

ЮРИДИЧЕСКИЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ БОЛЬШИХ ДАННЫХ (BIG DATA) В ГОСУДАРСТВЕННОМ УПРАВЛЕНИИ УЗБЕКИСТАНА

Абдимуминов Шермухаммад Журабекович

Студент 1-курса, группы А-1, факультет международного права и сравнительного
правоведения

zurabekhamraev@gmail.com

<https://doi.org/10.5281/zenodo.14173574>

Аннотация. В этой статье будет говориться об правовых аспектах использования больших данных (big data), а также глубокий анализ использования этих данных в государственном секторе Республики Узбекистан, включая международный опыт в этой сфере. Будут детально учтены нормативно- правовые акты законодательства Узбекистана, вместе с проблемами и их решениями в сфере использования больших данных.

Ключевые слова: большие данные, big data, обработка, IBM, объем, разнообразие, скорость данных, Государственный Налоговый Комитет, NADRA, персональные данные.

LEGAL ASPECTS OF USE OF BIG DATA IN PUBLIC ADMINISTRATION OF UZBEKISTAN

Abstract. This article will discuss the legal aspects of the use of big data, as well as a deep analysis of the use of this data in the public sector of the Republic of Uzbekistan, including international experience in this area. The regulatory legal acts of the legislation of Uzbekistan will be taken into account in detail, along with the problems and their solutions in the field of using big data.

Keywords: big data, big data, processing, IBM, volume, diversity, data speed, State Tax Committee, NADRA, personal data.

1. Введение

Термин «большие данные» применяется к наборам данных, которые становятся настолько большими, что с ними становится невозможно работать с использованием традиционных систем управления базами данных. Это наборы данных, размер которых превышает возможности широко используемых программных инструментов и систем хранения для сбора, хранения, управления, а также обработки данных в течение приемлемого прошедшего времени.¹

¹https://www.researchgate.net/publication/264555968_Big_Data_Analytics_A_Literature_Review_Paper

Большие объемы различных типов данных, полученных из различных типов источников, таких как люди, машины или датчики. Эти данные включают климатическую информацию, спутниковые снимки, цифровые фотографии и видео, записи переходов или сигналы GPS. Большие данные могут включать персональные данные: то есть любую информацию, относящуюся к отдельному лицу, и могут быть чем угодно: именем, фотографией, адресом электронной почты, банковскими реквизитами, сообщениями на сайтах социальных сетей, медицинской информацией или IP-адресом компьютера. Клиффорд Линч в 2008 году в своей работе «Nature» писал о взрывном росте объемов информации в мире. Кбольшим данным Линч отнес любые массивы неоднородных данных более 150Гб в сутки, однако единого критерия до сих пор не существует.² Объемы больших данных постоянно увеличиваются и в настоящее время варьируются от нескольких десятков терабайт (ТБ) до многих петабайт (ПБ) данных в одном наборе данных.³ По данным компании IBM к 2003 году мир накопил 5 эксабайтов данных, к 2008 году этот объем возрос до 0,18 зеттабайта, в 2013 году показатель составил 4,4 зеттабайтов. По данным «Statista», объем возрос до 64,2 зеттабайт в 2020 году и 79 зеттабайт в 2021 году, до 2025 года прогнозируется рост более чем до 180 зеттабайт. В 2035 году он будет расти вгеометрической прогрессии до более чем 2100 зеттабайт.⁴

2. Большие данные в компаниях

2.1 Структура больших данных

Большие данные характеризуют три основные характеристики: объем (volume), разнообразие (variety) и скорость (velocity), или три буквы V. Объемданных — это их размер и то, насколько они огромны. Разнообразие включаетв себя различные форматы и типы данных, а также различные виды использования и способы анализа данных. Наконец, скорость означает с которой данные изменяются или как часто они создаются.⁵ Некоторые ученые считают, что существуют и другие основополагающие больших данных, 4 иболее. Компания, работающая с большими данными «IBM» добавила и четвертый аспект — достоверность (veracity). Достоверность относится к качеству анализируемых данных. Данные с высокой достоверностью содержат много записей, которые ценны для анализа и вносят значимый вклад в общие результаты. С другой стороны, данные с низкой

² <https://trends.rbc.ru/trends/innovation/5d6c020b9a7947a740fea65c?from=copy>

³ Russom, P.: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40 (2011)

⁴ Periodica Journal of Modern Philosophy, Social Sciences and Humanities Volume 13 December 2022 (ДехкановШ.А.)

