
New and easier ways of working with aggregate data and geographies from UK censuses

Justin Hayes & Rob Dymond-Green
UK Data Service / Jisc

RGS-IBG Conference 2014
London
27 August 2014

UK Data Service
Census Support



Abstract

- The aggregate outputs from UK censuses provide detailed, high quality information on a wide range of the demographic and socio-economic characteristics of people and households across the UK. They are, however, very large and complex datasets, which can make them difficult to understand and use effectively. This presentation will describe and demonstrate how new methods of managing and providing access to the aggregate outputs from the UK 2011 Census via the InFuse interface (infuse.mimas.ac.uk) have made them easier to work with to exploit the valuable information they contain. InFuse offers several unique features including: searches for data across the entire collection of UK aggregate outputs by selecting combinations of variables rather than searching through hundreds of traditional tables; selection of data for equivalent geographical areas across the UK in a single operation rather than separate operations for different countries; and integration of data and metadata to support understanding and effective use.



Abstract

- The aggregate outputs from UK censuses provide detailed, high quality information on a wide range of the demographic and socio-economic characteristics of people and households across the UK. They are, however, very large and complex datasets, which can make them difficult to understand and use effectively. This presentation will describe and demonstrate how new methods of managing and providing access to the aggregate outputs from the UK 2011 Census via the InFuse interface (infuse.mimas.ac.uk) have made them easier to work with to exploit the valuable information they contain. InFuse offers several unique features including: **searches for data across the entire collection of UK aggregate outputs by selecting combinations of variables** rather than searching through hundreds of traditional tables; selection of data for equivalent geographical areas across the UK in a single operation rather than separate operations for different countries; and integration of data and metadata to support understanding and effective use.



Abstract

- The aggregate outputs from UK censuses provide detailed, high quality information on a wide range of the demographic and socio-economic characteristics of people and households across the UK. They are, however, very large and complex datasets, which can make them difficult to understand and use effectively. This presentation will describe and demonstrate how new methods of managing and providing access to the aggregate outputs from the UK 2011 Census via the InFuse interface (infuse.mimas.ac.uk) have made them easier to work with to exploit the valuable information they contain. InFuse offers several unique features including: **searches for data across the entire collection of UK aggregate outputs by selecting combinations of variables** rather than searching through hundreds of traditional tables; **selection of data for equivalent geographical areas across the UK in a single operation** rather than separate operations for different countries; and integration of data and metadata to support understanding and effective use.



Abstract

- The aggregate outputs from UK censuses provide detailed, high quality information on a wide range of the demographic and socio-economic characteristics of people and households across the UK. They are, however, very large and complex datasets, which can make them difficult to understand and use effectively. This presentation will describe and demonstrate how new methods of managing and providing access to the aggregate outputs from the UK 2011 Census via the InFuse interface (infuse.mimas.ac.uk) have made them easier to work with to exploit the valuable information they contain. InFuse offers several unique features including: **searches for data across the entire collection of UK aggregate outputs by selecting combinations of variables** rather than searching through hundreds of traditional tables; **selection of data for equivalent geographical areas across the UK in a single operation** rather than separate operations for different countries; and **integration of data and metadata** to support understanding and effective use.



Overview

- What is InFuse?
- InFuse demonstration
 - Global search using variable combinations
 - Integration of data and metadata
 - Global selection of UK 2011 census geographies
- Under the bonnet
 - Integrating models of characteristics and geographies
 - InFuse Web service and API
- Benefits and potentials
- What's next?



What is InFuse?

- Free public Web interface to UK census aggregate data
 - infuse.mimas.ac.uk
- Makes data easy to
 - Find (or not find!)
 - Understand
 - Use
- A work in progress
 - 2011 Census
 - UK data to local authority level
 - England and Wales to output area level
 - 2001 census data for England and Wales
 - More data from 2011, 2001 and other sources coming!
 - Lots of interface development planned



InFuse demonstration

- Global search using variable combinations
- Integration of data and metadata
- Simple model of UK 2011 census geographies



Integrated dissemination via InFuse

UK Data Service
Census Support



InFuse

UK Data Service Census Support | InFuse Home | Guide | Contact



- Home
- Guide
- Tutorial
- FAQs
- Citing data
- 2011 geographies
- 2011 topics
- Background
- Development

InFuse is a free service providing easy access to aggregate data from the UK 2011 and 2001 censuses

Access:

[2011 Census data](#)

[2001 Census data](#)

Data coverage

InFuse contains a selection of UK 2011 Census data down to Local Authority level, as well as England and Wales 2011 Census down to Output Area level. Find out more about our [2011 data release plans](#). We also have England and Wales 2001 Census aggregate data.

