



АНАЛИТИЧЕСКИЙ ЦЕНТР  
ПРИ ПРАВИТЕЛЬСТВЕ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

# Большие данные и официальная статистика



# Содержание

Большие данные и их признаки	3	Портал открытых данных РФ	11
Большие данные и официальная статистика	4	Назначение Портала открытых данных России	12
Основные игроки	5	BigData и аналитика в реальном времени	13
Всемирные конференции	6	Применение по отраслям	14
Ценность и целесообразность	7	Зарубежный опыт	15
Почему сейчас	8	Главные проблемы хранения больших объемов данных	18

# Большие данные и их признаки

**БОЛЬШИЕ ДАННЫЕ** - это изначально не систематизированные данные о социально-экономическом развитии и окружающей среде, генерируемые на основе информационно-коммуникационных технологий вне официальной статистики.

Большие данные характеризуются следующими признаками:

- **большой объем** информации
- **высокая скорость** изменения информации
- **разнообразие** и **разнородность** данных



# Большие данные и официальная статистика

## Примеры источников больших данных

- GPS навигаторы
- Спутниковые данные
- Электронная торговля
- Интернет запросы



Пример официальной статистики

ЕМИСС

<https://fedstat.ru>

## Пользовательские данные

- Данные носимых гаджетов
- Данные социальных сетей

**Данные** - действенный инструмент поддержки принятия государственных решений



Эффективность работы государственного аппарата

## Основные игроки

Проблема взаимодействия официальной статистики и больших данных уже несколько лет серьезно рассматривается в мире. В числе основных игроков (ведущих мировых центров) стоит выделить следующие:



Статистическая  
комиссия ООН



Глобальная рабочая группа по  
большим данным в  
официальной статистике



Проект ООН  
«Глобальный  
пульс»

# Всемирные конференции

Всемирные конференции по использованию больших данных в официальной статистике отражают признание необходимости выработки общих подходов и рассмотрения проблем больших данных.



**I-я Международная конференция** по использованию больших данных в официальной статистике состоялась в г. Пекине (КНР) 28-30 октября 2014 года.



**II Глобальная международная конференция** по использованию больших данных в официальной статистике прошла в г. Абу-Даби (ОАЭ) с 20 по 22 октября 2015 года.



**III Глобальная международная конференция** по использованию больших данных в официальной статистике состоится в г. Дублин (Ирландия) с 30 августа по 1 сентября 2016 года.



# Ценность и целесообразность

Стратегическую ценность больших данных для официальной статистики определяют следующие их характеристики:

- **большой объем данных** может способствовать получению более детализированных сведений по конкретным направлениям статистики;
- **высокая скорость изменения информации** способствует увеличению частоты статистических оценок;
- **разнообразие больших данных** может открыть возможности для производства статистики в новых областях, получения информации о не наблюдаемых официальной статистикой сферах;
- **разнородность происхождения больших данных** может способствовать многоаспектности измерения и, тем самым, повышению надежности статистических оценок.

Применение больших данных в официальной статистике сопряжено с решением проблем обеспечения конфиденциальности, финансирования работ, а также наличия специалистов и аналитиков



