

# Flexible dissemination software for the 2021 England & Wales and Northern Ireland Censuses

18th September 2023



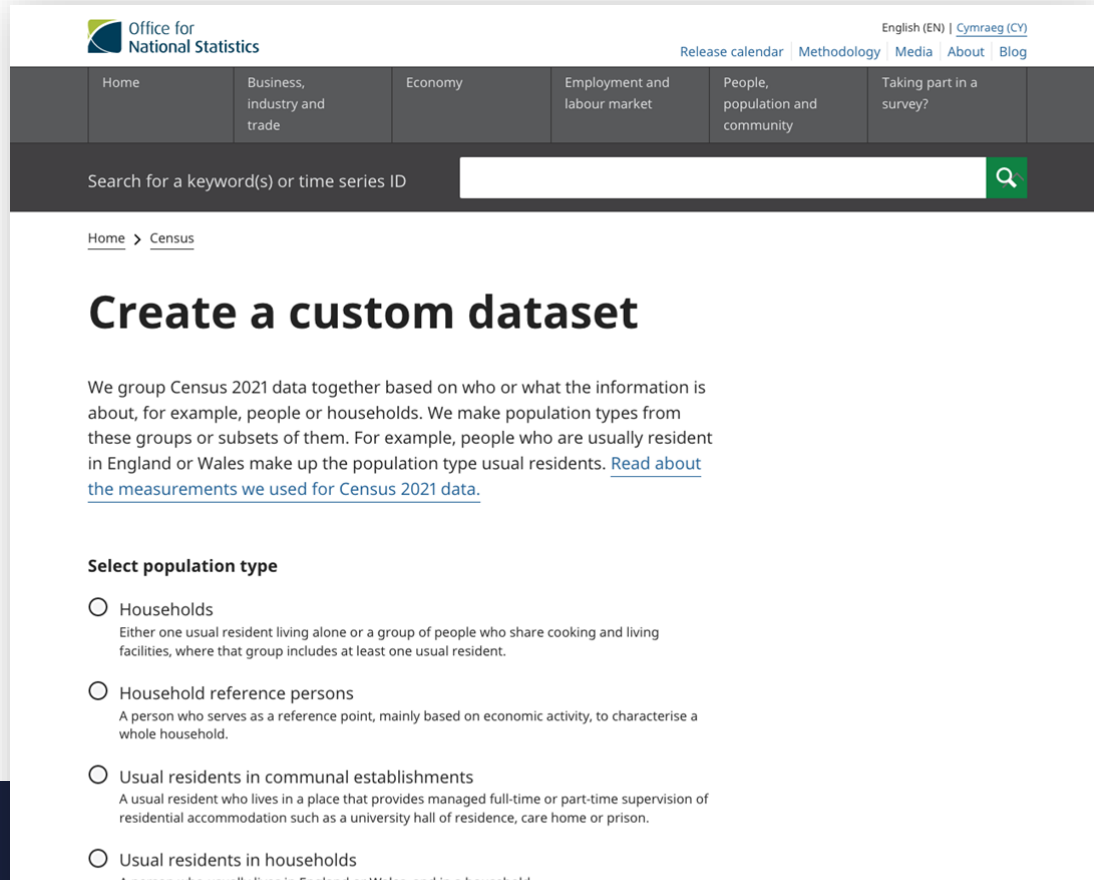
The Sensible  
Code Company

# Index

- What is Cantabular?
- Why automate data publication?
- Technical challenges and solutions
- What's next?
- Q&A

What is Cantabular?

# ONS: Create a custom dataset



The screenshot shows the ONS website interface. At the top left is the ONS logo. To the right are links for 'English (EN) | Cymraeg (CY)', 'Release calendar', 'Methodology', 'Media', 'About', and 'Blog'. A navigation bar contains links for 'Home', 'Business, industry and trade', 'Economy', 'Employment and labour market', 'People, population and community', and 'Taking part in a survey?'. Below this is a search bar with the text 'Search for a keyword(s) or time series ID' and a magnifying glass icon. The main content area has a breadcrumb 'Home > Census' and a large heading 'Create a custom dataset'. The text explains that Census 2021 data is grouped by who or what the information is about, such as people or households. It provides an example of 'usual residents' and includes a link to 'Read about the measurements we used for Census 2021 data.' Below this is a section titled 'Select population type' with four radio button options: 'Households', 'Household reference persons', 'Usual residents in communal establishments', and 'Usual residents in households'. Each option has a brief description.

Office for National Statistics

English (EN) | [Cymraeg \(CY\)](#)

[Release calendar](#) | [Methodology](#) | [Media](#) | [About](#) | [Blog](#)

[Home](#) | [Business, industry and trade](#) | [Economy](#) | [Employment and labour market](#) | [People, population and community](#) | [Taking part in a survey?](#)

Search for a keyword(s) or time series ID

[Home](#) > [Census](#)

## Create a custom dataset

We group Census 2021 data together based on who or what the information is about, for example, people or households. We make population types from these groups or subsets of them. For example, people who are usually resident in England or Wales make up the population type usual residents. [Read about the measurements we used for Census 2021 data.](#)

### Select population type

- Households  
Either one usual resident living alone or a group of people who share cooking and living facilities, where that group includes at least one usual resident.
- Household reference persons  
A person who serves as a reference point, mainly based on economic activity, to characterise a whole household.
- Usual residents in communal establishments  
A usual resident who lives in a place that provides managed full-time or part-time supervision of residential accommodation such as a university hall of residence, care home or prison.
- Usual residents in households  
A person who usually lives in England or Wales and in a household.



