The UK Data Service and census data

UKDS provides access to the latest as well as past census data from 1971 onwards

It brings expertise, knowledge, and vast experience on handling census data. Some of our colleagues have been working on census data for decades

Separate groups within UK Data Service work with different types of census data
Aggregate data

Aggregate data are the most widely used of census data: they are the aggregate observations for areas running from national level to small local units, and are what would most commonly be recognised as 'the census data'.

UK Data Service provides online access to aggregate data from the 1971 census onwards.

https://statistics.ukdataservice.ac.uk/

https://infuse.ukdataservice.ac.uk/
Flow data

Flow data enumerate movements of people from one place to another. They include migration and journeys to work, and travel to second residences. These data are characterised by their complexity, volume, and sparsity.

Data are available from censuses from 1981 onwards. Data from 2011 onwards have different levels of access. 2021 census data have not been released yet.

https://wicid.ukdataservice.ac.uk/
Microdata are anonymised records at individual level sampled from a single census. They were introduced in 1991, but since then samples have been released from censuses from 1961 onwards. They also have different levels of access.
Boundary data

Digital boundaries are used for mapping and spatial analysis, not just for census data but also for many other datasets. This data is provided by EDINA in Edinburgh, Scotland.

Boundary data are available from 1971 onwards in a number of different formats.

https://borders.ukdataservice.ac.uk
New systems for accessing UK census data
New systems for census data

CKAN for aggregate data

DCAT for improved data discovery

API-driven tools for flow data to supersede WICID
WICID was originally developed in the 1990s and went online in 2000 to provide a web-based interface to flow data. It was redesigned in early 2000s to support 2001 census outputs and was significantly upgraded in early 2010s to accommodate 2011 census data. WICID remains today the only available web tool in the UK dedicated to flow data with sub-setting capabilities.
API-driven tools for flow data

- Data discovery
- Data extraction with subsetting capabilities
- Integration
- Tailored interface
- User-centric development
- Secure authentication
Challenges and handling of sensitive variables in UK censuses
Peculiarities of the 2021/22 census

In England, Wales, and Northern Ireland the census took place in March 2021 at the height of the Covid-19 pandemic.

Scotland decided to move its census to March 2022 instead.
# Sensitive questions in UK census

<table>
<thead>
<tr>
<th>Religion</th>
<th>Religion</th>
<th>Sexual orientation</th>
<th>Gender identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern Ireland</td>
<td>England and Wales, Northern Ireland,</td>
<td>England and Wales, Northern Ireland, Scotland</td>
<td>England and Wales, Scotland</td>
</tr>
<tr>
<td></td>
<td>Scotland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre 2001</td>
<td>2001 onwards</td>
<td>2021 onwards</td>
<td>2021 onwards</td>
</tr>
</tbody>
</table>
Sensitive questions in census 2021/22

**England & Wales**

26. Which of the following best describes your sexual orientation?
   - Straight/Heterosexual
   - Gay or Lesbian
   - Bisexual
   - Other sexual orientation, write in

27. Is the gender you identify with the same as your sex registered at birth?
   - Yes
   - No, write in gender identity

**Scotland**

8. Which of the following best describes your sexual orientation?
   - Straight/Heterosexual
   - Gay or Lesbian
   - Bisexual
   - Other sexual orientation, please write in

4. Do you consider yourself to be trans, or have a trans history?
   - Yes
   - No, please describe your trans status (for example, non-binary, trans man, trans woman)

16. What is your religion?
   - No religion
   - Christian (including Church of England, Catholic, Protestant and all other Christian denominations)
   - Buddhist
   - Hindu
   - Jewish
   - Muslim
   - Sikh
   - Any other religion, write in

**Northern Ireland**

No question on gender identity

13. What religion, religious denomination or body do you belong to?
   - None
   - Church of Scotland
   - Roman Catholic
   - Presbyterian Church in Ireland
   - Methodist Church in Ireland
   - Other, write in

14. What religion, religious denomination or body were you brought up in?
   - None
Privacy and confidentiality in flow data

Dual geography identifiers

- Large matrices with low expected cell values

Flows often cluster towards short distance interactions

- The longer the distance between pairs the more likely to contain small or zero cell values
Protection of privacy

- Disclosure controls
- Access controls
- Legal implications
Disclosure controls

Targeted record swapping
- Swap people and households that stand out in the data with a similar record in the local area

Cell key perturbation
- Make slight changes to cell counts, which adds some noise to the data

Publishing rules
- Rules on publishing data with very low counts (usually less than 3)
Access controls

