

Открытые данные в государственном управлении в России и в мире

Иван Бегтин

Директор АНО “Информационная культура”

Председатель Ассоциации участников рынка данных

Кто я и почему об ЭТОМ рассказываю?

1. Много лет
проектировал
корпоративные и
государственные
data-rich проекты



2. Возглавляю АНО
Инфокультура.
Развиваю
открытые данные в
России

3. Веду блог/канал в телеграм
<https://t.me/begtin> посвящённый
работе с данными

4. Развиваю
технологический
стартап
APICrafter/DataCrafter
и общественный проект
Национальный цифровой архив
(ruarxive.org)

Largest Companies by Market Cap

companies: **4,755** total market cap: **\$85.664 T**

Rank by

Market Cap

Earnings

Revenue

Employees

Market Cap gain

Market Cap loss

Rank	Name	Market Cap	Price Today	Price (30 days)	Country
1	 Apple AAPL	\$2.057 T	\$122.54 -0.69%		us USA
2	 Saudi Aramco 2222.SR	\$1.872 T	\$9.36 -1.27%		sa S. Arabia
3	 Microsoft MSFT	\$1.792 T	\$237.58 0.67%		us USA
4	 Amazon AMZN	\$1.580 T	\$3,138 0.86%		us USA
5	 Alphabet (Google) GOOG	\$1.380 T	\$2,053 0.70%		us USA
6	 Facebook FB	\$827.61 B	\$290.63 -0.99%		us USA
7	 Tencent TCEHY	\$764.50 B	\$80.93 -2.22%		cn China
8	 Alibaba BABA	\$642.63 B	\$237.67 0.23%		cn China

Тезисы

- 1. Вся экономика цифровая**
- 2. Государство “догоняющий игрок”**
- 3. Датацентричность**
- 4. Новые риски**
- 5. Новые модели распределения полномочий**

**Из чего состоит работа с
данными в госсекторе**

С чего начинать

1. **Цели и задачи** и чёткое понимание того для чего данные необходимы. Это может быть госуслуга, система поддержки принятия решений, “умные системы” ИИ, перестройка отраслевых аналитических систем.
2. **Инвентаризация данных** и понимание того какие данные есть во внутреннем управлении и во внешних источниках.
3. **Технологии работы** с данными по сбору, обработке и созданию хранилищ и озер данных,
4. **Квалифицированные кадры** чтобы сделать всё выше обозначенное.

Стратегии работы с данными государственного сектора

Почему разработка стратегий это важно

- Использовать новые возможности для госуправления
- Поддержать инновационные компании
- Определить общие правила развития рынка данных
- Успокоить граждан (этика и мораль использования ИИ)
- Учесть новые риски
- Соответствовать новым ожиданиям граждан и бизнеса

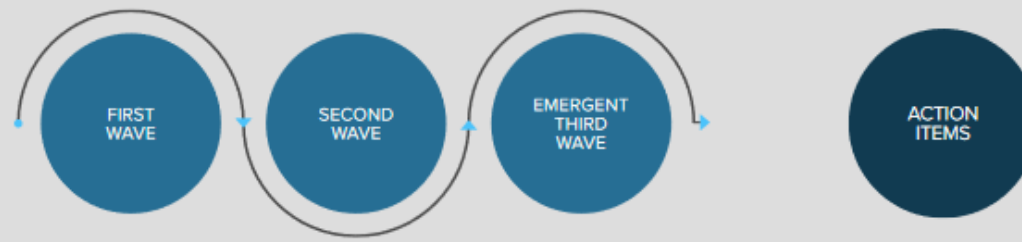
На какие вопросы мы ищем ответы:

- Что у нас есть?
- Того, что есть, достаточно?
- Собираем правильно?
- А имеем право собирать?
- Этические последствия понимаем?
- Кто у нас этим занимается?
- А они квалифицированы?
- Как находить и готовить специалистов?
- Какие технологии применять?
- А как это регулировать?

Третья волна открытых данных

Что такое третья волна открытых данных?

