Research challenges with the UK Census

Oliver Duke-Williams

UCL DIS Research Open Day, June 2024

DIS colleagues at IASSIST / CARTO 2024: Halifax, Nova Scotia May 28-31 2024

IASSIST: International Association for Social Science Information Service & Technology
CARTO: Association of Canadian Map Libraries and Archives
Association des cartothèques et archives cartographiques du Canada

- Challenges for a data service
- Unclear future: administrative data
- Challenges for longitudinal data
- Challenges for the 22nd century

UK Data Service

- ESRC funded service (2012-) for providing research ready (social science) data
  - Curation, metadata, documentation, training etc
  - Provides access to data in a number of repositories
- Census group is spread across multiple institutions
  - **UCL**
  - University of Manchester
  - University of Edinburgh
  - Jisc

Census data

- Census data can be divided in a number of ways
  - Types of data
    - Aggregate data; flow data; microdata; boundary data
  - Licenses
    - Open; Safeguarded; Secure
  - Historic / 'recent'
    - UKDS currently have (some types of) data from 1961 onwards
    - Separately, the Data Archive hold individual data from the 1851-1911 censuses, through the Integrated Census Microdata (I-CeM) project

The ‘UK’ Census

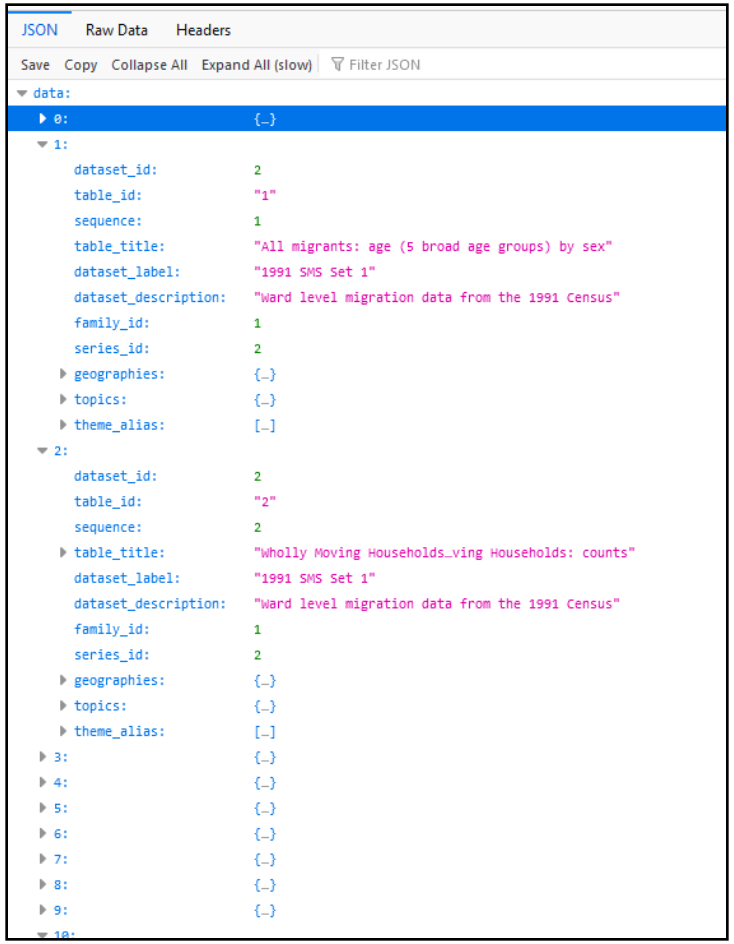
- We do not have a 'UK Census' but rather three separate decennial censuses, carried out be three agencies
  - England and Wales (Office for National Statistics, ONS)
  - Scotland (National Records of Scotland, NRS)
  - Northern Ireland (Northern Ireland Statistics and Research Agency, NISRA)
- There are also three Crown Dependencies (UKDS does not hold data from these)
  - Isle of Man
  - Bailiwick of Jersey
  - Bailiwick of Guernsey

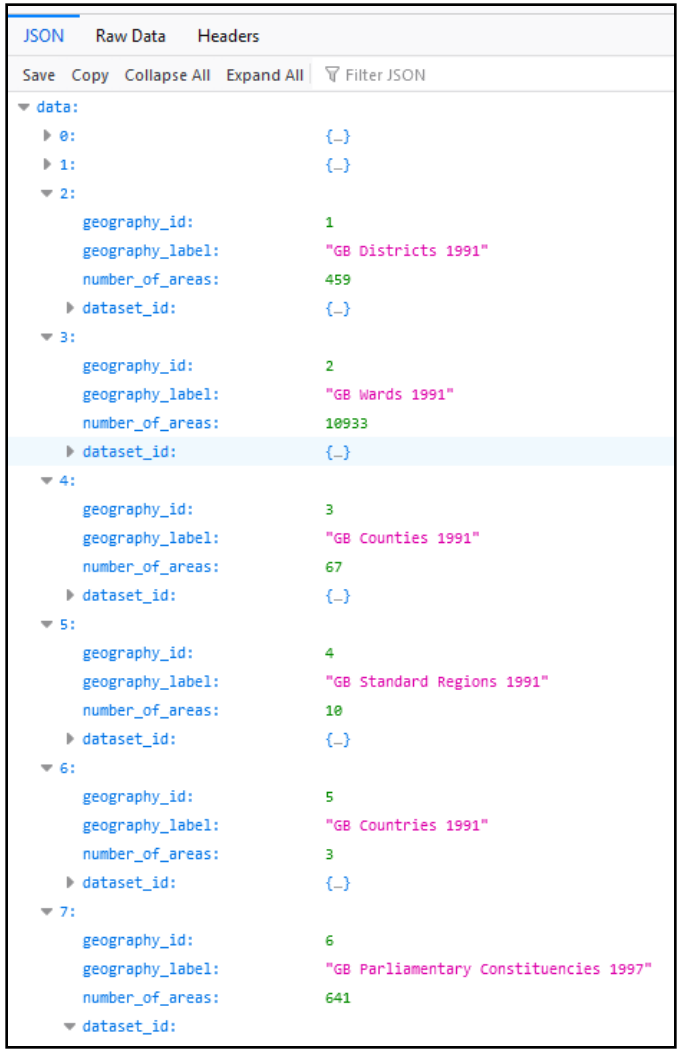
Practical challenges

- Our legacy systems which provided access to data from 2011 and earlier are outdated
- With the exception of microdata, they do not integrate well with other UKDS holdings
  - Many tables, very large tables – require complex (bespoke?) discovery and sub-setting
- New competitors
  - Custom datasets available through ONS
- Updating training materials etc

