

Flexible dissemination system for Census 2021

Introduction

- We are proposing to develop an online flexible dissemination system where users can specify the data they need and define their own queries to build tables.
- This presentation shows you some features of a prototype built to assess the feasibility of a flexible dissemination system.
- First, we'll explain the background to the flexible dissemination system.

The benefits of a flexible dissemination system

- The system would allow census data to be:
 - **flexible** – users will be able to define their own queries with opportunities for greater detail in the tables
 - **timely** – outputs will be available sooner than they were for the 2011 Census as they will be checked automatically online rather than manually by ONS
 - **accessible** – the system will be hosted on the ONS website, meaning the majority of census data will be available from one location

Disclaimer

- In this presentation, images are for illustrative purposes only. They don't reflect the final design of the dissemination system user interfaces or output table layout that might appear on the ONS website. Please read this disclaimer:
 - These slides have been produced to help you understand our proposals as outlined in the 'Initial view on 2021 Census output content design consultation document.
 - The demonstration tool uses artificial test data.
 - The output categories in tables are examples based on 2011 Census data. We'll be consulting with users closer to the 2021 Census to ensure output categories meet user needs and labels used accurately describe the population being counted.

Protecting confidentiality

- We have a legal obligation to protect the confidentiality of respondents and ensure detailed outputs don't reveal that information.
- Statistical disclosure control ensures statistical outputs provide as much value and utility to users as possible, whilst maintaining the confidentiality of individuals, households and organisations.

For 2021, we're proposing to protect the data by:

1. targeted record swapping
2. applying a small degree of perturbation to the data
3. applying a set of automated checks before releasing tables

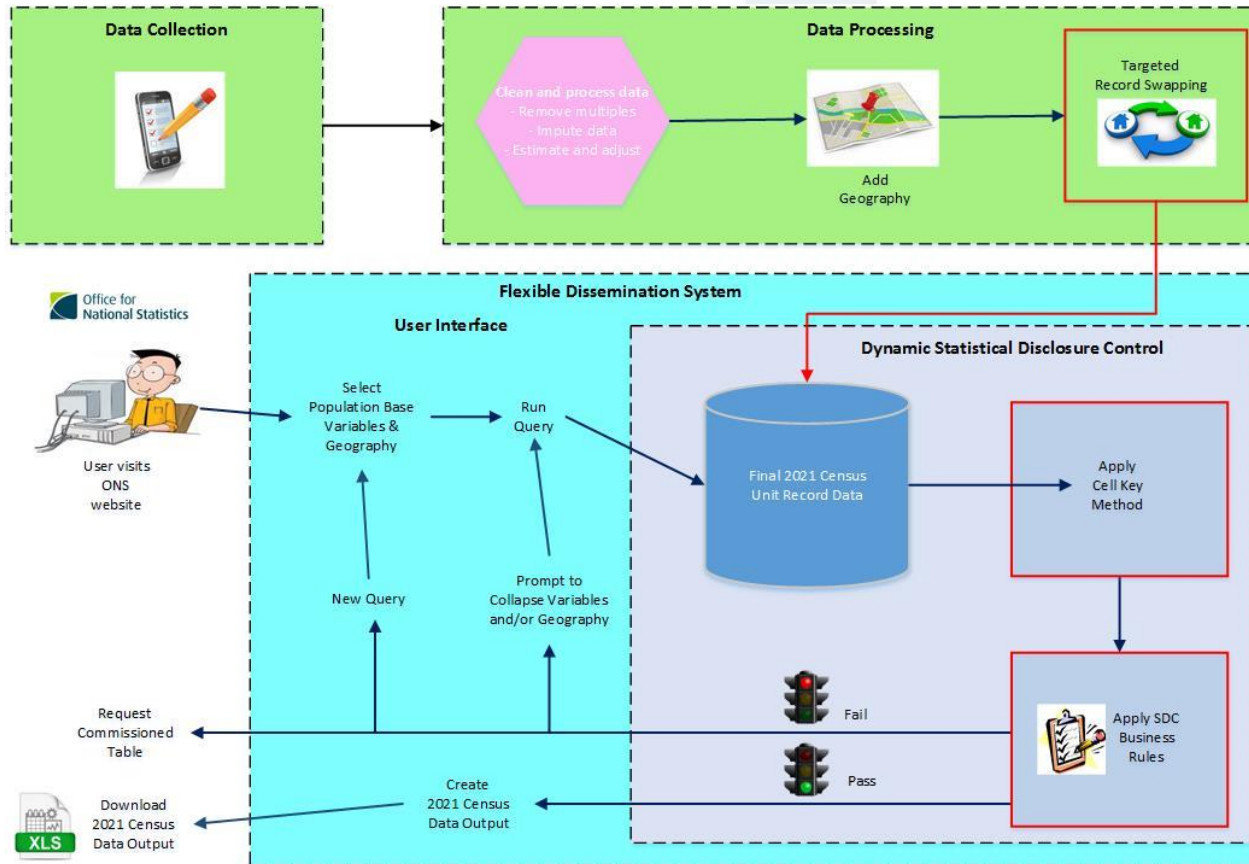
What is data perturbation?