- Новый подход к работе с открытыми данными
- Акцент на партнерство
- От открытости по умолчанию к публикации под заданные цели
- Развитие до регионального уровня



Concept:	Freedom of Information	Open Government Data	Re-use of Public and Private Data	▶ Foster institutional data capacity ▶ Increase Data Liquidity
Value Proposition:	Transparency	Transparency and Problem Solving	▶ Evidence based policy making ▶ Innovation and Entrepreneurship	▶ Articulate value and build an evidence base of the impact of data re-use for public interest purposes
Approach:	Data on Request (Right to Know)	Open by Default (Right to Share)	Publish with Purpose	▶ Create new data intermediaries ▶ Invest in a Data Demand Identification and Assessment Methodology ▶ Increase accessibility and findability of high-value datasets
Focus:	Pull Focused	Push Focused	Partnerships (Data Collaboratives)	▶ Establish governance frameworks and seek regulatory clarity ▶ Document and experiment with the number of operational models that are "fit for purpose" ▶ Develop common data sharing agreements and new licensing regimes
Geographical Emphasis	National	International and National	▶ Sub-national and Local ▶ Purpose-driven Cross-border Data Flow	▶ Create the technical infrastructure for re-use at the local level ▶ Invest in Subnational Capacity, Guidance, Legal Frameworks and Best Practice
Audience/ Demand:	▶ Journalists ▶ Lawyers and Activists	▶ Civic technologists and "Data Geeks" ▶ Government agencies ▶ Corporations, technology start-ups	▶ Community-based organizations ▶ NGOs, Human Rights Organizations, and Social Justice Advocates ▶ Academia and Research Institutions ▶ Small businesses and startups ▶ Government	▶ Bolster public competence ▶ Develop a data demand assessment and segmentation methodologies ▶ Strengthen accountability mechanisms to support rights-based data re-use.
Risks and Policies:	Secrecy and obfuscation	Privacy — Mosaic effect, demographically identifiable information (DII)	Data Responsibility and Data Rights Framework	▶ Track, monitor, and clarify decision and data provenance ▶ Invest in Data Responsibility Frameworks and Technologies ▶ Engage the public, especially vulnerable communities, to document public perceptions and opinions on data re-use
Institutional Responses:	Information Auditors	▶ Data Officer ▶ Open Data Portals	▶ Chief Data Steward ▶ Intermediaries	Create and empower (chief) data stewards

Ситуация в России с открытыми данными

GOVERNMENTS AND GROUPS	TOTAL SCORE (out of 100)	TOTAL SCORE CHANGE (since 1st Ed.)	READINESS (out of 100)	IMPLEMENTATION (out of 100)
Canada	76	18	86	87
UK	76	-4	83	89
Australia	75	17	79	84
France	72	17	84	77
South Korea	72	25	82	67
Mexico	69	33	79	67
Japan	68	24	78	68
New Zealand	68	5	79	72
USA	64	-11	79	76
Germany	58	2	76	72
Uruguay	56	23	71	70
Colombia	52	25	69	60
Russia	51	10	62	59
Brazil	50	15	63	56
Italy	50	8	61	61
India	48	16	64	49
Argentina	47	14	66	56
Ukraine	47	25	60	52
Philippines	42	19	54	42
Chile	40	2	54	55