⁵ EMC: Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508 (2012)

достоверностью содержат высокий процент бессмысленных данных. Не имеющее ценности в этих наборах данных называется шумом. Примером набора данных с высокой достоверностью могут служить данные медицинского эксперимента или испытания.

Существует три основных способа хранения данных: традиционный, в публичных облаках и в частных облаках. При традиционном («хранение у себя») способе информация записывается на диски, ленты или локальные хранилища. Хранение данных за определенную плату, зависящую от объема данных и сопутствующих услуг, предоставляют публичные облака. В пятерку стран, предоставляющих облачные технологии входят США, Китай, Великобритания, Германия и Япония (рис.1). Одной из первых компаний, которая стала зарабатывать совершенно новым образом, предоставляя свои неиспользуемые вычислительные ресурсы была Amazon. Сегодня крупнейшими облачными провайдерами являются Amazon Web Services (AWS), Google Cloud, IBM Bluemix, Oracle. Лидером на мировом рынке хранения больших данных является компания Amazon.

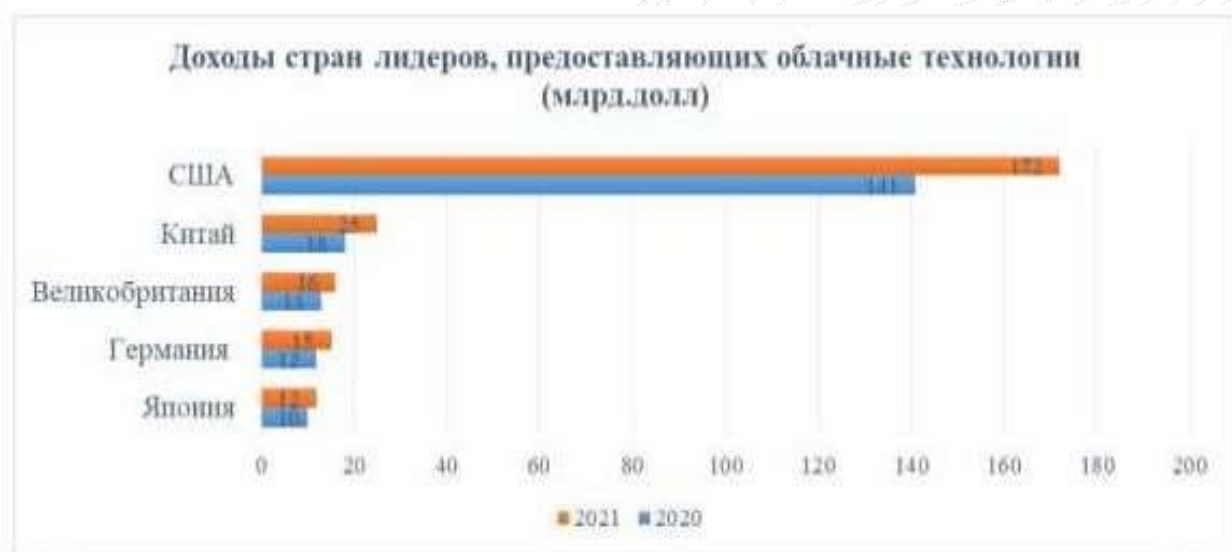


Рис.1 Доходы стран лидеров, предоставляющих облачные технологии⁶

2.2 Влияние больших данных гегемонам в сфере IT

Одним из таких примеров является африканская компания по финансовым технологиям «M-Pesa», созданная Кенийским оператором мобильной связи «Safaricom».

Использование готовой телекоммуникационной инфраструктуры и опираясь на знания о привычках и потребностях аудитории, компания предложила удобный платежный

⁶ <https://www.statista.com/study/105653/digital-economy-compass/>

инструмент — систему денежных переводов номеров телефона, для которого не требуется банковский счет. За этим следят и гиганты телекоммуникационного рынка T-Mobile и Orange. Грамотная работа с большими данными помогает амбициозным «start-up» программам достигать успеха. Например, «финтех-компания» Revolut с оценкой рыночной капитализацией 33 миллиардов долларов США, зависят от анализа больших данных. Компания управляет около 800 информационных панелей с различным набором данных. Большие данные позволяют персонализировать поиск в приложении, а алгоритмы машинного обучения предлагают клиентам индивидуальные программы на основе истории покупок.⁷ Такие крупные компании как Netflix и Spotify с рыночной капитализацией свыше 400 млрд. долларов США, на обоих, предоставляют свои услуги основывая на больших данных и их анализе. Например, Netflix изучая своих клиентов, их предпочтения, возраст, пол, даже различая сферу деятельности зрителя предоставляют такие фильмы и сериалы исключительно подходящей для определённой личности, они делают это посредством собирания и изучения данных с огромными размерами.