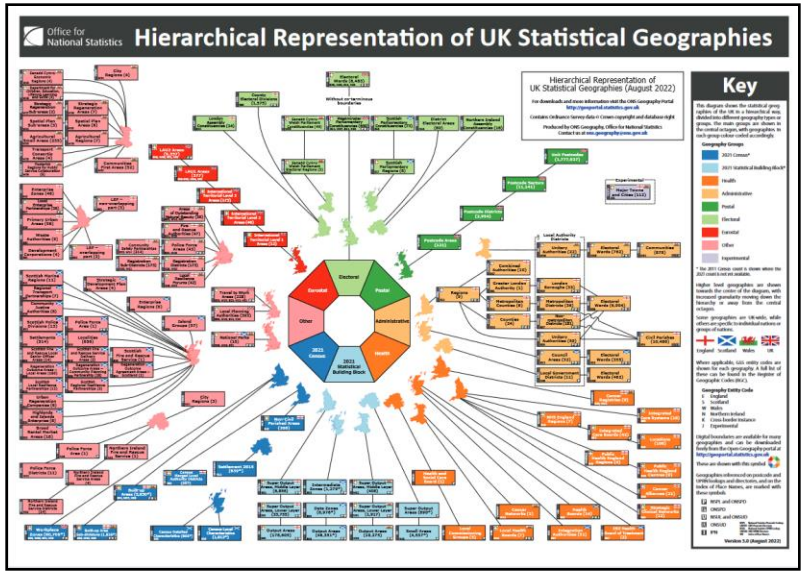
Challenges: legacy systems

https://wicid.ukdataservice.ac.uk/

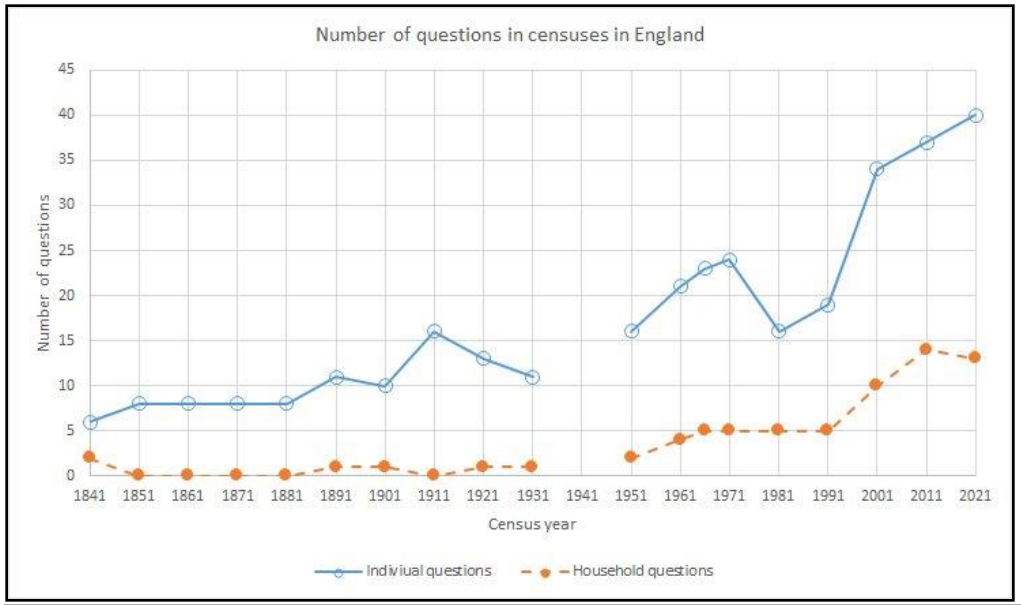




Challenges: data volume and complexity



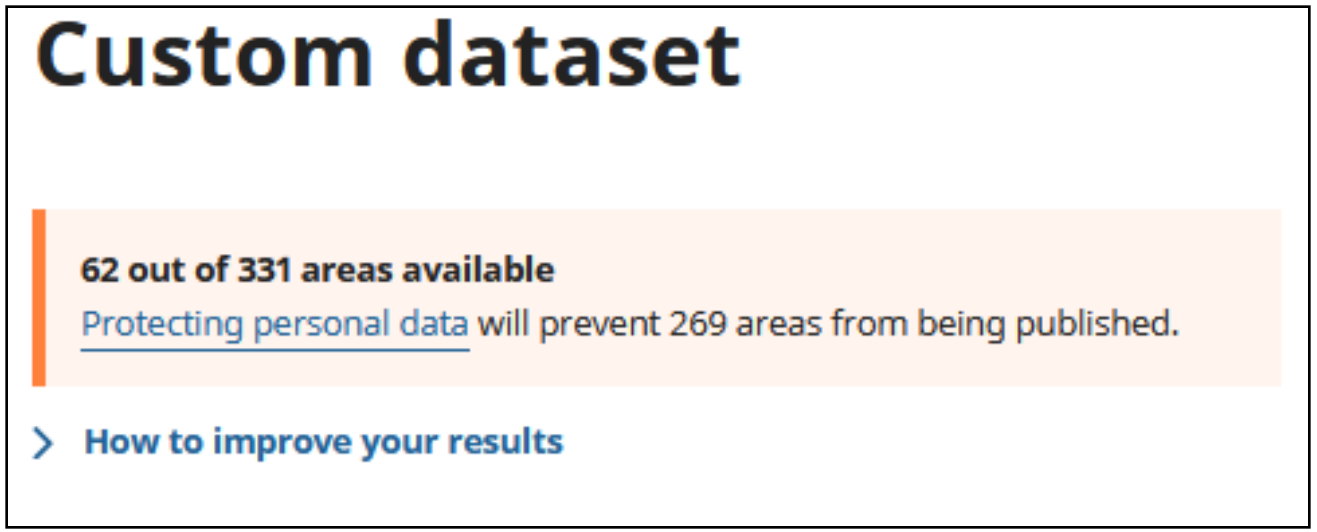
Source: https://geoportal.statistics.gov.uk/documents/ons::hierarchical-representation-of-uk-statistical-geographies-august-2022-1

