

## ХАРАКТЕРИСТИКИ БОЛЬШИХ ДАННЫХ, ЗНАЧИМЫЕ ДЛЯ ЦЕЛЕЙ ГРАЖДАНСКОГО ПРАВА\*

Аннотация. Большие данные можно рассматривать как имущество. Но российским юристам не совсем понятно, что следует понимать под термином «большие данные». И автор акцентирует внимание на сущности больших данных как информации (данных), как информационных технологиях и инструментах, как алгоритме. Особое внимание уделяется использованию больших данных.

Большие данные (далее – *big data*) сегодня рассматриваются как один из ценнейших активов компаний, что дает повод поднимать вопрос о необходимости их специального правового регулирования. Между тем подобные предложения делаются нередко без четкого понимания того, что собой представляют такие данные.

Дело в том, что большинство юристов рассматривают *big data* исключительно как огромные массивы разнообразной информации (данных), чему способствовало растиражированное во многих публикациях употребление Клиффордом Линчем<sup>1</sup> этого термина применительно к взрывному росту мировых объемов информации и многообразию данных.

Между тем огромный физический объем и разнообразие типов данных представляют собой только некоторые – наиболее известные – характеристики такого многоаспектного явления, как *big data*. И для того чтобы дать ему характеристику в контексте гражданского права, необходимо изучить и другие его аспекты.

### Хранилища *big data*: накопление, сбор, обработка и хранение данных

В п. 1 ст. 2 Федерального закона от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации»

---

\* Тезисы доклада на научном круглом столе «Трансформация институтов права интеллектуальной собственности в условиях цифровизации общества», проходившем в рамках XII Международного форума «Интеллектуальная собственность – XXI век».

<sup>1</sup> Редактор журнала “Nature”, которому приписывается введение этого термина в обиход в 2008 году. См.: <https://www.nature.com/articles/455001a>

(далее – Закон об информации) в легальной дефиниции понятия «информация» упоминаются в качестве ее разновидностей, во-первых, *сведения*, во-вторых, *сообщения*, в-третьих, *данные*.

В ГК РФ нашел отражение подход, согласно которому данные представляют собой некие информационные единицы, статично хранящиеся в электронных таблицах или базах данных. Правда, при разработке части четвертой Кодекса это понимание несколько модифицировалось: данные было предложено понимать в качестве «самостоятельных материалов (статей, расчетов, нормативных актов, судебных решений и иных подобных материалов)» (п. 2 ст. 1260). Это, вероятно, стало следствием того, что для разработчиков ГК РФ ориентиром служили исключительно базы известных справочно-правовых систем.

Применительно к *big data* нет никаких оснований говорить о статике: это динамичный, безостановочный процесс появления новых данных, часть которых изначально не структурирована и не обработана иным образом, а часть уже выступала предметом обработки. То есть, прежде всего, *big data* представляют собой постоянный поток огромных объемов информации, непрерывно поступающей из различных источников. Как отмечается во многих публикациях, к категории *big data* относится большинство потоков данных свыше 100 Гб в день.

Названный непрекращающийся поток данных поступает из различных источников, которые можно условно объединить в две основные группы:

– во-первых, технические источники, создающие порядка 90 % всей новой информации. Эта группа охватывает, в частности, интернет вещей (англ. Internet of Things, IoT), который «поставляет» информацию со всевозможных действующих датчиков, контроллеров, приборов учета потребления, устройств, устройств аудио- и видеорегистрации, измерительных комплексов и проч.; искусственный интеллект (англ. Artificial Intelligence, AI) и машинное обучение (англ. Machine Learning, ML);

– во-вторых, социальные источники, включающие, в частности, социальные медиа (англ. social media), охватывающие

разнообразные способы электронной коммуникации: социальные сети, виртуальные миры, специализированные форумы, профессиональные соцсети, блоги, фотохостинги, сайты отзывов, сайты знакомств и проч. (информация образуется из потока постов, комментариев, лайков, поисковых запросов, оценок, фото, аудио- и видеозаписей, отзывов и проч.); розничную торговлю (англ. retail), предоставляющую информацию о совершенных транзакциях, сведения из товарных чеков, из дисконтных карт и карт лояльности покупателей, из RFID-меток и проч.; здравоохранение, собирающее сведения о поставленных диагнозах и предложенных методиках лечения, восприимчивости пациентов к медицинским препаратам, об оценке эффективности этих препаратов и проч., что находит отражение в медицинских картах, результатах лабораторных исследований и т.д.

Следует заметить, что часть информации может поступать в виде структурированных, то есть упорядоченных определенным образом, данных (например, они могут быть собраны в таблицы, что позволяет применять в отношении них визуальный или автоматизированный анализ). Вместе с тем большая часть информации поступает в виде данных неструктурированных (англ. unstructured data), к которым относят, в частности, данные из соцсетей, видео- и аудиофайлы, данные GPS, спутниковые изображения, данные о перемещении мобильного абонента, данные с серверов, файлы PDF и проч. В отношении этих, а также полуструктурированных данных затруднено использование программ, предназначенных для работы со структурированными данными, – требуется разработка специальных технических решений, о которых будет сказано далее.