# ONS: Create a custom dataset

Office for National Statistics

English (EN) | Cymraeg (CY)

Release calendar | Methodology | Media | About | Blog

Home | Business, industry and trade | Economy | Employment and labour market | People, population and community | Taking part in a survey?

Search for a keyword(s) or time series ID

[Home](#) > [Census](#)

## Create a custom dataset

We group Census 2021 data together based on who or what the information is about, for example, people or households. We make population types from these groups or subsets of them. For example, people who are usually resident in England or Wales make up the population type usual residents. [Read about the measurements we use](#)

### Select population type

- Households  
Either one usual resident in facilities, where that group
- Household reference  
A person who serves as a reference, whole household.
- Usual residents in communal establishments  
A usual resident who lives in a place that provides managed full-time or part-time supervision of residential accommodation such as a university hall of residence, care home or prison.
- Usual residents in households  
A person who usually lives in a household in England, Wales and in a household

29th March 2023

5



# NISRA: Flexible Table Builder

The screenshot shows the NISRA Flexible Table Builder website. At the top left is the NISRA logo and the text 'FLEXIBLE TABLE BUILDER'. At the top right is the 'census 2021' logo. The main heading is 'Build or find Census 2021 tables', followed by a sub-heading: 'Explore the results of Census 2021 for Northern Ireland. Build your own table or search through and customise ready-made tables.' Below this are two main action buttons: 'Build a custom table' and 'Find ready-made tables', each with a right-pointing chevron and a brief description. Underneath is a 'Supporting information' section with three links: 'Explore our metadata', 'Help and support', and 'More information on Census 2021', each with a short explanatory text. The footer is dark blue and contains two columns of links: 'Links' (Census 2021 Area Explorer, NISRA, NISRA Data Portal, OpenDataNI) and 'Follow NISRA' (Facebook, Twitter, YouTube).

**NISRA** FLEXIBLE TABLE BUILDER census 2021

## Build or find Census 2021 tables

Explore the results of Census 2021 for Northern Ireland. Build your own table or search through and customise ready-made tables.

[Build a custom table](#) >  
Use our Flexible Table Builder to create your own Census 2021 table.

[Find ready-made tables](#) >  
Search and filter our selection of pre-built Census 2021 tables.

### Supporting information

[Explore our metadata](#)  
Browse our variable lists and classifications.

[Help and support](#)  
Browse our help page for further information and support (opens in a new window).

[More information on Census 2021](#)  
Visit the NISRA website for more information relating to Census 2021 (opens in a new window).

**Links**

- Census 2021 Area Explorer
- NISRA
- NISRA Data Portal
- OpenDataNI

**Follow NISRA**

- Facebook
- Twitter
- YouTube



# NISRA: Flexible Table Builder

**NISRA** FLEXIBLE TABLE BUILDER census 2021

## Build or find Census 2021 tables

Explore the results of Census 2021 for Northern Ireland. Build your own table or search through and customise ready-made tables.

[Build a custom table](#) > Use our Flexible Table Builder to create your own Census 2021 table.

[Find ready-made tables](#) > Search and filter our selection of pre-built Census 2021 tables.

### Supporting information

[Explore our metadata](#)  
Browse our variable lists and classifications.

[Help and support](#)  
Browse our help page for further information.

[More information on Census 2021](#)  
Visit the NISRA website for more information.

22nd June 2023

#### Links

[Census 2021 Area Explorer](#)  
[NISRA](#)  
[NISRA Data Portal](#)  
[OpenDataNI](#)

#### Follow NISRA

[Facebook](#)  
[Twitter](#)  
[YouTube](#)



# Powered by Cantabular

- Build your own cross-tabulation





# Powered by Cantabular

- Build your own cross-tabulation
- Created in real-time from microdata



# Powered by Cantabular

- Build your own cross-tabulation
- Created in real-time from microdata
- Using automated disclosure checks



# Powered by Cantabular

- Build your own cross-tabulation
- Created in real-time from microdata
- Using automated disclosure checks
- Hundreds of millions of possible tables



# Powered by Cantabular

- Build your own cross-tabulation
- Created in real-time from microdata
- Using automated disclosure checks
- Hundreds of millions of possible tables
- Metadata pieced together when a table is built



# Powered by Cantabular

- Build your own cross-tabulation
- Created in real-time from microdata
- Using automated disclosure checks
- Hundreds of millions of possible tables
- Metadata pieced together when a table is built
- All outputs generated on demand



# Powered by Cantabular

- Build your own cross-tabulation
- Created in
- Using auto
- Hundreds of millions of possible tables
- Metadata pieced together when a table is built
- All outputs generated on demand

*Automating data publication*



Why automate data publication?

