

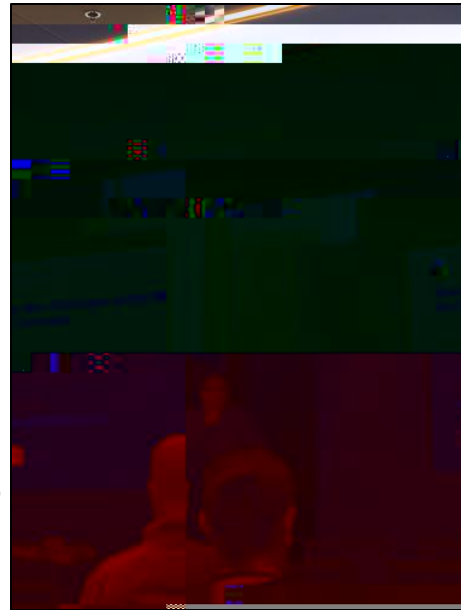
# Research challenges with the UK Census

Oliver Duke-Williams

UCL DIS Research Open Day, June 2024



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



## Challenges for a data service

Unclear future: administrative data

Challenges for longitudinal data

Challenges for the 22<sup>nd</sup> 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



We do not have a 'UK Census' but rather three separate decennial censuses, carried out by three agencies

- England and Wales (Office for National Statistics, ONS)

# 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Our legacy systems which provided access to data from 2011 and earlier are outdated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

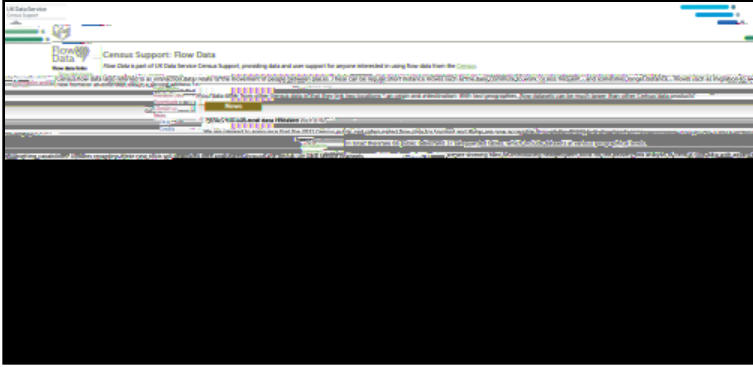
Many tables, very large tables require complex (bespoke?) discovery and sub-setting

New competitors

Custom datasets available through ONS

Updating training materials etc

# Challenges: legacy systems



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

```

JSON Raw Data Headers
[{"dataset_id": 2, "table_id": "1", "sequence": 1, "table_title": "All migrants: age (5 broad age groups) by sex", "dataset_label": "1991 SMS Set 1", "description": "Ward level migration data from the 1991 Census", "family_id": 1, "series_id": 2, "geography": {}, "topics": {}, "theme_aliases": []}, {"dataset_id": 2, "table_id": "2", "sequence": 2, "table_title": "Wholly Moving Households_ving Households: counts", "dataset_label": "1991 SMS Set 1", "description": "Ward level migration data from the 1991 Census", "family_id": 1, "series_id": 2, "geography": {}, "topics": {}, "theme_aliases": []}, {"dataset_id": 2, "table_id": "3", "sequence": 3, "table_title": "", "dataset_label": "", "description": "", "family_id": 1, "series_id": 2, "geography": {}, "topics": {}, "theme_aliases": []}, {"dataset_id": 2, "table_id": "4", "sequence": 4, "table_title": "", "dataset_label": "", "description": "", "family_id": 1, "series_id": 2, "geography": {}, "topics": {}, "theme_aliases": []}, {"dataset_id": 2, "table_id": "5", "sequence": 5, "table_title": "", "dataset_label": "", "description": "", "family_id": 1, "series_id": 2, "geography": {}, "topics": {}, "theme_aliases": []}, {"dataset_id": 2, "table_id": "6", "sequence": 6, "table_title": "", "dataset_label": "", "description": "", "family_id": 1, "series_id": 2, "geography": {}, "topics": {}, "theme_aliases": []}, {"dataset_id": 2, "table_id": "7", "sequence": 7, "table_title": "", "dataset_label": "", "description": "", "family_id": 1, "series_id": 2, "geography": {}, "topics": {}, "theme_aliases": []}, {"dataset_id": 2, "table_id": "8", "sequence": 8, "table_title": "", "dataset_label": "", "description": "", "family_id": 1, "series_id": 2, "geography": {}, "topics": {}, "theme_aliases": []}, {"dataset_id": 2, "table_id": "9", "sequence": 9, "table_title": "", "dataset_label": "", "description": "", "family_id": 1, "series_id": 2, "geography": {}, "topics": {}, "theme_aliases": []}, {"dataset_id": 2, "table_id": "10", "sequence": 10, "table_title": "", "dataset_label": "", "description": "", "family_id": 1, "series_id": 2, "geography": {}, "topics": {}, "theme_aliases": []}

