UK Data Service

Representing the users of UK demographic statistics

Professor Oliver Duke-William

o.duke-williams@ucl.ac.uk

Economic and Social Research Council

"Users of UK demographic statistics"

- A rich variety of data, not just census
- Not just data: data, metadata, documentation, training etc
- Users (and potential users) not just UK-based
- Users (and potential users) not just within 'usual' user groups
- UK Data Service is not the only story

UK Data Service

- ESRC funded
- A research infrastructure for curating and providing access to social science data
- Universities of Essex, Manchester, UCL, Edinburgh and Jisc
- Manages access to data in a number of repositories
- Engagement, impact, training etc

Browse data by theme or type

Ageing, COVID-19, crime, economics education, environment & energy, ethnicity, food, health housing, information & communication, labour, politics, poverty.

Repositories that are integrated by UK Data Service

- Various repositories, integrated to differing degrees
  - UK Data Archive
  - Economic and Social Data Service
  - Secure Data Service
  - Census Programme
  - Survey Question Bank
  - UKDS repositories

UKDS Licencing Framework

Open	• No real disclosure risk. Under open licences; almost no restrictions on reuse
Safeguarded	• Zero to low real disclosure risk. Requires authentication and authorisation e.g. registered user and End User Licence Agreement
Controlled	• Real disclosure risk. Requires project approval, user vetting and training; access via a safe setting; output checking

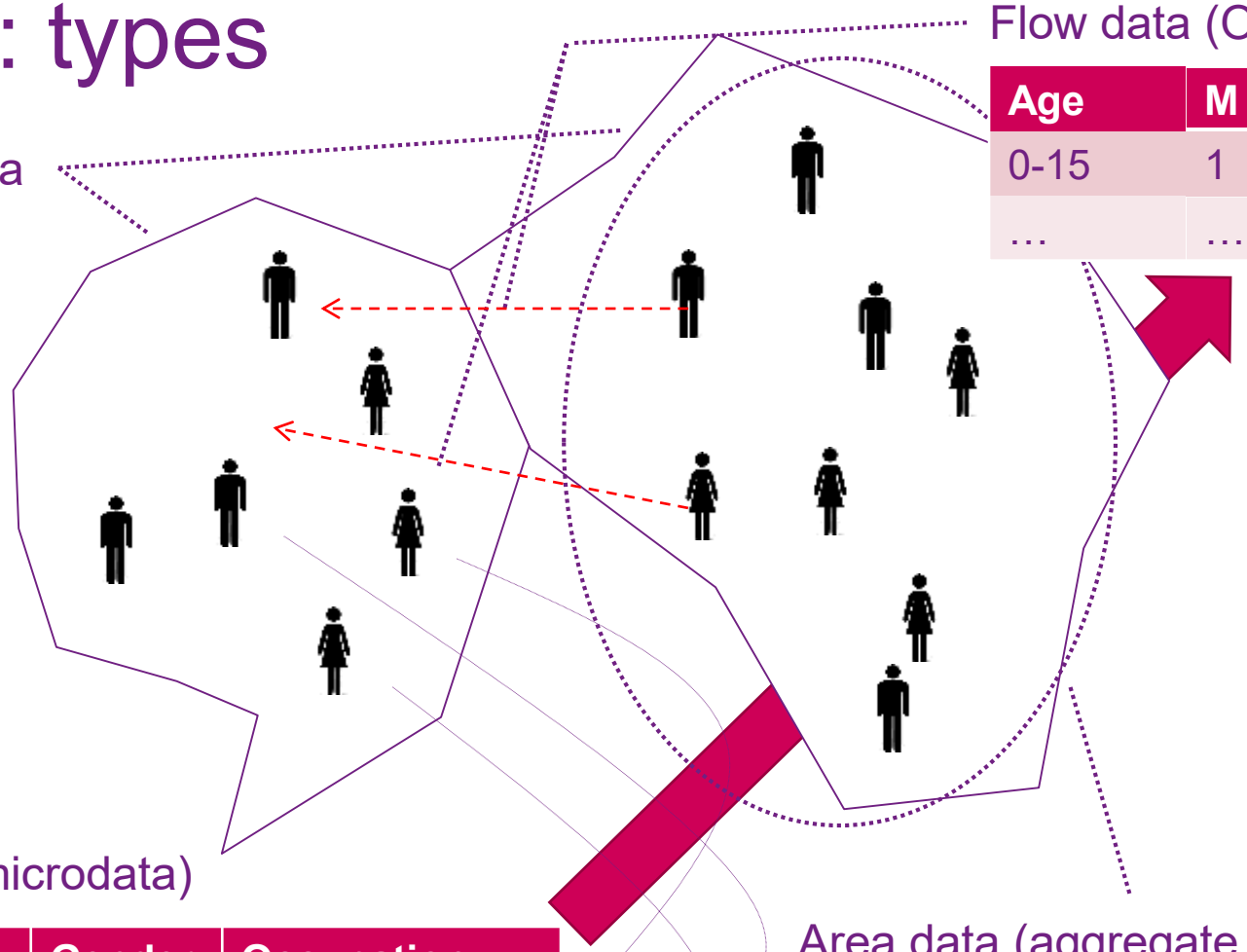
- ***User requirements***
  - ***Users want access to be as simple as possible***
  - ***Developers strive to make access look simple***
  - ***Many (most?) users recognise the need for controlling access to the most disclosive data***
    - ***The ability to access secure data remotely has improved the experience for some researchers***
    - ***There are still delays and hurdles***