Speed





“It allows us to release billions of anonymised census statistics far more quickly than ever before”

—Jen Woolford, Director of Population Statistics, ONS

Source: <https://blog.ons.gov.uk/2023/03/28/new-ways-to-access-census-2021-data/>

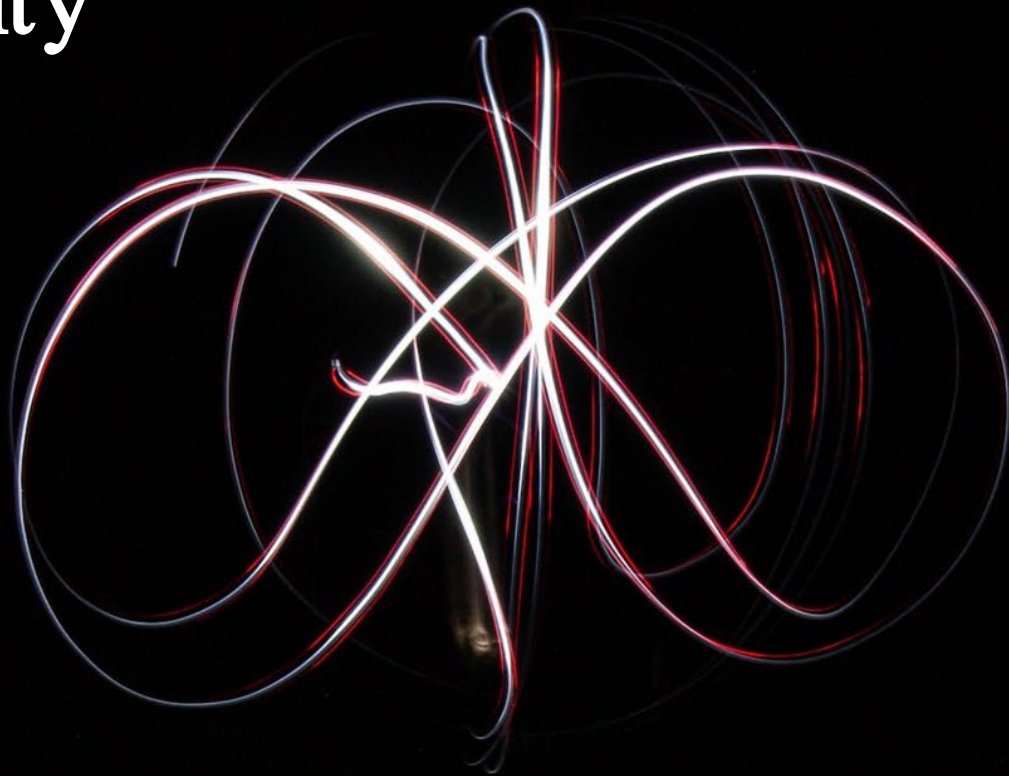


×2

faster



**Flexibility**



“It puts the power in the hands of users to find and download combinations of data they are most interested in.”

—Jen Woolford

Source: <https://blog.ons.gov.uk/2023/03/17/whats-next-for-census-2021/>



# Reproducibility & efficiency



“Reproducible analysis is more efficient,  
open, and easier to quality assure & re-use.”

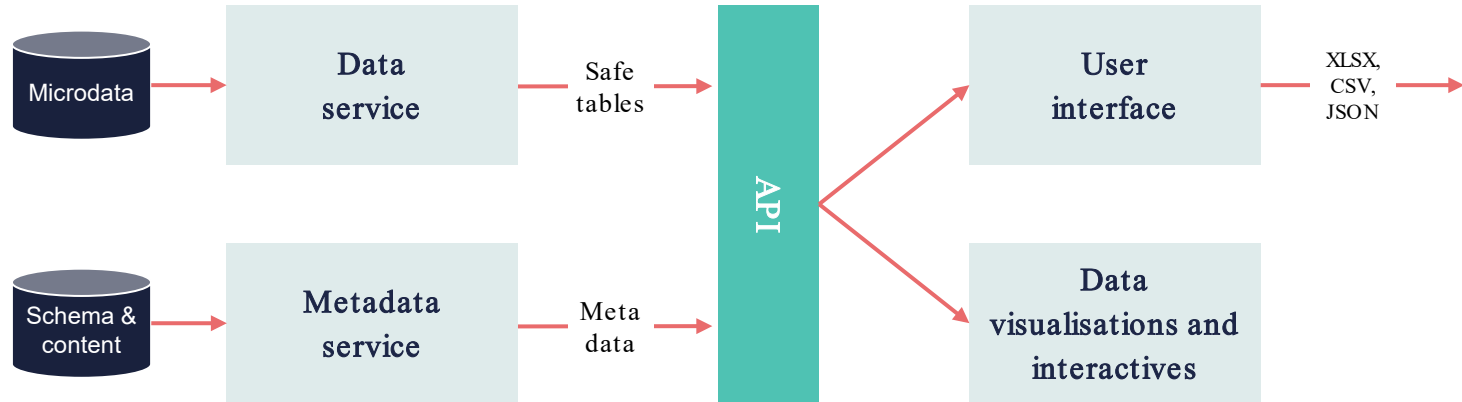
—UK Government Statistical Service

<https://analysisfunction.civilservice.gov.uk/policy-store/reproducible-analytical-pipeline-strategy/>



