

Школа молодого  
исследователя

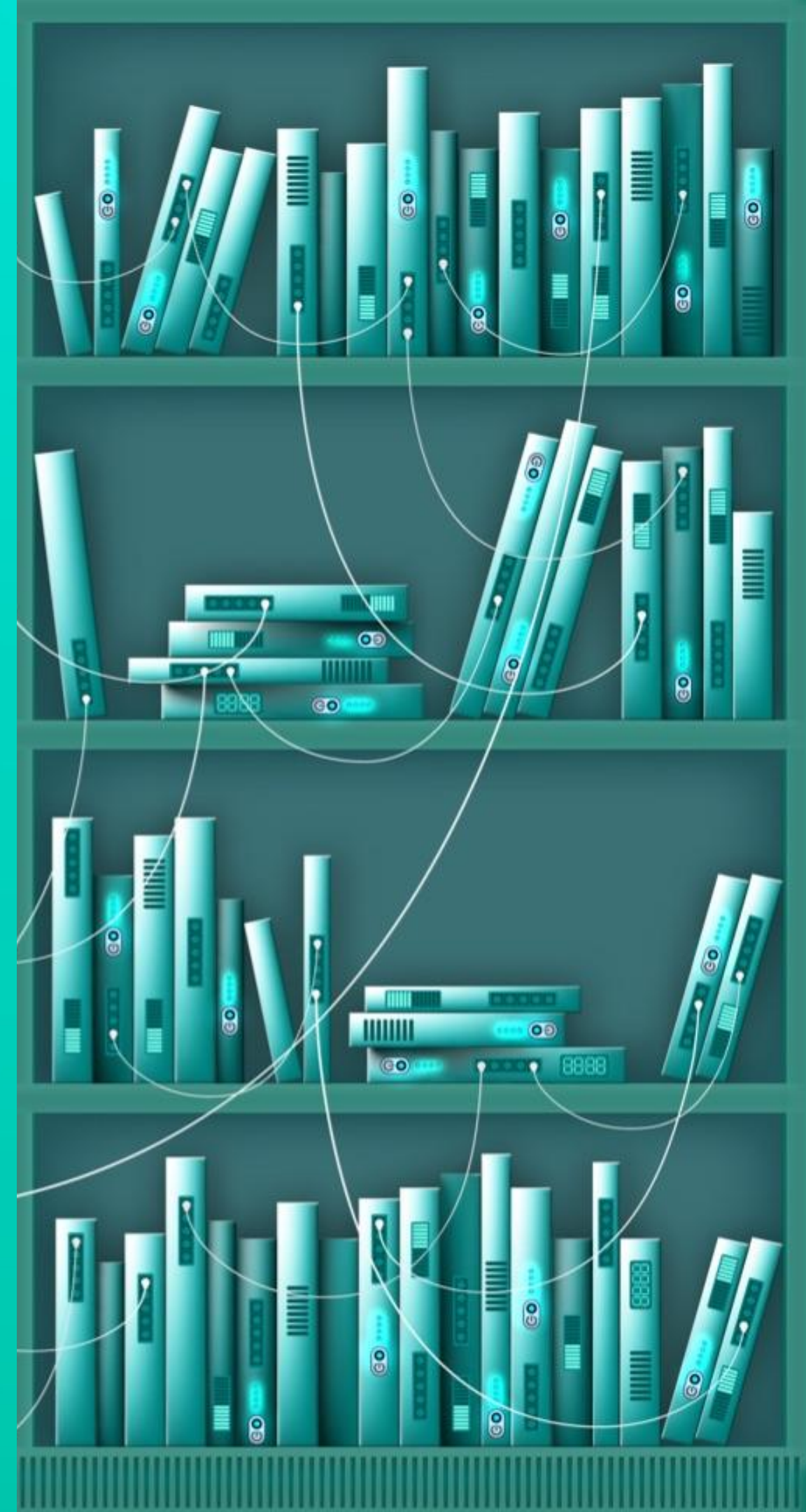
## Исследования на данных

Что можно изучать и где взять базы для исследований в социальных и экономических науках?



[data-in.ru](https://data-in.ru)

June in Che, 2022





# Сергей Тихонов

Старший эксперт

[s.tikhonov@cpur.ru](mailto:s.tikhonov@cpur.ru)

Магистр социальных наук (Европейский университет в Санкт-Петербурге, 2021). Работал в Центре институционального анализа науки и образования Европейского университета в Санкт-Петербурге (ЦИАНО).

Исследовательские интересы: методы идентификации причинных связей, статистика, оценка отдачи от образования





## **Инфраструктура научно-исследовательских данных — платформа доступа исследователей к данным о государстве и обществе**

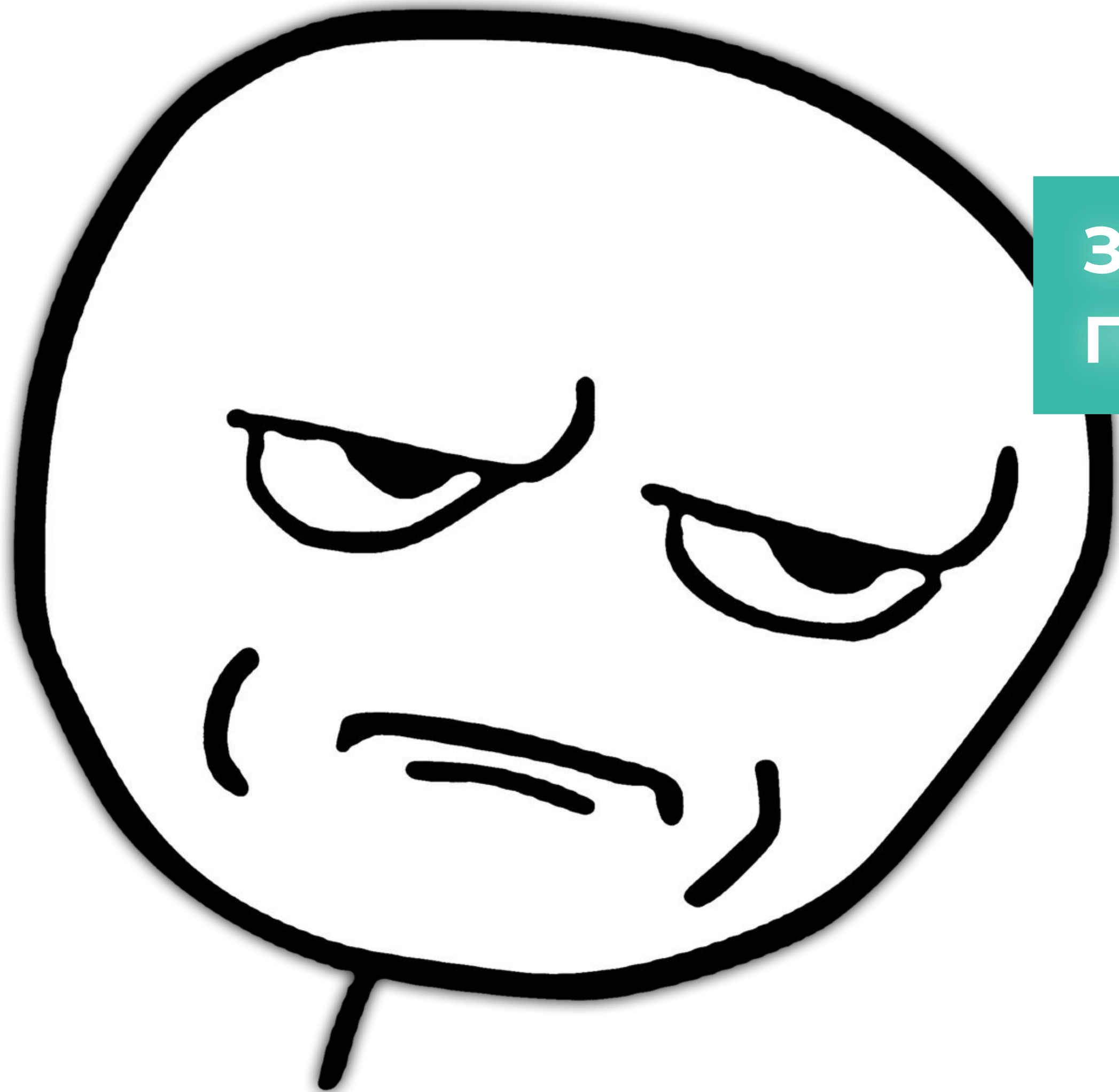
Наша миссия — обеспечивать продуктивное взаимодействие между владельцами данных о государстве и обществе и исследователями для принятия управленческих решений на основе данных и совершенствования государственной политики.

