,800</td>
</tr>
<tr>
<td>25-29</td>
<td>37,300</td>
<td>40,800</td>
<td>39,400</td>
<td>47,600</td>
<td>42,100</td>
<td>42,100</td>
<td>44,100</td>
<td>42,300</td>
<td>44,100</td>
<td>44,400</td>
<td>91%</td>
</tr>
<tr>
<td>30-34</td>
<td>33,700</td>
<td>36,400</td>
<td>35,700</td>
<td>43,100</td>
<td>38,600</td>
<td>38,600</td>
<td>39,800</td>
<td>36,900</td>
<td>38,600</td>
<td>40,900</td>
<td>93%</td>
</tr>
<tr>
<td>35-39</td>
<td>29,700</td>
<td>30,400</td>
<td>29,000</td>
<td>31,100</td>
<td>28,400</td>
<td>28,400</td>
<td>30,100</td>
<td>28,700</td>
<td>29,700</td>
<td>30,200</td>
<td>94%</td>
</tr>
<tr>
<td>40-44</td>
<td>27,000</td>
<td>28,300</td>
<td>27,600</td>
<td>29,000</td>
<td>26,600</td>
<td>26,600</td>
<td>28,600</td>
<td>26,800</td>
<td>28,600</td>
<td>28,900</td>
<td>95%</td>
</tr>
<tr>
<td>45-49</td>
<td>26,300</td>
<td>27,100</td>
<td>26,900</td>
<td>27,800</td>
<td>25,600</td>
<td>25,600</td>
<td>27,600</td>
<td>25,800</td>
<td>27,600</td>
<td>27,900</td>
<td>97%</td>
</tr>
<tr>
<td>50-54</td>
<td>22,800</td>
<td>23,300</td>
<td>23,700</td>
<td>23,600</td>
<td>22,200</td>
<td>22,200</td>
<td>22,200</td>
<td>22,200</td>
<td>22,200</td>
<td>22,200</td>
<td>97%</td>
</tr>
<tr>
<td>55-59</td>
<td>19,700</td>
<td>20,400</td>
<td>20,000</td>
<td>20,800</td>
<td>19,300</td>
<td>19,300</td>
<td>19,300</td>
<td>19,300</td>
<td>19,300</td>
<td>19,300</td>
<td>97%</td>
</tr>
<tr>
<td>60-64</td>
<td>19,400</td>
<td>20,100</td>
<td>19,600</td>
<td>20,200</td>
<td>19,000</td>
<td>19,000</td>
<td>19,000</td>
<td>19,000</td>
<td>19,000</td>
<td>19,000</td>
<td>97%</td>
</tr>
<tr>
<td>65-69</td>
<td>14,700</td>
<td>15,200</td>
<td>14,800</td>
<td>16,000</td>
<td>15,700</td>
<td>15,700</td>
<td>16,700</td>
<td>14,600</td>
<td>16,700</td>
<td>16,100</td>
<td>97%</td>
</tr>
<tr>
<td>70-74</td>
<td>12,100</td>
<td>12,600</td>
<td>12,200</td>
<td>12,800</td>
<td>13,100</td>
<td>13,100</td>
<td>13,100</td>
<td>12,900</td>
<td>13,100</td>
<td>13,100</td>
<td>97%</td>
</tr>
<tr>
<td>75-79</td>
<td>10,800</td>
<td>10,700</td>
<td>10,400</td>
<td>11,000</td>
<td>10,900</td>
<td>10,900</td>
<td>10,700</td>
<td>10,600</td>
<td>10,900</td>
<td>10,900</td>
<td>99%</td>
</tr>
<tr>
<td>80-84</td>
<td>6,800</td>
<td>6,500</td>
<td>6,300</td>
<td>8,800</td>
<td>8,000</td>
<td>8,000</td>
<td>8,800</td>
<td>8,600</td>
<td>8,800</td>
<td>8,600</td>
<td>99%</td>
</tr>
<tr>
<td>85+</td>
<td>850</td>
<td>5,500</td>
<td>5,800</td>
<td>8,000</td>
<td>6,800</td>
<td>6,800</td>
<td>8,000</td>
<td>8,000</td>
<td>8,000</td>
<td>8,000</td>
<td>99%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>388,200</strong></td>
<td><strong>428,200</strong></td>
<td><strong>420,500</strong></td>
<td><strong>456,000</strong></td>
<td><strong>429,800</strong></td>
<td><strong>429,800</strong></td>
<td><strong>468,100</strong></td>
<td><strong>40,200</strong></td>
<td><strong>468,100</strong></td>
<td><strong>410,700</strong></td>
<td><strong>93%</strong></td>
</tr>
</tbody>
</table>

Source: Office for National Statistics, National Health Service Information Centre, Department for Education, Welsh Government, Department for Work and Pensions/Her Majesty’s Revenue and Customs

Descriptive metadata for tables in this pack is provided on previous pages - Descriptive metadata for terms used.