# Technical challenges & solutions

# How it works





# Challenge #1

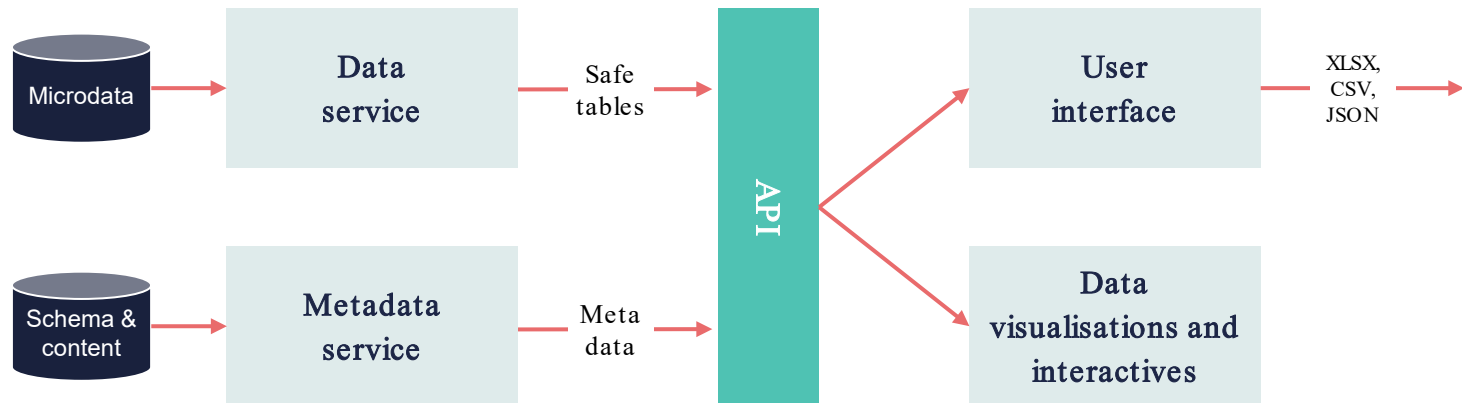
Automating tabulation and perturbation  
of outputs in real-time

# Need

Build cross-tabulations from confidential microdata and apply perturbation algorithms in real-time, in response to a user's query.



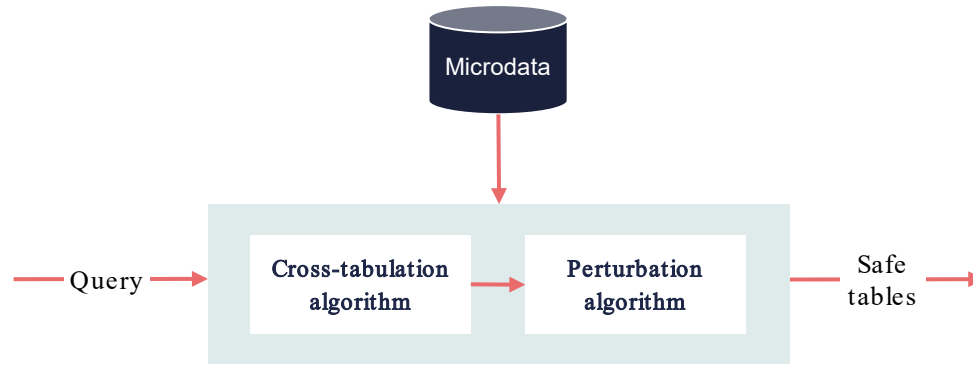
# How it works



# How it works



# How it works



# Making it fast and secure

- **Understand the speed killers:** all data transformations, every copy operation, network and I/O characteristics, memory footprint of data
- **Custom database:** implement only those features we need and optimise data structures
- **Minimal attack surface:** tiny API with tightly controlled inputs



# Perturbation approach

- Cell-key perturbation of frequency counts
- Independent perturbation of zeros in frequency counts
- Preservation of structural zeros

For more, see [“The methodological challenges of protecting outputs from a Flexible Dissemination System”](#) by Stephanie Blanchard.



# Use cases

- **Confidential microdata:** Cross-tabulation of confidential microdata with perturbation
- **Synthetic microdata:** Cross-tabulation of synthetic microdata (with no perturbation)
- **Static data:** Loading and serving of pre-computed static tables





## Challenge #2

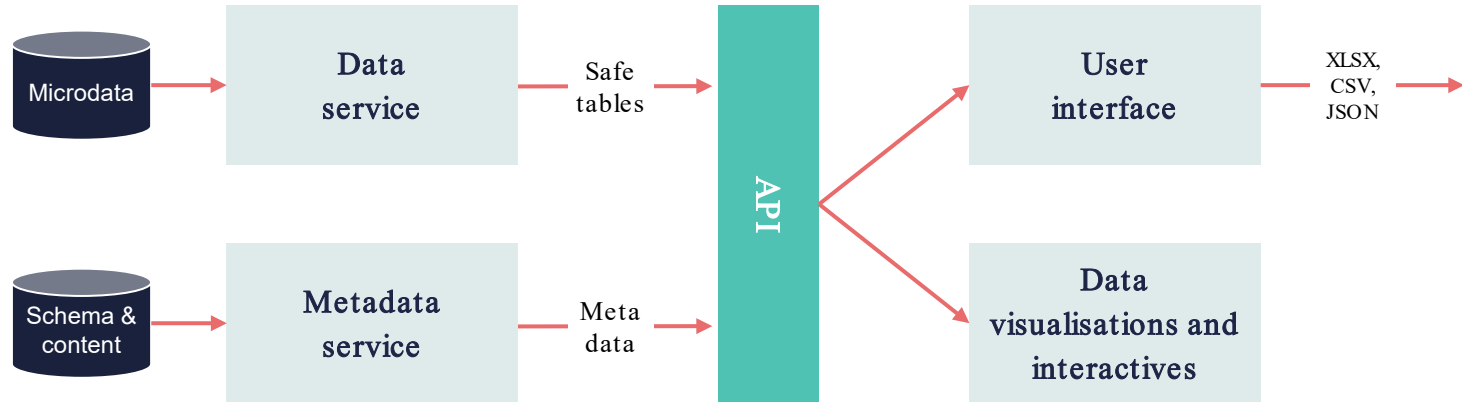
**Flexible data needs  
flexible metadata**