Census data

- UKDS Census provides access to a variety of outputs
- Census data can be divided in a number of ways
  - Types of data
  - Sources of data
  - Census years
  - Data licenses

Census data: types

Boundary data



Flow data (O/D data)

Age	M	F
0-15	1	2
…	…	…

Individual data (microdata)

ID	Age	Gender	Occupation
0001	23	F	Teacher
0002	17	M	Student
0003	45	F	Doctor
…	…	…	…

Area data (aggregate data)

Age	M	F
0-4	7	6
5-9	8	9
…	…	…

UKDS Census

- UKDS Census as an example of heterogeneous architecture
  - Different type of data, with different interfaces
  - Microdata fairly well integrated
  - Other types have bespoke interfaces querying separate databases
    - Newer systems being developed with API structure
    - APIs could in turn call upstream repositories
- A variety of other access routes as well, for some data types
  - Statistical agencies
  - NOMIS

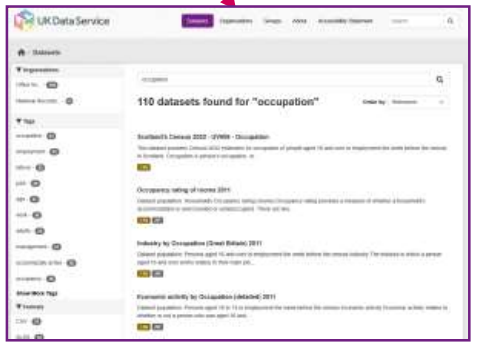
UKDS Census – heterogeneous architecture

UK Data Service website

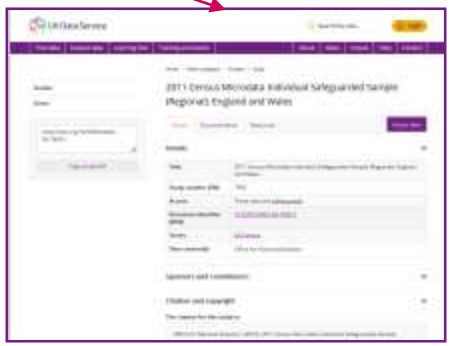
Catalogue queries

Boundary data

Flow data



Aggregate data



Microdata

Heterogeneous architecture

- Users probably do not want to have to learn new interfaces
- **BUT** there are substantial variations in data structures

- Newer developments
  - Data Explorer (aggregate data)
  - New Flow Data API and front end
  - CORDIAL-AI

- ***User requirements for access are diverse***
  - ***Different data types***
  - ***Different levels of user knowledge***
  - ***Different volumes of data***
  - ***Different levels of user technical expertise***
  - ***Different data formats***

Office for National Statistics

Release ca

Home | Business, industry and trade | Economy | Employment and labour market | Pe an

Search for a keyword(s) or time series ID

< Start again - Create a custom dataset

Custom dataset

176 out of 331 areas available

Protecting personal data will prevent 155 areas from being published.

> How to improve your results

Variables

Population type	All usual residents	
Area type	Lower tier local authorities	Change
Coverage	England and Wales	Change
Age (B)	4 categories • Aged 15 years and under • Aged 16 to 49 years • Aged 50 to 64 years • Aged 65 years and over	Change
Main language	11 categories	Change

NISRA | FLEXIBLE TABLE BUILDER

census 2021

CUSTOM TABLE

Main Language - 11 Categories by Age - 4 Categories

Download | View data

Summary

This table provides Census 2021 estimates that classify People by Main Language - 11 Categories by Age - 4 Categories for Northern Ireland. The table contains 44 counts.

