Putting Admin Data at the Core of Population and Migration Statistics

Kostas Loukas,
United Kingdom
1. Why do we want to transform population and migration statistics?
2. What is the ONS vision for population and migration statistics?
3. Where are we focusing:
   • Concepts and Definitions,
   • Data Sources,
   • Population and Migration statistics design
4. Challenges
5. Future work
Why do we want to transform pop and migration statistics?

- Keep up with the rate and scale of change from both technological and demographic perspectives
- To modernise, improve efficiency and reduce cost
- Survey response rates are falling
- There is a rapidly changing policy context (UK exiting the EU) – ONS need to inform a wide range of important decisions
- Complexity of user needs
The complexity of user needs

**PEOPLE**
- Where & how do they travel?
- What do they do?
- Where do they live?
- What are they like?
- Where do they work?
- How do they behave?
- What are they like?

**BUSINESSES**
- What do they produce?
- What do they buy?
- What do they do?
- What units do they have?
- Where are they located?
- Who do they employ?
- Where do they work?
- Change employment

**ADDRESSES**
- Where is it?
- What units do they have?
- Where are they located?
- What do they consume?
- What do they buy?

**ADDRESSES**
- What happens here?
- What size & condition is it?
- What are they like?
- What do they consume?
- What do they buy?

**Economy, Trade**
- Occupation
- Productivity, Hours worked
- Wealth, Debt
- Occupation

**Economy, Output**
- What do they produce?

**events**
- Weather events, floods, disasters, traffic events, service disruptions

**location**
- Sensor data
- Big data

**The environment**
- Environmental observations
- Remote sensing
- Landuse, air quality, pollution, water

**time & change throughout !**

**society**
- Population
- Demographics
- Education, health & life events
- Characteristics
  - Age, Sex, Ethnicity
  - Status, Health etc
- Transport

**the economy**
- Economy
- Trade

**address or location**
- Shops, offices
- Location
- CEs
- Classification of type & use
- Housing stock
- Residential, Business & Communal
- Planning, building, demolition
- Classification of type & use
- Housing stock
- Residential, Business & Communal
- Planning, building, demolition

**Office for National Statistics**
Vision

• Admin data first population, migration and social stats system for England and Wales but supplemented by surveys.

• Iterative transformation rather than binary decision between an admin data based system and a Census

• Recommendation on the future of population statistics in 2023 based on progress made, and how well new system is matured and meets user needs
Our transformation framework

Concepts
What concepts do we need to measure, and what definitions support these?
- UK Migration
  - UK arrivals
  - UK departures
  - Internal UK movement

Data
What data can we use to answer users’ questions?
- Administrative
- Census
- Survey
- Big/commercial/other data sources

What are the user needs?
- How is the population changing, including patterns of international migration?
- What impact is this having on the economy, labour markets and society?

Outputs
What outputs do we need to produce?
- Population and international migration outputs
- Migrants and migration analysis
- Population characteristics
- Demographic analysis
- Families and households statistics

Methods
What methods can we use to analyse administrative data?
- Estimate population flows
- Estimate population stocks
- Triangulate the results

Office for National Statistics
20 Sep 2019
Concepts and definitions: What do we want to measure?

• User needs inform the concepts and definitions that we need to provide statistics about.

• Existing definitions are based on UN standards, to allow international comparisons, however....

• ...people’s lives are complex, and their patterns of movement may not naturally fit the statistical definitions we use. To understand impacts, we need more flexibility in these.

• Our analysis of Home Office administrative data (July 2018) started to illustrate this.

Hypothetical example of an individual’s travel patterns during the period of their visa:
Data Sources

RTI (Income), Tax Credits, Benefits, Travel Events

Hospital Statistics, Patient Register

National Insurance, School Census, Higher Education Statistics, Births, Deaths, Visas

Continuous Activity

Registration-based Activity
Delivering stocks and flows of the population

**Flows approach**
- Derive population flows from admin data sources or adjust survey based flows using admin data
- Add flows to previous year's stocks to derive current year's usual resident population

**Stocks approach**
- The main focus is to get the best possible estimates for population change (i.e., migration)

**Hybrid Model**
Understanding international migration using data-driven rules

Time spent in UK, non-EU first year students studying at Higher Education Institutions in England and Wales