Благодаря платформе исследователи могут получать доступ к данным, а представители органов государственной власти ставить прикладные задачи и предоставлять для их решения наборы данных.



# Ключевые задачи проекта

- Публиковать ранее недоступные для исследований базы и наборы данных от государственных органов
- Предоставлять технические возможности исследователям для работы с данными и содействовать разработке сервисов на базе ИИ
- Обеспечивать юридические и методические инструменты для публикации данных
- Способствовать развитию сообщества исследователей, работающих с данными в России



**ЗАЧЕМ НУЖНЫ ТАКИЕ  
ПРОЕКТЫ КАК ИНИД?**

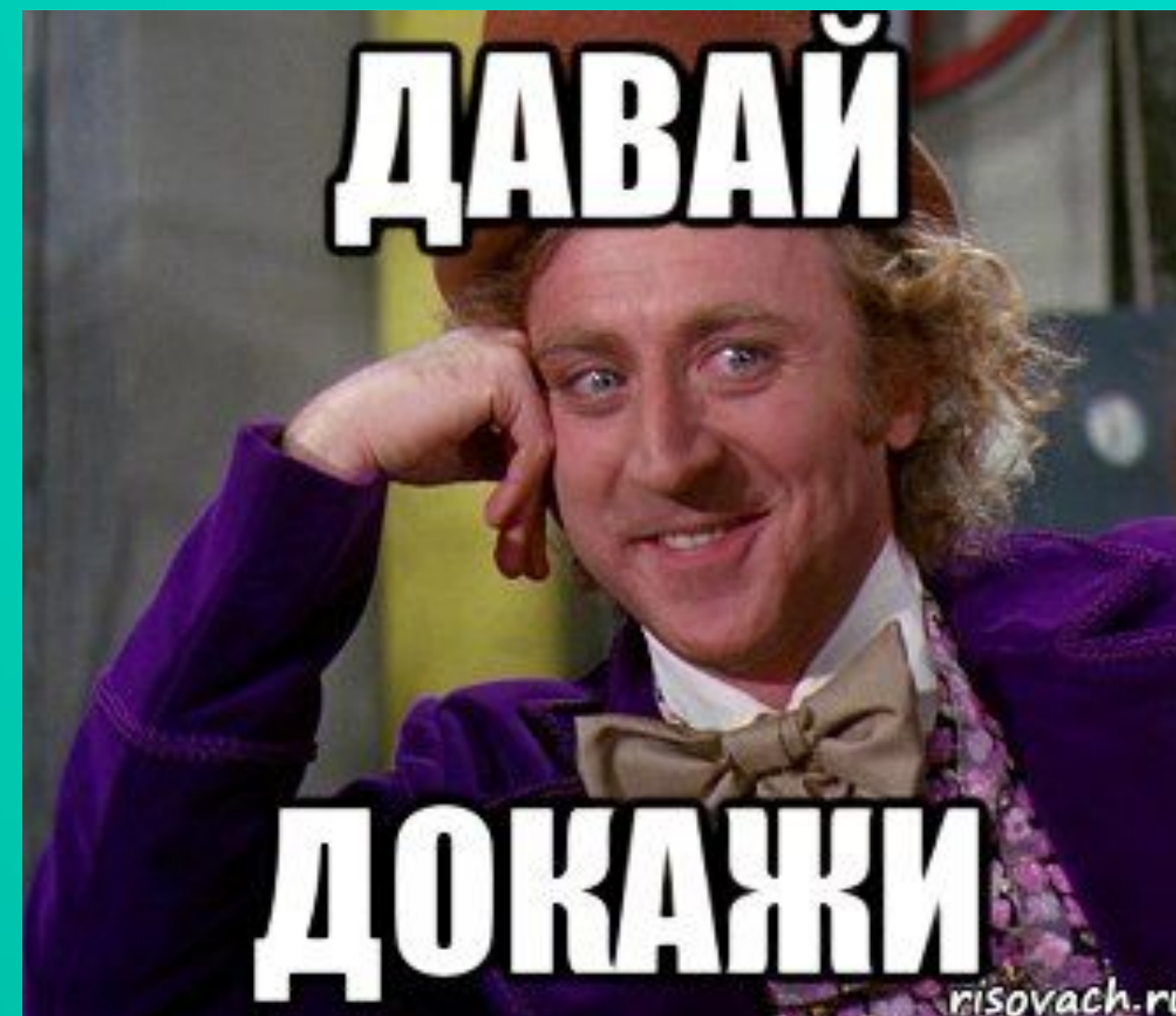


центр  
перспективных  
управленческих  
решений

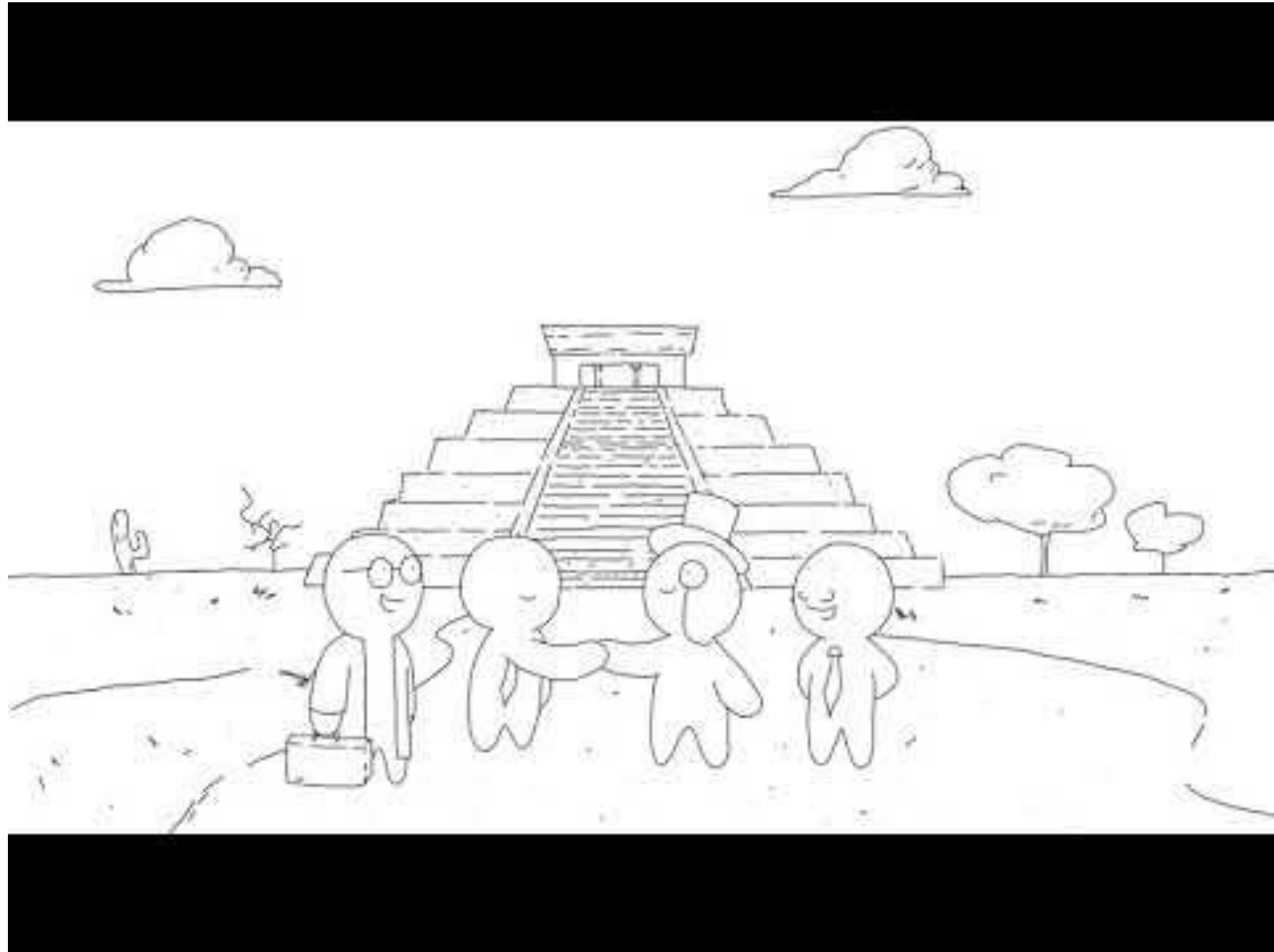


**ИНИИД**  
ИНФРАСТРУКТУРА  
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ  
ДАнных

# 1. Доказательная политика







портал  
«ДОКАЗАТЕЛЬНАЯ  
ПОЛИТИКА»



[ebpm.cpur.ru](http://ebpm.cpur.ru)

# Портал «Доказательная политика»

## КАК РАБОТАЕТ ДОКАЗАТЕЛЬНАЯ ПОЛИТИКА? (ФРЕЙМВОРК)

ЭТАПЫ (ЧТО?)

УЧАСТНИКИ (КТО?)

МЕТОДЫ (КАК?)



Кейсы  
Исследования  
Видеокорсы на  
русском языке  
Полезные ссылки



### Лекция 1. «А/В тесты»

Вводная лекция в курс, на которой разобрана базовая модель эксперимента, а также приведены решения проблемы несбалансированности контрольной и экспериментальной групп.



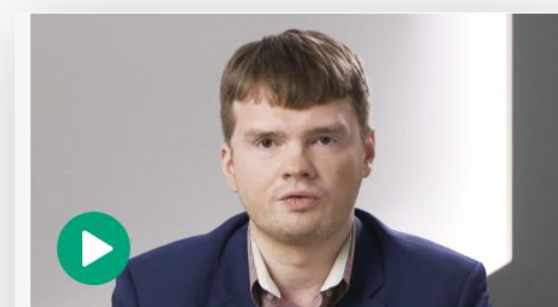
### Лекция «Доказательства с опорой на данные»

Вводная лекция про использование статистических методов, в том числе основанных на дизайнах квази-экспериментов, применяемых в экономической науке для оценки достижения эффекта гос.политики.