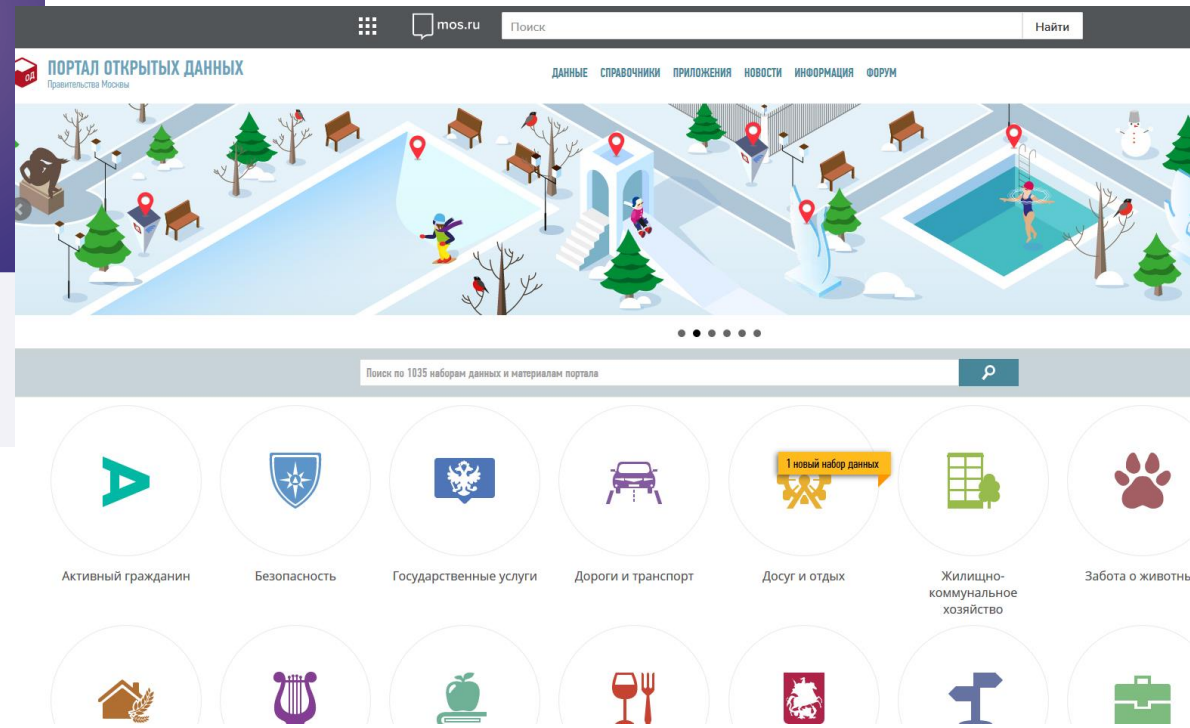
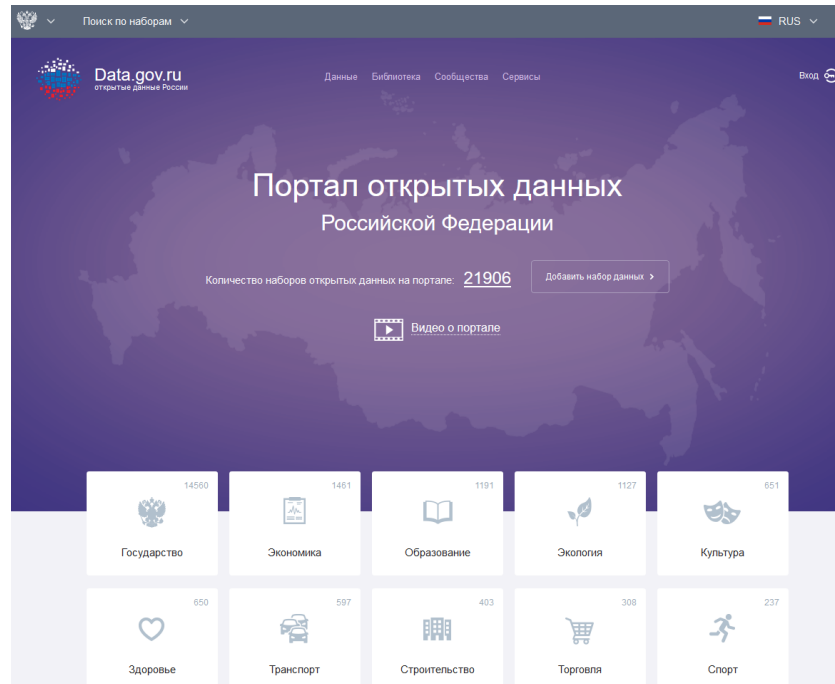
Open Data Barometer 2017

Россия страна со “средними” показателями. С сильной позицией в госфинансах и слабой позиции во всём что касается качества жизни.

