

# Progress towards a table builder with in-built disclosure control for 2021 Census

Keith Spicer

UNECE, 22 September 2017

# Contents

- Context: 2011 Census
- 2021 Census
  - Targeted Record Swapping
  - Outputs Package (Table Builder)
  - Cell key Method
  - Perturbing Zeros

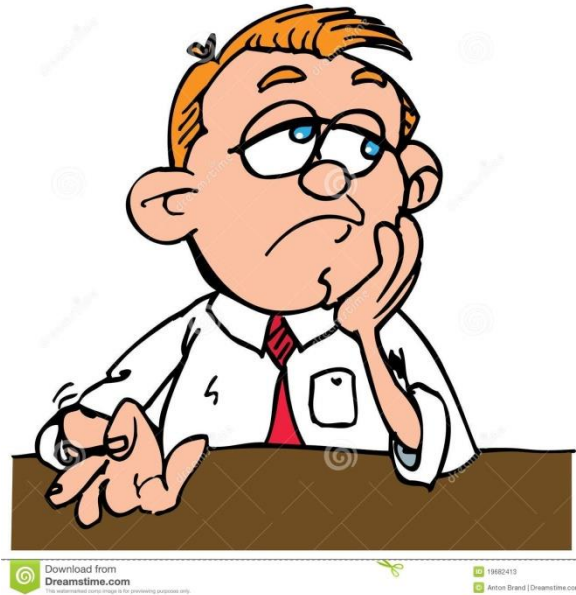
# 2011 Census

- Targeted record swapping
  - Targeted to “risky” records
- Table redesign
  - Criteria of % 1s that are “real” and attribute disclosures that are “real”
  - Sparsity gives higher chance of disclosure
  - Sparsity also gives perception of disclosure

# 2011 Census



Every table had to be checked for disclosure



Timing was affected

.....and sometimes.....

I wanted this table



They gave me this table



# 2021 Census

- UK Parliament discussed:
- Aim for 2021 to be the last traditional census in England and Wales
- Look to use administrative and other sources to replace traditional census
- Parallel running of traditional and admin censuses in 2021 census round

# 2021 Census

User concern from 2011 in three areas:

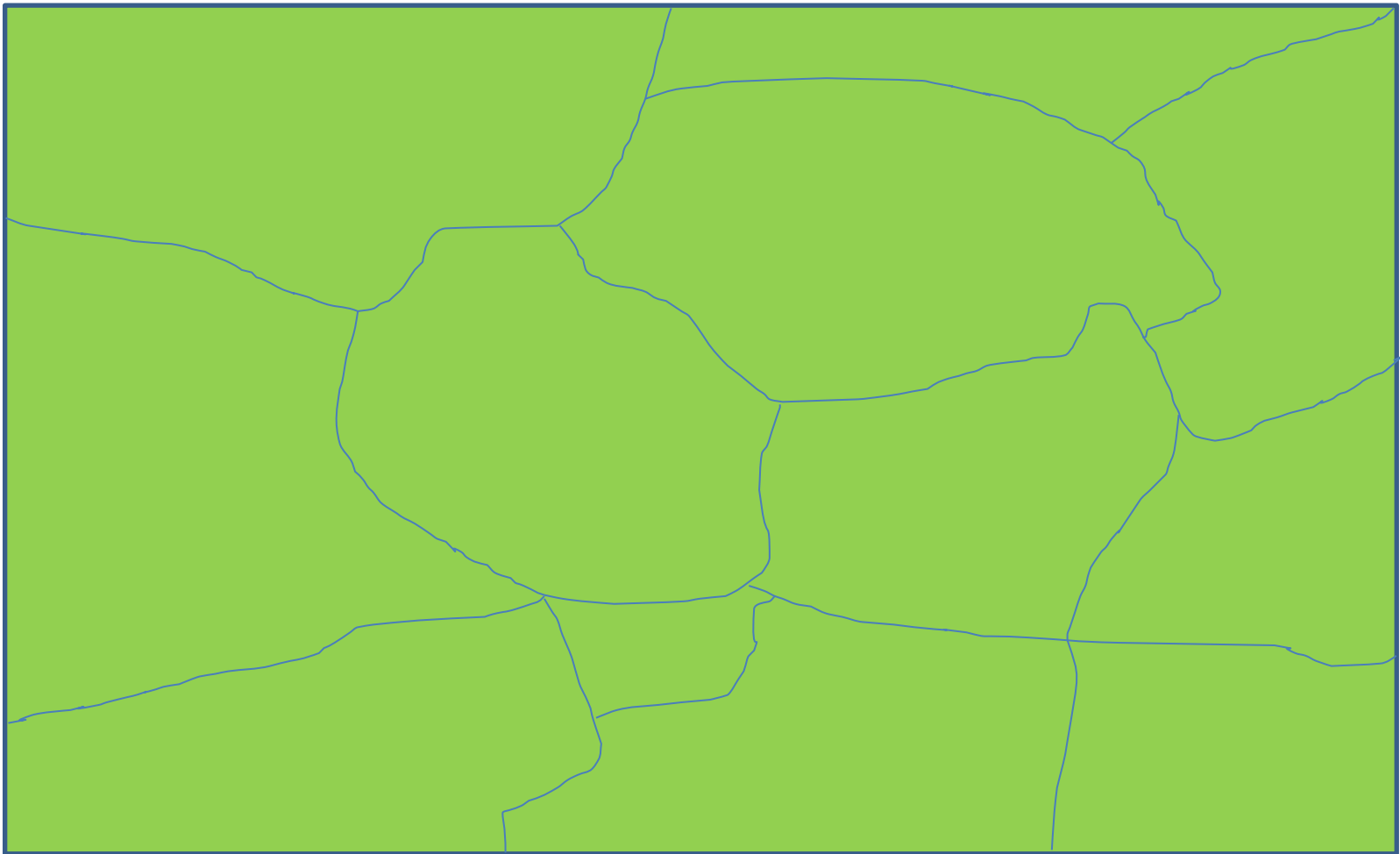
Flexibility

Accessibility

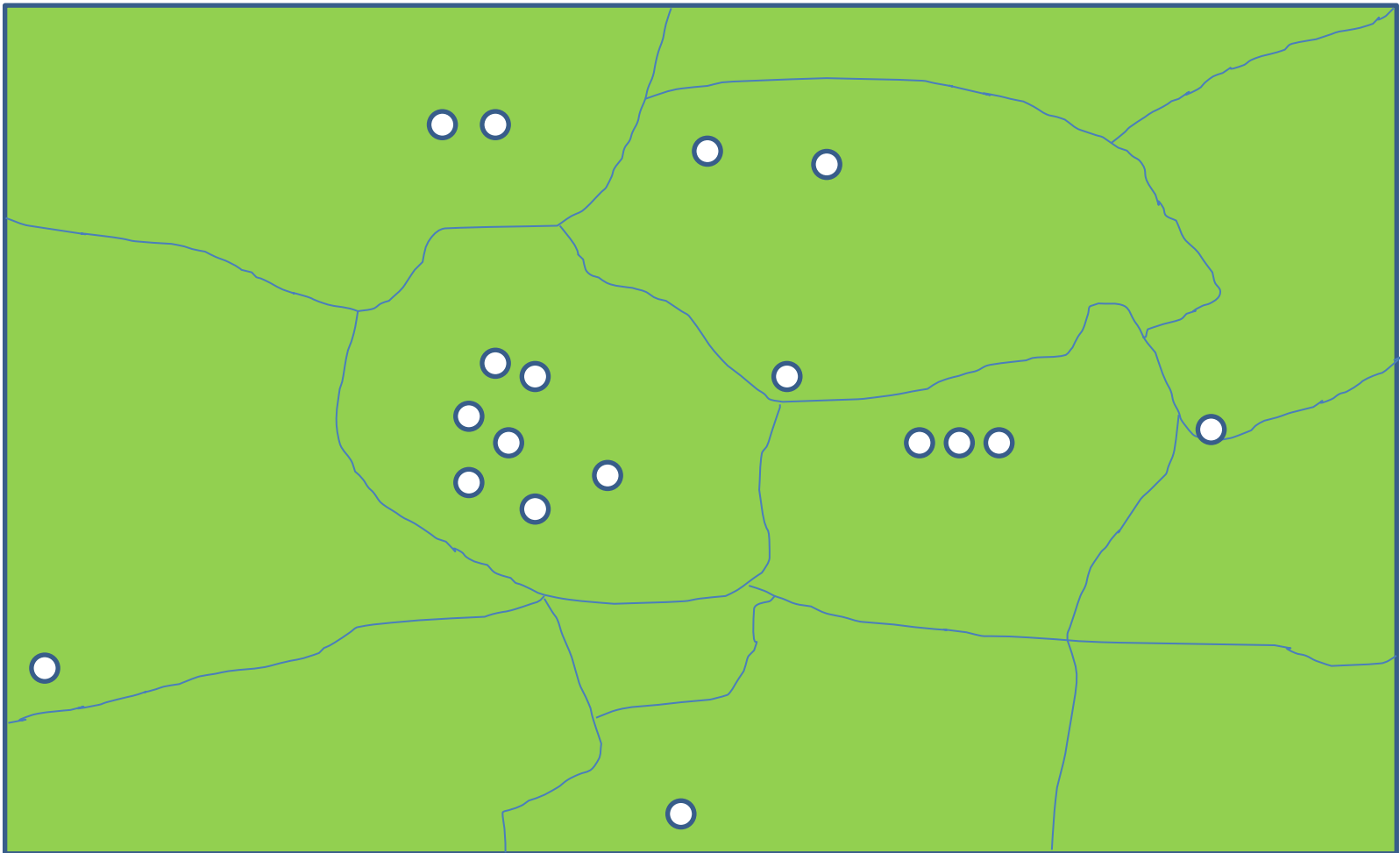
Timeliness



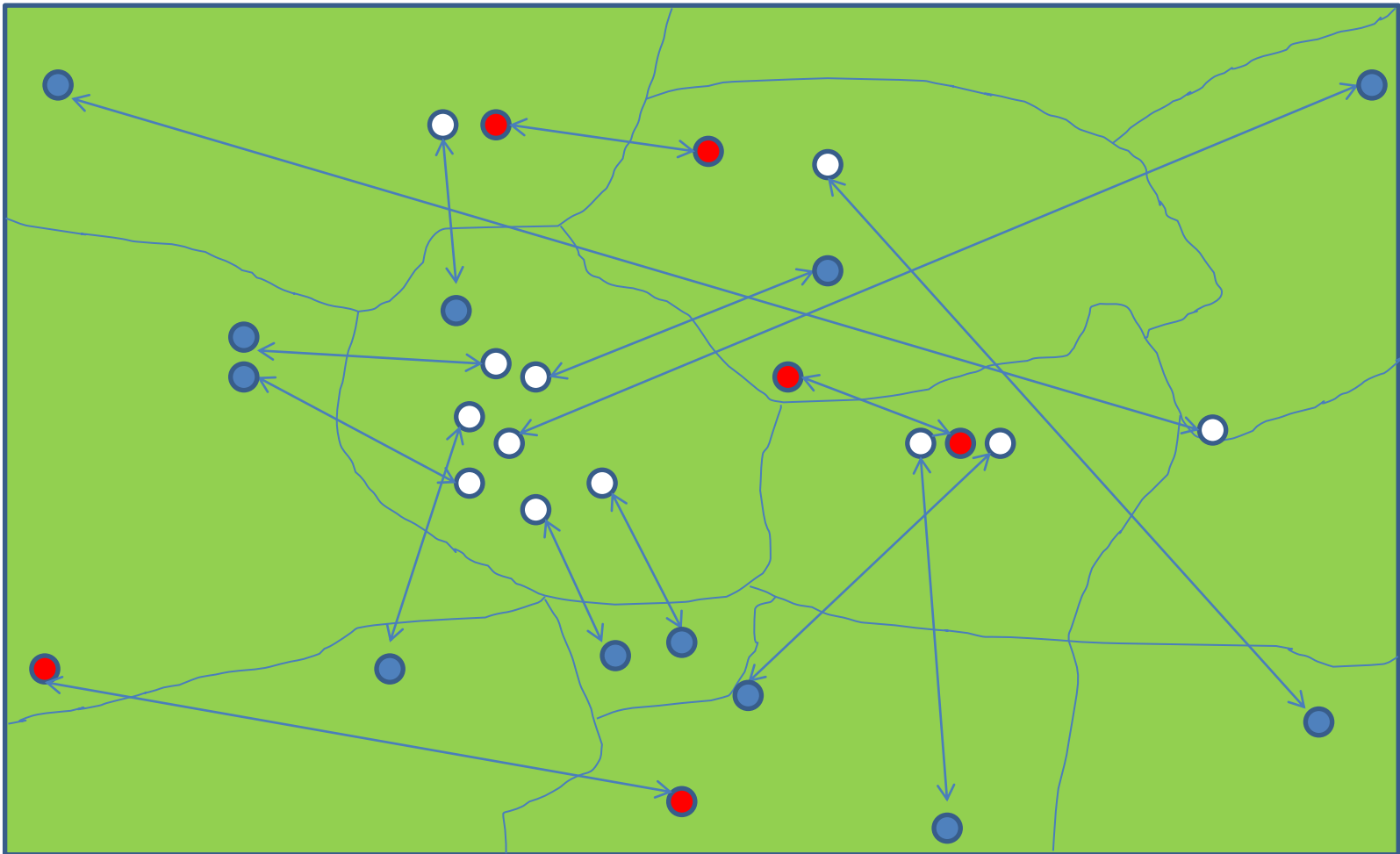
# Targeted Record Swapping I



# Targeted Record Swapping II



# Targeted Record Swapping III



# 2021 Census: Outputs Package

Aim to produce an outputs package

- Targeted record swapping **Swaps obviously identifiable people / households**
- Cell key method **Protects all by uncertainty, and differencing**
- User-defined tables from a table builder **Allows tables quickly**

# Cell Key Method

1 Assign each record a random number

Record	Rkey
$r_1 \rightarrow$	54
$r_2 \rightarrow$	4
$r_3 \rightarrow$	93
...	
$r_N \rightarrow$	26

2 For each cell, sum rkey and apply a function to get a cell key

Age by sex	Male	Female
0-15	.	.
16-24	.	4
25-34	.	.
...		

Record	Rkey
$r_2 \rightarrow$	4
$r_4 \rightarrow$	61