### Основано на научных исследованиях: мировой опыт док.политики

Колонка Максима Цыганкова на econs.online с конкретными кейсами применения доказательной политики в мире, и разбором того, что этот подход дал для граждан и государства.



### Спорт на доказательствах: результаты работы лаборатории НИУ ВШЭ

Дмитрий Дагаев о том, как с помощью анализа данных и математического моделирования можно улучшать ситуацию в российском спорте



# Что такое доказательная политика?



## Термин

Evidence-based policy-making  
Evidence-informed policy-making

Принятие решений органом государственной власти на основе доказательств, полученных с помощью научных методов

## Key words

- Государственная интервенция
- Эффекты (outputs-outcomes-impacts)
- Оценка
- Причинность
- Стейкхолдеры и бенефициары

## Вопросы доказательной политики

1. Какое **влияние** оказывает государственная политика на достижение социально-экономических эффектов?
2. Есть ли **причинно-следственная связь** между государственными интервенциями и итоговыми эффектами? Какие государственные интервенции / дизайны политик работают и эффективны?
3. Как реализуется политика и насколько она отражает **ожидания** бенефициаров?
4. Насколько **одинаково понимают** цели и процесс их достижения все участники процесса реализации государственной политики?

# Доказательная политика в государственном управлении

## ЭФФЕКТЫ

1

Создание **условий для сбора релевантных доказательств** и накопления данных об эффектах госполитики

Рост качества и полноты **данных** о государстве

2

Формирование **доказательств о результативности** мер госполитики (программ, проектов) через применение научных методов

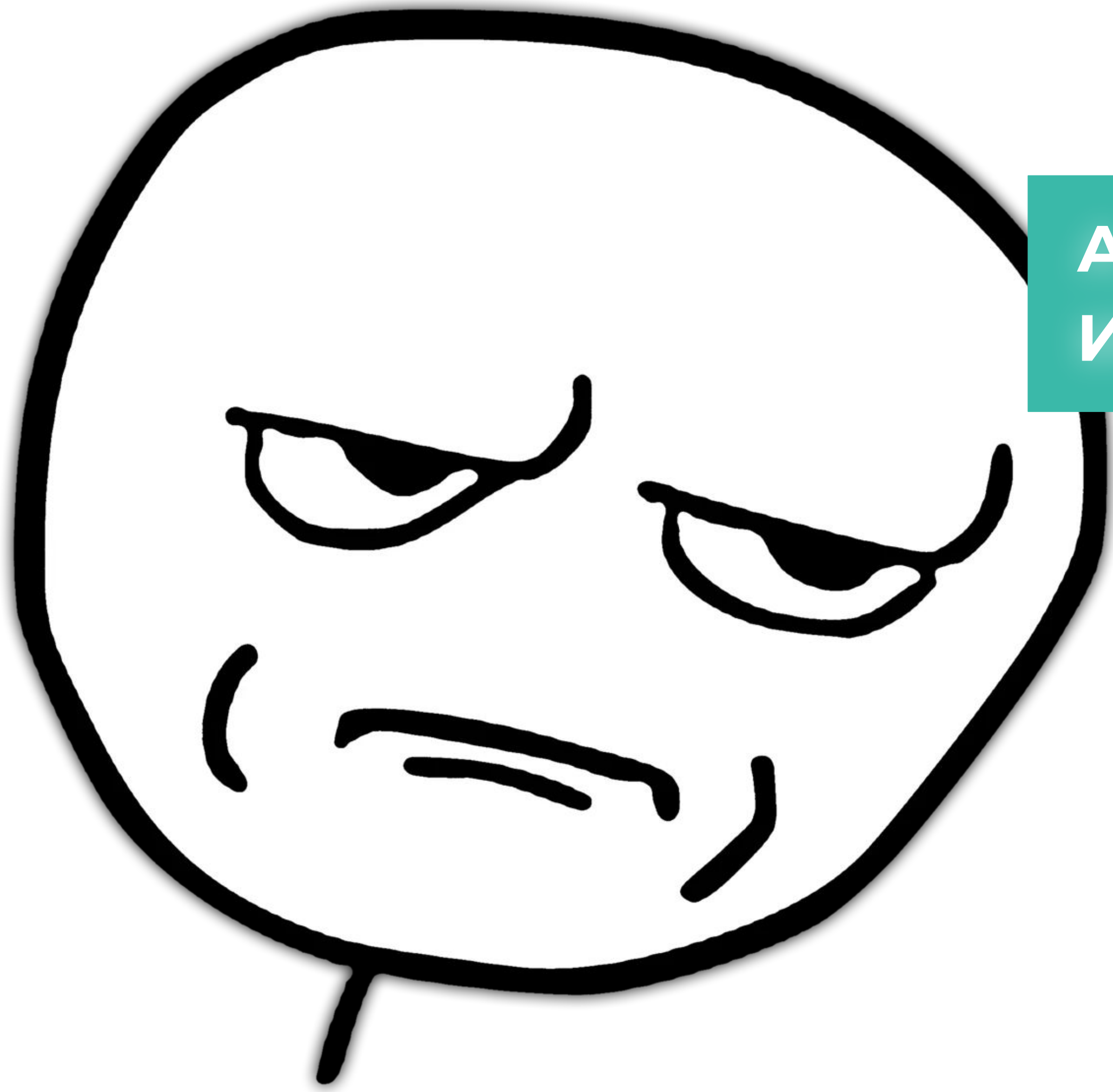
Повышение эффективности **мер госполитики**