Your table

Population	People	Start again
Geographic level	Northern Ireland	Change
Geographic area	Northern Ireland	
Variables	Main Language - 11 Categories, Age - 4 Categories	Change
Filters	None selected	Filter table
Pivot	No pivot applied	Pivot table

Download >

NISRA FLEXIBLE TABLE BUILDER census 2021

CUSTOM TABLE

Main Language - 11 Categories by Age - 4 Categories

Download | View data

Main Language - 11 Categories Code	Main Language - 11 Categories Label	Age - 4 Categories Code	Age - 4 Categories Label	Count
1	English	1	0-15 years	309,870
1	English	2	16-39 years	530,504
1	English	3	40-64 years	587,775
1	English	4	65+ years	323,361
2	Polish	1	0-15 years	3,026
2	Polish	2	16-39 years	9,253
2	Polish	3	40-64 years	7,386

NISRA FLEXIBLE TABLE BUILDER census 2021

← Back

Pivot your table

Rows: ⠿ Main Language - 11 Categories

Columns: ⠿ Age - 4 Categories Clear pivot

Save and return or cancel

Count	Age - 4 Categories			
Main Language - 11 Categories	0-15 years	16-39 years	40-64 years	65+ years
English	309,870	530,504	587,775	323,361
Polish	3,026	9,253	7,386	469
Lithuanian	1,155	4,557	3,024	242
Irish	1,798	2,300	1,433	440
Romanian	816	3,051	1,677	84
Portuguese	556	2,426	1,831	169