---

**All persons - Bristol, City of**

<table>
<thead>
<tr>
<th>Age group</th>
<th>Population Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>26,100</td>
</tr>
<tr>
<td>5-9</td>
<td>20,400</td>
</tr>
<tr>
<td>10-14</td>
<td>15,500</td>
</tr>
<tr>
<td>15-19</td>
<td>24,400</td>
</tr>
<tr>
<td>20-24</td>
<td>32,300</td>
</tr>
<tr>
<td>25-29</td>
<td>37,300</td>
</tr>
<tr>
<td>30-34</td>
<td>33,700</td>
</tr>
<tr>
<td>35-39</td>
<td>29,700</td>
</tr>
<tr>
<td>40-44</td>
<td>27,000</td>
</tr>
<tr>
<td>45-49</td>
<td>26,300</td>
</tr>
<tr>
<td>50-54</td>
<td>22,800</td>
</tr>
<tr>
<td>55-59</td>
<td>19,700</td>
</tr>
<tr>
<td>60-64</td>
<td>19,400</td>
</tr>
<tr>
<td>65-69</td>
<td>14,700</td>
</tr>
<tr>
<td>70-74</td>
<td>12,100</td>
</tr>
<tr>
<td>75-79</td>
<td>10,800</td>
</tr>
<tr>
<td>80-84</td>
<td>6,800</td>
</tr>
<tr>
<td>85+</td>
<td>850</td>
</tr>
</tbody>
</table>

5% Confidence Interval
### Beyond 2011 thinking: Statistical options

<table>
<thead>
<tr>
<th>Census options</th>
<th>Administrative data options</th>
<th>Survey option(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Census (long form to everyone)</td>
<td>Aggregate analysis</td>
<td>Address register + Survey</td>
</tr>
<tr>
<td>Rolling Census (over 5/10 year period)</td>
<td>(Intermediate) Sample linkage e.g. 1% of postcodes</td>
<td></td>
</tr>
<tr>
<td>Short Form (everyone), Long form (Sample)</td>
<td>100% linkage to create ‘statistical population spine’</td>
<td></td>
</tr>
<tr>
<td>Headcount + Annual Survey (US model)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Small area data
Looking towards a 2021 census
Administrative Data Census Research Outputs

These outputs are part of our research into whether we can use administrative data to produce population and census statistics in the future.

What the outputs show us

These outputs show what we could produce from an Administrative Data Census and give users an idea of their quality.

Our research so far

We will be focusing on the three key types of information that the census currently gives us. View everything we have published so far on:

- size of the population
- population characteristics
- households and families

What's next?

Each year, we aim to expand the breadth, depth and accuracy of our Research Outputs. This is dependent on the availability of the data and its statistical quality.
(Non-exhaustive) hierarchy of census entities

- Persons
- Families
- Households
- Household spaces
- Dwellings
- Addresses
- Communal Establishments
Before you start

Who should complete this questionnaire?
The householder is responsible for ensuring that this questionnaire is completed and returned.

The **householder** is the person who lives, or is present, at this address who:
- owns/rents (or jointly owns/rents) the accommodation; and/or
- is responsible (or jointly responsible) for paying the household bills and expenses

A household is:
- one person living alone; or
- a group of people (not necessarily related) living at the same address who share cooking facilities and share a living room or sitting room or dining area

What should you complete on this questionnaire?
- **Household questions** on pages 3-6 about this household and its accommodation
- **Individual questions** on pages 7-30 for every person who usually lives in this household. Every person who has been, or intends to be, in the UK for 3 months or more should be included in these questions at their usual UK address
- **Visitor questions on the back page** (page 32) for all other people staying overnight in this household on 27 March 2011

It is important to include visitors staying overnight in this household to make sure no-one is missed. Visitors who usually live elsewhere in the UK must also be included on a census questionnaire at their usual address.

You will find further information about who to include in this questionnaire on page 31.