- Data perturbation means slightly changing estimates where otherwise confidentiality would still be compromised in a table after the record swapping.
- The effect of perturbation is not statistically significant and doesn't damage the data quality, but the changes will result in a small amount of inconsistency between some related tables.
- We are currently investigating the minimum amount of perturbation required to protect confidentiality whilst ensuring accuracy.

Flexible dissemination system

- The diagram on the following slide illustrates how the future system might look.
- The green boxes at the top give a general picture of how we process the census data and the record swapping is shown in the top right.
- The blue boxes below set out the main processes involved in meeting a user request for data. These include applying the perturbation and the automated checks that apply the business rules to control access.

Overview of flexible dissemination system



Flexible dissemination system

Now we're going to show you some features of the current approach:

1. We'll show you a simplified version of the prototype we used to generate some tables.
2. We'll explain how users might get more data for some geographic areas than for others.
3. We'll show how the totals might differ due to perturbation.

The impact of automated checks

- We'll use the demonstration version of our prototype to show you how the automated checks affect which data are provided.
- On the next slide is the starting page for a query
- There are four selections to make:
 - population (later versions would have households and communal establishments)
 - geography
 - variables (sometimes called topics)
 - view or download the data

In this demo, the population is selected by default.

Step 1: Choose your population

1 Population

2 Geography

3 Variables

4 Data

Choose your population

People: Illustrative Demonstration

Save and continue

Your table

Cell count: 0

Population: People

Geography: None selected

Variables: None selected

Step 2: Choose your geography (Region)

- Our tool shows two geographies: London and the South West.
- We want to show data for an area with a diverse population so will choose London.

The screenshot displays the 'CENSUS DATA EXPERIMENTAL TABLEBUILDER' interface. At the top, the Office for National Statistics logo is visible. The main navigation bar shows four steps: 1 Population, 2 Geography (highlighted), 3 Variables, and 4 Data. The 'Choose geography' section includes a search box for 'Search available REGIONS', a 'Population' filter, and two options: 'London' and 'South West', each with a 'Browse 1 MERGED_LA' link. A 'Save and continue' button is located at the bottom of this section. On the right, the 'Your table' summary shows 'Cell count: 0', 'Population: People: Illustrative Demonstration', 'Geography: None selected', and 'Variables: None selected'. A footer at the bottom of the page reads: '© 2018 The Office for National Statistics | Version 9696621 built Thu 15 Feb 11:24:30 GMT 2018'.

Step 3: Choose a lower level of geography (local authority)

- We'll select the test local authority in London.

