

Программное обеспечение для гибкого распространения результатов переписей 2021 года в Англии и Уэльсе

18 сентября 2023 года



The Sensible
Code Company

Содержание

- Что такое Cantabular?
- Для чего нужно автоматизировать публикацию данных?
- Технические проблемы и решения
- Что дальше?
- «Вопрос-ответ»

Что такое Cantabular?

Национальная статистическая служба: Создать пользовательский набор данных

The screenshot shows the Office for National Statistics website. At the top, there is a navigation bar with the logo and name 'Office for National Statistics' on the left, and links for 'English (EN) | Cymraeg (CY)', 'Release calendar', 'Methodology', 'Media', 'About', and 'Blog' on the right. Below this is a dark navigation menu with categories: 'Home', 'Business, industry and trade', 'Economy', 'Employment and labour market', 'People, population and community', and 'Taking part in a survey?'. A search bar is located below the menu with the placeholder text 'Search for a keyword(s) or time series ID' and a magnifying glass icon. The main content area has a breadcrumb trail 'Home > Census' and a large heading 'Create a custom dataset'. The text below 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.



Национальная статистическая служба: Создать пользовательский набор данных

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 usual resident who lives in a household in England, Wales and in a household

29 марта 2023 года



NISRA: Гибкий конструктор таблиц

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: Гибкий конструктор таблиц

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'. A large teal banner is overlaid on the bottom right of the screenshot, containing the date '22 июня 2023 года'.

Links

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

Follow NISRA

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



На платформе Cantabular

- Создайте собственную перекрестную таблицу



На платформе Cantabular

- Создайте собственную перекрестную таблицу
- Создано в режиме реального времени на основе микроданных



На платформе Cantabular

- Создайте собственную перекрестную таблицу
- Создано в режиме реального времени на основе микроданных
- Использование автоматических проверок раскрытия информации



На платформе Cantabular

- Создайте собственную перекрестную таблицу
- Создано в режиме реального времени на основе микроданных
- Использование автоматических проверок раскрытия информации
- Сотни миллионов возможных таблиц



На платформе Cantabular

- Создайте собственную перекрестную таблицу
- Создано в режиме реального времени на основе микроданных
- Использование автоматических проверок раскрытия информации
- Сотни миллионов возможных таблиц
- Метаданные объединяются при построении таблицы

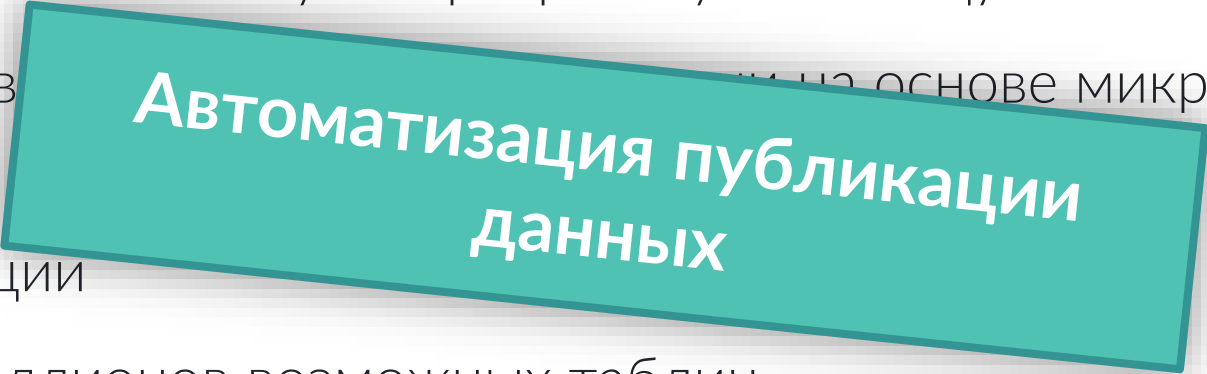


На платформе Cantabular

- Создайте собственную перекрестную таблицу
- Создано в режиме реального времени на основе микроданных
- Использование автоматических проверок раскрытия информации
- Сотни миллионов возможных таблиц
- Метаданные объединяются при построении таблицы
- Все результаты генерируются по запросу



На платформе Cantabular

- Создайте собственную перекрестную таблицу
- Создано в  на основе микроданных
- Используйте информацию
- Сотни миллионов возможных таблиц
- Метаданные объединяются при построении таблицы
- Все результаты генерируются по запросу