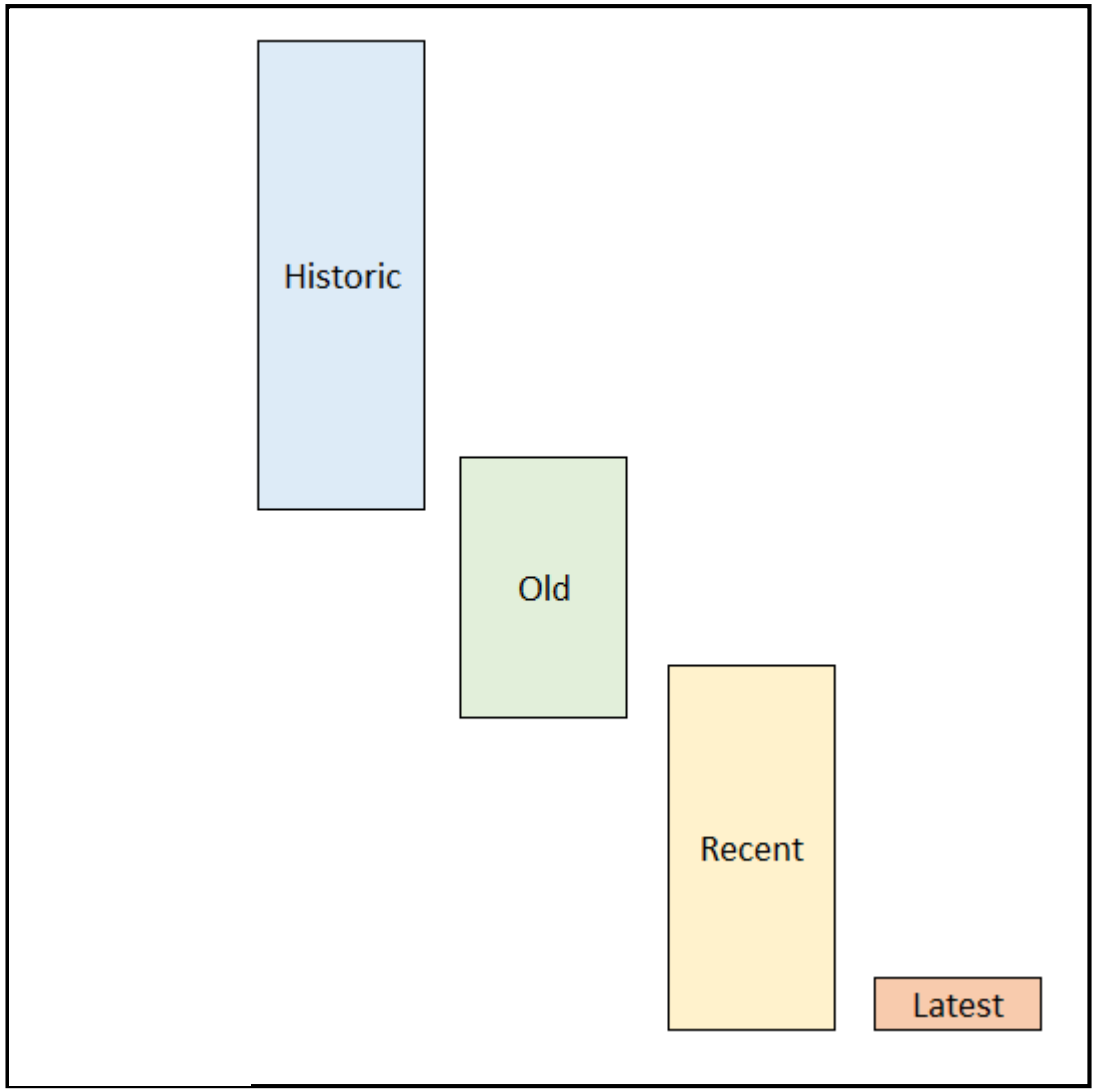
Census data: sources

- We have three censuses
  - England and Wales
  - Scotland
  - Northern Ireland
- ***User requirements***
  - ***Some users work at a UK level, others at national / sub-national level***
  - ***Where research is done at UK level, users typically want harmonised data***
  - ***Some users only want population data (whether census or other) as a denominator; their research focus is the numerator***

Census data licenses

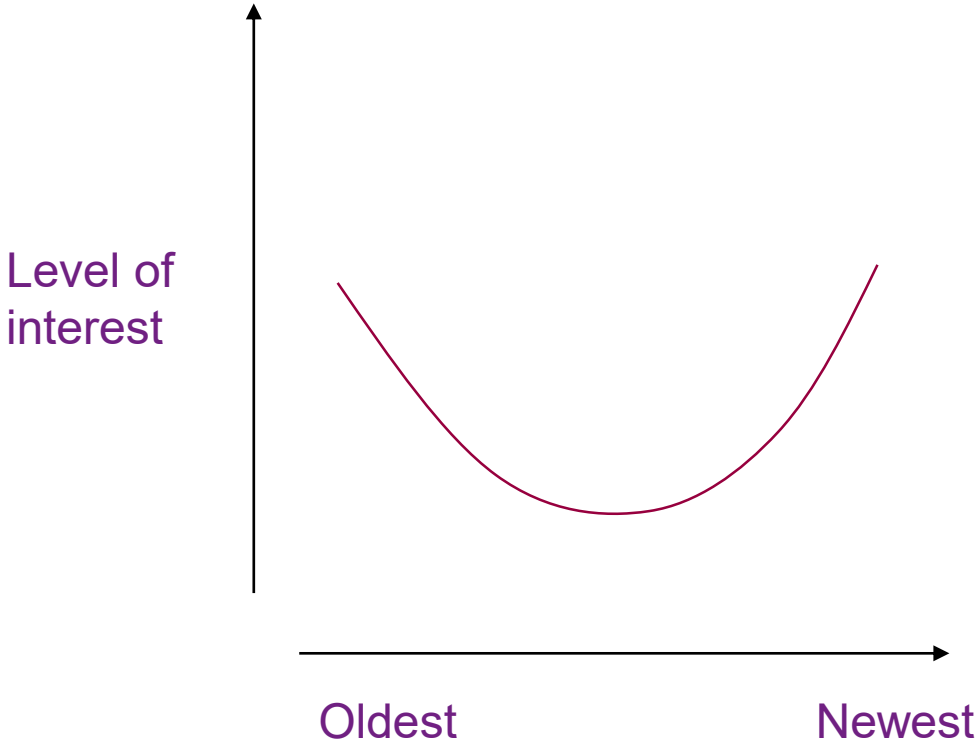
	Open	Safeguarded	Secure
Aggregate data	✓		
Flow data	✓	✓	✓ (via SRS/IDS)
Boundary data	✓		
Microdata			
Cross-sectional (UKDS)	✓	✓	✓ (mostly via SRS/IDS; some via Secure Lab)
Longitudinal (not UKDS)			✓ (via SRS/IDS)