The screenshot shows the 'CENSUS DATA EXPERIMENTAL TABLEBUILDER' interface. At the top, the Office for National Statistics logo is on the left, and the title 'CENSUS DATA EXPERIMENTAL TABLEBUILDER' is on the right. Below this is a navigation bar with four steps: 1 Population, 2 Geography (highlighted), 3 Variables, and 4 Data. The main content area is titled 'Choose geography' and features a search box with the placeholder 'Search available MERGED_LAs'. Below the search box, the breadcrumb 'Population > London' is shown, along with 'Add all' and 'Remove all' links. A list of results shows 'London MERGED_LA #1' with a checked checkbox and a 'Browse 10 OAs' link. A green 'Save and continue' button is at the bottom left. On the right, a 'Your table' summary box displays: 'Cell count: 1', 'Population: People: Illustrative Demonstration', 'Geography: London MERGED_LA #1', and 'Variables: None selected'. A dark footer bar at the bottom contains the text: '© 2018 The Office for National Statistics | Version 9696621 built Thu 15 Feb 11:24:30 GMT 2018'.

Step 4: Choose lower level of geography (Output Area)

- We'll now select the Output Areas in our test local authority.

The screenshot shows the 'CENSUS DATA EXPERIMENTAL TABLEBUILDER' interface. At the top, there are four steps: 1 Population, 2 Geography (highlighted), 3 Variables, and 4 Data. The 'Choose geography' section includes a search bar for 'Search available OAs', a breadcrumb trail 'Population > London > London MERGED_LA #1', and links for 'Add all' and 'Remove all'. A list of ten London Output Areas (OA #01 to OA #09) is shown, each with a checked checkbox. On the right, a 'Your table' summary box displays: 'Cell count: 10', 'Population: People: Illustrative Demonstration', 'Geography: London OA #01, London OA #02, London OA #03, London OA #04, London OA #05, London OA #06 ... and 4 more.', and 'Variables: None selected'.

Step 5: Select variables

- We've selected the 19 category ethnic group.
- In this prototype, a description appears as the cursor hovers over the option.

Office for National Statistics | CENSUS DATA EXPERIMENTAL TABLEBUILDER

1 Population > 2 Geography > 3 Variables > 4 Data

Choose your variables

🔍 Search available variables

Age of individual +

Ethnic group -

8 categories

19 categories

Religion

Sex

Save and continue

Contains

White: English/Welsh/Scottish/Northern Irish/British
White: Irish
White: Gypsy or Irish Traveller
White: Other White

Black/African/Caribbean/Black British: Caribbean
Black/African/Caribbean/Black British: Other Black
Other ethnic group: Arab
Other ethnic group: Any other ethnic group

Your table

Cell count: 190

Population: People: Illustrative Demonstration

Geography: London OA #01, London OA #02, London OA #03, London OA #04, London OA #05, London OA #06 ... and 4 more.

Variables: Ethnic group

Data confidentiality

100% **10 out of 10 areas** pass confidentiality checks.

© 2018 The Office for National Statistics | Version 9696621 built Thu 15 Feb 11:24:30 GMT 2018

Step 6: Select more variables

Office for National Statistics | CENSUS DATA EXPERIMENTAL TABLEBUILDER

1 Population 2 Geography 3 Variables 4 Data

Choose your variables

Search available variables

Age of individual +

Ethnic group -

8 categories

19 categories

Religion

Sex

Save and

Contains

- No religion
- Christian (including Church of England, Catholic, Protestant and all other Christian denominations)
- Buddhist
- Hindu
- Jewish
- Muslim
- Sikh
- Other religion
- Not stated

Your table

Cell count: 1900

Population: People: Illustrative Demonstration

Geography: London OA #01, London OA #02, London OA #03, London OA #04, London OA #05, London OA #06 ... and 4 more.

Variables: Ethnic group, Religion

Data confidentiality

100% 10 out of 10 areas pass confidentiality checks.