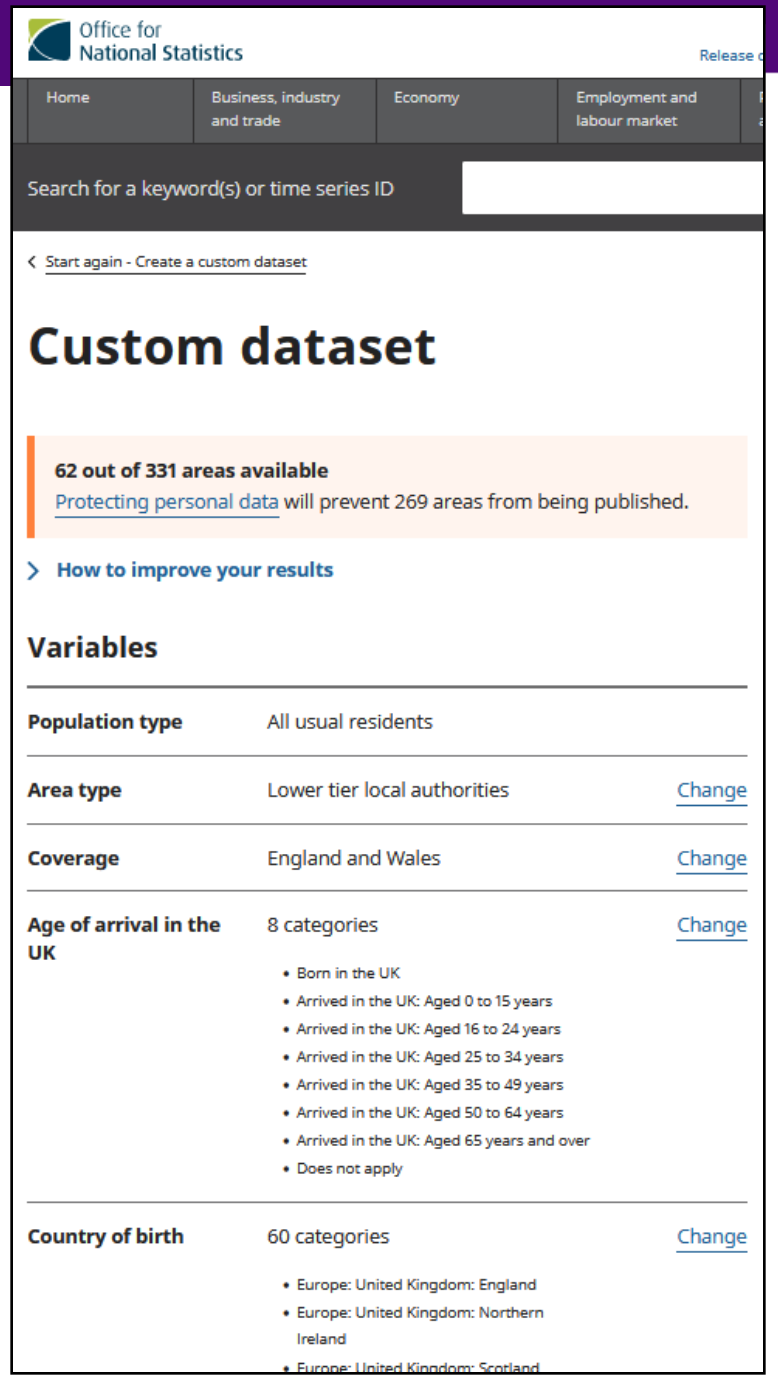


Source: https://x.com/oliver_dw/status/1164521314360733701?s=20

Challenge: new competition

- ONS custom dataset builder
  - Easy to use
  - Long requested
  - Well designed





https://www.ons.gov.uk/datasets/create

Research challenges

- Merging data to provide UK level results
  - Are questions and concepts the same?
  - Are data classifications compatible?
- Providing time series data
  - How to handle change in variables over time
  - Are definitions really consistent?
  - Are naming schemes appropriate?

- Challenges for the Data Service
- Unclear future: administrative data
- Challenges for longitudinal data
- Challenges for the 22nd century

An unclear future: the last census?

Source: BBC, https://www.bbc.com/news/uk-51468919

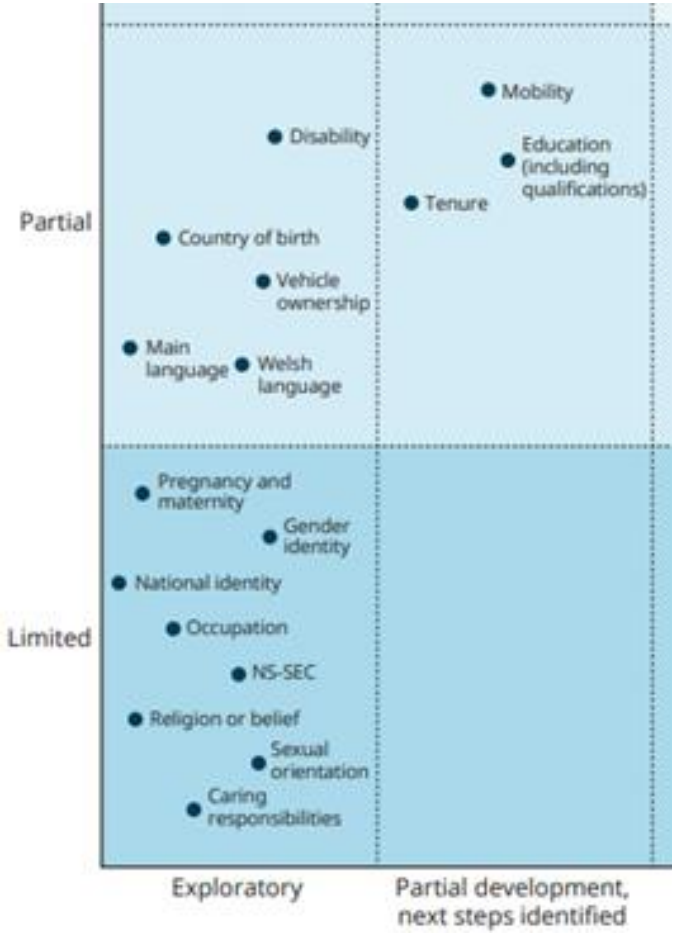
Source: BBC, https://www.bbc.com/news/10584385