About the data

UK censuses are surveys of the entire population of the UK carried out every ten years. Their primary purpose is to provide the UK government with high quality information about demographic and socio-economic characteristics on which to base policy and spending decisions. The information is also used for a wide variety of other purposes across all sectors of society.



service
ort

Variable combination selection

Filters [Clear filters](#) Showing 277 topic combinations Page 1 of 18 [First](#) < [1](#) [2](#) [3](#) > [Last](#)

Topics

- Accommodation type
- Adult lifestage (alternative adult definition)
- Adults in household, number of
- Age
- Alternative population
- Arrival in the UK, age upon
- Bedrooms, number of
- Bedrooms, number of persons per bedrooms in household
- Bedrooms, occupancy rating
- Car or van availability
- Care (unpaid), provision of
- Central heating in household, type of
- Children, age of
- Communal establishment management and type
- Communal establishment

Information about

- Persons
- Households
- Rooms
- Cars or vans
- Hectares
- Dwellings
- Bedrooms
- Communal

<ul style="list-style-type: none">• Accommodation type Select	<ul style="list-style-type: none">• Accommodation type• Bedrooms, number of persons per bedrooms in household• Tenure Select	<ul style="list-style-type: none">• Age• Sex Select	<ul style="list-style-type: none">• Age• Main language (detailed) Select
<ul style="list-style-type: none">• Accommodation type• Household spaces Select	<ul style="list-style-type: none">• Accommodation type• Dependent children, age of• Bedrooms, number of• Usual resident population Select	<ul style="list-style-type: none">• Age• Care (unpaid), provision of Select	<ul style="list-style-type: none">• Age• Marital and civil partnership status Select
<ul style="list-style-type: none">• Accommodation type• Usual resident population Select	<ul style="list-style-type: none">• Accommodation type• Central heating in household, type of• Tenure Select	<ul style="list-style-type: none">• Age• Economic activity Select	<ul style="list-style-type: none">• Age• Long-term health problem or disability Select



Variable combination selection

Filters [Clear filters](#)

2 topic combination(s) found for topic(s) (Economic activity, Religion)

Page 1 of 1 [First](#) < 1 > [Last](#)

Topics

- Accommodation type
- Adult lifestage (alternative adult definition)
- Adults in household, number of
- Age
- Alternative population
- Arrival in the UK, age upon
- Bedrooms, number of
- Bedrooms, number of persons per bedrooms in household
- Bedrooms, occupancy rating
- Car or van availability
- Care (unpaid), provision of
- Central heating in household, type of
- Children, age of
- Communal establishment management and type
- Communal establishment

Information about

- Persons
- Households
- Rooms
- Hectares
- Dwellings
- Bedrooms
- Communal

• Age
• Economic activity
• Religion

Select

• Age
• Economic activity
• Religion
• Sex

Select

ce

Census Support



Category combination selection



Age

- Age 16 and over

Economic activity

- All categories: economic activity
- Economically active
- Economically inactive
- ...Unemployed

Religion

- All categories: Religion
- Christian
- Buddhist
- Hindu
- Jewish
- Muslim
- Sikh
- Other religion
- No religion
- Religion not stated

Unit

- Persons

Selected category combinations

77996 - Age : Age 16 and over - economic activity : All categories: economic activity - Religion : Christian - Unit : Persons
78008 - Age : Age 16 and over - economic activity : All categories: economic activity - Religion : Jewish - Unit : Persons
78012 - Age : Age 16 and over - economic activity : All categories: economic activity - Religion : Muslim - Unit : Persons

ce



Area selection

- United Kingdom
 - England and Wales (1 areas)
 - Countries (4 areas)
 - England
 - Regions (9 areas)
 - East Midlands
 - East of England
 - London
 - North East
 - North West
 - Counties (4 areas)
 - Cumbria
 - Greater Manchester
 - Lancashire
 - Merseyside
 - Local Authorities (39 areas)
 - Wards and Electoral Divisions (933 areas)
 - Middle Super Output Areas and Intermediate Zones (924 areas)
 - Lower Super Output Areas and Data Zones (4497 areas)
 - Output Areas and Small Areas (23343 areas)
 - Merging Wards and Electoral Divisions (4 areas)
 - South East
 - South West
 - West Midlands
 - Yorkshire and The Humber
 - Counties (35 areas)
 - Local Authorities (324 areas)