Идентичную систему использует в своей деятельности Spotify – исходя из нашего настроения потребностей выбирает музыку, соответствующую нашему запросу, опять же посредством сбора и обработки больших данных.

3. Facebook и Узбекистан

Согласно внесённым изменениям в 2021 году 16 апреля в Закон «О персональных данных» от 2 июня 2019 года, касательно статьи – 27¹, которая гласит следующим образом, собственник и (или) оператор при обработке персональных данных граждан Республики Узбекистан с использованием информационных технологий, в том числе во всемирной информационной сети Интернет, обязан обеспечить их сбор, систематизацию и хранение в базах персональных данных на технических средствах, физически размещённых на территории Республики Узбекистан и зарегистрированных в установленном порядке в Государственном реестре баз персональных данных.⁸ После вступления в силу настоящего пункта социальные сети, такие как Facebook, Telegram, Instagram, TikTok, Twitter, Gmail, почтовые сервисы Mail.ru и интернет-сервисы, такие как Google, Яндекс, базы данных (серверы), содержащие персональную информацию граждан Узбекистана, их необходимо будет привезти в Узбекистан и оставить себе. Пресс-секретарь Министерства информационных технологий и развития связи Шерзод Ахмедов сообщил, что головные

⁷ Tuychieva O.N. (2022). MODERN TRENDS IN THE DEVELOPMENT OF FINTECH AND BIG DATA IN UZBEKISTAN. Экономика и социум, (11-1 (102)), 321-330.

⁸ Закон Республики Узбекистан "О персональных данных" №ЗРУ-547 от 02.07.2019.

офисы всех социальных сетей, которыми пользуются граждане Узбекистана, были официально проинформированы о новом законодательстве и до вступления закона в силу переместили серверы. По информации представленной BBC,⁹ эта новость в Узбекистане была негативно воспринята компанией Facebook. По информации, поступившей в издание, Facebook обеспокоен тем, что такие страны, как Россия, Китай и Узбекистан пытаются таким образом ограничить свободу слова. Ответ секретаря Министерства информационных технологий Республики Узбекистан был таким: «Если персональные данные граждан Узбекистана хранятся на территории иностранного государства, государство не может контролировать выполнение требований закона «О персональных данных» иностранной компанией в отношении данных своих граждан. Кроме того, даже если узбекские компании хранят персональные данные на территории другого государства, государство не может контролировать выполнение установленных правил, поскольку не имеет возможности проводить проверки на территории другого государства», - было сказано в издании kun.uz.¹⁰ В Узбекистане одним из первых государственных органов использующим большие данные современного типа, стал Государственный Налоговый Комитет, которая как сообщалось станет основой для автоматизированной системы контроля за уплатой НДС. В системе Комитета существует собственный Центр обработки данных (ЦОД), целью которого является, сокращение теневого сектора экономики, повышение прозрачности экономики, создание равных конкурентных условий. Интенсивность обмена информацией составляет порядка 15 тыс. транзакций в секунду, в результате чего отмечается ежедневный рост объема информации ЦОД более чем на 9,5 Гб и общий объем налоговой информации превышает 22 Тб

– информация на февраль 2019 год – следовательно эти данные расширились в разы.

4. Международный опыт

4.1. История создания идеального электронного правительства

В современном мире, о том, что технологии способны перевернуть мир, а также вероятность робопостапокалипсиса, волнует не малое количество людей и речь идет уже не о фантастических проектах Голливуда.

Пытаясь отыскать примеры использования государством больших данных, мы чаще всего обращаемся к опыту так называемых «развитых стран» или G7 (Большая семерка), предполагая, что только им сегодня доступны подобные решения.