3

Принятие **управленческих решений** с использованием доказательств о результативности мер госполитики

Рост обоснованности и прозрачности принятия **государственных решений**





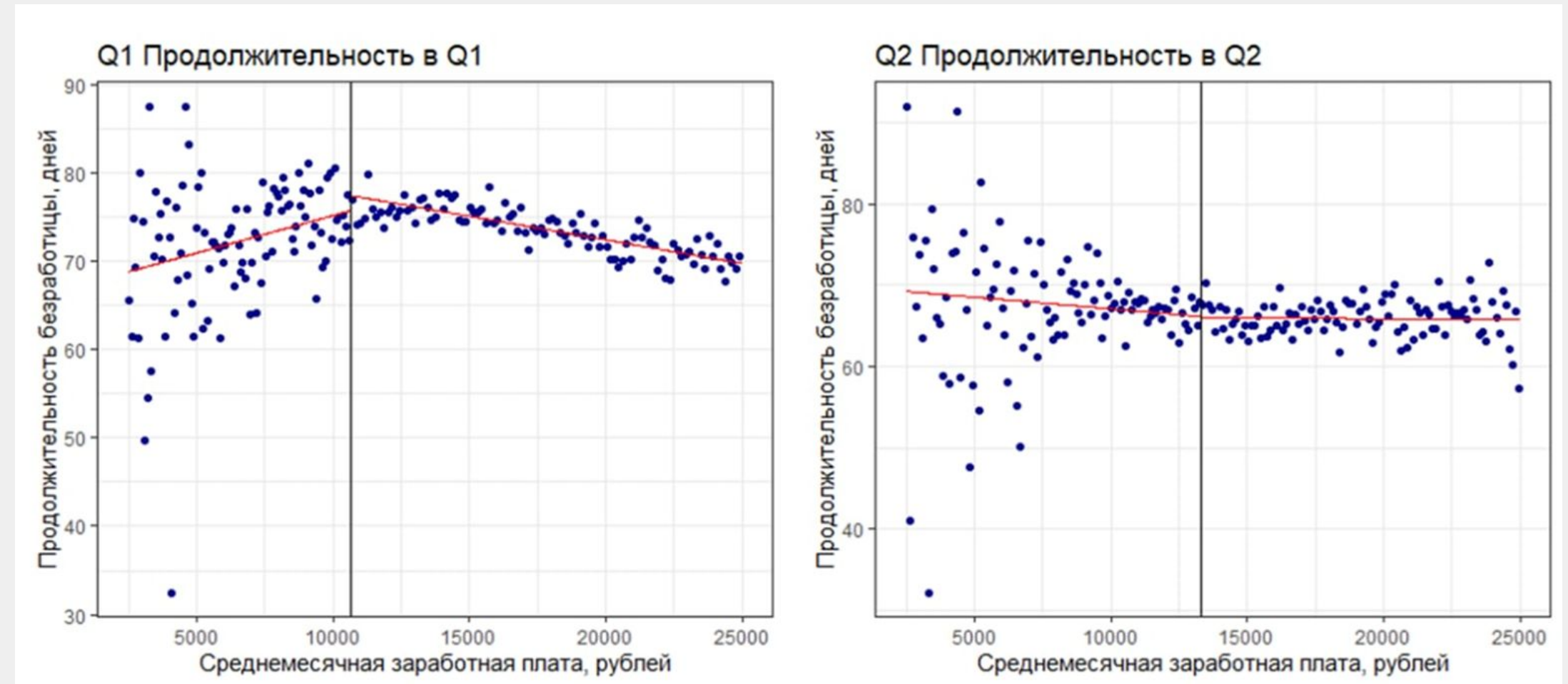
**А ПРИ ЧЕМ ЗДЕСЬ  
ИССЛЕДОВАНИЯ?**

# Оценка политики поддержки безработных

## Данные из каталога ИНИД

[Регистрируемая безработица в России: обезличенные микроданные о характеристиках граждан и полученных услугах за 2017-2021 гг.](#)

[Сведения об обращениях граждан за пособием по безработице](#)



Средняя заработная плата и продолжительность безработицы

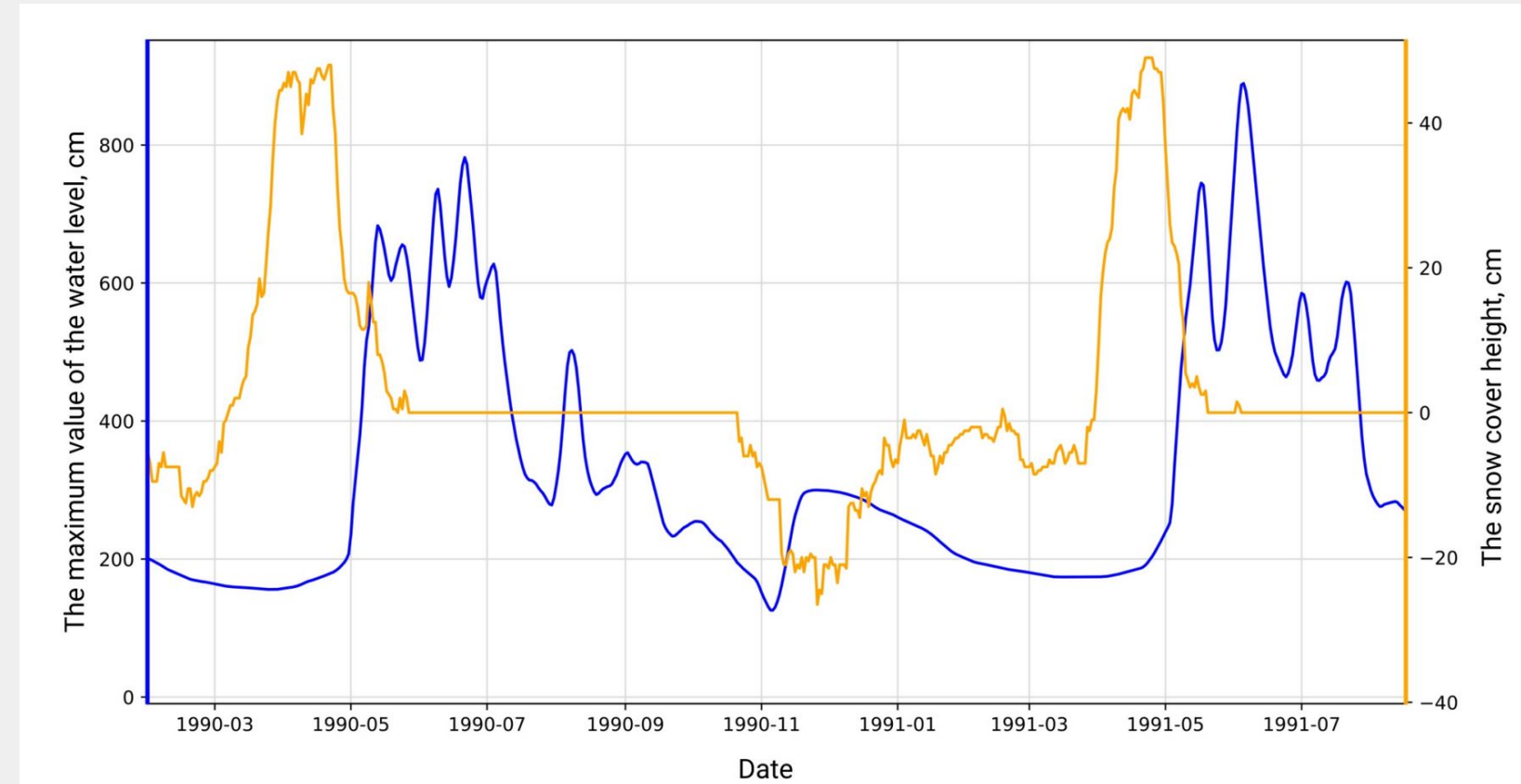
**Что:** оценены эффекты величины пособия по безработице на продолжительность безработицы, вероятность трудоустройства и другие характеристики.