Data download

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	CDU_ID	GEO_COD	GEO_LABEL	GEO_TYPE	GEO_TYP2	F77992	F78000	F78004	F78012	F78016	F77472	F77480	F77484	F77492	F77496	
2						Age : Age 16 and over - Ec										
3	4	E92000001	England	Countries	CTRY	42989620	210693	656851	1779655	334049	317724	2509	9121	37096	4387	
4	8	E12000001	North East	Regions	RGN	2134449	5797	6429	31698	4608	15810	125	181	703	47	
5	9	E12000002	North West	Regions	RGN	5727629	18589	31524	236175	6759	44689	225	535	4897	143	
6	10	E12000003	Yorkshire	Regions	RGN	4285941	12817	19910	212576	17703	34155	187	394	4005	214	
7	11	E12000004	East Midlands	Regions	RGN	3694767	11392	74465	95003	35040	26204	146	1030	1834	482	
8	12	E12000005	West Midlands	Regions	RGN	4507405	14639	58892	240641	105078	34587	185	1173	4746	1382	
9	13	E12000006	East of England	Regions	RGN	4738333	19634	43137	100241	14623	26967	142	495	1900	193	
10	14	E12000007	London	Regions	RGN	6549173	71983	337030	691023	102514	66860	937	4404	15904	1417	
11	15	E12000008	South East	Regions	RGN	6992666	38140	72278	137839	43144	44247	393	763	2502	433	
12	16	E12000009	South West	Regions	RGN	4359257	17702	13186	34459	4580	24205	169	146	605	76	
13																
14																
15																
16																
17																
18																



Traditional aggregate outputs

- Conceived as sets of tables defining characteristics
 - Combinations of variables, categories, universes, footnotes
 - Per-table variations
 - Extended metadata separate
 - Not all tables are available for all areas
- Complexity and inconsistency in tables
 - Effectively separate datasets
 - Not amenable to global, automated operations
 - Heavyweight, inflexible dissemination applications



Integrating model of characteristics

- Deconstruct tables
 - Collect and rationalise variables and categories from all tables
 - *Variable-ise* table universes and footnotes
 - Create a library of standardised variables and categories to describe all data
 - Re-insert observation values at category intersections
 - Enables global operations/queries
 - Framework for Attachment of extended metadata
 - Facilitates description and transfer using standards



InFuse 2011 release 2: Raw data

- England and Wales Local and Detailed Characteristics to output area level
- UK harmonised data to local authority level
- 422 tables, mainly multivariate
- 31 geography types
- 11,311 files
- 15Gb volume
- Contextual metadata separate



InFuse 2011 release 2: Processed data

- 97 variables
- 2,501 categories
- 281 variable combinations
- 140 thousand category combinations
- 4.6 billion values
 - A 460Km high stack of sticky notes!
 - Anticipating approximately 10 billion values in all