**Автоматизация публикации
данных**

**Для чего нужно автоматизировать
публикацию данных?**

Скорость



«Позволяет нам публиковать миллиарды анонимных статистических данных переписи населения гораздо быстрее, чем когда-либо прежде»

– **Джен Вулфорд, директор отдела статистики населения, НСС**

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

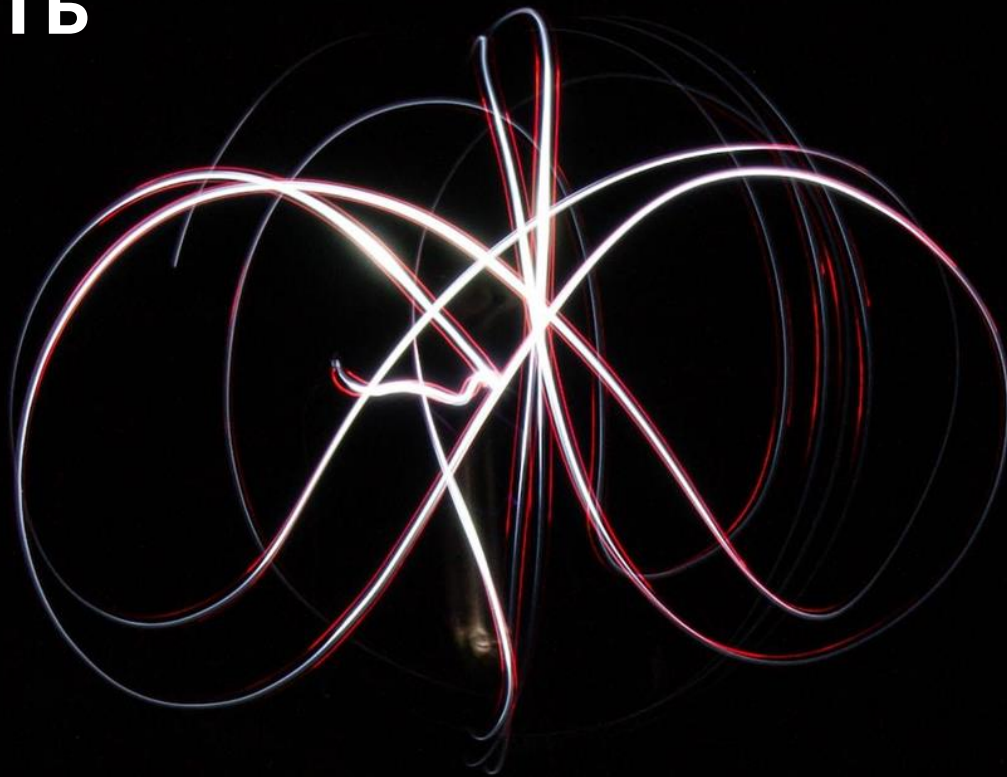


x2

быстрее



Гибкость



«Дает пользователям возможность
находить и загружать комбинации
данных, которые им наиболее
интересны»

– Джен Вулфорд

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



Воспроизводимость и эффективность



«Воспроизводимый анализ более эффективен, открыт, легче обеспечить его качество и повторно его использовать» ”