⁹ https://www.youtube.com/watch?v=DND_4tD1NpM

¹⁰ <https://kun.uz/news/2021/02/20/yangi-qonun-ozbekistonda-ijtimoiy-tarmoqlar-bloklanishiga-olib-keladimi-vazirlilik-izohi>

Однако, волею судьбы в числе лидеров в этой области оказался Пакистан – государство, занимающее 44 место в мире по объему ВВП.

Создание одной из крупнейших в мире базы биометрических данных, началось по инициативе Беназир Бхутто - National Database & Registration Authority (NADRA) – так называется проект, заинтересованная в появлении технологий, способствующее укреплению демократии, помогающее бороться с коррупцией, терроризмом и стихийными бедствиями, эффективно проводить реформы. Это связано с несколькими причинами – в 1971 году после того как Бангладеш разделился, в Пакистан прибыло много бенгальцев, а в 1979 году – афганские беженцы, к тому же в стране всегда проживало огромное количество иранцев. Многие из них смогли получить национальные удостоверения личности, предоставив поддельные документы, в то время как граждане страны их не имели.

4.2. Успех проекта

Для того, чтобы навести порядок в системе учета, было решено создать базу биометрических данных, включающую в себя отпечатки 10 пальцев и цифровую фотографию. Работа началась в 2000 году, чтобы стимулировать граждан пройти регистрацию, было выпущено распоряжение правительства, запрещающее открывать счета в банке без предъявления ID. Электронное удостоверение личности необходимо было также при любом взаимодействии с государством – регистрации брака, рождения и смерти, оформлении пособий, подключении электро- или газоснабжения, включая оплату коммунальных услуг. Уже на первых этапах проекта она дала свои плоды, согласно анализу по собранным данным и сопоставлению их с другими государственно-информационными системами были выявлены люди, которые получали по 2 и даже 3 пенсии, граждане, которые за небольшую плату отбывали наказание за настоящих преступников. При проверке списков избирателей выяснилось, что 45% или 37 млн записей в них дублируются – некоторые граждане зарегистрированы более 20 раз, а некоторые уже давно умерли. Дальше – еще больше. По имеющимся на тот момент данным, из более чем 120 млн граждан Пакистана платили налоги только около 800 тысяч. Насегоднешний день, на базе NADRA создана другая система, которая продолжает свою деятельность для таких государств, как Кения, Шри-Ланка, Нигерия, Бангладеш, Судан и другие. При поддержке NADRA также была создана биометрическая база данных беженцев для ООН. Сегодня NADRA – полностью самостоятельная организация, за

последние 5 лет ее доходы увеличились на 300%, а зарплаты сотрудников – на 121%.¹¹

5. Узбекистан в этой сфере

5.1 Достижения Узбекистана в интеграции big data

Одной из самых популярных и возможно лидирующих проектов по количеству использования данных граждан является – my.gov.uz. В современном мире технологий, популярна фраза: «Данные – новая нефть», это объясняется тем, что данные могут повлиять на мир на прямую, как нефть влияет на политику.

В Узбекистане за последние годы достигнуто многое по интеграции системы обмена данными в процессе оказания государственных услуг. Через Единый портал интерактивных государственных услуг (my.gov.uz) более 670 услуг оказываются онлайн. Кроме того, 165 государственных организаций подключены к межведомственной интеграционной платформе (G2G), где создана возможность обмена более 150 видами данных. К системе электронного правительства также подключился и частный сектор, 137 организаций из числа банков и компаний. Другим словами, сотни систем уже «общаются» друг с другом ежедневно. Согласно Министерству цифровых технологий, с 2021 года обмен данными между ведомствами был произведён 1,8 миллиарда раз. Благодаря этому были отменены истребования около 110 справок и разрешительных документов у населения и бизнеса. Это само по себе создаёт огромный пласт данных, который нужен для принятия решений.

Существуют некоторые вызовы в этой системе:

- какие государственные структуры управляют какими данными (кто является владельцем данных);
- на каком основании это делается;
- какие права имеют ведомства на использование данных других госорганов;
- совместимость — стандарты, классификации;
- связывание данных (посредством общих идентификаторов).