Number of areas for which data will be released

© 2018 The Office for National Statistics | Version 305802 | Built Thu 15 Feb 14:50 GMT 2018

Selecting variables

- As we choose the variables, you'll see a reading showing the number of areas for which data will be released for that selection.
- If we select too many variables, we won't get any data for that area.
- We select the religion variable and see we get data for all areas. Users in urban areas are likely to find they can request more detail than was available in ONS tables from the 2011 Census, for most areas.

Step 7: Download data

- Lastly, we can choose whether to download the data or view on screen.

Office for National Statistics | CENSUS DATA EXPERIMENTAL TABLEBUILDER

1 Population 2 Geography 3 Variables 4 Data

Download your table

Cell count: 1900

Population: People: Illustrative Demonstration

Geography: London OA #01, London OA #02, London OA #03, London OA #04, London OA #05, London OA #06 ... and 4 more.

Variables: Ethnic group, Religion

[Download CSV](#) [View data](#)

Data confidentiality

100% 10 out of 10 areas pass confidentiality checks.

Perturbation: in order to simulate the control of confidentiality that is proposed for the 2021 Census outputs, some values in tables are perturbed; record swapping will remain the main source of protection.

Table preview

Displaying the first 10 rows only

OA code	OA label	Ethnic group code	Ethnic group label	Religion code	Religion label
LONDON_OA_01	London OA #01	1	White: English/Welsh/Scottish/Northern Irish/British	1	No religion
LONDON_OA_01	London OA #01	1	White: English/Welsh/Scottish/Northern Irish/British	2	Christian (inclu
LONDON_OA_01	London OA #01	1	White: English/Welsh/Scottish/Northern Irish/British	3	Buddhist

New query

- Now we'll create another query, but for a different part of the country.
- We'd like to show the effect of selecting an area with a more homogeneous population, so we choose the South West region.

Step 1: Choose your population

1 Population

2 Geography

3 Variables

4 Data

Choose your population

People: Illustrative Demonstration

Save and continue

Your table
Cell count: 0
Population: People
Geography: None selected
Variables: None selected

Step 2: Choose your geography (Region)

1 Population

2 **Geography**

3 Variables

4 Data

Choose geography

Population

[Add all](#) | [Remove all](#)

London [Browse 1 MERGED_LA](#) >

South West [Browse 1 MERGED_LA](#) >

Save and continue

Your table
Cell count: 0
Population: People: Illustrative Demonstration
Geography: None selected
Variables: None selected

Step 3: Choose your geography (local authority)

- We select our test rural local authority in the South West.

The screenshot shows the 'CENSUS DATA EXPERIMENTAL TABLEBUILDER' interface. At the top, the Office for National Statistics logo is on the left, and the title 'CENSUS DATA EXPERIMENTAL TABLEBUILDER' is on the right. Below the title is a navigation bar with four steps: 1 Population, 2 Geography (highlighted), 3 Variables, and 4 Data. The main content area is titled 'Choose geography' and features a search box for 'MERGED_LAs'. Below the search box, the breadcrumb 'Population > South West' is shown, along with 'Add all' and 'Remove all' links. A list of results shows 'South West MERGED_LA #1' with a checked checkbox and a 'Browse 10 OAs' link. A green 'Save and continue' button is at the bottom. On the right, a 'Your table' summary box displays: 'Cell count: 1', 'Population: People: Illustrative Demonstration', 'Geography: South West MERGED_LA #1', and 'Variables: None selected'.

Step 4: Choose lower level of geography (Output Area)

- We then select all the test Output Areas in that local authority.

[Population](#) > [South West](#) > South West MERGED_LA #1

[Add all](#) | [Remove all](#)

<input checked="" type="checkbox"/>	South West OA #01
<input checked="" type="checkbox"/>	South West OA #02
<input checked="" type="checkbox"/>	South West OA #03
<input checked="" type="checkbox"/>	South West OA #04
<input checked="" type="checkbox"/>	South West OA #05
<input checked="" type="checkbox"/>	South West OA #06
<input checked="" type="checkbox"/>	South West OA #07
<input checked="" type="checkbox"/>	South West OA #08
<input checked="" type="checkbox"/>	South West OA #09
<input checked="" type="checkbox"/>	South West OA #10

[Save and continue](#)

Your table

Cell count: 10

Population: People: Illustrative Demonstration

Geography: South West OA #01, South West OA #02, South West OA #03, South West OA #04, South West OA #06, South West OA #07 ... and 4 more.