$r_{56} \rightarrow$	7
$r_{72} \rightarrow$	90
Sum =	162

e.g. take last two digits  $\rightarrow$  **Ckey = 62**

3 Use a look up table to get perturbation value

		Cell Key $\longrightarrow$								
		1	2	3	...	61	62	63	...	99
Cell Value $\downarrow$	1		+1							
	2			+1				-1		
	3									+1
	4	-1						<b>+1</b>		
	5			-1		-1				
	...									

4 Apply pvalue to cell

Age by sex	Male	Female
0-15	.	.
16-24	.	5
25-34	.	.
...		

# Notes for Cell Key Method

- Adapted from “ABS method”
- Method primarily for protecting against differencing
- We are looking at a light touch (record swapping still the primary approach)
- Considering the need to retain 1s and 2s in outputs
- Introduces another layer of uncertainty for intruder
- Consistency in same cell across tables
- Some inconsistencies in breakdowns

# Perturbing Zeros I

- Work in progress
- Additional  $n$  1s perturbed to 0s
- Can balance by perturbing  $n$  0s to 1s
  - Counts therefore unbiased
  - Increases protection through uncertainty
  - Means that a 1 does not necessarily represent a record in the microdata
  - Must ensure not changing 'structural zeros'

# Perturbing Zeros II

- Methods for this are being assessed.
- Example:
- Zero cells need a 'cell key' but there are no records or record keys
- Each variable has a set of category keys
- Combination (sum) of category keys = cell key
- (any structural zero; cell key set to 0)
- n highest cell keys are perturbed from 0 to 1



# Key Points

- Aim to have Table Builder for users
- Targeted Record Swapping
- Cell Key Method
- Try to retain small cell counts
- Benefits of this approach to other collections
  
- Work continuing.....
- Other areas: microdata products, origin-destination tables, admin census

# Questions and Discussion