## Метеорология района реки Лена и притоков: месячные, ежедневные и восьмисрочные погодные характеристики за 1985–2020 гг.

### Данные из каталога ИНИД

[Метеорология района реки Лена и притоков: месячные, ежедневные и восьмисрочные погодные характеристики за 1985–2020 гг.](#)



Зависимость между высотой снежного покрова и количеством воды в Лене

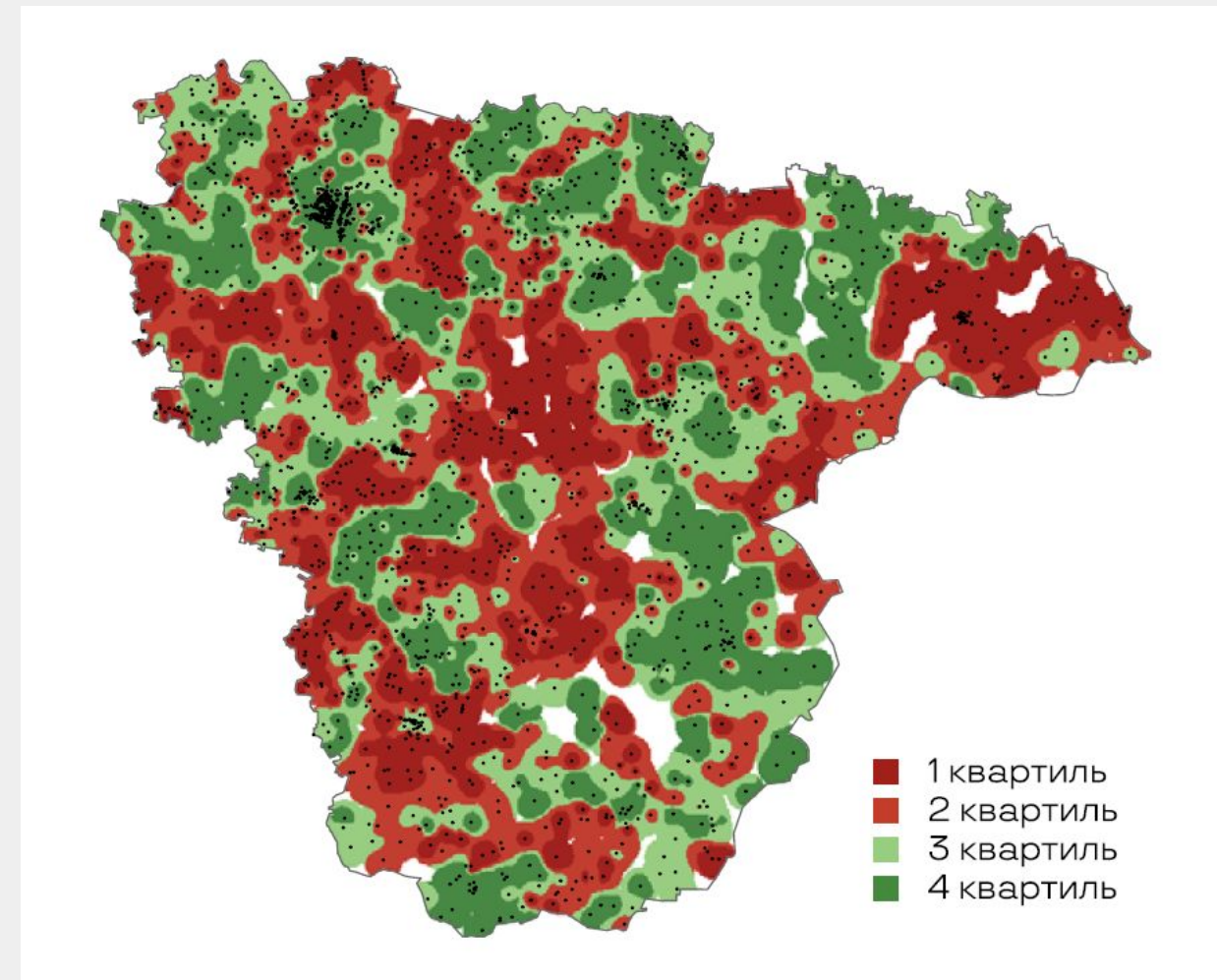
**Что:** представлен гибридный подход к краткосрочному прогнозированию подъёма уровня воды на реках. Он основан на объединении данных из различных источников (метеостанций, датчиков уровня воды, данных дистанционного зондирования). Для повышения эффективности прогнозирования методы машинного обучения и физическая модель SnowmeltRunoff объединены в композитную модель (пайплайн) с использованием методов автоматического машинного обучения.

## Аналитический доклад: Доступность первичной медпомощи в России модель оценки и результаты ее применения на допандемийных данных

### Данные из каталога ИНИД

[Федеральный реестр медицинских организаций;](#)  
[Населенные пункты России:](#)  
[численность населения и географические координаты](#)

[Населённые пункты России:](#)  
[численность населения и географические координаты](#)



Доступность первичной медпомощи в Воронежской области

**Что:** оценена модель пространственной доступности первичного звена здравоохранения. Показано неравномерное распределение медицинской помощи по территории страны, а также ее большая доступность для городов с населением более 100 тыс.