```

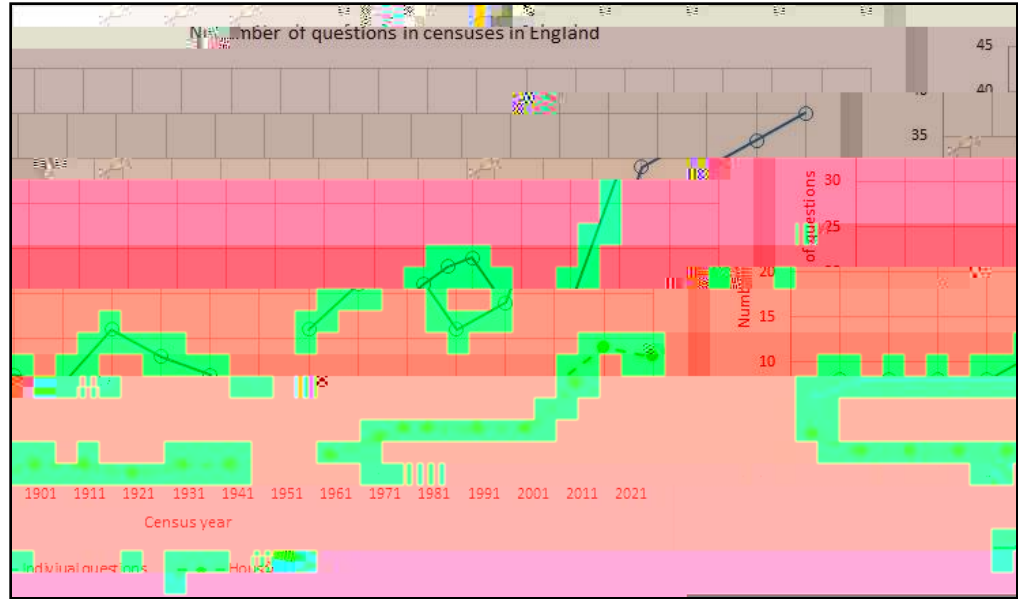
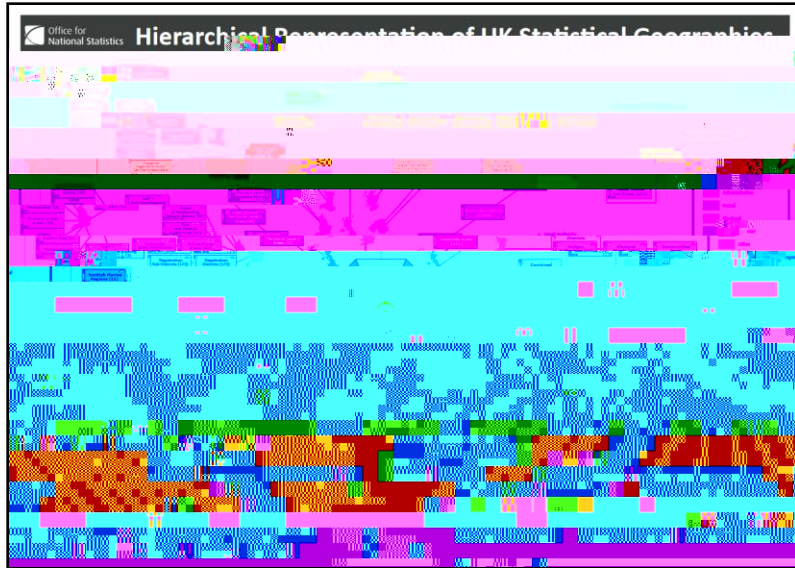
```

JSON Raw Data Headers
Save Copy Collapse All Expand All Filter JSON
data:
  0: {}
  1: {}
  2: {
    dataset_id: 2,
    table_id: "1",
    sequence: 1,
    table_title: "All migrants: age (5 broad age groups) by sex",
    dataset_label: "1991 SMS Set 1",
    description: "Ward level migration data from the 1991 Census",
    family_id: 1,
    series_id: 2,
    geography: {},
    topics: {},
    theme_aliases: []
  },
  3: {
    dataset_id: 2,
    table_id: "2",
    sequence: 2,
    table_title: "Wholly Moving Households_ving Households: counts",
    dataset_label: "1991 SMS Set 1",
    description: "Ward level migration data from the 1991 Census",
    family_id: 1,
    series_id: 2,
    geography: {},
    topics: {},
    theme_aliases: []
  },
  4: {
    dataset_id: 2,
    table_id: "3",
    sequence: 3,
    table_title: "",
    dataset_label: "",
    description: "",
    family_id: 1,
    series_id: 2,
    geography: {},
    topics: {},
    theme_aliases: []
  },
  5: {
    dataset_id: 2,
    table_id: "4",
    sequence: 4,
    table_title: "",
    dataset_label: "",
    description: "",
    family_id: 1,
    series_id: 2,
    geography: {},
    topics: {},
    theme_aliases: []
  },
  6: {
    dataset_id: 2,