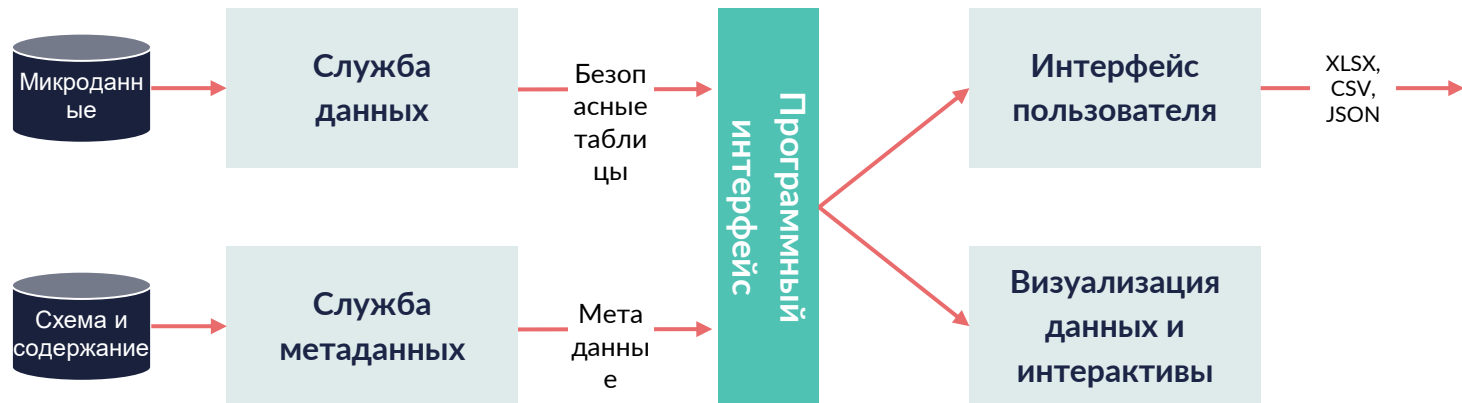
— Государственная статистическая служба Великобритании

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



Технические сложности и решения

Как это работает



Проблема 1

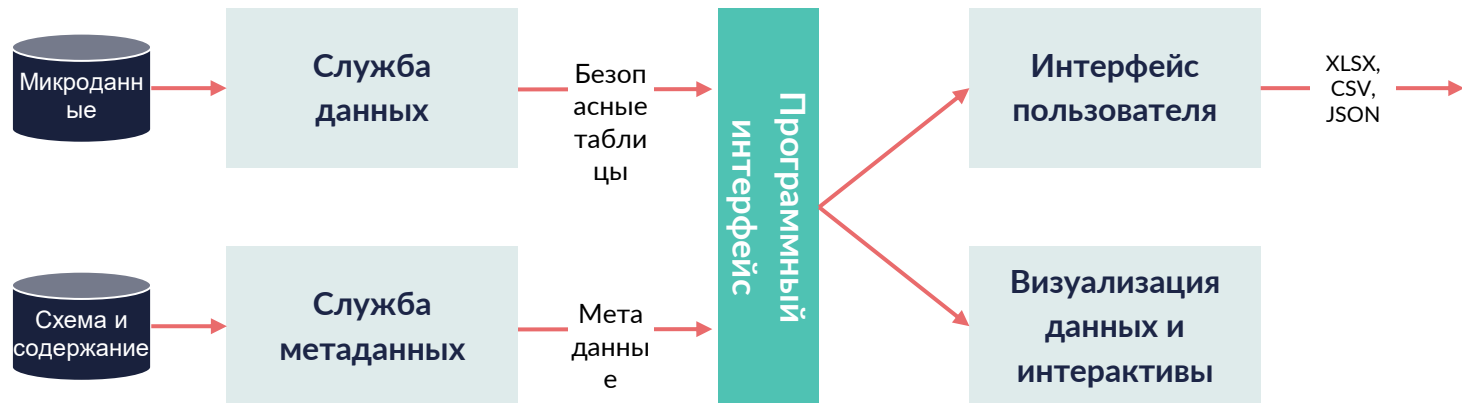
Автоматизация табулирования и
пертурбации результатов в режиме
реального времени

Потребность

Создавать перекрестные таблицы на основе конфиденциальных микроданных и применять алгоритмы пертурбации в режиме реального времени в ответ на запрос пользователя.



Как это работает



Как это работает



Как это работает



Обеспечить скорость и надежность

- **Понимать, что является тормозом:** все преобразования данных, каждая операция копирования, характеристики сети и ввода-вывода, объем памяти, требуемой для данных
- **Пользовательская база данных:** реализовать только те функции, которые нам нужны, и оптимизировать структуры данных
- **Минимальная поверхность атаки:** крошечный программный интерфейс приложений с жестко контролируемыми входными данными



Пертурбативный подход

- Пертурбация отсчетов частоты по ячейковому ключу
- Независимая **пертурбация нулей** в отсчетах частот
- Сохранение **структурных нулей**

Дополнительную информацию см. в статье Стефани Бланшар [«Методологические проблемы защиты результатов гибкой системы распространения»](#).