# Почему сейчас

Динамичное проникновение телекоммуникаций в частную жизнь и бизнес



*\*Источник данных: Омнибус GfK, 2015, вся Россия 16+*



# Почему сейчас

## ПРОНИКНОВЕНИЕ ИНТЕРНЕТА В РОССИИ

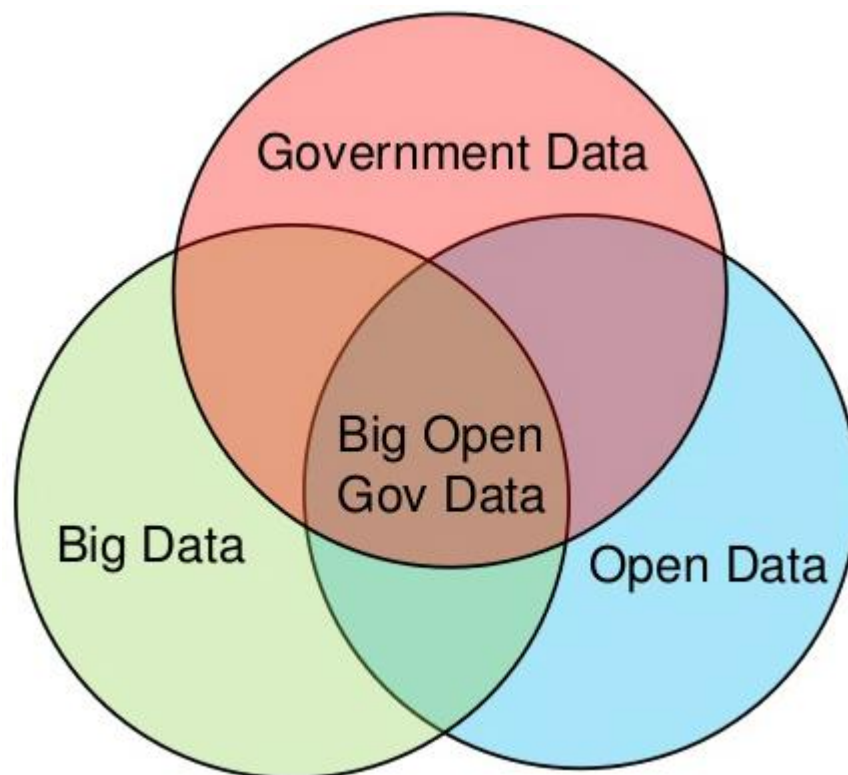
**84** миллиона человек в возрасте от 16 лет и старше являются пользователями Интернета в России