## Данные из каталога ИНИД

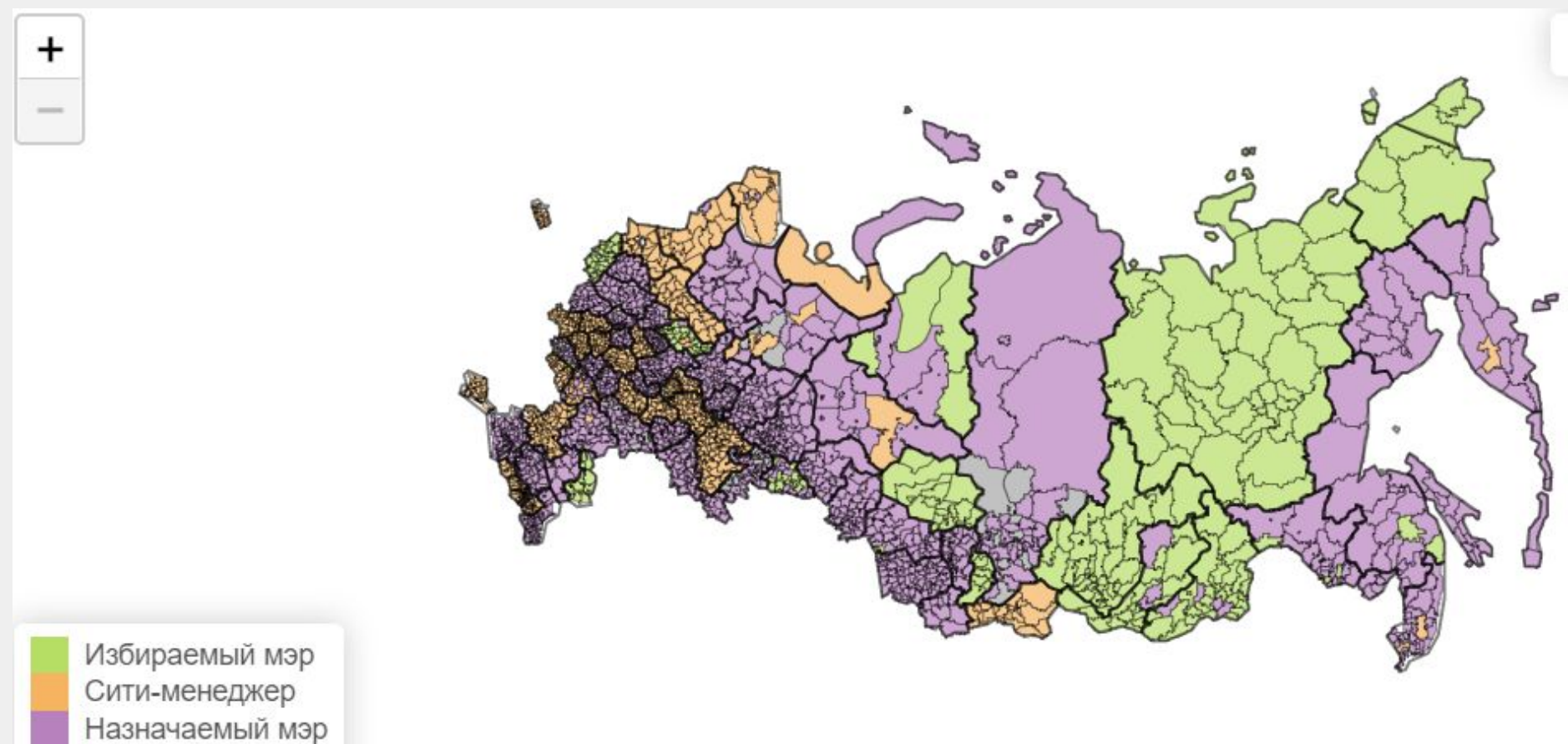
[База данных муниципальных образований Росстата](#) (в начале 2022 — большой апдейт: синхронизация справочников, 2006-2020)

[База моделей управления муниципальными образованиями РФ](#)

[Населенные пункты России:](#) численность населения и географические координаты

[История изменений муниципалитетов в России](#) + библиотеки [susanin](#) и [ovrazhki](#)

## От избрания к назначению. Оценка эффекта смены модели управления муниципалитетами в России



Модели выборы мэров по муниципалитетам

**Что:** с 2003 года в России идет муниципальная реформа: переход от прямых выборов мэра населением к различным непрямым моделям назначения для встраивания муниципалитетов в вертикаль власти с одновременным повышением эффективности их работы.

Выводы:

- Муниципалитеты лишились автономии, но не стали эффективнее;
- Повышение зависимости муниципальных образований от регионального уровня власти;
- Снижение общего объема расходов муниципального образования.

## Данные из каталога ИНИД

[Демография в России:  
объединенные панельные  
данные за 1990-2019 годы](#)

В публичном [репозитории](#) ЦПУР выложен код, с помощью которого можно обработать этот датасет и получить визуализации временных рядов по статистическим показателям.



Коэффициент естественного прироста по регионам России в 2018 году, %

**Что:** ведения по демографическим показателям Росстата, которые были опубликованы в ежегодных статистических сборниках ведомства за период 1990-2019 гг. Включает в себя 101 показатель, охватывающих 4 крупных блока: группировка городов по численности населения; рождаемость, смертность и естественный прирост; браки и разводы; миграция (прибытие, выбытие).

ЗДЕСЬ БУДЕТ АБЗАЦ ПРО ВИЗ



## Данные из каталога ИНИД

[Пожары в России:](#)  
[данные о местах и типах](#)  
[природных пожаров за](#)  
[2012–2021 гг.](#)



**Что:** визуализация к датасету с данными о местах и типах природных пожаров за 2012–2021 гг. в России. На карту нанесены все лесные пожары, случившиеся в России за этот период. Можно заметить, что лесные пожары:

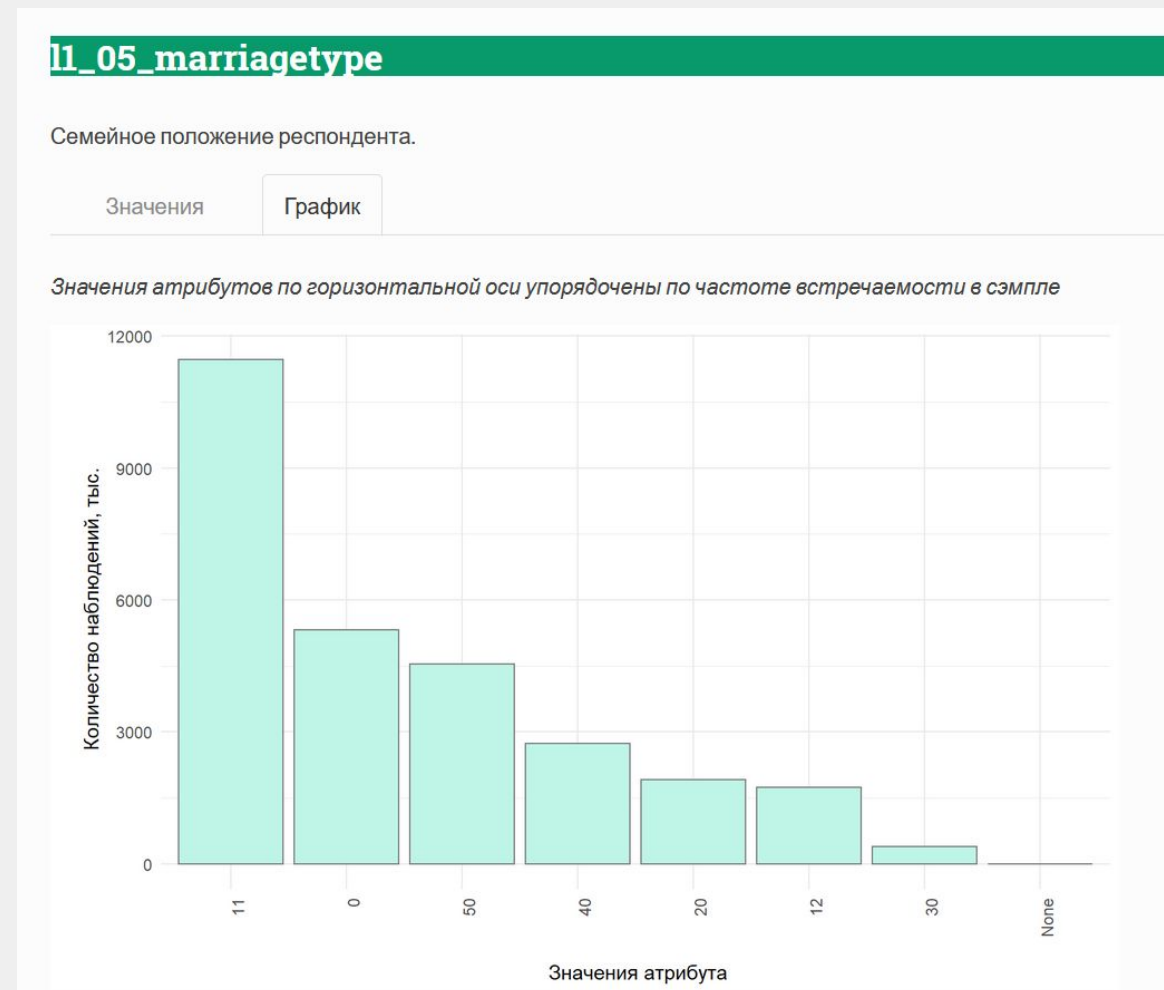
- В большей степени охватывают Сибирь и Дальний Восток;
- Распространены в высоких широтах.