Public
- Available to any user via the Open Government Licence

Safeguarded
- Available via UKDS to members of academia, local and central government, NHS, and UK parliaments and assemblies via End User Licence (EUL) so need to register and login

Secured
- Available to approved researchers via the Approved Researcher Scheme using the ONS Secure Research Service (SRS)
Flow data can be disclosive

Up to 37% of people in the GB/UK migration data, have unique combinations of origin and destination at the most detailed level of geography

Up to 58% of people in the GB/UK journey to work data, have unique combinations of origin and destination at the most detailed level of geography

https://doi.org/10.1007/s12061-017-9247-1
Significant fall in migration between 2011 and 2021 in England and Wales
Handling of sensitive variables for flow data in previous censuses

Religion

Journey to work
- 2001 – commissioned, LA level, univariate

Migration
- 2001 – commissioned, complex
- 2011 – planned, LA level, univariate
One-year migrants in England and Wales, 2021

Source: 2021 Census, England and Wales
https://www.ons.gov.uk/datasets/create/filter-outputs/44d65627-fb10-444f-a6de-9e7d39a4be8a
LGB+ population by age in England and Wales

Figure 1: Young people were the most likely to have identified as lesbian, gay, bisexual or another minority sexual orientation (LGB+)

Percentage of usual residents aged 16 years and over who identified as LGB+ by age, England and Wales, 2021

- 16 to 24 years: 6.91%
- 25 to 34 years: 5.63%
- 35 to 44 years: 3.50%
- 45 to 54 years: 2.39%
- 55 to 64 years: 1.59%
- 65 to 74 years: 0.84%
- 75 years and over: 0.37%

Source: Office for National Statistics - Census 2021

92.5% answered 3.2% identified as LGB+
Estimated numbers of LGB+ migrants in England and Wales

Applying % LGB+ identification by age to internal migrants by age, we get an estimate of c. 265K LGB+ migrants in England and Wales, 2020-21.

However, this assumes that LGB+ and non-LGB people have the same propensity to migrate.

We need data on migrants by age by sexual orientation to confirm or reject this.
Population sizes are comparable with past sensitive data
Trans population by age in census 2021

Figure 1: 1 in 100 people aged 16 to 24 years identified as trans

Percentage of usual residents aged 16 years and over who identified as trans by age, England and Wales, 2021

- 16 to 24 years: 0.54%
- 25 to 34 years: 0.77%
- 35 to 44 years: 0.64%
- 45 to 54 years: 0.47%
- 55 to 64 years: 0.33%
- 65 to 74 years: 0.25%
- 75 years and over: 0.22%

Source: Office for National Statistics – Census 2021
Gender identity

The starting estimate (0.54% of migrants) would give around 35,000 trans migrants. Applying age specific migration and trans identification rates would give a revised estimate of about **40,000 trans migrants**.

Again, this relies on an untested assumption that propensity to migrate is the same in trans and cis populations.

Gender identity would likely need stronger access controls.
Journey to work data

Far more people travel to work than migrate. However, the disclosure risks posed by location tracing are far more pertinent in journey to work data, and fine scale geography is problematic.

Applying similar calculations we reach estimates of:

- **568,000** people travelling to work in the UK (not WFH) and LGB+
- **92,000** people travelling to work in the UK (not WFH) and trans

Again we assume LGB+/trans people have similar likelihood of travelling to work as the whole population.
### What bespoke tables might be feasible with 2021 data?

<table>
<thead>
<tr>
<th></th>
<th>Sexual orientation</th>
<th>Gender identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Migration</td>
<td>No public tables</td>
<td>No public tables</td>
</tr>
<tr>
<td></td>
<td>Safeguarded at LA level with grouped responses</td>
<td>No safeguarded tables</td>
</tr>
<tr>
<td></td>
<td>Secured at LA/wards level with detailed responses</td>
<td>Secured at regions/LA level with detailed responses</td>
</tr>
<tr>
<td>Journey to work</td>
<td>No public tables</td>
<td>No public tables</td>
</tr>
<tr>
<td></td>
<td>Safeguarded at LA level with grouped responses</td>
<td>No safeguarded tables</td>
</tr>
<tr>
<td></td>
<td>Secured at LA/MSOA level with detailed responses</td>
<td>Secured at regions/LA level with grouped or detailed responses</td>
</tr>
</tbody>
</table>
Thank you.

Oliver Duke-Williams, o.duke-Williams@ucl.ac.uk
Vassilis Routsis, v.routsis@ucl.ac.uk