    table_id: "5",
    sequence: 5,
    table_title: "",
    dataset_label: "",
    description: "",
    family_id: 1,
    series_id: 2,
    geography: {},
    topics: {},
    theme_aliases: []
  },
  7: {
    dataset_id: 2,
    table_id: "6",
    sequence: 6,
    table_title: "",
    dataset_label: "",
    description: "",
    family_id: 1,
    series_id: 2,
    geography: {},
    topics: {},
    theme_aliases: []
  },
  8: {
    dataset_id: 2,
    table_id: "7",
    sequence: 7,
    table_title: "",
    dataset_label: "",
    description: "",
    family_id": 1,
    series_id: 2,
    geography: {},
    topics: {},
    theme_aliases: []
  },
  9: {
    dataset_id: 2,
    table_id: "8",
    sequence: 8,
    table_title: "",
    dataset_label: "",
    description: "",
    family_id: 1,
    series_id: 2,
    geography: {},
    topics: {},
    theme_aliases: []
  },
  10: {
    dataset_id: 2,
    table_id: "9",
    sequence: 9,
    table_title: "",
    dataset_label: "",
    description: "",
    family_id: 1,
    series_id: 2,
    geography: {},
    topics: {},
    theme_aliases: []
  },
  11: {
    dataset_id: 2,
    table_id: "10",
    sequence: 10,
    table_title: "",
    dataset_label: "",
    description: "",
    family_id: 1,
    series_id: 2,
    geography: {},
    topics: {},
    theme_aliases: []
  }
}