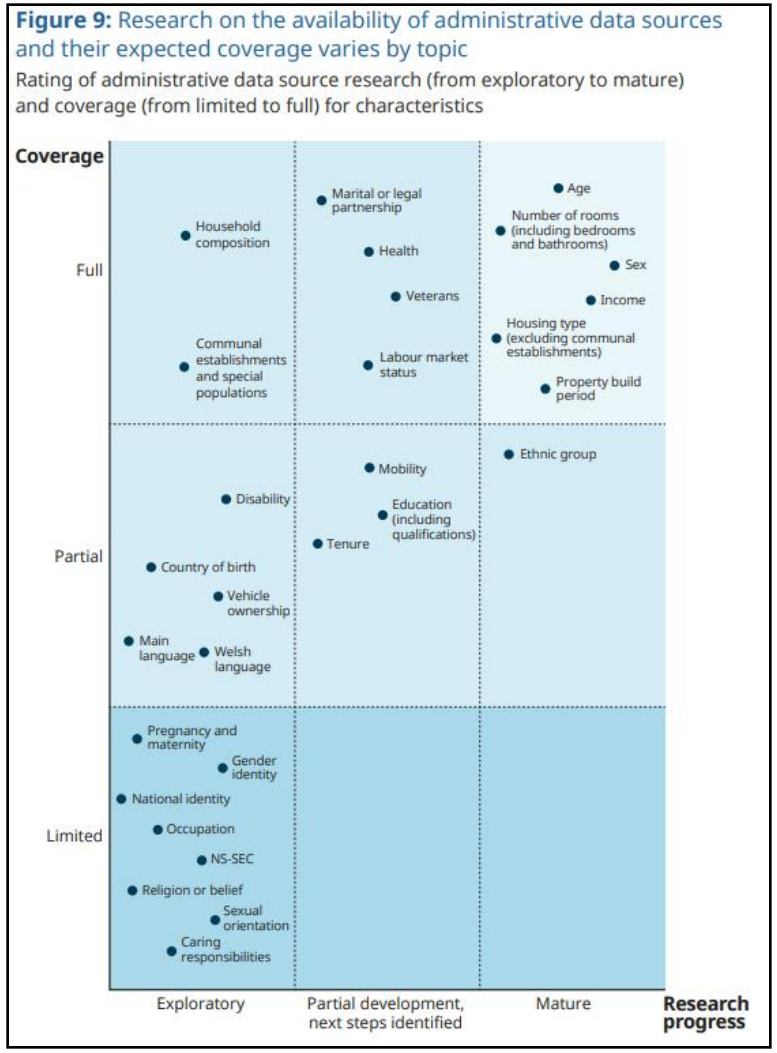
Administrative censuses

- ‘Spine’ constructed by matching records in different admin sources
  - Heath data
  - Education data
  - Tax and benefits data
    - Typically, need at least two sources to ‘confirm’ identity
  - Sources need cleaning
    - Remove / rationalise duplicates
- Basic age / sex / area counts can be derived, but will need to allow for undercount
- More detailed observations can be assembled by attaching attribute data

Challenges for data providers

- Linkage of records
- Range of variables
- Low level geography





Source: https://consultations.ons.gov.uk/ons/futureofpopulationandmigrationstatistics/

Research challenges

- What will 'census' data look like in the future
  - 2021 may be the last census
  - Future products may be based on administrative data
    - What will be the best tools to disseminate these?
    - What will be the best approach to archiving data?
    - What assumptions can / should we make?
- How will future researchers investigate change over time?
- Research on value of census
- Gauging costs and benefits is difficult

Research challenges

- How can variables of concern be estimated?
  - Aggregate estimates from surveys?
  - Partial / rolling linkage?
  - Use of non-traditional sources?

- Challenges for the Data Service
- Unclear future: administrative data
- Challenges for longitudinal data
- Challenges for the 22nd century