# Микроданные Всероссийской переписи населения

## Данные из каталога ИНИД

Выборки с  
микроданными  
переписей 2002 и 2010  
годов



Распределение индивидов по типу брачного статуса

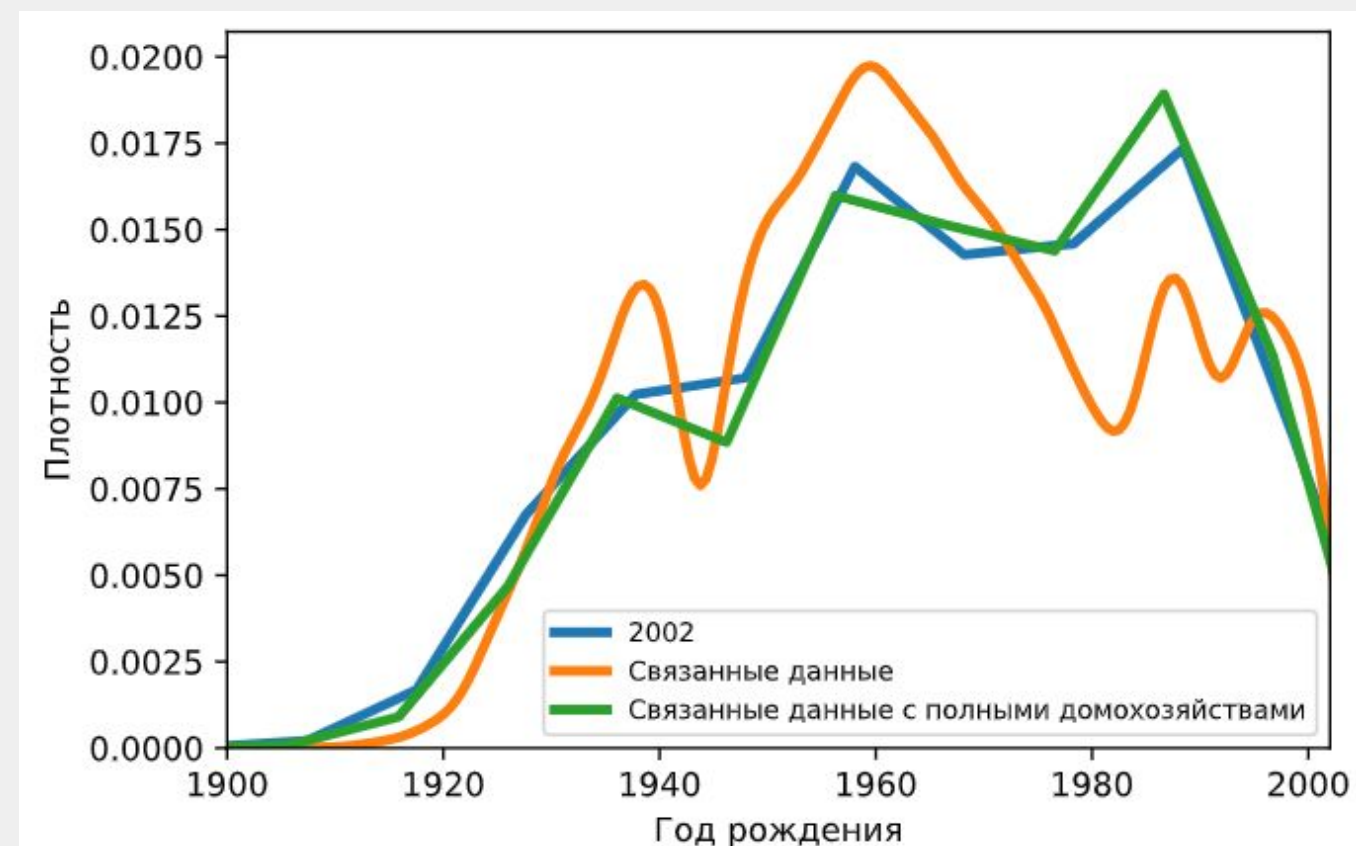
- Что:** подготовлены микроданные (об индивидах домохозяйствах в муниципальном разрезе) за 2002 и 2010 года. Получилось сделать:
- Доступны академическим исследователям после регистрации
  - Практически полный набор данных (внесены только некоторые дополнения и упрощения, связанные с необходимостью анонимизации данных)
  - Расширена информация о родстве
  - Нормальный **кодбук** с описанием переменных и их распределениями

## Данные из каталога ИНИД

Связанные микроданные переписей 2002 и 2010 годов



## Связанные данные Всероссийской переписи населения



Плотность распределения между различными формами связанных данных

**Что:** в мире для подготовки связанных данных основывается на данных об имени, адресе и обязательности заполнения данных; В России перепись **общественно** обязательна и **анонимна**, поэтому приходится прибегать к методам косвенного связывания. Использовались два типа стратегий:

- Вероятностные (допущения, которые позволяют нам идентифицировать);
- Детерминистические (если совпадает какая-то комбинация признаков).

# Susanin: geonorm + ovrzhki

**Susanin** — сервис для упрощения работы с геоданными. Надёжный путеводитель по административно-территориальным и муниципальным данным. Состоит из двух частей

1

## geonorm

Декомпозиция адресов и их привязка к административным геоданным.

Проставляет также информацию о муниципальном районе.

