

# Микроданные для науки: зарубежный опыт организации инфраструктуры доступа

Витовт Копыток

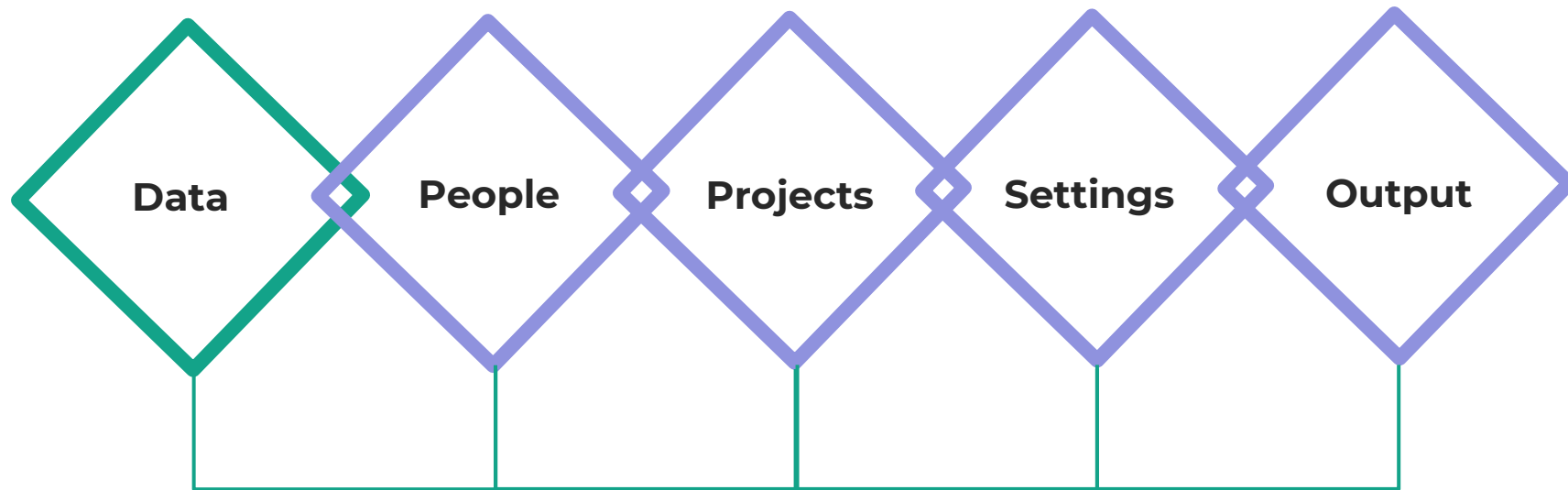


Международный день открытых данных  
Москва, 6 марта



Статистически контроль раскрытия информации (Statistical Disclosure Control) – не единственный способ обеспечить безопасность микроданных

## 5 аспектов защиты («5 SAFES»)



баланс «полезность-безопасность»: обработка данных + механика и права доступа



## «Продвинутый контур» ЦПУР

### Сведения о регистрируемой безработице в России

Набор содержит обезличенные микроданные о характеристиках граждан, которые обращались в органы занятости населения за государственной услугой по содействию в поиске подходящей работы. В рамках оказания этой услуги в России осуществляется официальная регистрация безработных граждан и принимается решение о назначении им пособия по безработице.

- ✓ Верифицированные исследователи, аффилированные с российскими учебными или исследовательскими организациями
- ✓ Доступ через RDP без возможности выгрузить данные
- ✓ Проверка результатов перед отгрузкой исследователю

Размещено: 25.02.2021

Обновлено: 01.02.2021

Размер набора: 12,6 ГБ

[Описание набора](#) ↓

[Источник](#) ↗

Форматы:

CSV

Версия датасета: ?

Последний обновленный ▾

Доступ: ? Продвинутый

Добавить в корзину



# Как публикуют данные переписей: кейс Великобритании



## Census 2011

- учебная выборка (1%) с усеченным числом атрибутов для свободного скачивания
- две исследовательские выборки (5%) для верифицированных исследователей через the Virtual Microdata Laboratory (VML)
- две расширенные выборки (10%) для верифицированных исследователей с доступом через дата-центры

**Safe Data**

**Safe Data**  
**Safe People**  
**Safe Settings**  
**Safe Project**  
**Safe Output**



## Как публикуют данные переписей: кейс Германии



### административные микроданные

- файлы для научного использования (анонимизированные микроданные) для исследовательских организаций
- удаленное выполнение программного кода с последующим контролем раскрытия статистической информации
- физический доступ к микроданным через гостевые RDC (в университетах и исследовательских центрах)