```



# Challenges: data volume and complexity



Source: [https://x.com/oliver\\_dw/status/1164521314360733701?s=20](https://x.com/oliver_dw/status/1164521314360733701?s=20)

Source: <https://geoportal.statistics.gov.uk/documents/ons::hierarchical-representation-of-uk-statistical-geographies>

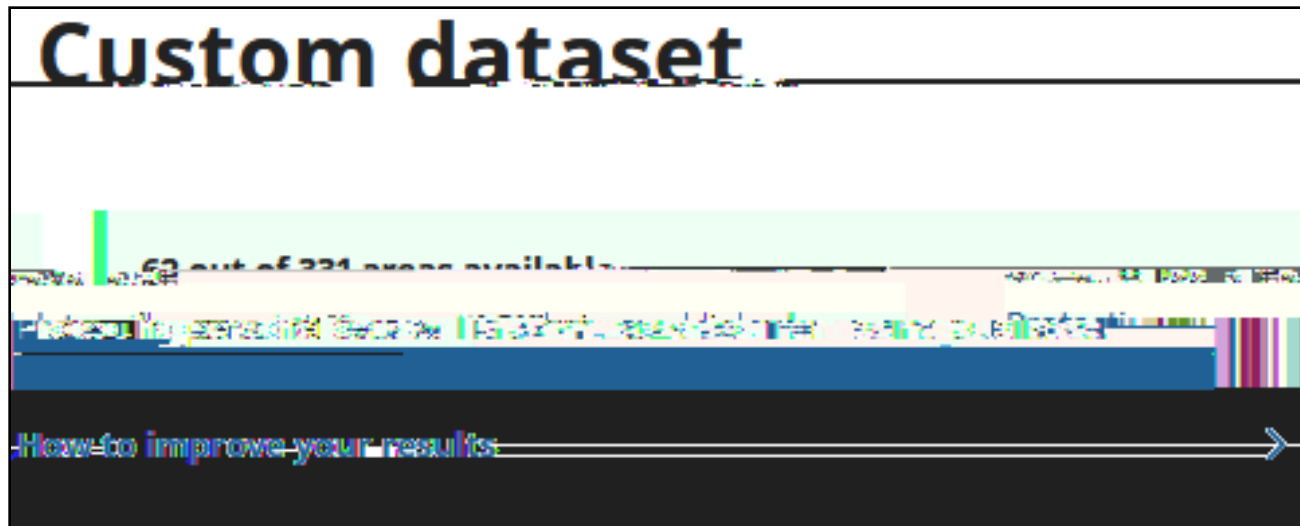
# Challenge: new competition

## ONS custom dataset builder

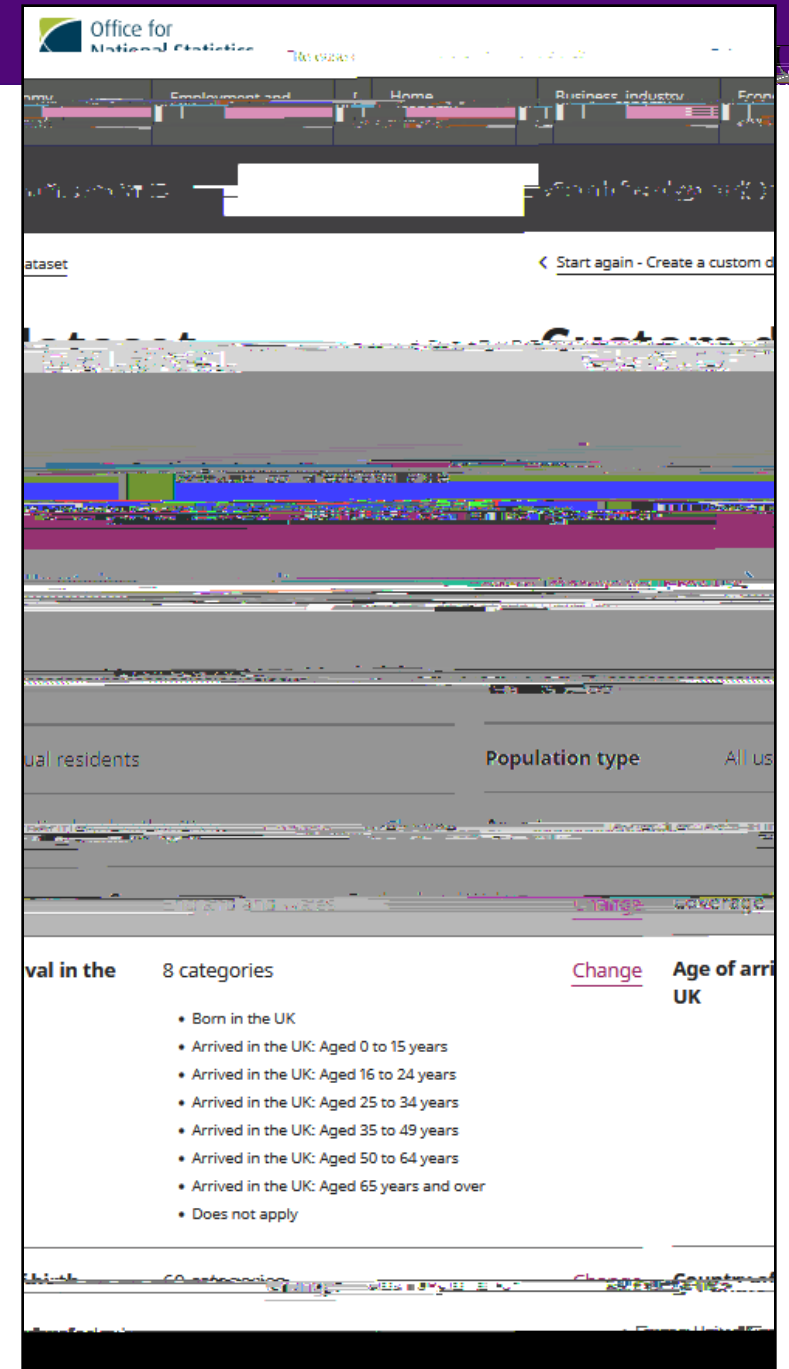
Easy to use

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?