Number of days in UK within 16 month period of interest
Delivering stocks and flows of the population

**Flows approach**

**Stocks approach**

- Use data-driven rules to develop usual resident population stocks (admin based population estimates)
- Extract migration components by differencing current and previous year’s admin-based population stocks
- Focus on producing population stocks with some undercoverage but minimal overcoverage

**Hybrid Model**
Comparison of Census and Admin based stocks
England and Wales, Males, 2011

Latest Publication: Admin Based Population Stocks using data-driven rules
Delivering stocks and flows of the population

- Flows approach
- Stocks approach
- Hybrid Model
  - Evaluate differences between components of change and the usually resident population estimates they inform
  - Triangulate estimates using a coverage survey
Bias in Admin Data Population Estimates

**Under-coverage**
- Persons not registered with an admin system (e.g., recent migrants, people de-registered due to inactivity) but are residents in the country
- A person living in area 1 but are incorrectly registered in another area
- Census statistical framework established for measuring under-coverage (dual system estimation)

**Over-coverage**
- Persons that have emigrated or moved out of area but their admin record is not updated to reflect the move
- A person registered in area 1 but they now live in another area
- Under and over-coverage to be measured and adjusted from the same survey collection
Social Survey Transformation

• Transformation of the social survey portfolio
• **Admin data first** approach
• A requirement for **large scale household surveys** will remain though the purposes of these surveys are likely to change
  • **to fill evidence gaps and fulfil user needs** for info that admin data cannot provide
  • to assess the **coverage** of admin data sources
• Statistical re-design and rationalisation of our surveys required
• More flexible survey collection framework for users
• Development and testing of an **Integrated Population and Characteristics Survey** (IPACS) for the future sourcing of resident population, migration & social data
Concept of the IPACS ‘Master Wave’

• Sample size - circa 500,000 households per annum

• Mixed mode collection

• Full household enumeration

• Integrate questions from LFS and Population Coverage Survey

• New systems / operational teams needed to support survey

• Sample design requires stratification for admin data coverage (under-coverage and over-coverage)
Parallel research into producing the most successful Census in 2021 and a potential admin data census.

Timeline of Transformation

- **Now**
  - Census coverage survey
  - Integrated Population and Characteristics Survey

- **2021**
  - 2021 Census
  - Admin Data
  - CCS1
  - IPACS2

- **2023**
  - Integrating different data sources to produce prototype dataset for multivariate and longitudinal analysis.

- **The Future**
  - Recommendation on the future of the population statistics

---

1 Census coverage survey
2 Integrated Population and Characteristics Survey
Challenges

• Administrative data captures behaviours but our existing measures are based on intentions. How do these come together?
• How do we ensure international comparability of the new concepts?
• How do we define the population of interest (i.e. is it at a point in time, real time, quarterly, or updates of change)
• How can we construct uncertainty/quality measures for admin based estimates? What parts in the statistical process can produce errors and how do we communicate these to users?
• Record level linkage: moving to an integrated data framework while maintaining confidentiality and data security
Future Work

• Start IPACS collection in 2020,
• Produce migration estimates supported by admin data,
• Continue development of admin-based population stocks
• Stakeholder engagement: find out what users want from migration and population statistics, what their priorities are
• Assessing the quality of our data sources and developing a sources of error framework
• Working with census teams to understand how this work can feature in QA and contingency work
Links to further research

More in-depth slides on our stock and flows research and international migration case studies:
Transforming population and migration statistics: Research into developing an alternative approach to producing administrative data-based population stocks and flows
Transforming population and migration statistics: Emigration patterns of non-EU students
Transforming population and migration statistics: International student employment activity
Transforming population and migration statistics: Benefits and income activity patterns
Transforming population and migration statistics: NINo and NHS registration lags
Transforming population and migration statistics: Patterns of circular movement into the UK

Migration survey data sources workplan (Feb 2019):
Understanding different migration data sources workplan

How ONS look after and use data for public benefit:
Principles and policies for how we use, manage and secure data

Other Admin Data Census research:
Admin Data Census Project pages
Research outputs: An update on developing household statistics for an Admin Data Census
Research Outputs: Using mobile phone data to estimate commuting flows