**Safe Data**  
**Safe People**

**Safe Data**  
**Safe People**  
**Safe Settings**  
**Safe Project**  
**Safe Output**



# Институционализация предоставления доступа к микроданным

## Кто (организация)

	<i>Госорган самостоятельно</i>	<i>Специальная подведомственная организация</i>	<i>Партнерство с университетом фондом, НКО</i>
<i>Safe Settings</i>			
<i>Прямая передача данных</i>	National Center for Education Statistics (США), Statistics Sweden (Швеция); Statistics Netherlands (Нидерланды)	RDC-IAB (Германия): специальные научные файлы, ADR UK (Великобритания): в рамках проектов	Ohio and the Longitudinal Data Archive (США): передача через SFTP на рабочий ПК
<i>Удаленный защищенный доступ</i>	Secure Research Service (Великобритания), National Center for Education Statistics (США)	RDC-IAB (Германия), CCP Data Science unit (ЮАР), ADELE (Италия), ADR UK (Великобритания)	The Stanford-San Francisco Unified School District Partnership (США)
<i>Защищенная физическая локация</i>	Secure Research Service (Великобритания)	RDC-IAB (Германия), NZ Data Lab (Новая Зеландия), ADR UK (Великобритания), ADELE (Италия)	The Stanford-San Francisco Unified School District Partnership (США), NB-IRDT (Канада), Federal Statistical Research Data Centers (США), Statistics Finland (Финляндия)

Как (механика)



# Институционализация предоставления доступа к микроданным для России

## Кто (организация)

	<i>Госорган самостоятельно</i>	<i>Специальная подведомственная организация</i>	<i>Партнерство с университетом фондом, НКО</i>
<i>Safe Settings</i>			
<i>Прямая передача данных</i>	National Center for Education Statistics (США), Statistics Sweden (Швеция); Statistics Netherlands (Нидерланды)	RDC-IAB (Германия): специальные научные файлы, ADR UK (Великобритания): в рамках проектов	Ohio and the Longitudinal Data Archive (США): передача через SFTP на рабочий ПК
<i>Удаленный защищенный доступ</i>	Secure Research Service (Великобритания), National Center for Education Statistics (США)	RDC-IAB (Германия), CCP Data Science unit (ЮАР), ADELE (Италия), ADR UK (Великобритания)	<b>Доступ к микроданным переписи через партнерскую сеть: удаленный доступ и доступ через защищенные рабочие места</b>
<i>Защищенная физическая локация</i>	Secure Research Service (Великобритания)	RDC-IAB (Германия), NZ Data Lab (Новая Зеландия), ADR UK (Великобритания), ADELE (Италия)	

Как (механика)



# 5 шагов для расширения доступности и применимости данных переписей России

**1.** Выборки микроданных с ограниченным числом атрибутов для переписей **2002, 2010, (2015)** в **свободном доступе**

**2.** Расширенные выборки микроданных переписей **2002, 2010, (2015)**, доступные **верифицированным исследователям** через **защищенную инфраструктуру**

**3\*.** Связанные микроданные переписей **2002, 2010, (2015)**, доступные **верифицированным исследователям** через **защищенную инфраструктуру**

**4.** Ограниченная выборка в свободном доступе и расширенные выборки переписи **2021**, доступные **верифицированным исследователям** через **защищенную инфраструктуру**

**5\*.** Связанные микроданные переписей **2002, 2010, (2015), 2021** и административные данные (ФНС, ПФР), доступные через **защищенную инфраструктуру**

# Приложение

# Решения для ВПН-2021 (+ 2002, 2010, 2015)

## План-минимум

Опубликовать выборки с микроданными переписей **с ограниченным числом атрибутов** в свободном доступе

## План-максимум

Реализовать доступ к **расширенным выборкам с микроданными** переписей через защищенную инфраструктуру для верифицированных исследователей:

- открытые и равные правила получения доступа;
- публичный перечень исследовательских проектов и верифицированных исследователей, получивших доступ к данным

## Экстра-план

Связать данные переписей за разные годы в панель, а также с источниками административных данных (Deterministic + Probabilistic Record Linkage). Примеры:

- [AUSTRALIAN CENSUS LONGITUDINAL DATASET](#);
- [Opportunity Insights](#)