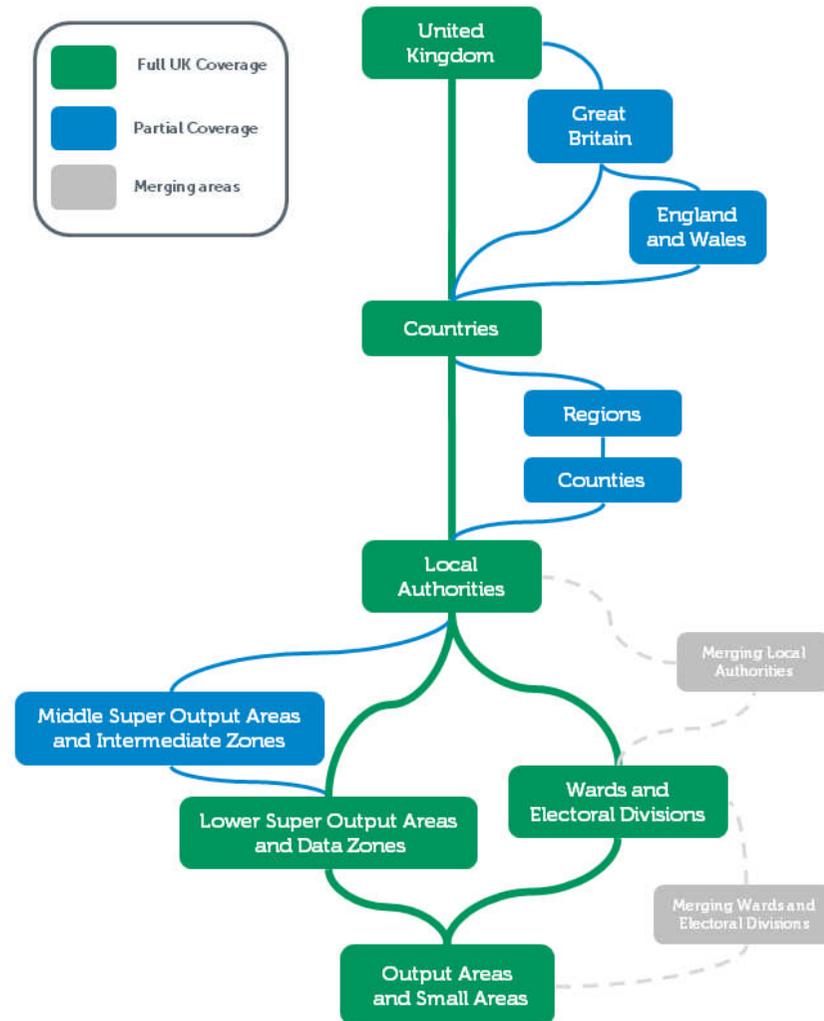


Integrating model of geographies

- Assembly and integration of raw geographies
 - 31 geography types
 - 241,334 areas (anticipating ~ 2 million including postcodes)
 - Direct and indirect hierarchies
- Simplified presentational model
 - 11 composite geography layers
 - Simplification of merged geographies in England and Wales
- Model enables easy calculation of 'missing' data
- Linkage between descriptive and geography models
 - Partial availability of data for geographies and extents



Admin and statistical geography layers



Benefits and potentials

- Dissemination
 - Self-describing, standards-compliant data via APIs is easily operable enabling lightweight, generic applications
 - API delivery provides information on use and contact with users
- Data linkage
 - Extensible (universal?) models of characteristics and geographies provide framework for linkage and comparability of data from disparate sources
- Data production
 - Models provide clear framework for production of new data
 - New data becomes part of a 'data continuum' and is immediately accessible via generic applications



What's next?

- More data
 - 2011 and previous censuses
 - Boundary and flow data
 - Detailed comparison of data between censuses
- More functionality
 - Personalisation
 - Analysis and visualisation
- Public version of InFuse API
 - Weird and wonderful alternatives to InFuse
- Work with data producers
 - Operable data from source (ditch the tables)
 - Flexible generation of aggregate data unit data with automated disclosure control



UK Data Service Census Support

- Specialist unit of UK Data Service
- Access and support services for outputs from UK censuses
- Add value by making census outputs easy to **find, understand** and **use**
- Engagement with UK census agencies
- Long history of technological innovation in service development
- **Main services now freely available to everyone!**
- census.ukdataservice.ac.uk





Census Support

Who we are

We provide access to and user support for 1971, 1981, 1991, 2001 and 2011 UK census data

[Read more](#)

START

LOGIN / REGISTER

DISCOVER UK DATA SERVICE

Search and browse the UK Data Service

Census Support data collection:

Census Data Website

Want to search wider? [Search the whole of the UK Data Service](#)

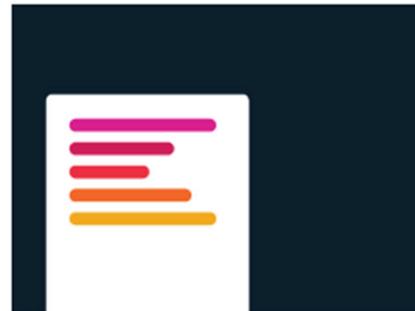
LATEST TWEETS

UKDataService Are you going to the GESIS Summer School in Survey Methodology? Cologne, 8-30 August <http://t.co/IV9rDiN1R>

UKDataService @CLOSER_UK If you don't know, LSYPE is Long. Study of Young People in England. Data and documentation available here <http://t.co/vMUGMj42A>

UKDataService RT @CLOSER_UK: @usociety has alcohol data from age 10

LATEST NEWS



Book now for an introduction to using UK census data geographically

Webinar week at UK Data Service: Five free online sessions

Updated 2011 Census Output Area

UK DATA SERVICE CENSUS SUPPORT



We are the UK Data Service Census Support. This site replaces the old ESRC Census Portal and provides access to a range of census-related resources.

We do not hold the census longitudinal data. However, these can be found at

QUICK ACCESS TO

- UK Data Service
- Aggregate data
- Flow data
- Microdata
- Boundary data



Census Support at Manchester

- Aggregate component of census outputs



Justin Hayes



Richard Wiseman



Rob Dymond-Green



Jamey Hart



Census Support at Manchester

- Aggregate component of census outputs



Justin Hayes



Richard Wiseman



Rob Dymond-Green



Jamey Hart



Give InFuse a go!

infuse.mimas.ac.uk

- Comments, questions and ideas welcome
- ukdataservice.ac.uk/help