The ONS Longitudinal Study

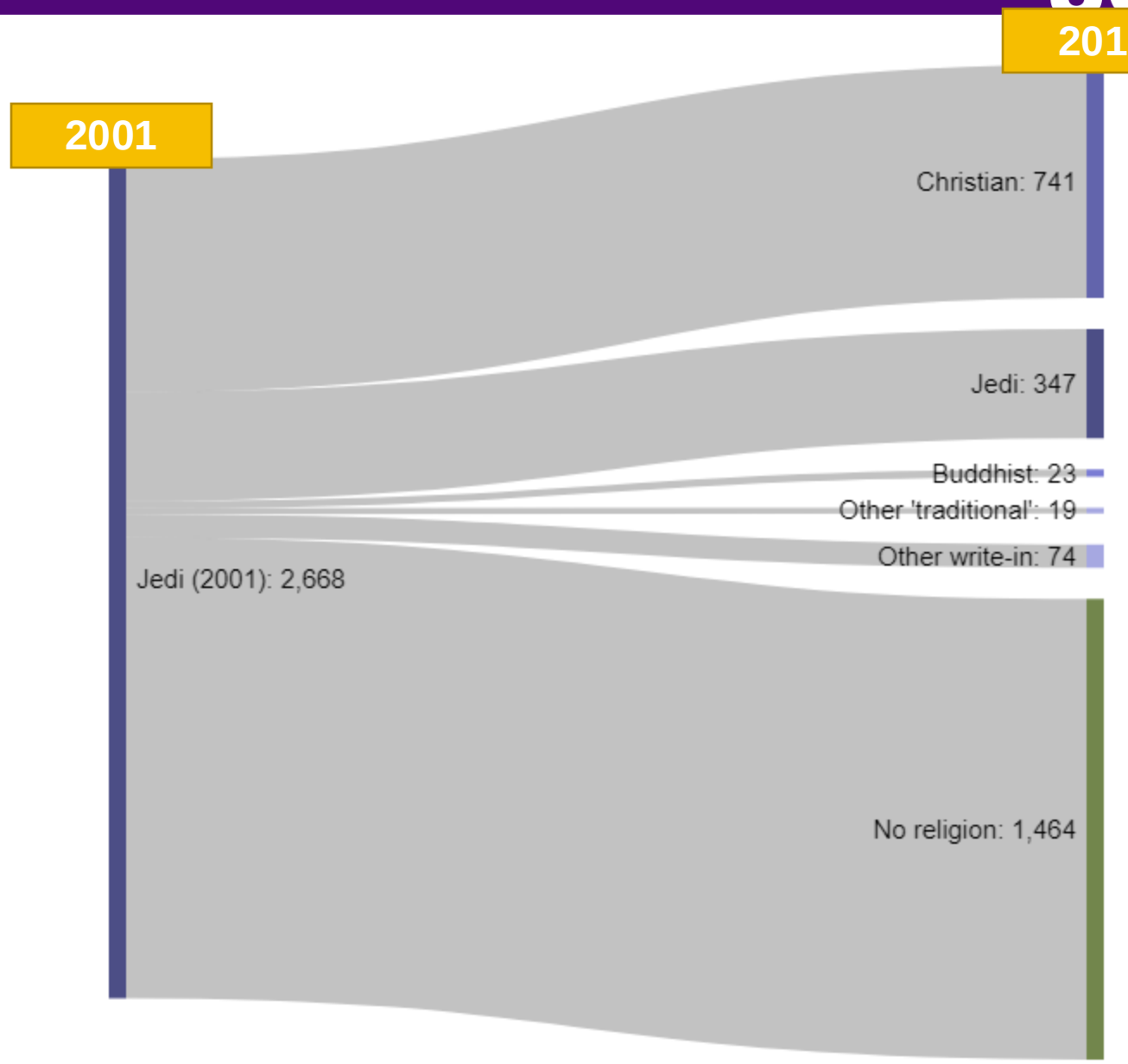
- The ONS LS is based on 4 sample birth dates
  - Approximately 1% of population (4/365)
  - Sample members do not know that they are in the sample
- The data for England and Wales include both census data (1971-2011) and administrative data, and some vital events data
- Access to the data is controlled

Who changes religion?

- Religion is not an immutable characteristic; it can change over time
- The LS allows us to explore changes in religion
- We select sample members who were present at both 2001 and 2011