# Need

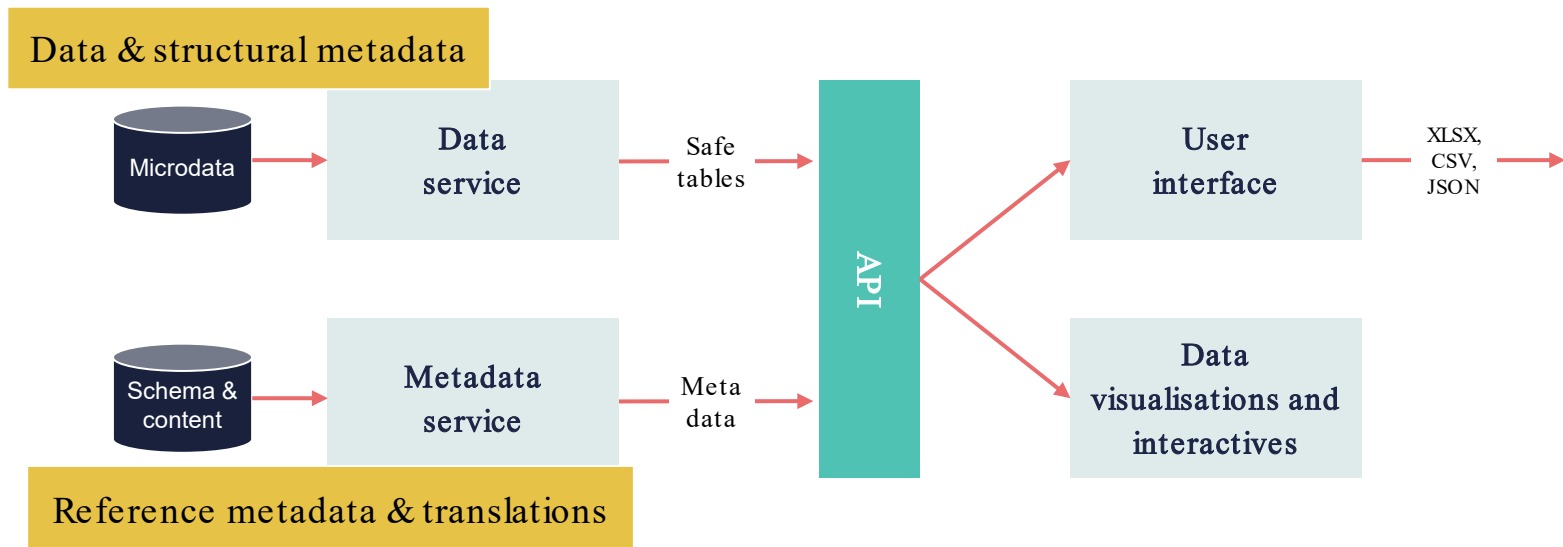
Develop a capability to allow metadata and translations to be associated with flexibly created outputs.



# How it works



# How it works



# Technical approach

- **User-defined schema:** allow specification of arbitrary custom metadata using a user-defined schema parsed when the software is run
- **Support multiple languages:** Census outputs need to be in English and Welsh; use metadata service to store translations of metadata

# Data dictionary

[Home](#) > [Metadata explorer](#) > [People](#) > Ethnic Group

## Ethnic Group

### Description

A person's ethnic group as stated in the census. The ethnic groups reported in this classification are those for which there was a tick-box on the census questionnaire, plus any ethnic group with more counts than the lowest frequency tick-box count.

[Build a table with this variable](#)

### Variable categories

Showing all 13 categories from this variable.

Label	Code
White	1
Irish Traveller	2
Roma	3
Indian	4
Chinese	5
Filipino	6
Pakistani	7
Arab	8

[See more](#)

### Related variables

Derived variables: [Ethnic Group - 2 Categories](#), [Ethnic Group - 5 Categories](#)

