



ИГОРЬ АШМАНОВ

БОЛЬШИЕ ПОЛЬЗОВАТЕЛЬСКИЕ ДАННЫЕ

ПОДГОТОВЛЕНО
«АШМАНОВ И ПАРТНЕРЫ»
НОЯБРЬ 16



БОЛЬШИЕ ПОЛЬЗОВАТЕЛЬСКИЕ ДАННЫЕ — НОВАЯ РЕАЛЬНОСТЬ

- Накоплены большие объёмы БПД и технологии обработки
- Возник рынок, в условиях правового вакуума
- Происходит массовая утечка БПД за рубеж
- Есть риски и возможности (в том числе риск упустить направление)
- Есть поручение Президента №10 от 29.01.2016 разобраться, нужно создать дорожную карту.



ЧТО ТАКОЕ «БОЛЬШИЕ ДАННЫЕ»

- Модное, «шумовое» слово
- Не научное, а скорее медийное понятие
- Не новое понятие (ему примерно 30 лет, раньше называлось Data Mining и т.п.)
- Не алгоритмы обработки, а сами данные
- Большие массивы данных со специфическими свойствами



«БОЛЬШИЕ ДАННЫЕ» — ОПРЕДЕЛЕНИЕ

- Большие (нельзя обработать вручную)
- **Целое больше суммы частей** (сверху видны закономерности)
- Данные из разных каналов (соцсети, поисковики, камеры, телефоны)
- Данные разного формата (текст, связи, картинки, видео, голос)
- С ретроспективой (накопление признаков и маркеров)
- Данные о людях



ЧТО ТАКОЕ БПД

Неформальное определение:

Большие пользовательские данные — это данные, собираемые в сети различными сервисами, устройствами и приложениями о поведении пользователей, понимаемых как пользовательские профили.

- Данные о профилях пользователей
- Бывают первичные (сырые) и производные (вычисляемые)
- Не технология, а данные



КАТЕГОРИИ БПД С ТОЧКИ ЗРЕНИЯ ПОЛЬЗОВАТЕЛЯ

- **Явно отдаваемые.** Адреса почты, посты/комментарии, профили, фото, лайки и т.п. «Регулируются» пользовательскими соглашениями.
- **Ограниченного распространения.** Переписка, закрытые обсуждения, география, биометрия.
- **Косвенные данные.** Поведенческие, география, биометрия, фото, видео, привычки, взгляды, граф связей.
- **Воруемые данные.** Фото, видео, СМС, голос, клавиатурный ввод, пароли, переписка (смартфоны, смарт-ТВ, keyloggers, трояны)



ЧЕМ ОТЛИЧАЮТСЯ ОТ ПД

Персональные данные

- **Идентифицирующие.** ФИО, номер кредитки, телефон, адрес, прочее.
- **Есть осознание проблемы** государством и обществом, есть процедуры хранения и передачи.
- **Собираются операторами ПД,** с уведомлением пользователя.
- **Регулируются:** 152-ФЗ, проверки Роскомнадзора.

БПД

- **Анонимные / агрегированные / сырые.** Вычисления отделены от сбора, сырые данные от обработанных
- **Позволяют вычислить ПД:** правило 4 точек в разных сферах, сличение векторов из разных каналов
- **Собираются и продаются всеми:** реклама, скоринг, коллекторы
- **Не регулируются.** Нет осознания проблемы обществом и государством



КТО СОБИРАЕТ БПД

- **Навигаторы:** поисковики и браузеры (запросы, интересы, клики, посещения третьих сайтов, география)
- **Счётчики** разных типов (до 50 на странице, клики, география, интересы, запросы)
- **Рекламные сети** (интересы, клики, география)
- **Коммуникационные сервисы:** социальные сети, публичные почты, мессенджеры (соцдем, адреса, фото, интересы, граф связей)
- **Платёжные системы:** финансовая и экономическая информация
- **Платформы:** ОС, провайдеры, платёжные системы (запросы, видео, аудио, тексты, сообщения, аккаунты, пароли)



НОВЫЕ СУЩНОСТИ

- **Смартфоны и приложения** (тексты, СМС, снимки, геоположение, контакты, ...)
- **Смарт-ТВ** отправляют поток данных из квартиры (видео, жесты, трёхмерные сцены, звук, потребительское поведение).
- **Носимые устройства, фитнес-браслеты** — изначально «в облаке» (биометрия, медицинское состояние, перемещения)
- **Анализ ДНК** — данные о национальности, родственниках, болезнях
- **Автомобили** (видео, звук, география, поведенческие данные)
- **«Интернет вещей»** — данные об окружении, устройствах, производстве, ремонте



КАКИЕ ДАННЫЕ СОБИРАЮТСЯ

- **Поведенческие:** поисковые запросы, клики, посещения сайтов, география, перемещения, точки пребывания
- **Социальные:** посты, комменты, соцдем, фото, видео, политические взгляды, работа, привычки, хобби
- **Биологические:** фото, видео, ДНК, отпечатки пальцев, пульс, давление, физическая активность, сигнатура голоса и т.п.
- **Финансово-экономические:** счета, активы, покупки, транзакции, уровень доходов, активы
- **Личные, интимные:** семья, связи, круг общения, привычки, половая ориентация, развлечения, психология, взгляды



МЕТОДЫ СБОРА И АНАЛИЗА

- **Анализ по всей линейке сервисов:** поиск, новости, рекламная система, почта, социальная сеть, карта, мессенджер, платёжная система, браузер, фотохостинг, счётчик, ...
- **Кросс-канальный анализ:** сравнение векторов с разных источников, правило 4 точек для разных сфер, сведение времён выхода в сеть и постов в соцсетях и т.п.
- **Кросс-платформенный анализ** — сведение воедино разных устройств пользователя (отпечатки браузера, ультразвук и т.п.)



РИСКИ БПД

- **Слежка**, утечка данных граждан РФ к частным компаниям и за рубеж. Большинство платформ — американские
- **Нарушение приватности**, продажа пользовательских данных, навязчивая реклама, раскрытие интимных данных
- **Влияние и пропаганда**, персональное и массовое информационное давление, создание «информационных пузырей»



ЭТИЧЕСКИЕ И ПРАВОВЫЕ АСПЕКТЫ

- Проблема информированности пользователей
- Проблема безопасности
- Коммерческий интерес собирателей БПД
- Проблема «наблюдающих сервисов»
- Отсутствие этического кодекса и саморегуляции отрасли
- Отсутствие регулирования оборота, правовой вакуум
- Отсутствие отраслевого регулятора и оператора БПД



КОМУ ДОЛЖНЫ ПРИНАДЛЕЖАТЬ БПД?

- **Пользователю?** Он не знает, кто собирает, что собирают, не может контролировать или запретить
- **Площадкам?** Они по факту владеют БПД, но их пользовательские соглашения незаконны; право использования и оборота надо регулировать.
- **Вывод: это собственность нации,** нужна защита государством.
- **Метафора «цифровых недр».** Владелец площадки не владеет недрами под ней. Для разработки и оборота «цифровой нефти» нужно регулирование.



ВОЗМОЖНОСТИ БПД

- **Большой рынок:** профиль пользователя стоит денег, при упорядоченном обороте БПД — большой рынок
- **Алгоритмы анализа:** промышленная революция, нельзя упустить, чтобы потом не импортировать
- **Знания о нации:** медицина, образование, настроения, выборы, занятость, развлечения, отдых и т.п.
- **Знания об инфраструктуре и производстве:** транспорт, дороги, ЖКХ, электросети, производства.
- **Знания об экономике:** доходы, структура расходов и сбережений, страхование, льготы, планирование развития, прочее



ЧТО ДЕЛАТЬ: ГОСУДАРСТВО

- **Законотворчество:** регулирование оборота БПД, аналогия «цифровых недр», категории операторов БПД и их права
- **Создание рынка:**
 - стандарты доступа к данным,
 - равные условия для всех игроков отрасли (регулирование зарубежных игроков)
 - Государственный оператор БПД (?)
- **Развитие:** кадры, образование, институты развития, доступ к данным
- **Работа с пользователями:** просвещение, информирование, создание сервисов управления своими БПД



ЧТО ДЕЛАТЬ: РАБОТА С ОТРАСЛЬЮ

- Саморегулирование:
 - Стандартные пользовательские соглашения
 - Отраслевая СРО
 - Этический кодекс
- Стандарты информирования пользователей и государства об обороте БПД
- Отраслевой оператор БПД (?)
- Стандарты доступа к данным



СПАСИБО

ИГОРЬ АШМАНОВ

Управляющий партнер
«Ашманов и партнеры»

+7 495 269 0630

+7 985 222 5451

igor@ashmanov.com