Случаи использования

- **Конфиденциальные микроданные:** Перекрестная таблица конфиденциальных микроданных с пертурбацией
- **Синтетические микроданные:** Перекрестная таблица синтетических микроданных (без пертурбации)
- **Статические данные:** Загрузка и обслуживание предварительно рассчитанных статических таблиц



Проблема 2

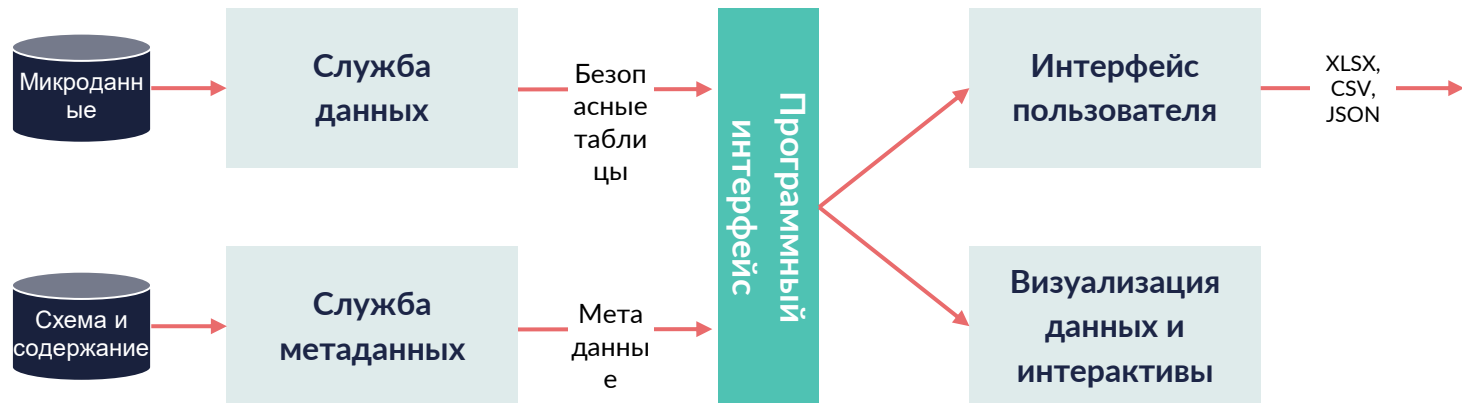
Гибким данным нужны
гибкие метаданные

Потребность

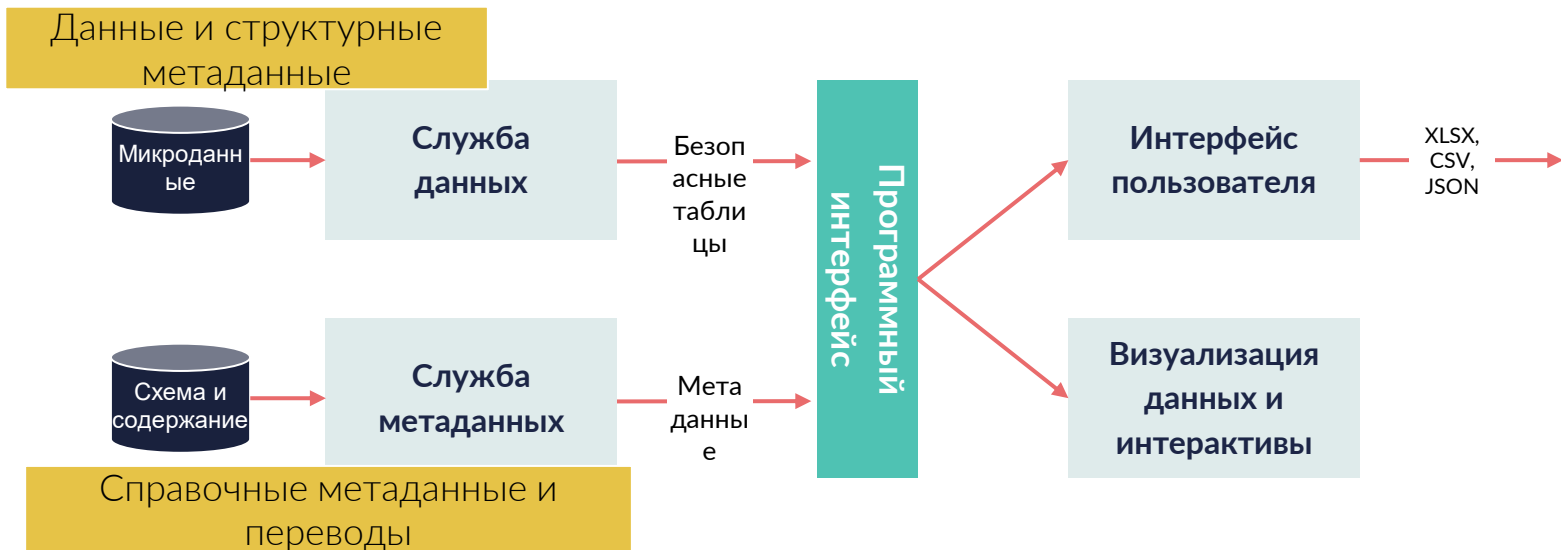
Разработать возможность связывать метаданные и переводы с гибко создаваемыми результатами.