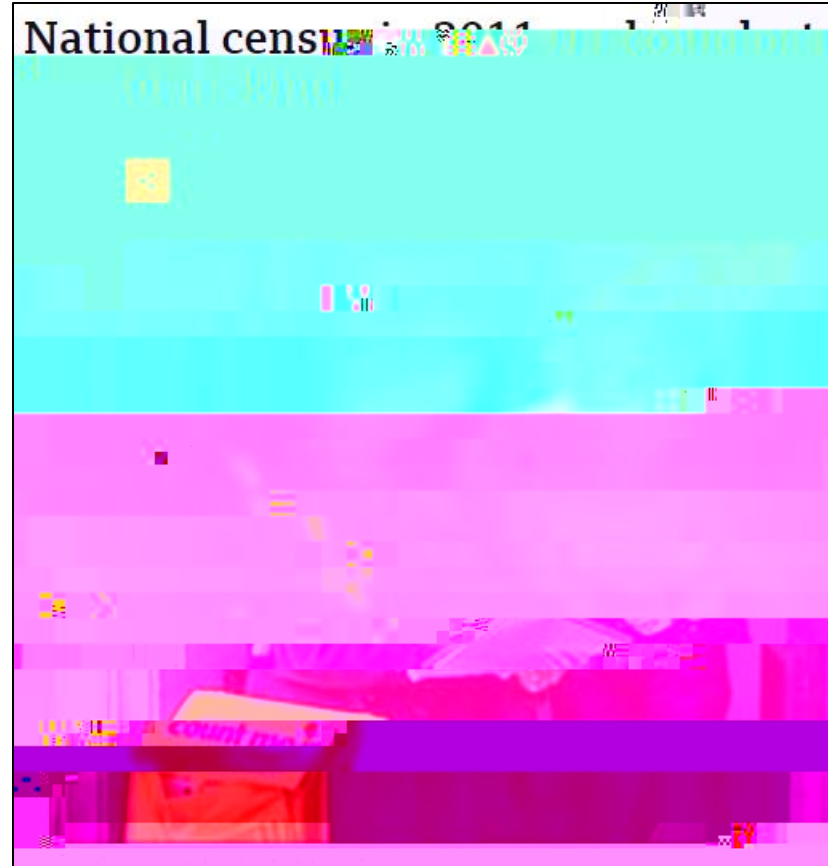
Challenges for the Data Service

**Unclear future: administrative data**

Challenges for longitudinal data

Challenges for the 22<sup>nd</sup> century

# An unclear future: the last census?



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

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

# Administrative censuses

## sources

Health data

Education data

Tax and benefits data

## Sources need cleaning

Remove / rationalise duplicates



# Research challenges



# 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

# 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

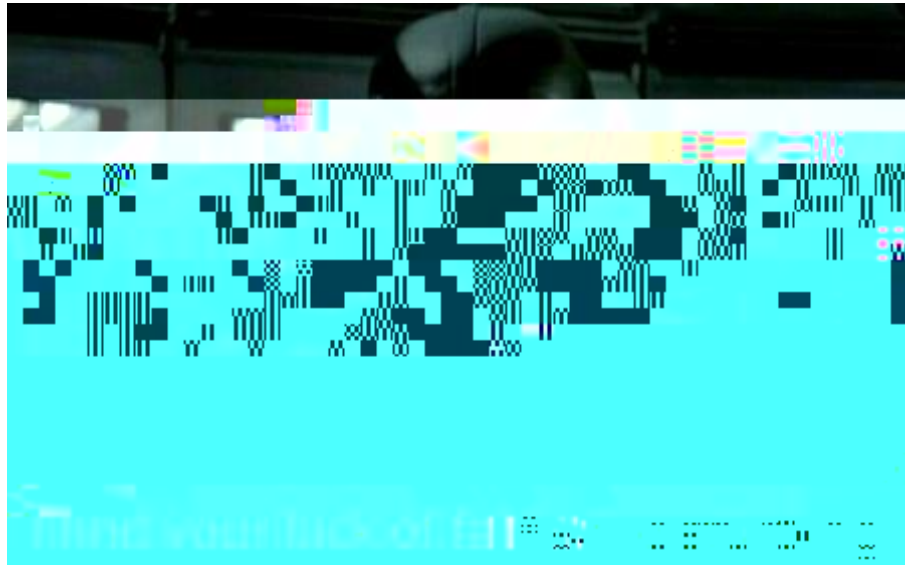
# Who changes religion?