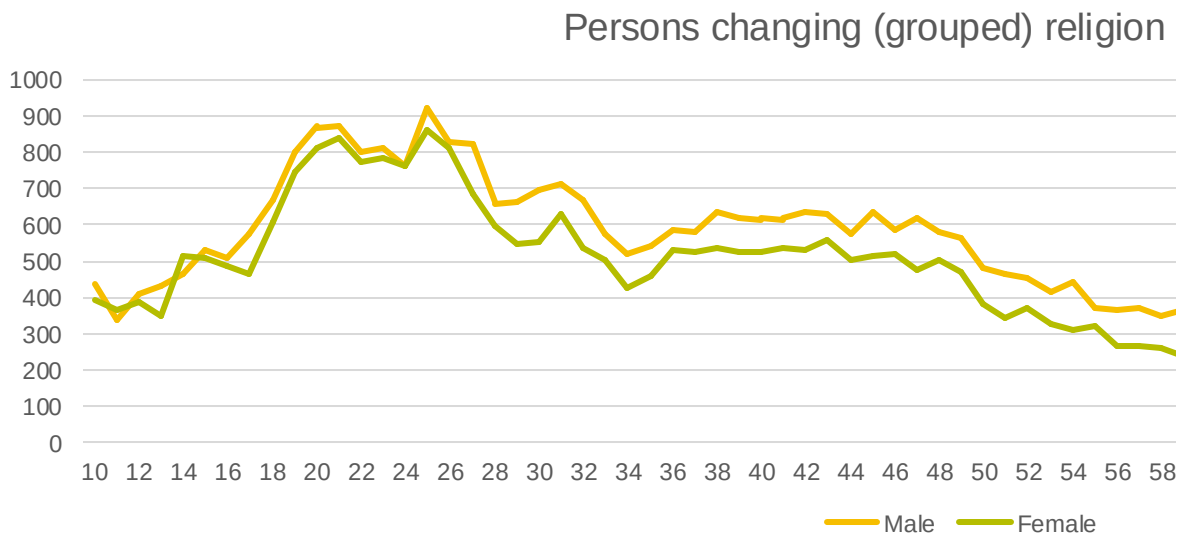
Jedi exodus

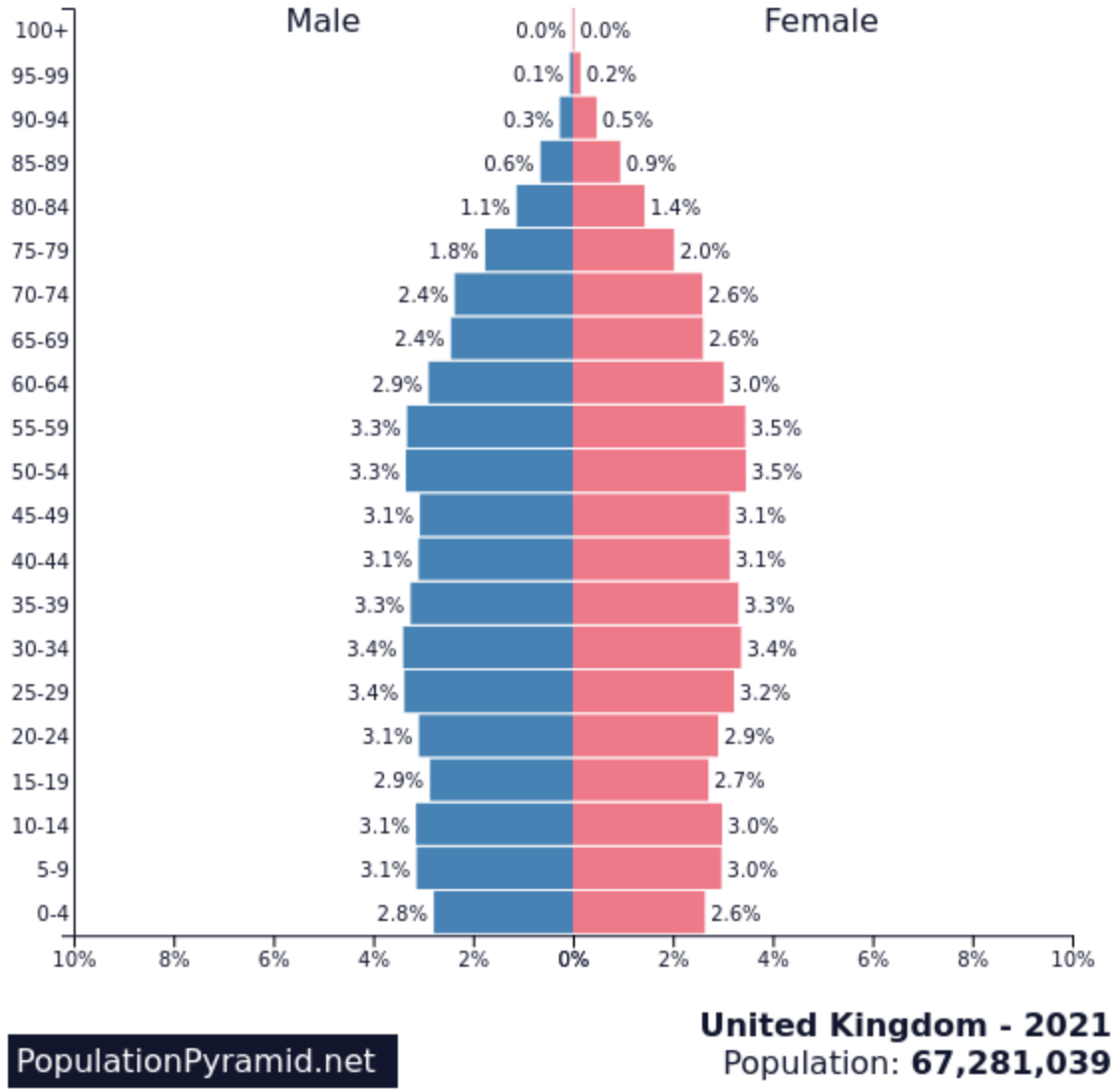
Star Wars (1977), Twentieth Century Fox



Source: ONS Longitudinal Study

Who changes religion?