Census data: time frame

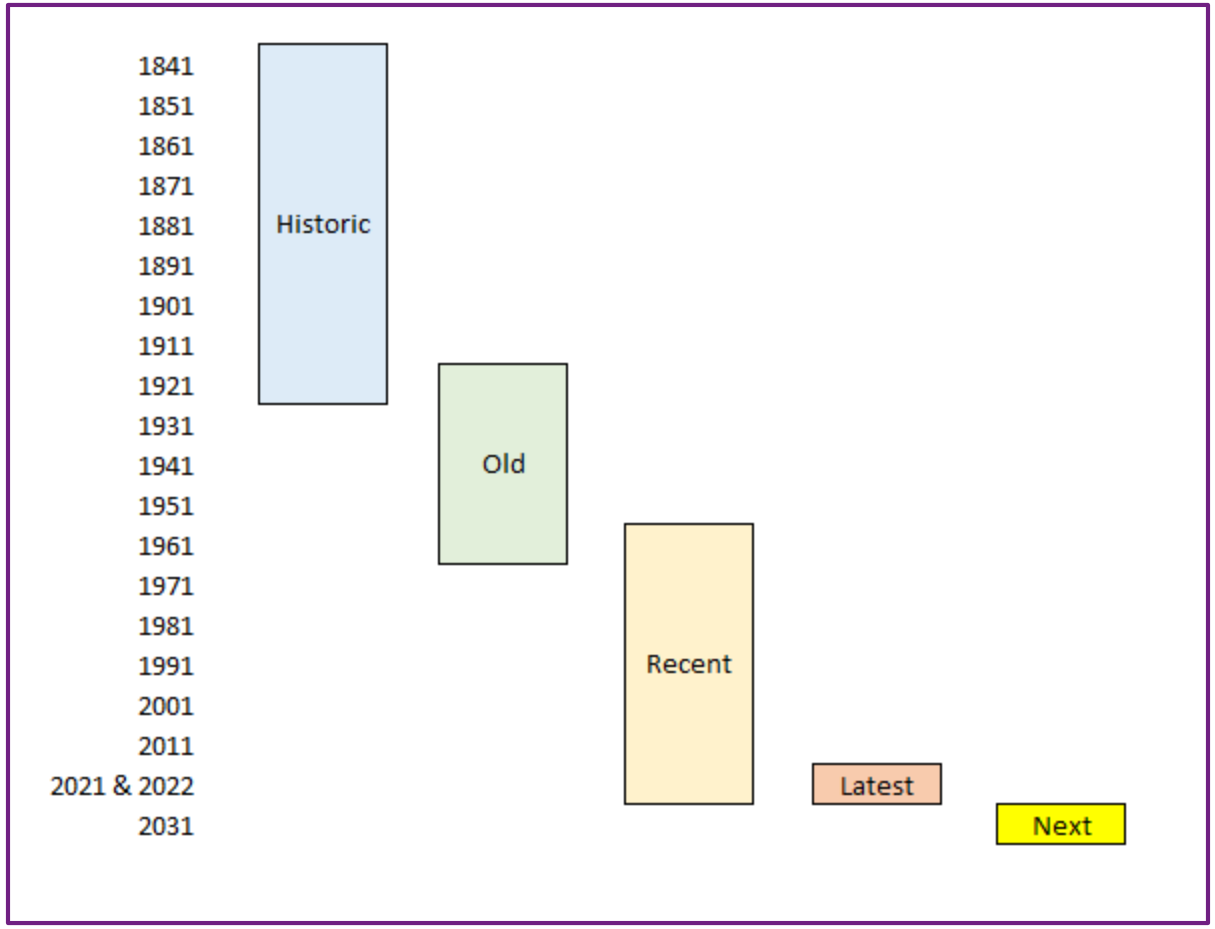


- ***User requirements***
  - ***Users would prefer to get similar data from the same location***
  - ***Users would like to minimise distinctions between types of data***
  - ***As with harmonisation between countries, users think comparability between years should be easy***

(Speculative) user interest



2031 User requirements



A recommended census in 2031

- Some users have a keen interest in question development
  - Consultation has started in Scotland; is awaited elsewhere
- There are a variety of users with different needs
  - Consultation should be diverse
  - Solutions at different technical levels are needed
  - We should consider access needs
- Replicability of data
  - Can users re-play analysis?
- Historic data
  - The 2021 Census as an archival proposition
  - Future demographic data as an archival proposition