Religion is not an immutable characteristic; it can change over time

The LS allows us to explore changes in religion

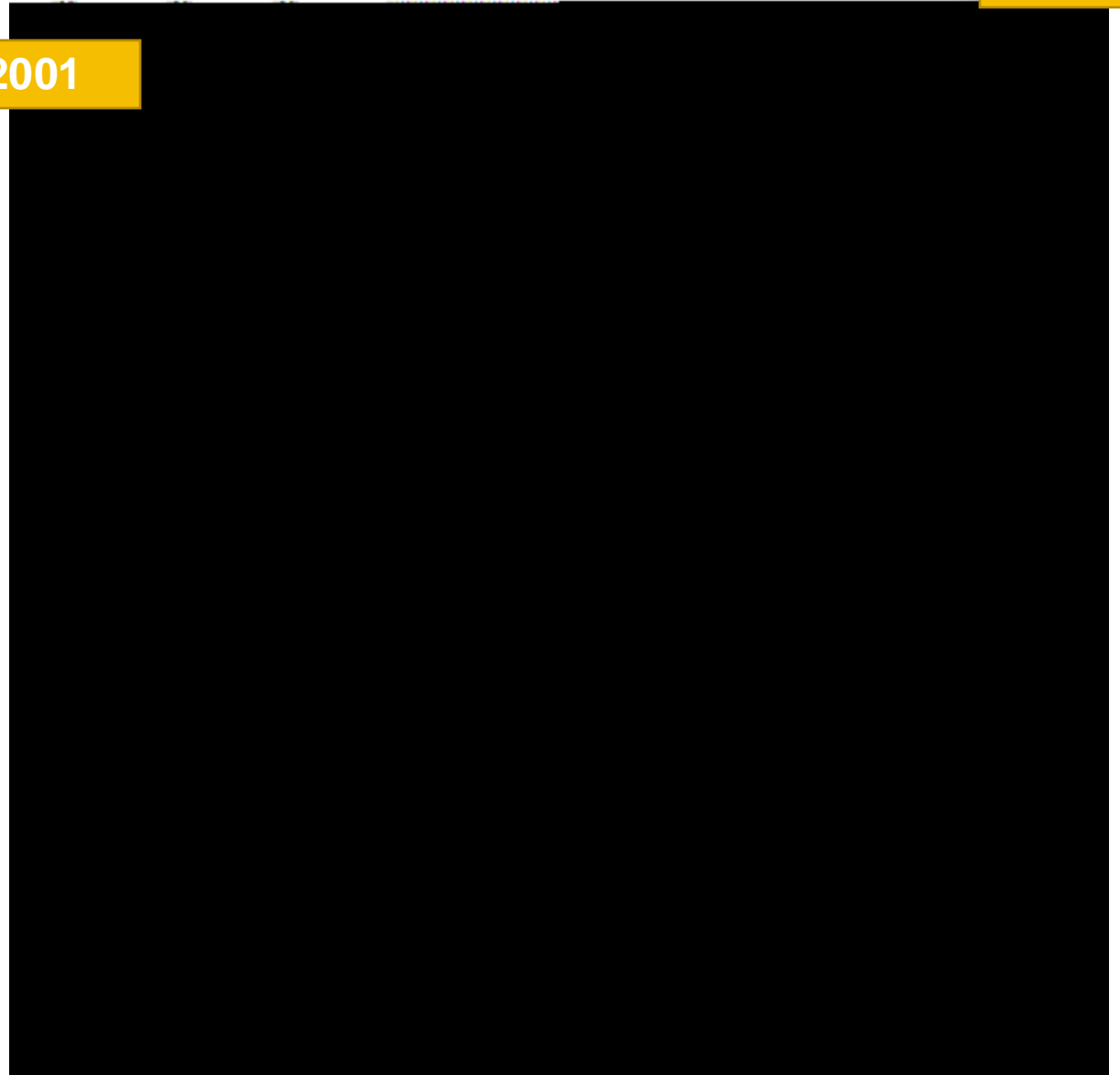
We select sample members who were present at both 2001 and 2011