Will you need extra questionnaires?
- If there are more than six people in this household, or there are more than three visitors staying overnight, you can choose either to complete the entire questionnaire online, or fill in this questionnaire and contact us to request one or more **Continuation Questionnaires**
Potential hierarchy of administrative entities

- Persons
  - New construct A
  - New construct B
  - New construct C
  - Communal Establishments
- Addresses
Internet enumeration

- 2010/11 Censuses – internet completion an optional extra
  - England and Wales 2011 – 16%
- Higher quality from internet responses
- Post-2011 increased emphasis
  - New Zealand 2013 – 34% (24.5%-38.7%)
  - Australia 2016 – 10.6% (target 65%)*
  - Canada 2016 – 68.3% (39.5%- 71.2%)
- Most 2020/2021 enumerations aiming for internet as primary channel (New Zealand 2018)
- Deliver code, invite online completion, follow-up/support those that do not complete

*DDOS attack on night of Census (!)
Household Questionnaire

Return to:
FREEPOST 2011 Census,
Processing Centre, UK

A message to everyone - act now
Everyone should be included in the census - all people, households and overnight visitors.
It is used to help plan and fund services for your community - services like transport, education and health.
Please complete your census questionnaire on 27 March 2011, or as soon as possible afterwards. You can fill it in online or on paper.

Taking part in the census is very important and it’s also compulsory. You could face a fine if you don’t participate or if you supply false information.
Your personal information is protected by law. Census information is kept confidential for 100 years.
So help tomorrow take shape and be part of the 2011 Census.

J. N. Macken
Director
Office for National Statistics

Where can you get help?
- www.census.gov.uk
- Census hotline 0300 0201 101
- Text Relay 18001 0300 0201 160

Help is available in large print and Braille

H1 - Household Questionnaire - England

Your progress

Household questions

H1 - H2
H3
H4 - H5
H6
H7 - H8
H9 - H10
H11
H12
H13
H14

Individual questions
Visitor questions
Submission

What type of accommodation is this?
A whole house or bungalow that is:
- detached
- semi-detached
- terraced (including end houses)
- a flat, maisonette or apartment that is:
- in a purpose-built block of flats or tenement
- part of a converted or shared house (including bedsits)
- in a commercial building (for example, in an office building, hotel, or even a shop)
- a mobile or temporary structure:
- a caravan or other mobile or temporary structure

Is this household's accommodation self-contained?

This means that all the rooms, including the kitchen, bathroom and toilet, are behind a door that only this household can use.

Yes
No

Declaration
This questionnaire has been completed to the best of my knowledge and belief.

Signature

Date
Telephone number

We may contact you if we need to collect missing information.

If you have lost your envelope, please return to FREEPOST 2011 Census, Processing Centre, UK
BIG data shift?

- Data about things and events as proxies for population presence and characteristics
- Population locations at home/work, travel, inc. entry/exit checks for international migration...
- Potential replacement of (some) census and administrative sources, but limited attributes and VERY limited linkage
- Challenges of acceptability, coverage, bias, calibration and stability of messy data
- Context is everything!
- Big Data could augment good admin data, but will still be demand for large coverage surveys
Mobile phone tracking data 'could replace census questions'

Thousands of people have had their movements tracked by the Office for National Statistics to see if they can find out where they live and work.
Anonymised, imputed flows >15 at LA-LA level, for 3 LAs over 4 weeks, on one network!

Thousands of people have had their movements tracked by the Office for National Statistics to see if they can find out where they live and work.
2011 Census processing model

- Traditional
- Internet

Census database
- Coverage survey
- Admin checks

Adjustment and estimation
- Small area aggregate
- Interaction data
- Microdata
- Outputs

Enumeration
2021 Census processing model

- Internet
- Traditional
- Admin sources

Enumeration

Census/ statistical population database

Coverage survey
Admin checks

Adjustment and estimation

Outputs

- Small area aggregate
- Interaction data
- Microdata
Some questions for you

• Which census variables might be of most use to your research?
• How might they be affected by 2021 internet enumeration?
• Could they be obtained without conducting a census?
• What would the biggest challenges be?
• (Think: concepts, definitions, data sources –admin/big data? data quality and matching issues)
Discussion