Как это работает



Как это работает



Технический подход

- **Пользовательская схема:** позволяет указывать произвольные пользовательские метаданные с использованием пользовательской схемы, анализируемой при запуске программного обеспечения.
- **Поддержка нескольких языков:** Результаты переписи должны быть на английском и валлийском языках; использовать службу метаданных для хранения переводов метаданных



Словарь данных

[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](#)

Автоматическое создание файла в формате Excel

	A	B	C	D	E	F	G
1	DT-0024: Economic Activity (9 Categories) by Age (11 Categories) by Sex						
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	Count			Age - 11 Categories		Sex	
5				25-29 years		30-34 years	
6	Local Government District 2014 Code	Local Government District 2014 Label	Economic Activity - 9 Categories	Female	Male	Female	Male
7	N09000001	Antrim and Newtownabbey	Economically active: Employee (including full-ti	3,431	3,338	3,712	3,147
8	N09000001	Antrim and Newtownabbey	Economically active: Self-employed (including I	135	280	214	214
9	N09000001	Antrim and Newtownabbey	Economically active: Unemployed (including fu	109	200	91	91
10	N09000001	Antrim and Newtownabbey	Economically inactive: Retired	0	3	1	1
11	N09000001	Antrim and Newtownabbey	Economically inactive: Student	90	98	89	89
12	N09000001	Antrim and Newtownabbey	Economically inactive: Looking after home or fa	338	43	502	199
13	N09000001						199
14	N09000001						150
15	N09000001						0
16	N09000002						4,976
17	N09000002						6,255
18	N09000002						304
19	N09000002						1
20	N09000002						94
21	N09000002						82
22	N09000002						297
23	N09000002						263
24	N09000002						0
25	N09000003						9,235
26	N09000003						733
27	N09000003						780
28	N09000003						14
29	N09000003						793
30	N09000003						211
31	N09000003						747
32	N09000003						728
33	N09000003						0
34	N09000004						2,550
35	N09000004						503
36	N09000004						189
37	N09000004						2
38	N09000004						67
39	N09000004						52
40	N09000004						222
41	N09000004						233
42	N09000004						0
43	N09000005						2,850
44	N09000005						376
45	N09000005						325
46	N09000005						0
47	N09000005						119
48	N09000005						88
49	N09000005						261
50	N09000005						259
51	N09000005						0
52	N09000006						2,287
53	N09000006						407
54	N09000006						144
55	N09000006						0
56	N09000006						52
57	N09000006						41

	A	B	C
1	DT-0024: Economic Activity (9 Categories) by Age (11 Categories) by Sex		
2	This dataset provides Census 2021 estimates that classify usual residents in Northern Ireland by Economic Activity, Age, and Sex.		
3	National Statistics Theme		
4	Census 2021		
5	Coverage		
6	Local Government District 2014		
7	Contact name		
8	Census Customer Services		
9	Contact number		
10	028 9025 5156		
11	Contact email		
12	census@nirsra.gov.uk		
13	National Statistics Data		
14	Yes		
15	Responsible Statistician		
16	Dr David Marshall		
17	Year of data		
18	2021		
19	Variables		
20	Economic Activity - 9 Categories		
21	Age - 11 Categories		
22	Sex		
23	Description of data		
24	This dataset provides Census 2021 estimates that classify usual residents in Northern Ireland by Economic Activity, Age, and Sex.		
25	Abstract		
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		



[← 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

→ 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



Примеры преимуществ

- **Контактный телефон неправильный?** Измените его в одном месте и во всех загрузках в формате Excel он изменится автоматически.
- **Нужно поменять метку категории?** Смените ее в одном месте и она появится автоматически в пользовательском интерфейсе, словаре данных, визуализациях и загрузках Excel.