# Jedi exodus



Star Wars (1977), Twentieth Century Fox

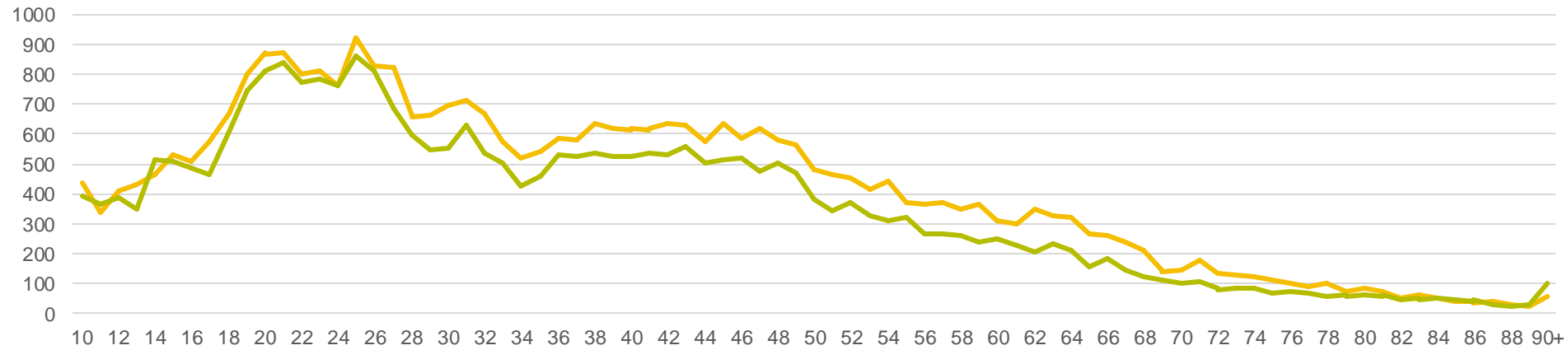
2001



Source: ONS Longitudinal Study

# Who changes religion?

Persons changing (grouped) religion 2001



# 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

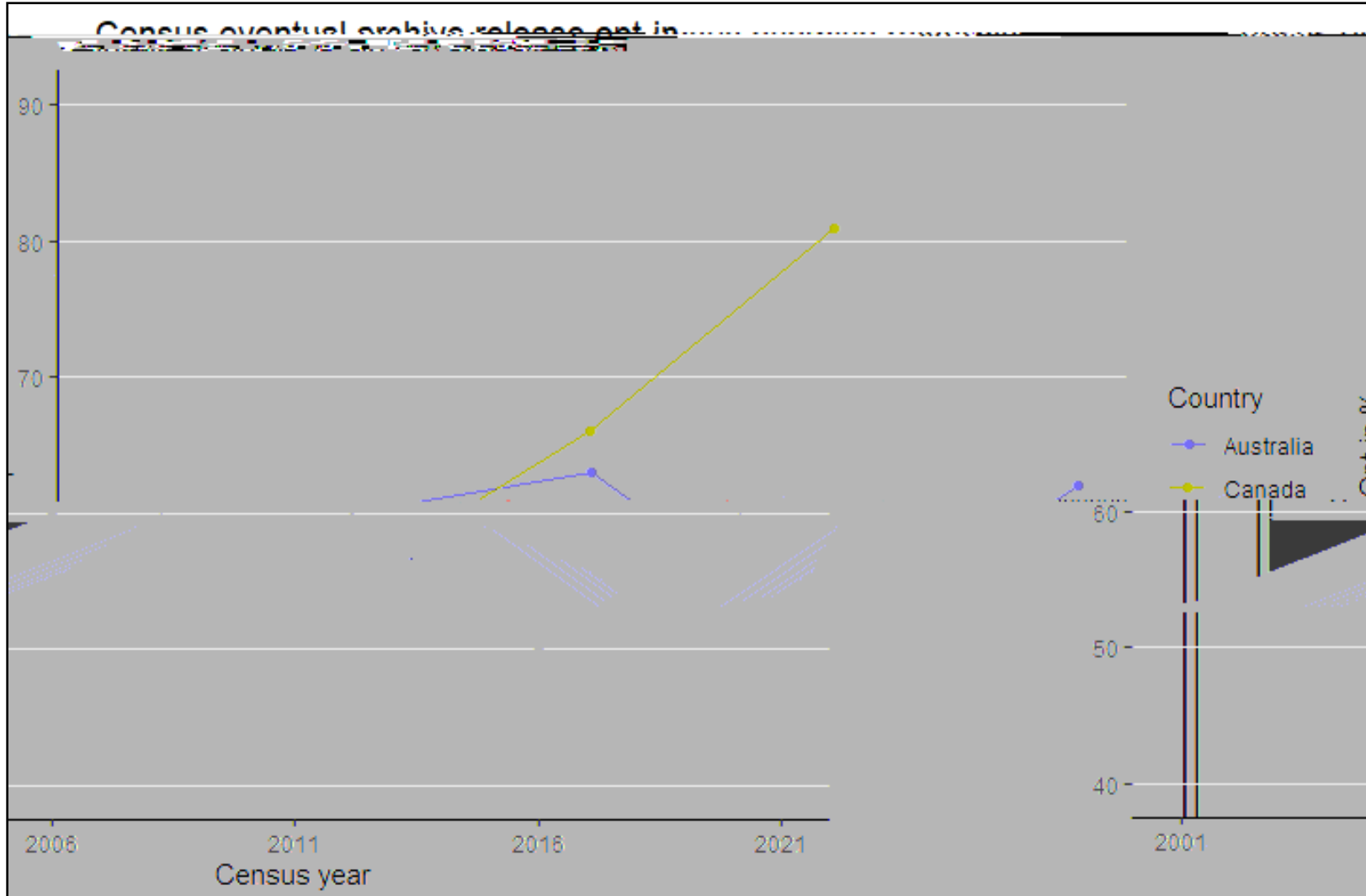