*\*Источник данных: Омнибус GfK, 2015, вся Россия 16+*

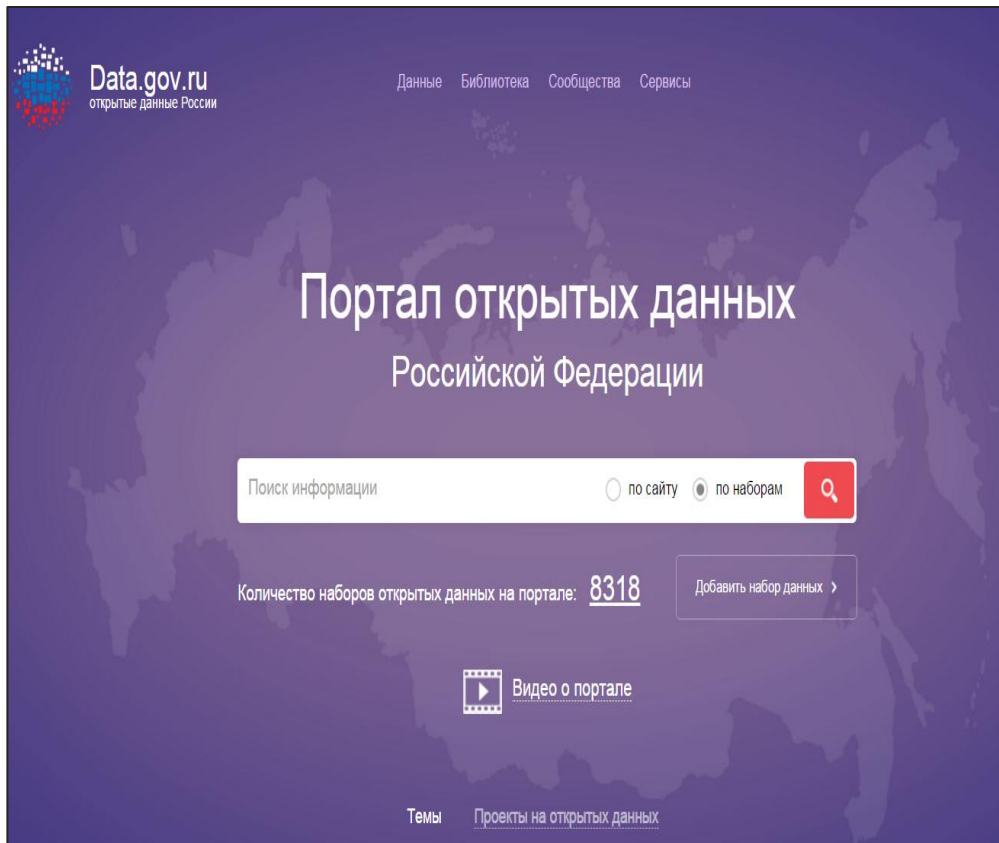
# Большие данные и открытые данные

Открытые данные - неотъемлемая часть больших данных

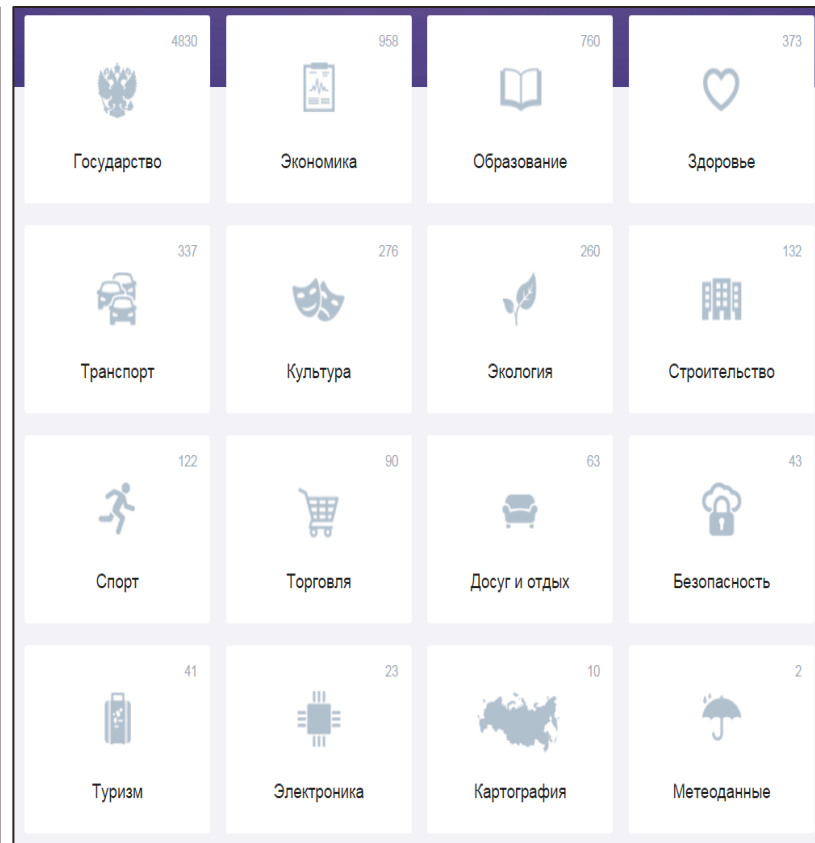


# Портал открытых данных РФ

Главный агрегатор структурированных государственных данных



The screenshot shows the homepage of Data.gov.ru. At the top left is the logo with the text "Data.gov.ru" and "открытые данные России". Navigation links include "Данные", "Библиотека", "Сообщества", and "Сервисы". The main heading reads "Портал открытых данных Российской Федерации". A search bar contains the text "Поиск информации" and has radio buttons for "по сайту" and "по наборам", with a red search icon. Below the search bar, it states "Количество наборов открытых данных на портале: 8318" and a "Добавить набор данных" button. At the bottom, there is a "Видео о портале" link with a play icon and "Темы" and "Проекты на открытых данных" links.




4830 Государство	958 Экономика	760 Образование	373 Здоровье
337 Транспорт	276 Культура	260 Экология	132 Строительство
122 Спорт	90 Торговля	63 Досуг и отдых	43 Безопасность
41 Туризм	23 Электроника	10 Картография	2 Методанные

A grid of 16 data categories, each with an icon and a count. The categories are: Государство (4830), Экономика (958), Образование (760), Здоровье (373), Транспорт (337), Культура (276), Экология (260), Строительство (132), Спорт (122), Торговля (90), Досуг и отдых (63), Безопасность (43), Туризм (41), Электроника (23), Картография (10), and Методанные (2).