Проблема 3

Автоматические проверки раскрытия информации

Потребность

Обеспечить возможность определять и автоматизировать проверки раскрытия информации, а также создавать новые проверки без необходимости внесения изменений в программное обеспечение.



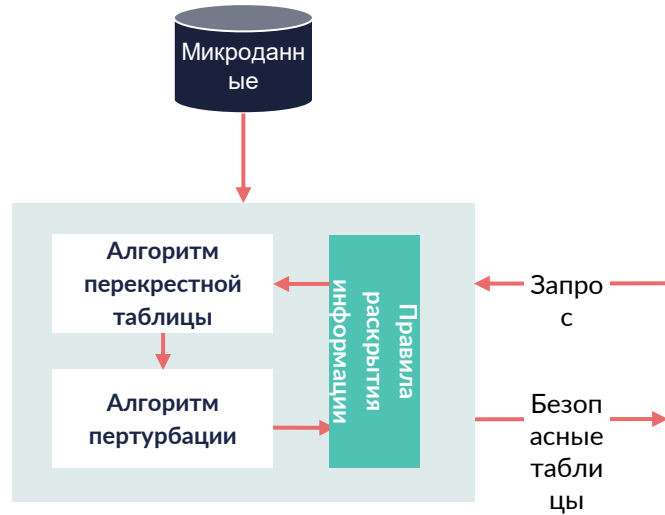
Как это работает



Как это работает



Как это работает



Доклад конференции «Новые методы и технологии для статистики» 2021 года

Язык правил раскрытия информации для принятия решения о возможности публикации таблиц частот

2. ОБОСНОВАНИЕ

Хотя для определения правил контроля раскрытия информации можно использовать существующие компьютерные языки, использование такого подхода будет иметь следующие недостатки:

- Основной программный код и данные должны быть защищены от кода правил, созданных пользователем, чтобы обеспечить устойчивость к ошибкам. Это влечет за собой значительные потери связи и производительности.
- Использование тесной связи для повышения производительности создает трудности с управлением версиями между пользовательским кодом и основным кодом программного обеспечения (например, изменения в представлении данных).
- Язык общего назначения имеет больше возможностей для непреднамеренного поведения и побочных эффектов, например, путем непреднамеренного добавления глобального состояния.
- Программы на языках общего назначения, как правило, не поддаются параллельному выполнению, если не проявить особую осторожность и если авторы не обладают знаниями соответствующих методов.
- Языки общего назначения обычно имеют большой объем и, таким образом, представляют собой большую трудность в обучении для автора программы.
- Код, написанный пользователем на языке общего назначения, обычно имеет более разнообразные способы определения данного правила и, таким образом, представляет собой большую проблему для читателя.



Проверки запросов

- **Установить максимальное количество переменных:** блокировать запросы, которые приведут к слишком разреженным выходным данным, до их выполнения.
- **Установить максимальное количество ячеек:** блокировать запросы, содержащие слишком много наблюдений.
- **Чувствительные переменные:** ограничить запросы к конфиденциальным или подробным переменным.



Проверки результата

- **Раскрытие атрибутов:** раскрытие индивидуальных или групповых атрибутов в таблице может быть обнаружено и подавлено.
- **Раскрытие личности:** таблицы, содержащие слишком много значений одного, также могут быть заблокированы.
- **Разреженность и доминирование:** блокировать таблицы, которые имеют пропорционально слишком много нулей или слишком смещены в сторону одной конкретной категории