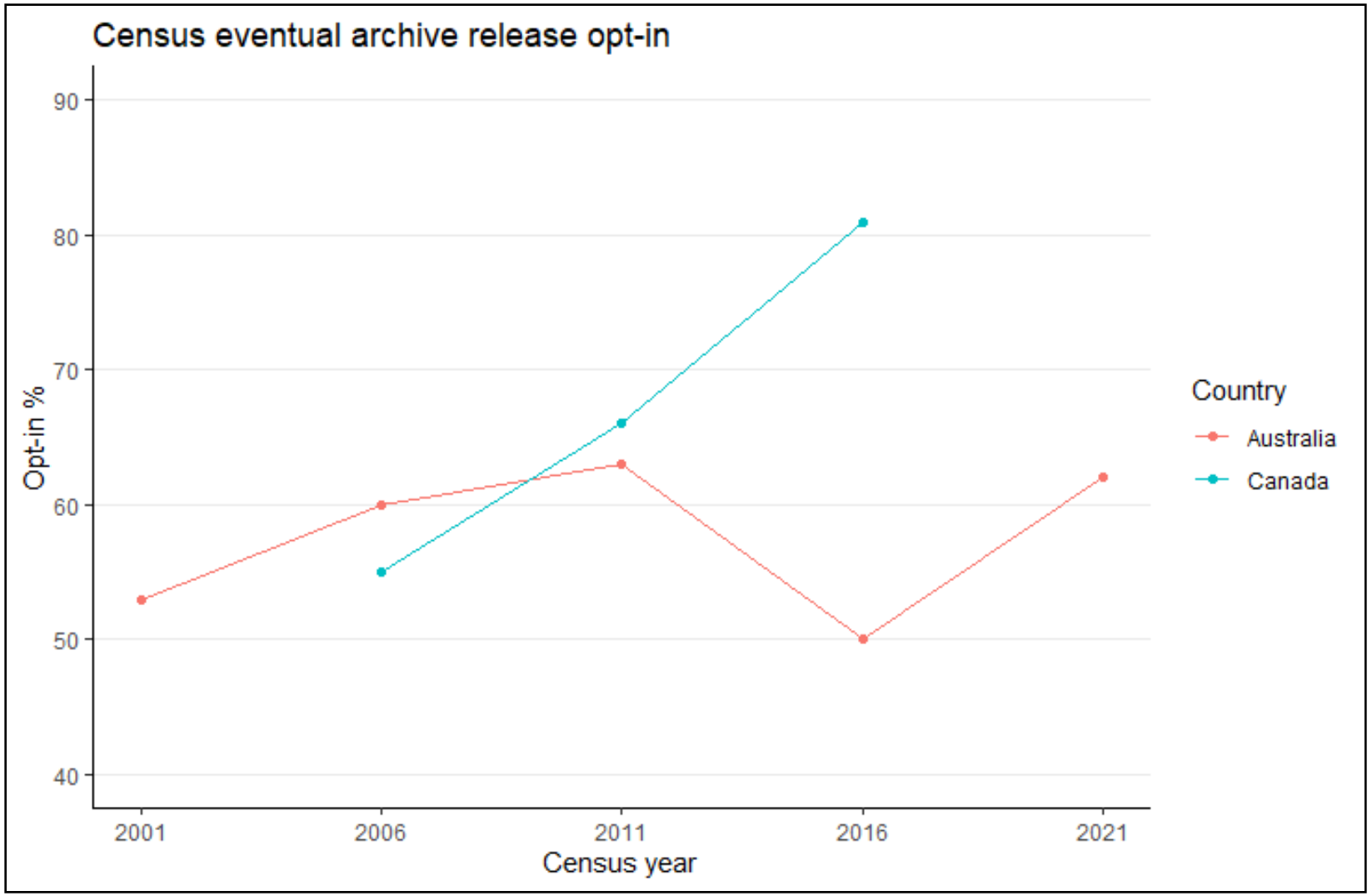
Source: ONS Longitudinal Study

Research challenges

- How to proceed without new census data?
  - Continue to add admin data?
  - What will this contain? Will it be consistent?
  - Will it be useful as time moves on?
- How to proceed with admin census data?
  - How important are variables that might not be included?

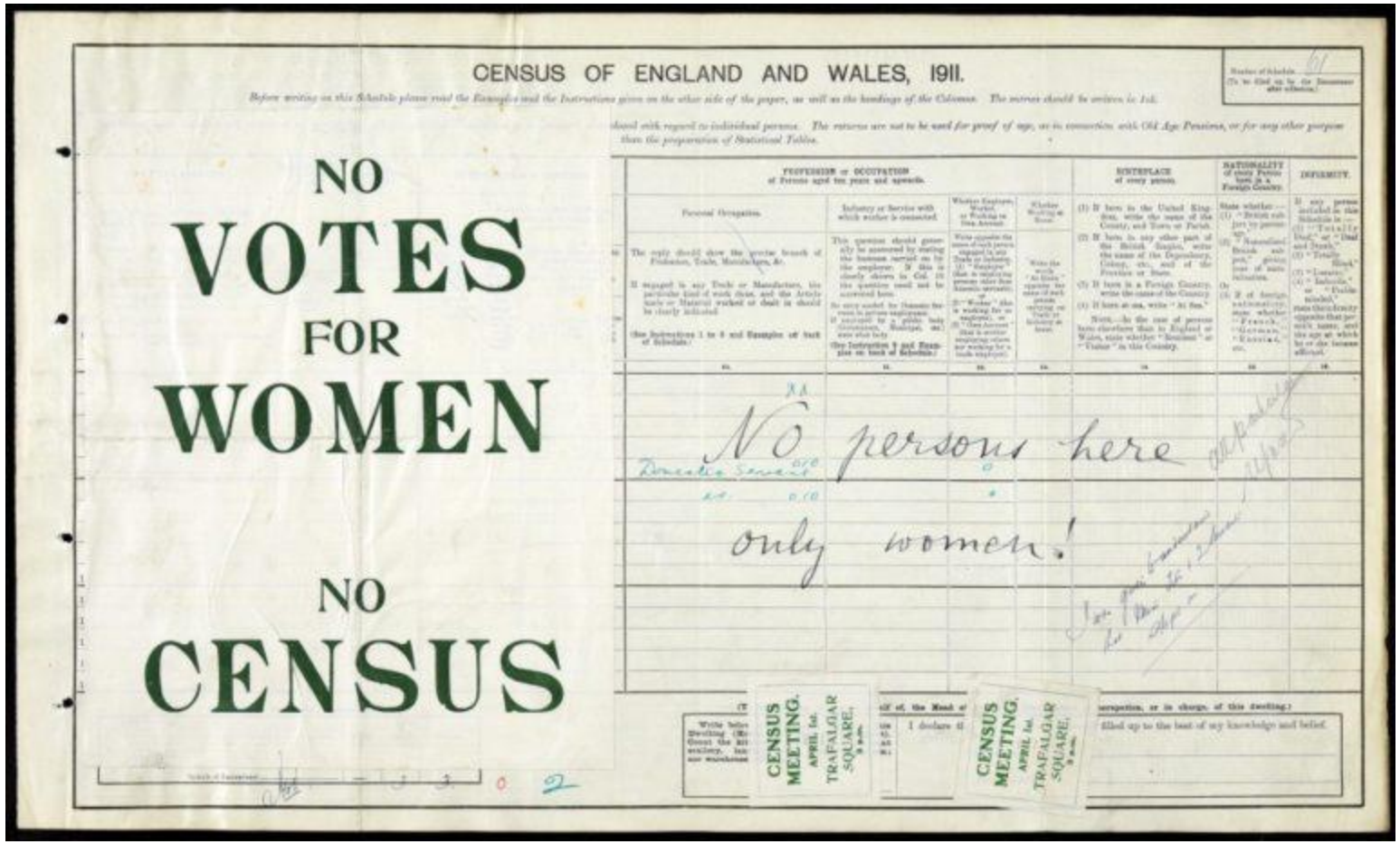
- Challenges for the Data Service
- Unclear future: administrative data
- Challenges for longitudinal data
- Challenges for the 22nd century