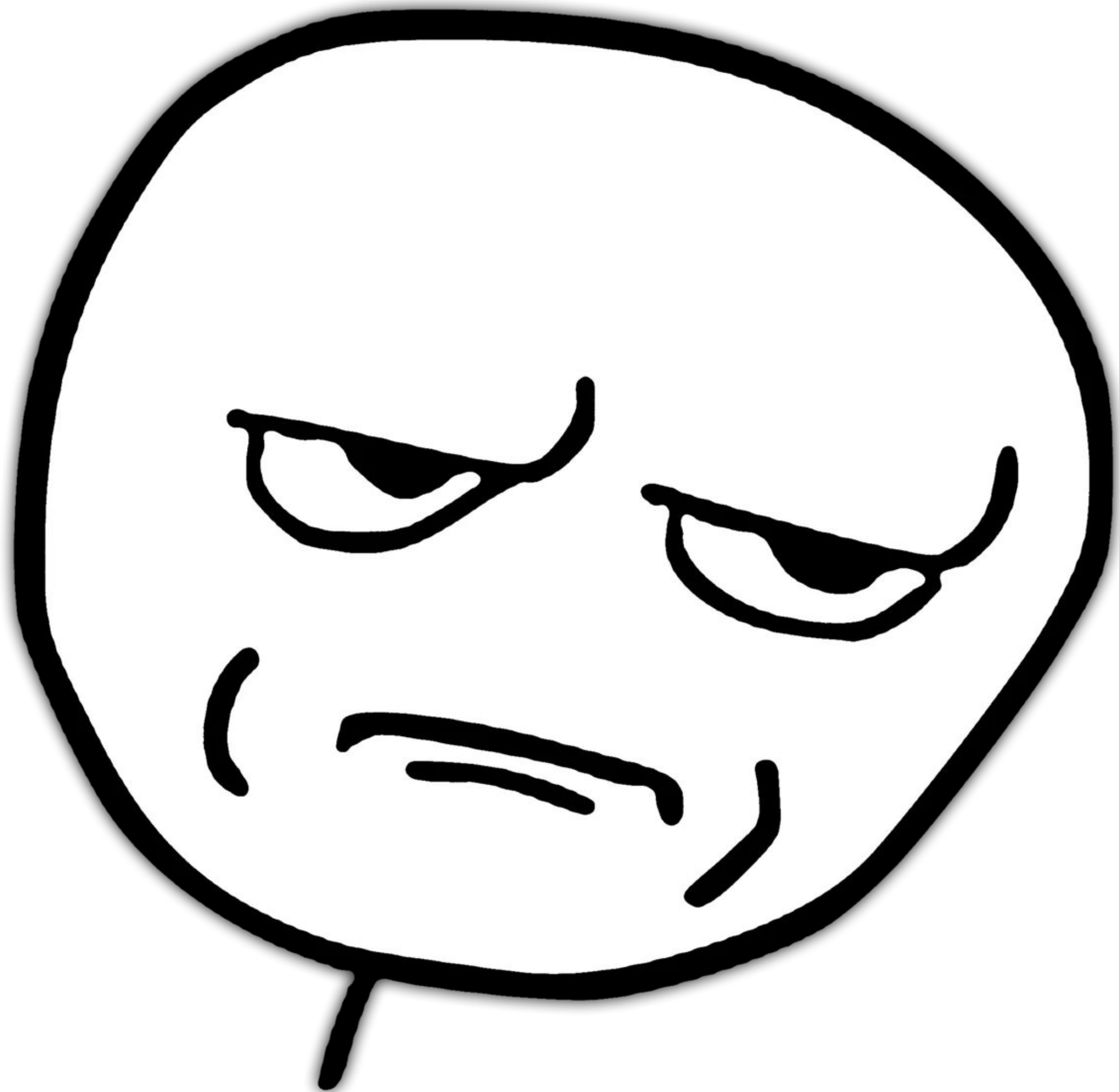
2

## ovrazhki

По паре (муниципалитет, дата) проставляет их ОКТМО и собственный внешний идентификатор, соответствующий составу поселений, входящих в муниципалитет на эту дату.

Ссылка на [GitHub](#)





Поставьте лайки

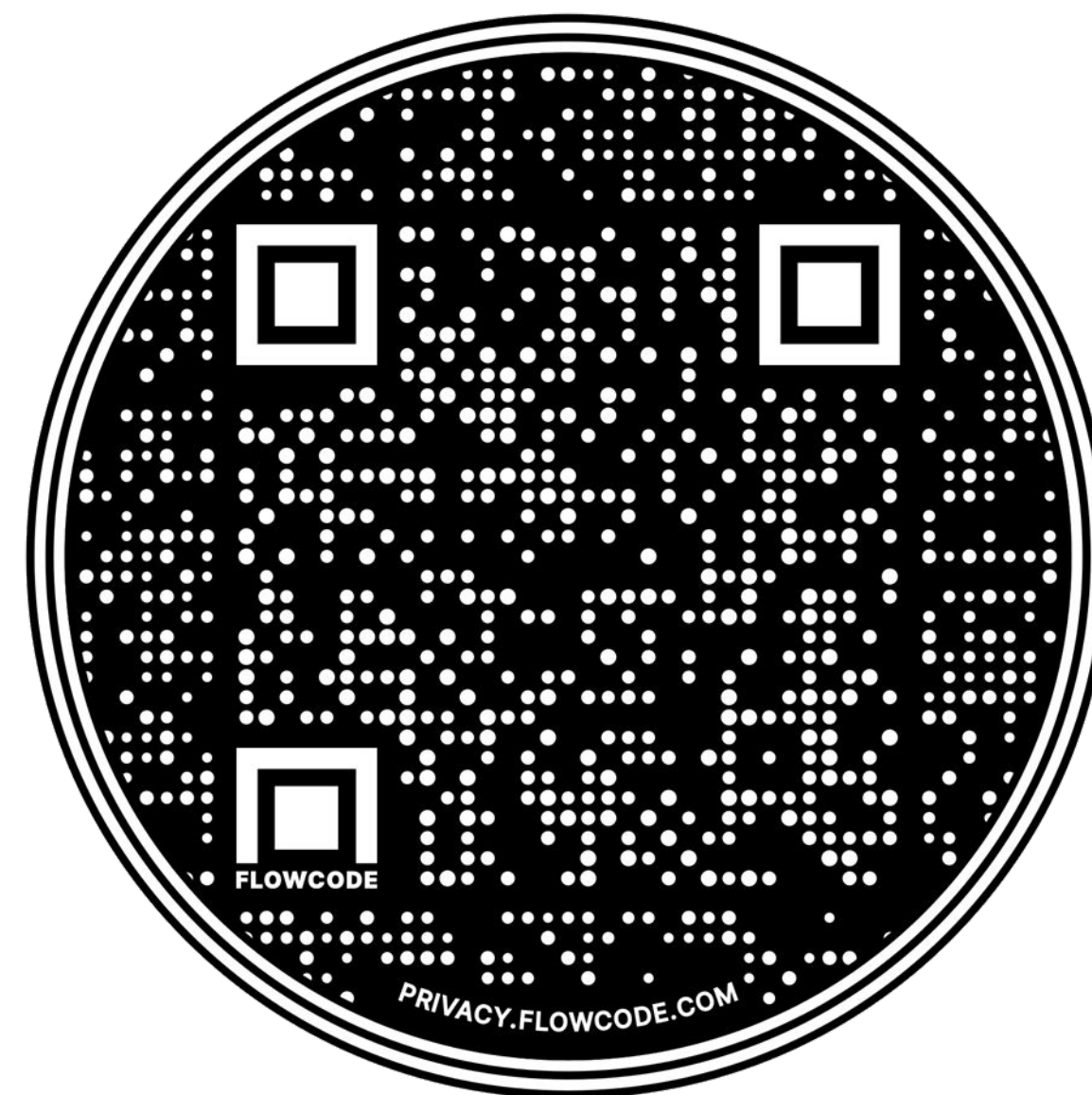
ИНИД на  
YOUTUBE



ИНИД в  
VK



ИНИД в  
Telegram



ТЕЛЕГРАМ-КАНАЛ  
«ДОКАЗАТЕЛЬНЫЙ ПРОБЕЛ»  
<https://t.me/evidencespace>

# Контакты

**Сергей Тихонов**

 @stikhonov  
[s.tikhonov@cpur.ru](mailto:s.tikhonov@cpur.ru)



[Платформа ИНИД](#) / [ВКонтакте](#) / [Telegram](#) / [YouTube](#)