Где искать открытые государственные данные?

- Порталы открытых данных
- Государственные информационные системы
- Государственные информационные ресурсы
- Официальные сайты

Порталы открытых данных



Итоговая ситуация

1. **Нет ответственного** органа власти и высшего должностного лица.
2. **Данные продолжают публиковаться**, хотя и в худшем качестве.
3. **Государственные информационные системы** содержат большую часть данных собираемых государством, некоторые из них открытых.
4. **Разнонаправленные тренды** открытия и закрытия данных.

Текущая ситуация с регулированием данных в России

Государственная политика работы с данными

Сильные стороны

Национальная стратегия ИИ

Слабые стороны

Отсутствие национальной стратегии данных
Застой в федеральном портале открытых данных
Сужение задач в Национальной системе управления управления данными (НСУД)

S

W

O

T

Возможности

Разработка национальной стратегии данных
Перезапуск обмена лучшими практиками с другими странами

Угрозы

Значительная самоизоляция от мировых процессов
Влиятельные “ цари-данных” не желающие предоставлять к ним доступ

Реальная ситуация

- Нет стратегий управления данными, ни на уровне регулирования, ни на уровне госуправления
- Низкое качество данных ГИСах
- Недокументированные программные интерфейсы
- Утечки персональных данных (включая легальные!)

Ключевые инициативы

1. **Начало инвентаризации данных** в государственных информационных системах с конца 2020 года по инициативе Минцифры России.
2. **НСУД** развивается для поддержки цифрового профиля, портала госуслуг и других чётко определённых задач.
3. **Ведомственные инициативы** по созданию хранилищ и озер данных.
4. **Массовое обучение CDTO и CDO** на уровне базовых цифровых навыков необходимых государственным служащим.

Управление [мета]данными

Почему метаданные это важно

- Data Scientist'ы не любят не творческую работу: собирать, чистить и документировать данные
- ИТ команды регулярно меняются, ротация в 2-3 года
- ИТ системы регулярно меняются, от 2 до 5 лет
- невозможно найти того кто понимает как устроен тот или иной набор исторических данных
- новые специалисты долго погружаются в работу
- данные должны соответствовать требованиям регуляторов (комплаенс)

DMS – [Meta]Data Management Systems

Открытые

- Hortonworks Atlas
- DataHub
- Amundsen
- Marquez
- Uber Databook
- Airbnb's Data Portal
- Metacat
- Metamapper

Коммерческие

- Atlan
- Colibra
- Alation
- Redgate

Облачные

- Aristotle
- Google Cloud
- Azure Data Catalog
- Data.World
- Data Catalog Oracle Cloud Infrastructure (OCI)

Как происходит инвентаризация данных

По шагам (1/3)

1. Сформулировать цели и задачи инвентаризации
2. Определить один из подходов:
 - изучаем чем владеем
 - ищем где найти нужное у других

По шагам (2/3)

3. Провести инвентаризацию информационных ресурсов

- анализ полномочий
- опросные листы
- порталы открытых данных
- реестры информационных систем
- каталоги веб-сайтов

По шагам (3/3)

4. Запросить консультации у экспертов/ассоциаций
5. Сопоставить цели и задачи с имеющимися данными
6. Спроектировать обновление систем от задач.
7. Определиться с итоговым результатом (каталог, целевая система, озеро данных)

В качестве завершения. С чего начинать

1. **Цели и задачи** и чёткое понимание того для чего данные необходимы. Это может быть госуслуга, система поддержки принятия решений, “умные системы” ИИ, перестройка отраслевых аналитических систем.
2. **Инвентаризация данных** и понимание того какие данные есть во внутреннем управлении и во внешних источниках.
3. **Технологии работы** с данными по сбору, обработке и созданию хранилищ и озер данных,
4. **Квалифицированные кадры** чтобы сделать всё выше обозначенное.

Спасибо за внимание!

Иван Бегтин, директор
АНО «Инфокультура»

Эл. почта: ibegtin@infoculture.ru

Сайт: <https://www.infoculture.ru/>