# Automatic Excel generation

	A	B	C	D	E	F	G
1	<b>DT-0024: Economic Activity (9 Categories) by Age (11 Categories) by Sex</b>						
2	This worksheet contains one table.						
3	Freeze panes are turned on. To turn off freeze panes please consult the help documentation for your software.						
4	<b>Count</b>			<b>Age - 11 Categories</b>		<b>Sex</b>	
5				<b>25-29 years</b>		<b>25-29 years</b>	
6						<b>30-34 years</b>	
7	<b>Local Government District 2014 Code</b>	<b>Local Government District 2014 Label</b>	<b>Economic Activity - 9 Categories</b>	<b>Female</b>	<b>Male</b>	<b>Female</b>	<b>Male</b>
8	N09000001	Antrim and Newtownabbey	Economically active: Employee (including full-ti	3,431	3,338	3,712	3,147
9	N09000001	Antrim and Newtownabbey	Economically active: Self-employed (including I	135	280	214	214
10	N09000001	Antrim and Newtownabbey	Economically active: Unemployed (including fu	109	209	91	91
11	N09000001	Antrim and Newtownabbey	Economically inactive: Retired	0	3	1	1
12	N09000001	Antrim and Newtownabbey	Economically inactive: Student	90	98	89	89
13	N09000001	Antrim and Newtownabbey	Economically inactive: Looking after home or fa	338	43	502	199
14	N09000001					198	199
15	N09000001					150	147
16	N09000002					0	0
17	N09000002					4,976	5,365
18	N09000002					625	416
19	N09000002					304	167
20	N09000002					1	8
21	N09000002					94	152
22	N09000002					82	865
23	N09000002					297	260
24	N09000002					263	221
25	N09000003					0	0
26	N09000003					9,235	9,311
27	N09000003					733	592
28	N09000003					780	381
29	N09000003					14	6
30	N09000003					793	372
31	N09000003					211	1,934
32	N09000003					747	689
33	N09000003					728	491
34	N09000004					0	0
35	N09000004					2,550	2,875
36	N09000004					503	263
37	N09000004					189	119
38	N09000004					2	0
39	N09000004					67	88
40	N09000004					52	599
41	N09000004					222	208
42	N09000004					233	142
43	N09000005					0	0
44	N09000005					2,850	3,449
45	N09000005					376	233
46	N09000005					325	183
47	N09000005					0	1
48	N09000005					119	79
49	N09000005					88	776
50	N09000005					261	338
51	N09000005					259	198
52	N09000006					0	0
53	N09000006					2,287	2,478
54	N09000006					407	198
55	N09000006					144	83
56	N09000006					0	2
57	N09000006					52	40
						41	401


	A	B	C
1	<b>DT-0024: Economic Activity (9 Categories) by Age (11 Categories) by Sex</b>		
2	This dataset provides Census 2021 estimates that classify usual residents in Northern Ireland by Economic Activity, Age, and Sex.		
3	<b>National Statistics Theme</b>		
4	Census 2021		
5	<b>Coverage</b>		
6	Local Government District 2014		
7	<b>Contact name</b>		
8	Census Customer Services		
9	<b>Contact number</b>		
10	028 9025 5156		
11	<b>Contact email</b>		
12	<a href="mailto:census@nirsra.gov.uk">census@nirsra.gov.uk</a>		
13	<b>National Statistics Data</b>		
14	Yes		
15	<b>Responsible Statistician</b>		
16	Dr David Marshall		
17	<b>Year of data</b>		
18	2021		
19	<b>Variables</b>		
20	Economic Activity - 9 Categories		
21	Age - 11 Categories		
22	Sex		
23	<b>Description of data</b>		
24	This dataset provides Census 2021 estimates that classify usual residents in Northern Ireland by Economic Activity, Age, and Sex.		
25	<b>Abstract</b>		
26	The census collected information on the usually resident population of Northern Ireland on Census Day (21 March 2021). Initial contact letters or questionnaire packs were delivered to every household and communal		



# Translations

[← Zurück](#)

## Wählen Sie Ihre Variablen

co 

Alles

4 passende Ergebnisse gefunden

[Suche löschen](#)

### [Art der Unterkunft](#)

3 Klassifizierungen verfügbar >

### [Geburtsland der Referenzperson für den Haushalt](#)

1 Klassifizierung verfügbar >

### [Land der Geburt](#)

4 Klassifizierungen verfügbar >

### [Wirtschaftliche Tätigkeit](#)

3 Klassifizierungen verfügbar >

### Ihre ausgewählten Variablen

Noch keine Variablen ausgewählt

[Speichern und weitermachen](#)

### Ihr Datensatz

#### Vertraulichkeit der Daten



348 von 348 Bereichen bestehen die Vertraulichkeitsprüfungen.

**Anzahl der Zellen:** 348

**Bevölkerung:** Personen- und Haushaltspseudodaten

**Geografische Ebene:** Lokale Behörde

**Geografisches Gebiet:** Die gesamte Bevölkerung


**Variablen:** Keine ausgewählt

**Filter:** Keine ausgewählt



# Translations

العربية [English](#)



[→ إلى الخلف](#)

## اختر المتغيرات الخاصة بك

ابحث عن المتغيرات المتاحة

الكل

تم العثور على 4 نتائج مطابقة

[مسح البحث](#)

[بلد الميلاد](#)  
4 تصنيفات متاحة

[بلد ميلاد الشخص المرجعي للأسرة](#)  
تصنيف واحد متاح

[نشاط اقتصادي](#)  
3 تصنيفات متاحة

[نوع الإقامة](#)  
3 تصنيفات متاحة


### المتغيرات التي اخترتها

لم يتم تحديد متغيرات حتى الآن

[احفظ وتابع](#)

### مجموعة البيانات الخاصة بك

سرية البيانات

348 من أصل 348 مناطق تجتاز اختبارات السرية:  100%

عدد الخلايا: 348

تعداد السكان: البيانات الزائفة للأفراد والأسرة

المستوى الجغرافي: سلطة محلية

منطقة جغرافية: جميع السكان

المتغيرات: لم يتم تحديد أي شيء

الفلتر: لم يتم تحديد أي شيء



# Example benefits

- **Contact phone number wrong?** Change it in one place and all Excel downloads will automatically use it.
- **Need to change a category label?** Change it in one place and it will appear automatically in the UI, data dictionary, visualisations and Excel downloads.