D.J.Martin@soton.ac.uk

Photo: Dave Martin
<table>
<thead>
<tr>
<th>Category</th>
<th>RED – Limited or no data available</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Demographics</td>
</tr>
<tr>
<td></td>
<td>Educational background</td>
</tr>
<tr>
<td></td>
<td>Ethnicity, identity, language &amp; religion</td>
</tr>
<tr>
<td></td>
<td>Health</td>
</tr>
<tr>
<td></td>
<td>Housing</td>
</tr>
<tr>
<td></td>
<td>Labour market</td>
</tr>
<tr>
<td></td>
<td>Migration</td>
</tr>
<tr>
<td></td>
<td>Travel</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>GREEN – Some data available to ONS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Demographics</td>
</tr>
<tr>
<td></td>
<td>Educational background</td>
</tr>
<tr>
<td></td>
<td>Ethnicity, identity, language &amp; religion</td>
</tr>
<tr>
<td></td>
<td>Health</td>
</tr>
<tr>
<td></td>
<td>Housing</td>
</tr>
<tr>
<td></td>
<td>Labour market</td>
</tr>
<tr>
<td></td>
<td>Migration</td>
</tr>
<tr>
<td></td>
<td>Travel</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>AMBER – Some data available but ONS currently doesn’t have access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Demographics</td>
</tr>
<tr>
<td></td>
<td>Educational background</td>
</tr>
<tr>
<td></td>
<td>Ethnicity, identity, language &amp; religion</td>
</tr>
<tr>
<td></td>
<td>Health</td>
</tr>
<tr>
<td></td>
<td>Housing</td>
</tr>
<tr>
<td></td>
<td>Labour market</td>
</tr>
<tr>
<td></td>
<td>Migration</td>
</tr>
<tr>
<td></td>
<td>Travel</td>
</tr>
</tbody>
</table>
Research Outputs: Estimating the size of the population in England and Wales, 2017 release

Research and population statistics from a methodology different to that currently used in the production of official population statistics. These Research Outputs are not official statistics on the population.

In this section

1. Disclaimer
2. Main points
3. Things you need to know about this release
4. Our aim
5. How do the 2016 estimates perform at a national (England and Wales) and country level?
6. How do the 2016 estimates perform at the local authority level?
7. Further developments and next steps
8. Feedback
Possible new construct A: “household-dwelling unit” (Statistics Finland)

- Consists of the permanent occupants of a dwelling
  - Related concepts include: building, dwelling, consumption unit, residential home, structure of household-dwelling unit
- Concept adopted in 1980 census. In earlier years the concept of household was used, which consisted of family members and other persons living together who made common provision for food
Admin Census processing model

Input admin data sources

Health/demography
Income/employment
Housing

Statistical population database

Coverage survey
Big Data

Adjustment and estimation

Outputs

Small area aggregate
Interaction data
Microdata
Future Statistical population processing model

Diverse data sources
- Health/demography
- Income/employment
- Housing
- Big data sources

Statistical population database

Coverage survey/sample census

Adjustment and estimation
- Small area aggregate
- Interaction data
- Microdata

Outputs
What future for the census?

- 2020/21 – conventional census questionnaires, primarily internet enumeration, administrative data integration
- New risks!
- Beyond 2021 – routinisation of administrative data integration, new data types from big data sources
- Ongoing calibration and extensive public debate re. acceptability of methods and value of data
- 2030/31 – diverse data sources but widespread concern over reliability – demand for large coverage surveys and integrated adjustment and estimation systems for official statistics
International context 2021

- Traditional+Internet: USA, France
- Rolling Survey: Ireland, Portugal
- Traditional Census: Malta
- “Leapfroggers”

- Register+Linked Admin data: Sweden, Finland, Netherlands, Belgium
- Traditional+Internet+Admin data: England and Wales, New Zealand, Canada, Australia, Northern Ireland

https://statswiki.unece.org/display/censuses/2020+Population+Census+Round
International context beyond 2021

- Traditional+Internet: USA, France, Ireland, Portugal
- Register+Linked Admin data: Sweden, Finland, Netherlands, Belgium
- England and Wales, New Zealand, Canada, Australia
- Traditional Census: Malta
- “Leapfroggers”
International context beyond 2021

Big data sources

- Traditional+Internet
  - USA, France
  - Ireland, Portugal

- Traditional Census
  - Malta
  - “Leapfroggers”

- Register+Linked Admin data
  - Sweden, Finland, Netherlands, Belgium

- Traditional+Internet+Admin data
  - England and Wales, New Zealand, Canada, Australia

https://statswiki.unece.org/display/censuses/2020+Population+Census+Round
International context beyond 2021

Big data sources

Sweden, Finland, Netherlands, Belgium

England and Wales, New Zealand, Canada, Australia

Traditional+Census

Traditional+Internet

Traditional+Internet+Admin data

Register+Linked Admin data

"Leapfroggers"

https://statswiki.unece.org/display/censuses/2020,+Population+ Census+Round
Thank you!

D.J. Martin@soton.ac.uk