Variables: None selected

Step 5: Select variables

- As we choose the variables, you'll see a reading showing the number of areas for which data will be released for that selection.
- We won't get data for all the Output Areas.

The screenshot shows the 'CENSUS DATA EXPERIMENTAL TABLEBUILDER' interface. The navigation bar has four steps: 1 Population, 2 Geography, 3 Variables (active), and 4 Data. The main area is titled 'Choose your variables' and includes a search bar and several variable selection options: 'Age of individual' (unchecked), 'Ethnic group' (checked, with radio buttons for '8 categories' and '19 categories'), 'Religion' (checked), and 'Sex' (unchecked). A 'Save and' button is visible. A 'Your table' summary on the right shows 'Cell count: 1900', 'Population: People: Illustrative Demonstration', 'Geography: South West OA #01, South West OA #02, South West OA #03, South West OA #04, South West OA #05, South West OA #06 ... and 4 more.', and 'Variables: Ethnic group, Religion'. A 'Data confidentiality' section features a 70% progress indicator and the text '7 out of 10 areas pass confidentiality checks.' A 'Contains' dropdown menu is open, listing categories such as 'No religion', 'Christian (including Church of England, Catholic, Protestant and all other Christian denominations)', 'Buddhist', 'Hindu', 'Jewish', 'Muslim', 'Sikh', 'Other religion', and 'Not stated'. A black arrow points from the '7 out of 10 areas' text to the right.

Number of areas for which data will be released

Impact of automated checks

- These queries showed we're getting data for fewer areas in the South West than in London. This is illustrated in the red and green circle.
- The automated checks in the flexible dissemination system are making data for some areas unavailable. This is to protect information about some people in those areas.
- Some queries may result in no data at all. There will be guidance on the ONS website about the protection and checks on the data. This will explain that census data for the less diverse areas of the country will be limited in detail to protect confidentiality.

Step 6: Download data

- We'll download the table anyway, despite not getting all the data for that area.

Office for National Statistics | CENSUS DATA EXPERIMENTAL TABLEBUILDER

1 Population > 2 Geography > 3 Variables > **4 Data**

Download your table

Cell count: 1900


Population: People: Illustrative Demonstration

Geography: South West OA #01, South West OA #02, South West OA #03, South West OA #04, South West OA #05, South West OA #06 ... and 4 more.

Variables: Ethnic group, Religion

[Download CSV](#) [View data](#)

Data confidentiality

 **7 out of 10 areas** pass confidentiality checks.

Perturbation: in order to simulate the control of confidentiality that is proposed for the 2021 Census outputs, some values in tables are perturbed; record swapping will remain the main source of protection.

Table preview

Displaying the first 10 rows only

OA code	OA label	Ethnic group code	Ethnic group label	Religion code	Relig
SOUTH WEST_OA_01	South West OA #01	1	White: English/Welsh/Scottish/Northern Irish/British	1	No re
SOUTH WEST_OA_01	South West OA #01	1	White: English/Welsh/Scottish/Northern Irish/British	2	Chris
SOUTH WEST_OA_01	South West OA #01	1	White: English/Welsh/Scottish/Northern Irish/British	3	Budd

Minor inconsistencies in the output data

- We've just looked at an important feature of the flexible dissemination system. This is to make more data available than in previous censuses, but this won't be possible for some areas.
- Another feature of the flexible dissemination system is that the minor **perturbation of data** to protect confidentiality causes minor differences between tables.