# Challenge #3

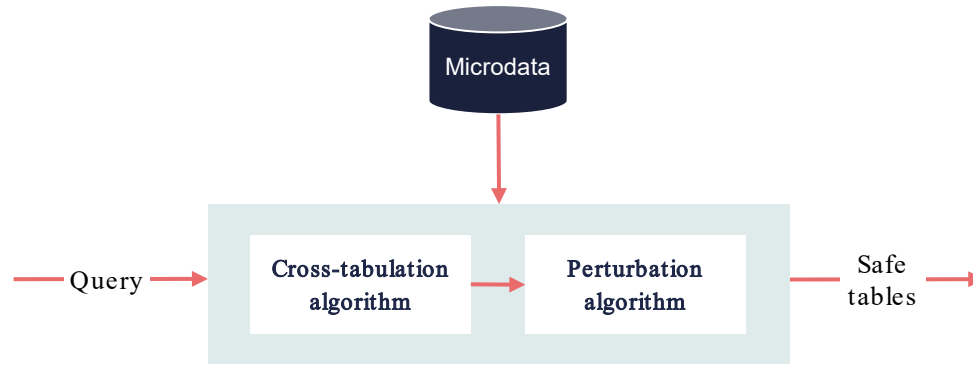
## Automating disclosure checks

# Need

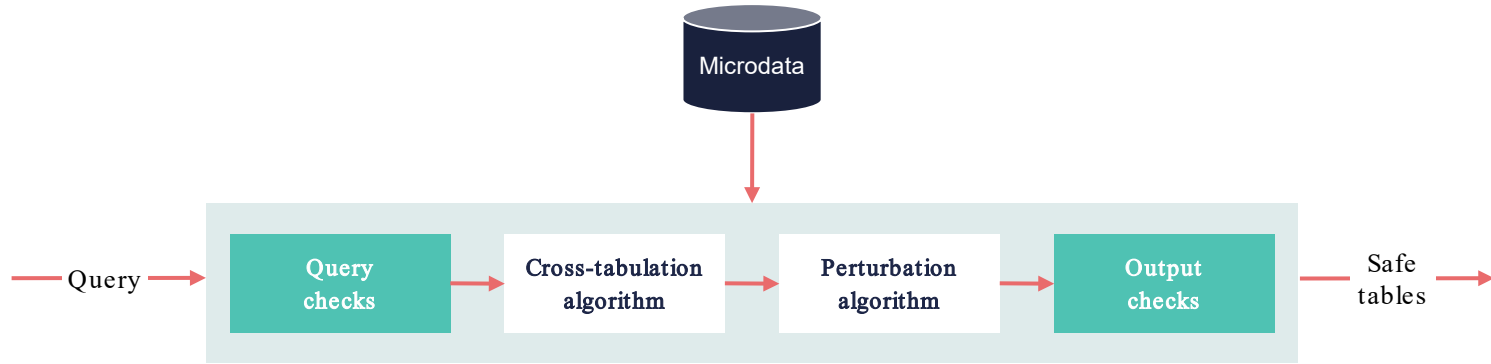
Create the capability to allow disclosure checks to be specified and automated, and for new checks to be created without requiring software changes.



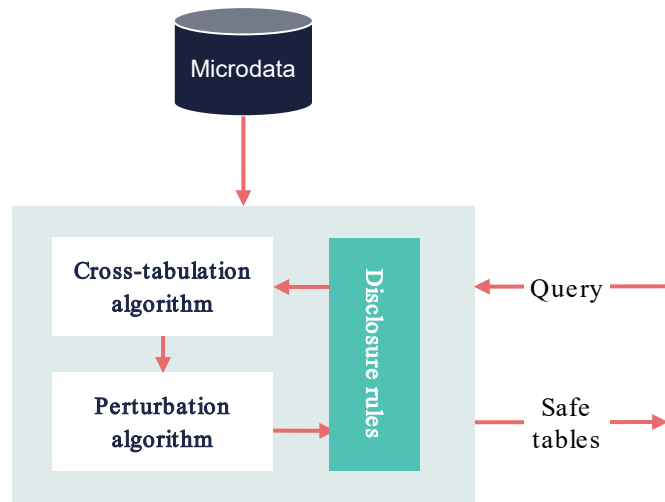
# How it works



# How it works



# How it works



# NTTS 2021 conference paper

## A Disclosure Rules Language for Deciding Publishability of Frequency Tables

### 2. RATIONALE

Whilst it is possible to use existing computer languages to specify the disclosure control rules, using such an approach would have the following disadvantages:

- The core software code and data would need protecting from user authored rules code in order to maintain resilience in the face of bugs. This incurs significant communication and performance penalties.
- Using close coupling for performance reasons creates difficulties with version management between user code and core software code (e.g. changes in data representation).
- A general purpose language has more scope for unintended behaviour and side effects, e.g. by unintentionally adding global state.
- Programs in general purpose languages are typically not amenable to parallel execution unless great care is taken and authors are knowledgeable in appropriate techniques.
- General purpose languages are typically large in scope and thus present more of a learning challenge for a program author.
- User authored code in a general purpose language would typically have more diverse ways of specifying a given rule and thus present more of a challenge for a reader.