5.2. Нормативно-правовые акты касательно больших данных

ЗРУ «О персональных данных»

ЗРУ «О кибербезопасности»

ПП-4996 «О мерах по созданию условий для ускоренного внедрения технологий искусственного интеллекта»

УП-5598 «О дополнительных мерах по внедрению цифровой экономики,

¹¹ Наталья Рудычева/CNews Analytics

электронного правительства, а также информационных систем в государственном управлении Республики Узбекистан»

УП-6079 «Об утверждении Стратегии «Цифровой Узбекистан-2030» и мерах по ее эффективной реализации»

6. Проблемы и решения больших данных

Минусы больших данных можно рассмотреть подробнее по этим пунктам:

1. Обмен данными и участие: Они требуют готовности организаций делиться важной информацией. Это часто вызывает сложности, так как не все компании готовы обмениваться данными из-за конкурентных соображений, а также из-за потенциальных рисков утечки информации.

2. Нехватка талантов и навыков: Работа с большими данными требует квалифицированных специалистов, таких как аналитики данных, и инженеры данных, которые могут управлять огромными объемами информации и извлекать из них ценные инсайты. Однако таких профессионалов часто не хватает.

3. Опасения по поводу безопасности: Безопасность данных — одна из главных проблем в области больших данных. Хранение и обработка большого объема информации может сделать систему уязвимой для утечек и кибератак, что может привести к потере конфиденциальной информации.

4. Сбор данных в обезличенной форме: Для соблюдения конфиденциальности компании обязаны собирать данные в обезличенной форме. Это может усложнить анализ, так как определенные параметры теряются, а также увеличивает расходы на соблюдение стандартов конфиденциальности.

Для решения перечисленных проблем в работе с большими данными можно рассмотреть следующие стратегии:

Решением может быть:

1. Обмен данными и участие - Разработать юридически обоснованные соглашения, которые бы защищали интересы компаний и гарантировали безопасность данных при обмене. Ограничение и контроль доступа к данным, чтобы минимизировать потенциальные риски для конкурентоспособности компаний.

2. Нехватка талантов и навыков- Программы обучения и стажировки: Организации могут инвестировать в программы обучения и стажировки для развития талантов и подготовки сотрудников. Использование доступных инструментов анализа данных и автоматизации некоторых процессов может снизить зависимость от высококвалифицированных специалистов.

3. Опасения по поводу безопасности - Шифрование данных: Внедрение методов шифрования данных на всех этапах их обработки и передачи. Внедрение стандартов кибербезопасности, то есть использование передовых стандартов безопасности, таких как ISO/IEC 27001, помогает улучшить защиту данных от угроз.

4. Сбор данных в обезличенной форме - Соблюдение стандартов конфиденциальности: Автоматизация процесса соблюдения стандартов конфиденциальности (например, GDPR) для снижения затрат и упрощения работы с обезличенными данными.

Эти подходы помогут улучшить качество работы с большими данными, минимизируя риски и сложные моменты, связанные с безопасностью и управлением информацией.

«Страна не будет развиваться, если каждая отрасль не будет идти в ногу с цифровыми технологиями» — заявил Президент Узбекистана Шавкат Мирзиёев.¹²

REFERENCES

1. https://www.youtube.com/watch?v=DND_4tD1NpM
2. <https://www.bigdataframework.org/the-four-vs-of-big-data/>
3. <https://trends.rbc.ru/trends/innovation/5d6c020b9a7947a740fea65c>
4. <https://www.nature.com/nature/volumes/455/issues/7209>
5. https://www.researchgate.net/publication/264555968_Big_Data_Analytics_A_Literature_Review_Paper
6. <https://www.statista.com/study/105653/digital-economy-compass/>
7. <https://kun.uz/news/2021/02/20/yangi-qonun-ozbekistonda-ijtimoiy-tarmoqlar-bloklanishiga-olib-keladimi-vazirlik-izohi>
8. <https://www.spot.uz/ru/2019/02/07/big-data/>
9. https://www.cnews.ru/articles/bolshie_dannye_v_gossektore_opyt_pakistana
10. <https://president.uz/ru/lists/view/6937>
10. Russom, P.: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40 (2011)
11. Periodica Journal of Modern Philosophy, Social Sciences and Humanities Volume 13 December, 2022 (Ш.А. Дехканов)
12. Tuychieva O.N. (2022). MODERN TRENDS IN THE DEVELOPMENT OF FINTECH AND BIG DATA IN UZBEKISTAN. Экономика и социум, (11-1 (102)), 321-330

¹² <https://president.uz/ru/lists/view/6937>

13. Закон Республики Узбекистан "О персональных данных" №ЗРУ-547 от 02.07.2019
14. Государственный Налоговый Комитет Республики Узбекистан.