Output table totals

- The next slides demonstrate the differences between output table totals.
- Repeating the same query for the same geographic area will always result in the same estimates.
- However, there will be some inconsistencies in the totals between different tables due to the perturbation applied by the cell-key method. This can occur by:
 - 1) geography
 - 2) variables

Comparing summed Output Area data with local authority tables

- The two tables on the next slide illustrate the small differences in table totals when data is aggregated by geography.
- The sum of estimates for small areas that constitute a larger area may not exactly match data in an identical table for the larger area, but the differences are likely to be small.
 - For example - total for local authority may not equal the sum of output areas within that local authority.

Geography

- Local authority total vs sum of aggregated output areas.

Ethnicity by occupation at local authority level

LA=1	Occupation									
Ethnicity:	1	2	3	4	5	6	7	8	9	
White	3885	6014	4325	5726	5202	4172	4394	4282	6784	44784
Mixed	27	55	64	44	47	53	70	34	70	464
Asian	198	357	134	168	171	135	292	233	294	1982
Black	23	93	42	49	34	76	44	38	97	496
Chinese	23	40	11	6	25	15	23	15	33	191
	4156	6559	4576	5993	5479	4451	4823	4602	7278	47917

Ethnicity by occupation at Output Area level aggregated to local authority level

LA=1	Occupation									
Ethnicity:	1	2	3	4	5	6	7	8	9	
White	3886	6012	4324	5730	5205	4171	4391	4276	6784	44779
Mixed	27	57	66	43	48	52	68	35	70	466
Asian	196	359	133	166	172	137	292	231	291	1977
Black	23	93	41	48	33	76	44	41	94	493
Chinese	22	40	11	5	25	15	24	15	32	189
	4154	6561	4575	5992	5483	4451	4819	4598	7271	47904

For White ethnic group category:

- total for row at local authority level is 44,784
- total for row in Output Area aggregation is 44,779
- Difference of 5 (0.015%)

Variables

- The next two tables illustrate the small differences in table totals that may occur when a common variable is cross-tabulated with another variable.
- In multivariate tables, which share a common variable, there may be small differences in the row totals when cross-tabulated with other variables, but the differences are likely to be small.
 - For example - age by ethnic group vs age by marital status.

Variables

Age by ethnic group at local authority level

LA=1

	N.A.	White	Mixed	Asian	Black	Chinese	Total:
0-15	89	10146	420	1179	233	107	12174
16-24	568	6946	186	679	106	67	8552
25-34	23	6895	145	822	158	106	8149
35-45	3	7810	87	581	178	57	8716
45-54	3	8378	83	393	96	34	8987
55-64	0	7518	37	227	29	20	7831
65-74	0	5612	15	111	26	8	5772
75+	0	4856	13	54	22	4	4949
Total:	686	58161	986	4046	848	403	65130

Age by marital status at local authority level

LA=1

	Single	Married	Separated	Divorced	Widowed	Total:
0-15	12176	0	0	0	0	12176
16-24	8313	207	11	13	8	8552
25-34	5249	2439	241	199	20	8148
35-44	2886	4468	403	914	44	8715
45-54	1517	5355	432	1521	160	8985
55-64	613	5312	223	1257	426	7831
65-74	308	3769	105	659	930	5771
75+	274	2030	38	251	2357	4950
Total:	31336	23580	1453	4814	3945	65128

Small difference in totals

Guidance for users

- ONS will publish guidance about the protection and checks on the data.
- Users will be advised that data won't always match in some tables and be reassured that the effects are minimal.
- They will also see the guidance that mentions that any repeat of the same query always produces the exact same estimates.

Summary

- This presentation was intended to inform uses of the two main features associated with a flexible dissemination system for the 2021 Census.
 - 1) There could be greater detail in tables, but not for all areas.
 - 2) Estimates may differ between similar tables for the same areas but these differences are not statistically significant.
- If you would like to provide feedback, please contact us at census.outputs@ons.gov.uk.