# Query checks

- **Set maximum variables:** block queries that will lead to overly sparse outputs before they are run.
- **Set maximum cell count:** block queries that contain too many observations.
- **Sensitive variables:** limit queries against sensitive or detailed variables.



# Output checks

- **Attribute disclosure:** individual or group attribute disclosure in a table can be detected and suppressed.
- **Identity disclosure:** tables containing too many values of one can also be blocked.
- **Sparsity and dominance:** block tables that have proportionally too many zeros or are too skewed towards one particular category.



# Challenge #4

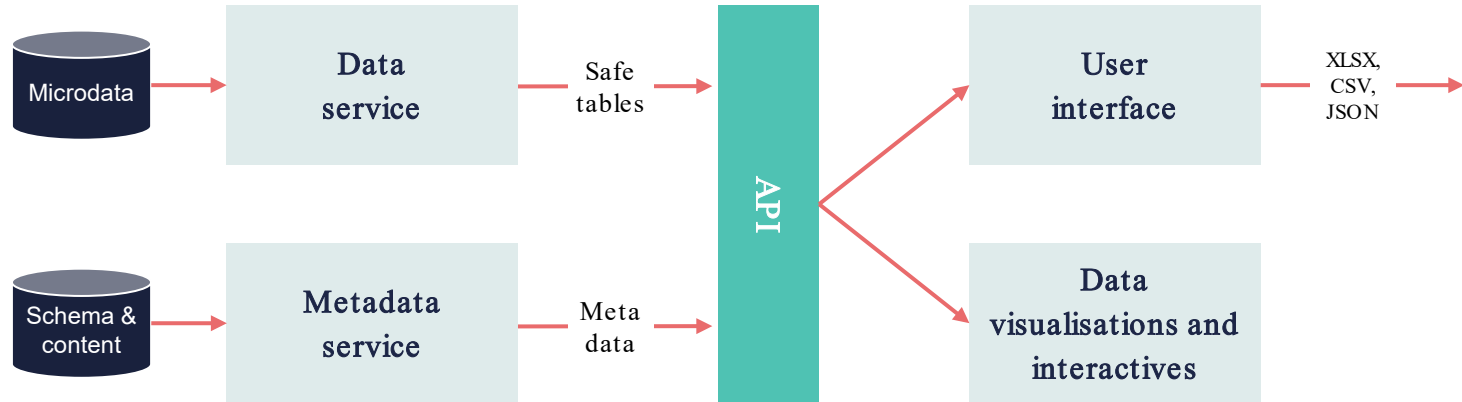
Helping users build their own tables

# Need

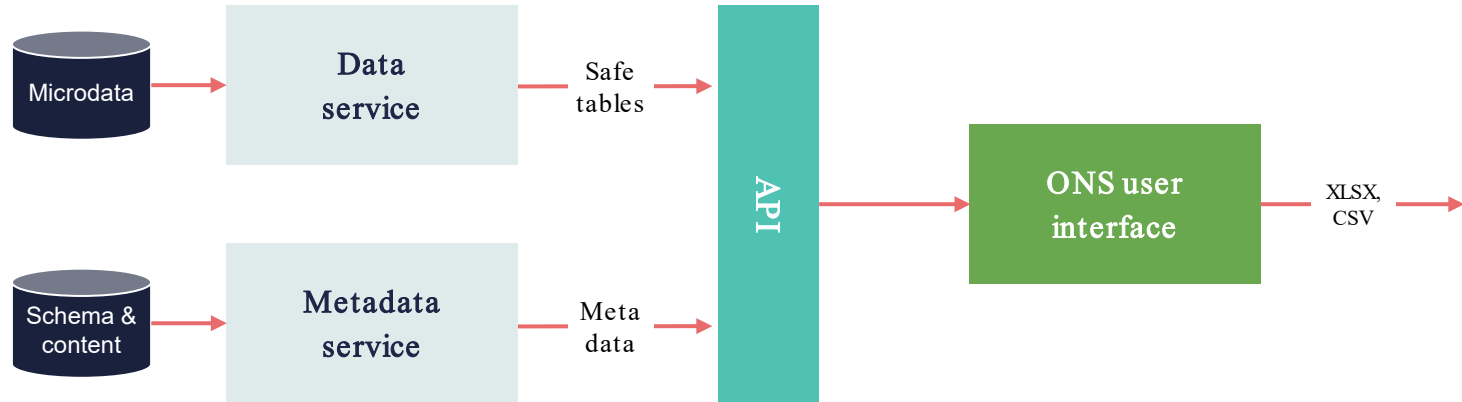
Design a user interface that helps users build their own table from scratch from a microdata-based dataset.



# How it works



# How it works



Demo:

<https://build.nisra.gov.uk>

What's next?



Protect privacy

Improve efficiency

Deliver more data, faster, with  
automated data publishing

Multiple language support

Easy-to-use & accessible



# What's next?

- **Supporting magnitude data:** improving tools for loading and finding pre-built magnitude data tables
- **Improving administration features:** adding features and control around publication of data
- **Adding visualisation tools:** allowing exploratory visualisation and mapping in the user interface



# Thanks!

mike@sensiblecode.io

<https://build.nisra.gov.uk>

<https://sensiblecodeio.github.io/ni-census-maps/>

<https://www.ons.gov.uk/datasets/create>

<https://cantabular.com>