Unclear future: administrative data

Challenges for longitudinal data

**Challenges for the 22<sup>nd</sup> century**



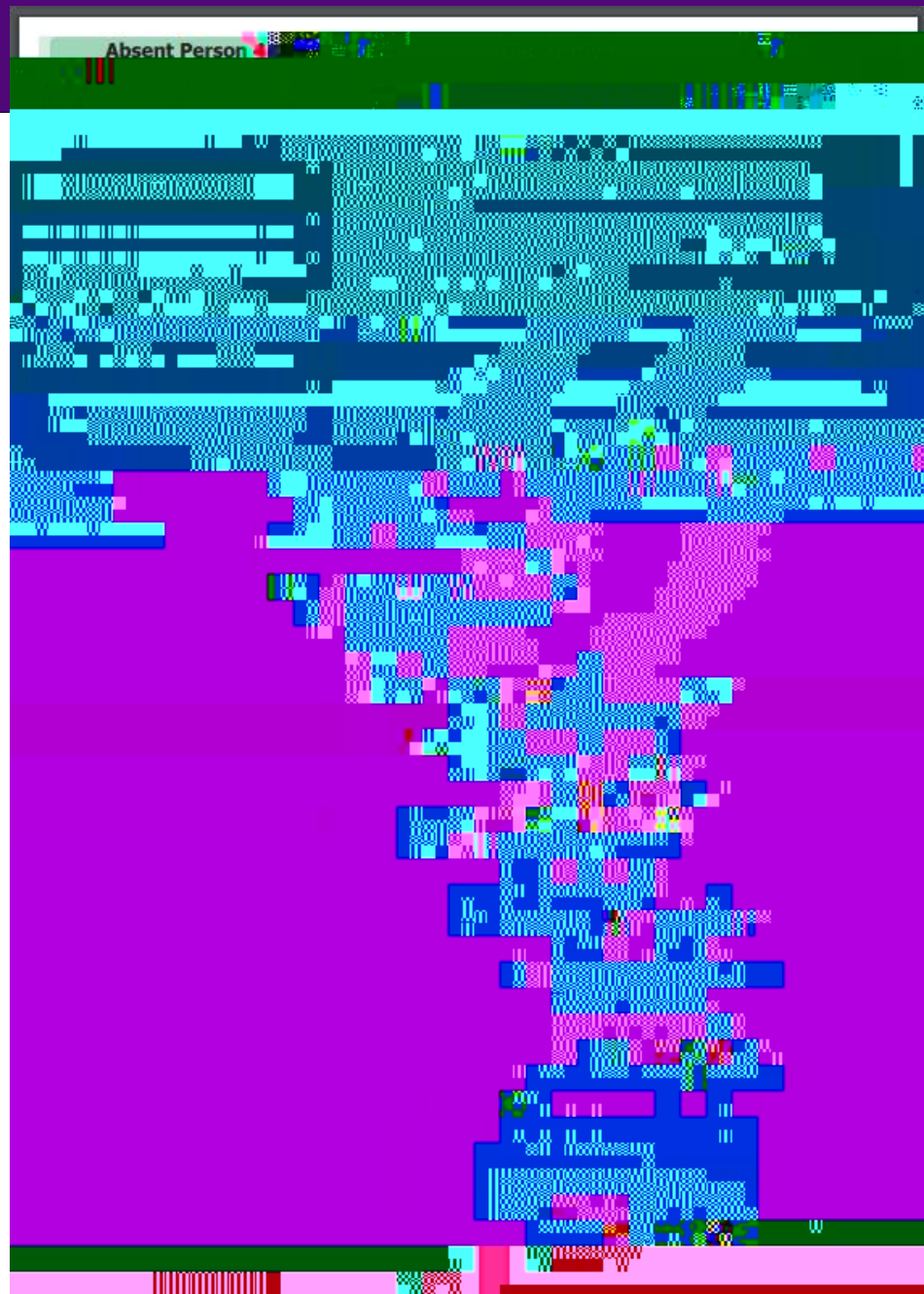
# Privacy/trust and data retention



# Challenges for the future



text response  
Ireland, 2022



## **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