Privacy/trust and data retention



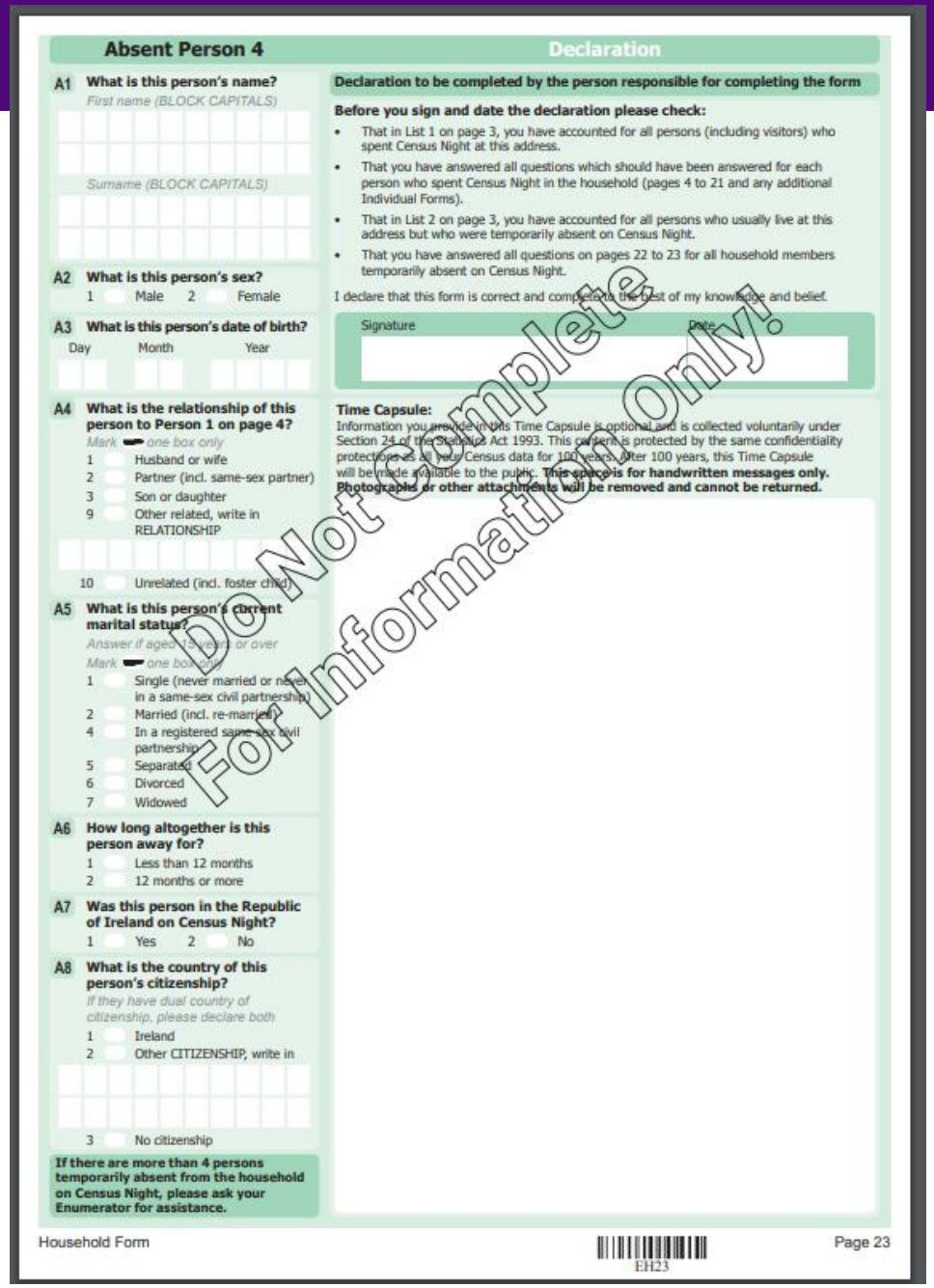
Challenges for the future

- In the UK: all data retained, released after 100 years
  - What data should be retained for future historians, historical demographers, genealogists etc?
  - Census schedules make it fairly obvious (most of the time) what to keep
  - How do we select administrative data to retain?
  - How important is emotional engagement with the retained data? How excited will our descendants be about transactional records of routine occurrences?



Source: https://blog.nationalarchives.gov.uk/no-vote-no-census-the-1911-suffrage-census-protests/

- ‘Time capsule’ free text response
- Ireland, 2022



Absent Person 4

A1 What is this person's name?
First name (BLOCK CAPITALS)

Surname (BLOCK CAPITALS)

A2 What is this person's sex?
1 Male 2 Female

A3 What is this person's date of birth?
Day Month Year

A4 What is the relationship of this person to Person 1 on page 4?
Mark one box only
1 Husband or wife
2 Partner (incl. same-sex partner)
3 Son or daughter
9 Other related, write in RELATIONSHIP
10 Unrelated (incl. foster child)

A5 What is this person's current marital status?
Answer if aged 15 years or over
Mark one box only
1 Single (never married or never in a same-sex civil partnership)
2 Married (incl. re-married)
4 In a registered same-sex civil partnership
5 Separated
6 Divorced
7 Widowed

A6 How long altogether is this person away for?
1 Less than 12 months
2 12 months or more

A7 Was this person in the Republic of Ireland on Census Night?
1 Yes 2 No

A8 What is the country of this person's citizenship?
If they have dual country of citizenship, please declare both
1 Ireland
2 Other CITIZENSHIP, write in
3 No citizenship

If there are more than 4 persons temporarily absent from the household on Census Night, please ask your Enumerator for assistance.

Declaration

Declaration to be completed by the person responsible for completing the form

Before you sign and date the declaration please check:

- That in List 1 on page 3, you have accounted for all persons (including visitors) who spent Census Night at this address.
- That you have answered all questions which should have been answered for each person who spent Census Night in the household (pages 4 to 21 and any additional Individual Forms).
- That in List 2 on page 3, you have accounted for all persons who usually live at this address but who were temporarily absent on Census Night.
- That you have answered all questions on pages 22 to 23 for all household members temporarily absent on Census Night.

I declare that this form is correct and complete to the best of my knowledge and belief.

Signature Date

Time Capsule:
Information you provide in this Time Capsule is optional and is collected voluntarily under Section 24 of the Statistics Act 1993. This content is protected by the same confidentiality protections as all 2022 Census data for 100 years. After 100 years, this Time Capsule will be made available to the public. **This space is for handwritten messages only. Photographs or other attachments will be removed and cannot be returned.**

Do Not Complete. For Information Only.

Source:https://www.cso.ie/en/media/csoie/census/census2021/Sample_Census_2022_Household_Form_English.pdf

Acknowledgements

The permission of the Office for National Statistics to use the Longitudinal Study is gratefully acknowledged, as is the help provided by staff of the Centre for Longitudinal Study Information & User Support (CeLSIUS). CeLSIUS is funded by the ESRC under project ES/V003488/1. The authors alone are responsible for the interpretation of the data.

This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates