



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

«Влияние современных информационных технологий по обработке «Больших Данных» (Big Data) на защиту частной жизни при трансграничной передаче персональных данных: Зарубежный и российский опыт»

Савельев Александр Иванович, старший научный сотрудник
НИУ "ВШЭ", к.ю.н., юрисконсульт IBM Россия/СНГ

Понятие «Больших данных»

- 1) В 2013 г. общее количество хранящейся в мире информации составило 1,2 зеттабайта (≈ 1.2 триллиона гигабайт).
- 2) По прогнозам компании IDC общее количество информации будет удваиваться каждые 2 года и составит к 2020 г. порядка 40 зеттабайт.
- 3) Большая часть данных, которая будет произведена в период с 2012 по 2020 годы, будет сгенерирована не людьми, а различного рода устройствами в ходе их взаимодействия друг с другом и сетями данных.

Как обработать такой объем разнородной информации, подверженной постоянным изменениям?

- **Big Data («Большие данные»)** - совокупность инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов из различных источников, подверженных постоянным обновлениям, в целях повышения качества принятия управленческих решений, создания новых продуктов и повышения конкурентоспособности.

Преимущества «Больших данных»

Данные – это новая нефть. В «сыром» виде они имеют мало ценности, но будучи переработанными способны становиться ценным активом. Технологии «Больших данных» – это высокотехнологичный способ переработки данных.

Что это даёт бизнесу?

- 1) Конкурентные преимущества;
- 2) Возможность создавать новые бизнес-модели и сервисы
- 3) Оптимизацию процессов принятия управленческих решений

Примеры использования «Больших данных»

- «Таргетинг»: Формирование направленной рекламы и высокоперсонализированных сервисов
- Системы предотвращения мошенничества в банках и иных финансовых организациях
- Анализ платежеспособности и страховых рисков (расчет индивидуальной страховой премии, кредитного лимита и т.д.)
- Здравоохранение: прогнозирование эпидемий, индивидуальный подход к лечению пациента на основе анализа его генома, анализа новейших научных исследований и опыта лечения иных пациентов;
- Совершенствование системы предотвращения преступлений, в том числе на основе прогнозной аналитики

1. Несовместимость технологий «Больших данных» с базовыми принципами законодательства о персональных данных → дополнительные угрозы неприкосновенности частной жизни граждан
2. Персонализация = Индивидуальный подход = ...
Дискриминация

«Большие данные» Vs. законодательство о персональных данных (2)

1. Несовместимость технологий «Больших данных» с принципами минимизации персональных данных и ограничения их обработки заранее определенными целями.

Основная ценность данных в эпоху «Больших данных» определяется возможностью их:

- повторного использования
- в самых различных сочетаниях с иными данными

→ Непредсказуемость является неотъемлемой характеристикой технологий «Больших данных».

Коммерческая ценность данных, жесткая конкуренция, а также значительное удешевление технологий их обработки и хранения стимулирует организации собирать данные «про запас».

В итоге заявленные в законе ограничения и принципы, связанные с конкретными целями обработки обречены на декларативность.

«Большие данные» Vs. законодательство о персональных данных (2)

2. Концепция «информированного, конкретного и сознательного» согласия субъекта персональных данных на их обработку в Интернет-среде и эпоху «Больших данных» - фикция и иллюзия.

- Сложность восприятия текстов политик конфиденциальности. «Парадокс прозрачности».
- Отсутствие возможности что-то поменять
- Невозможность анализа возможных рисков от раскрытия персональных данных
- Взаимобусловленность согласия с заявленными Интернет-сервисом способами обработки данных и получения соответствующего сервиса
- Огромное множество Интернет-сервисов и веб-сайтов, обрабатывающих персональные данные конкретного пользователя

«Большие данные» Vs. законодательство о персональных данных (3)

3. Обезличивание персональных данных не является панацеей.

- рост производительности и доступности вычислительных мощностей
- огромный массив доступной в сети Интернет личной информации
- наличие в обезличенных данных какого-либо атрибута, потенциально относимого к личности (ибо тотальное обезличивание лишает их какой-либо ценности при обработке)

→ В эпоху «Больших данных» всегда существует техническая возможность деанонимизации даже тщательно обезличенных данных.

Ряд зарубежных исследований подтверждают это:

1. Narayanan A., Shmatikov V., De-Anonymizing Social Networks. <http://userweb.cs.utexas.edu/~shmat/shmat-oak09.pdf>
2. Narayanan A., Shmatikov V. Robust De-Anonymization of Large Sparse Datasets // The University of Texas. 2008 IEEE Symposium on Security and Privacy. https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf
3. Ohm P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization // UCLA Law Review No. 57. 2010.

Возможные пути решения проблемы

1. Осознание ее наличия и несовершенства существующих в законодательстве способов ее решения, в том числе посредством так называемой «локализации» хранения персональных данных
2. Концентрация регуляторных усилий на крупных операторах персональных данных (социальные сети, поисковые системы, информационные брокеры).
3. Особое внимание должно быть уделено регулированию обработки «метаданных» (информации о трафике и пр.)
4. Повышение уровня наглядности соответствующих условий политик конфиденциальности
5.

Изменение подхода к адресатам регулирования

Концентрация регуляторных усилий на крупных операторах персональных данных (социальные сети, поисковые системы, информационные брокеры).

+

Предоставление им гибкости при обмене данными с «доверенными лицами» (Trusted parties) за выбор и действия которых они будут нести ответственность (*Accountability approach*)

Нечто подобное имеет место применительно к трансграничной передаче данных в европейском праве. *Adequate safeguards*: 1) Обязательные корпоративные правила (*Corporate binding rules*); 2) стандартные условия, включенные в договор между оператором-экспортером и получателем персональных данных.

Особое внимание к регулированию «метаданных»

Особое внимание должно быть уделено регулированию обработки «метаданных» (т.н. «информации об информации»): данных о трафике, геолокационных данных и т.п.

Возможные направления регулирования:

- ограничение срока их хранения определенным сроком (6 месяцев для поисковых запросов. См.: EU Article 29 Data Protection Working Party, Opinion 1/2008 on data protection issues related to search engines. 00737/EN WP 148. April 2008.)
- Обеспечение самоуничтожения данных по прошествии определенного срока техническими средствами (*Privacy by design*)

Повышение уровня наглядности соответствующих условий политик конфиденциальности

	Не осуществляется сбора избыточных персональных данных, которые не требуются для достижения заявленных целей обработки	
	Не осуществляется хранения персональных данных по достижении целей обработки	
	Персональные данные не обрабатываются в целях иных, чем были заявлены при их сборе	
	Персональные данные не предоставляются третьим лицам, осуществляющим предпринимательскую деятельность	
	Доступ к персональным данным не продается и не предоставляется по подписке	
	Персональные данные хранятся исключительно в зашифрованном виде	

ст. 13а Проекта
Регламента ЕС «О
защите данных»



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Спасибо за внимание!

101000, Россия, Москва, Мясницкая ул., д. 20

Тел.: (495) 621-7983, факс: (495) 628-7931

www.hse.ru