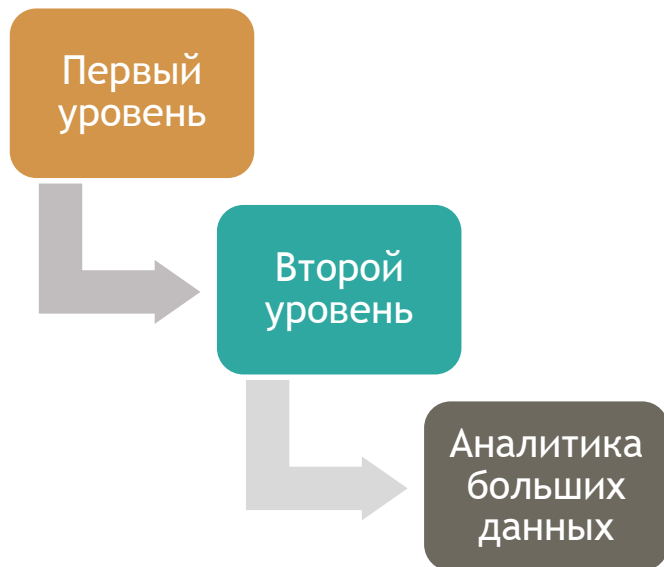
# Назначение Портала открытых данных России

Ключевым назначением портала открытых данных России является обеспечение централизованного доступа к информационным ресурсам, представленным в форме открытых данных

Целями создания и использования портала открытых данных являются:

- Предоставление на едином информационном ресурсе наиболее полных перечней и сведений о наборах открытых данных, публикуемых органами государственной власти и органами местного самоуправления Российской Федерации, а также иных востребованных наборах открытых данных
  - Создание информационной площадки для взаимодействия с широкой общественностью по вопросам формирования, публикации и использования открытых государственных данных, включая тематическое информирование и получение обратной связи по их качеству и востребованности
  - Формирование и реализация единой технологической политики в области открытых государственных данных
- 

# BigData и аналитика в реальном времени



Специалисты утверждают, что в 2016 году мы познакомимся с новой главой истории развития **аналитики «больших данных»** по мере развития двухуровневой модели обработки.

## Опыт Dunkin` Donuts и Big Data в реальном времени

По кассовым чекам компания получает данные, какие именно предложения получили наибольший отклик у покупателей. Данный подход обработки данных позволил **увеличить прибыль и оборачиваемость товаров** на складе.



# Применение по отраслям


**В энергетике** - повышение точности распределения мощностей генераторов

**В горнодобывающей и нефтяной промышленности** - эффективность разработки месторождения, отслеживание графика капитального ремонта и состояния оборудования, прогнозирование спроса на продукцию и цены

**Правоохранительные органы** - система ALPR (автоматическое распознавание номерных знаков), карта преступности



# В здравоохранении

- Smart технологии
  - Предрасположенность к заболеваниям по анализу крови пациентов и доноров
  - Снижение затрат государства и повышения эффективности лечения
  - Борьба с заболеваниями и отслеживания роста эпидемий
  - Предсказания потенциально возможных эпидемий
- 


# В логистике и транспорте

- Отслеживание парка автомобилей
- Учет расходов на топливо
- Мониторинг заявок клиентов

## Опыт РЖД


- Сокращение срока подготовки отчетности
- Повышение точности распределения затрат

## Опыт ФНС России


- Создание аналитического слоя федерального хранилища данных
  - Единое информационное пространство и технология доступа к налоговым данным
- 



## Зарубежный опыт

- **Австралия** - Big Data аналитика используются для устранения последствий чрезвычайных ситуаций
  - **Испания** - сервис Madrid iNTeligente (Умный Мадрид)
  - **Канада** - Big Data используется для развития здравоохранения
  - **Япония** - министерство внутренних дел планирует с апреля 2016 года начать применять большие данные для своевременного предоставления информации об эвакуации во время стихийных бедствий
  - **Китай** - создание smart-городов
- 

# Главные проблемы хранения **больших объемов данных**

- Безопасность вычисления в распределенных программных системах
  - Безопасность нереляционных баз данных
  - Безопасное хранение данных
  - Проверка достоверности
  - Мониторинг безопасности в режиме реального времени
  - Шифрование управления доступом и обеспечение безопасности соединений
  - Фрагментарный контроль доступа
  - Происхождение данных
- 

Благодарю за внимание