Проблема 4

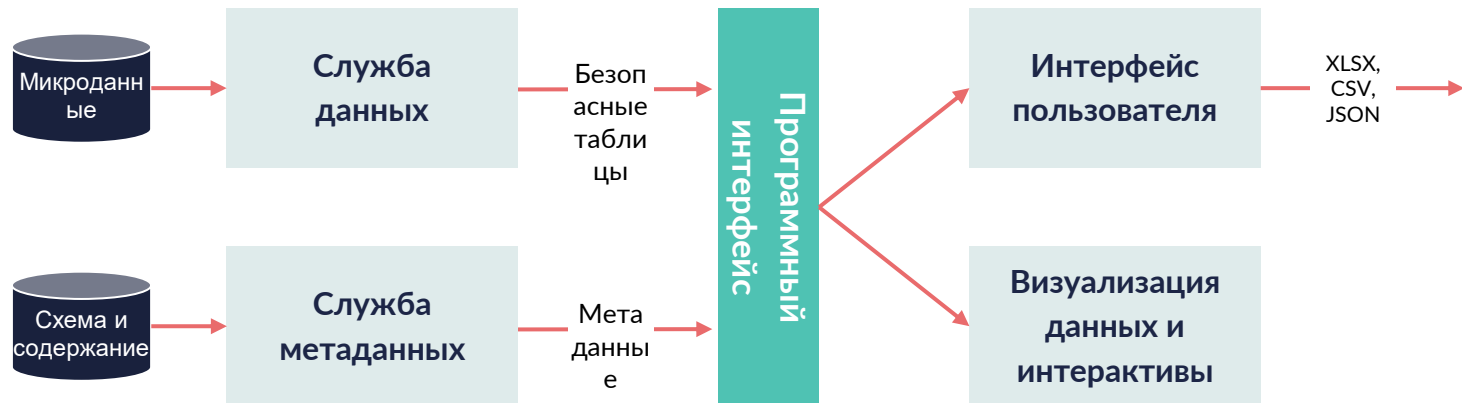
Помощь пользователям в построении
их собственных таблиц

Потребность

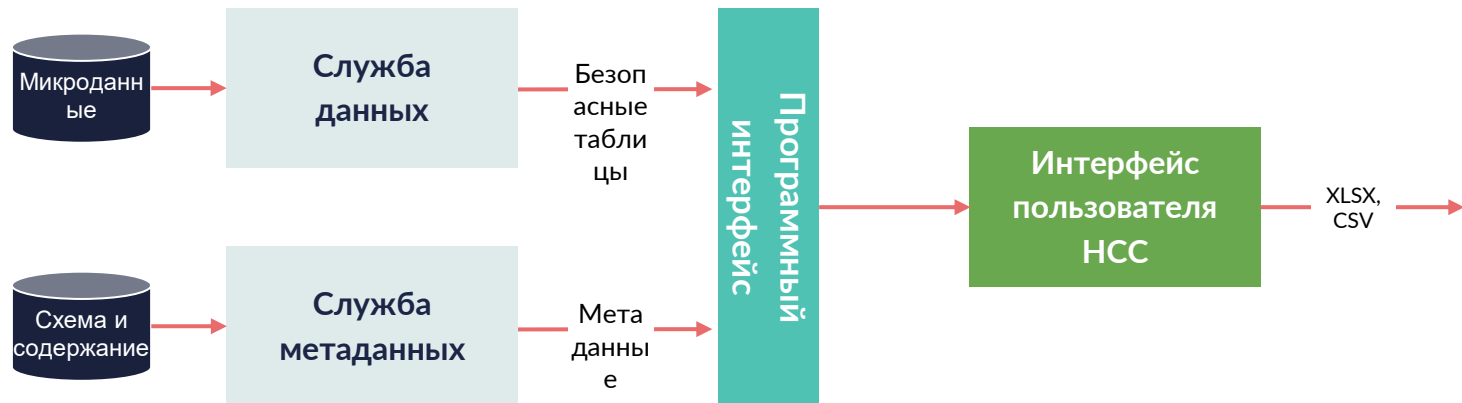
Разработать пользовательский интерфейс, который помогает пользователям строить свои собственные таблицы с нуля на основе набора данных на основе микроданных.



Как это работает



Как это работает



Демонстрационная
версия:

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

Что дальше?

Защитить
конфиденциальность

Повысить
эффективность

Обеспечить больше данных, быстрее, с
автоматической публикацией данных

Поддержка нескольких
языков

Простота в
использовании и
доступность



Что дальше?

- **Поддержка данных величин:** совершенствование инструментов для загрузки и поиска готовых таблиц данных величин
- **Улучшение функций администрирования:** добавление функций и контроля публикации данных
- **Добавление инструментов визуализации:** разрешение исследовательской визуализации и картографирования в пользовательском интерфейсе



Спасибо!

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>