Непрерывное появление – постоянным потоком – огромных объемов обозначенной информации повлекло формирование принципиально других методов сбора, накопления, обработки и хранения, нежели для обычных баз данных, подразумевающих упорядоченную («по строкам и столбцам») организацию собранных сведений. Кроме того, в отличие от традиционных баз данных, в которых данные измеряются количеством информационных единиц

(«материалов» в терминологии разработчиков ГК РФ), применительно к *big data* традиционно говорят о трех V: *volume* (физическом объеме); *velocity* (скорости прироста); *variety* (многообразие самих данных, их источников и проч.).

Обозначенные различия выявили потребность в использовании принципиально иных технологий, нежели используемые для создания и управления базами данных. Результатом стало то, что подходы, инструменты и методы сбора, накопления, обработки и хранения *big data* характеризуются как альтернативные традиционным системам управления базами данных. И вполне закономерно, что возможность распространения на хранилища *big data* положений ГК РФ о базах данных является проблематичным: например, авторско-правовая охрана баз данных, предполагающая наличие творческого подхода при подборе и компоновке составляющих базы, не может быть предоставлена хранилищу данных, содержащему как структурированную, так и неструктурированную информацию.

Важен еще один момент: с технической точки зрения именно новые технологии сбора, накопления, хранения и обработки данных рассматриваются в качестве *big data*. И таким образом, когда речь идет о сборе, накоплении, обработке и хранении данных, термин “*big data*” может быть использован для обозначения, во-первых, огромных объемов разнообразной информации и, во-вторых, технологий и инструментов сбора, накопления, обработки и хранения, альтернативных тем, которые используются для баз данных.

✓ Применительно к пониманию *big data* как информации следует обратить внимание на то, что всякая имеющая действительную или потенциальную экономическую ценность информация может становиться объектом гражданских прав (даже в отсутствие на это прямого указания в тексте ст. 128 ГК, поскольку содержащийся в ней перечень объектов гражданских прав является открытым) [1, с. 16]. Вместе с тем следует иметь в виду, что, как и любой нематериальный объект, сама информация в силу естественных свойств не является оборотоспособной – в гражданский оборот могут быть введены имущественные права на информацию (данные).

Проблематичность признания имущественных прав на *big data* обусловлена тем, что в их состав нередко входят данные, относящиеся к числу персональных, а в этих условиях дискуссия вызывает субъект правообладания такими данными (в отношении данных, поступающих из технических источников, подобные споры практически не ведутся). Это связано с тем, что за субъектом персональных данных традиционно признают права на эти данные, но лишь *личные неимущественные* права. В отношении допустимости признания за субъектом персональных данных *имущественных* прав на эти сведения эксперты высказывают различные, иногда диаметрально противоположные мнения. При этом на имущественные права претендуют и компании, накапливающие такие данные (в частности, соцсети, в которых аккумулируется множество личных сведений пользователей), и компании, собирающие эти данные, в том числе из открытых источников: наличие таких прав у этих компаний создавало бы для них неограниченные возможности по использованию этих данных и распоряжению правами на них по своему усмотрению. Эта проблема проявилась при рассмотрении судами известного дела «ВКонтакте против Дабл» (№ А40-18827/17) и со всей очевидностью требует разрешения.

✓ Ценность технологий и инструментов сбора, накопления, обработки и хранения *big data* объясняется сложностью решаемой задачи: в частности, Uber накопил данные о более чем 5 млрд поездок, а Tesla – о вождении на совокупной дистанции более чем 2 млрд км.

*Big data* в понимании их как технологий и инструментов сбора, накопления, обработки и хранения данных представляют интерес с позиций права интеллектуальной собственности. Это обусловлено тем, что подобные ИТ-решения включают в себя объекты, охраняемые в качестве объектов авторского или патентного права либо секрета производства (ноу-хау). В частности, правовую охрану могут получить аппаратные средства, алгоритмы, базы данных, исходный код, интерфейс и т.п.

И здесь усматривается иной (по сравнению с предыдущим случаем) субъектный состав возможных правоотношений: исключительные (имущественные) права на соответствующие

объекты возникают у первоначальных правообладателей (как правило, разработчиков) и могут быть переданы заинтересованным лицам на условиях лицензии (ст. 1235 ГК РФ) или отчуждены в полном объеме (ст. 1234 ГК РФ). Здесь же можно упомянуть и возможность договорных отношений заинтересованных компаний с системными интеграторами, внедряющими такого рода IT-решения, правовым основанием которых становятся договоры оказания услуг.

С учетом сказанного проблематика *big data* в разбираемом аспекте – это, по сути, частный случай проблем, связанных с правовой охраной и защитой IT-решений.

### ***Big data analytics*: анализ данных для цели принятия оптимального решения**

Настоящую часть статьи, вероятно, было бы правильным предварить эпиграфом «Данные – это новая нефть!» (англ. “Data is the new oil!”) – фразой, прозвучавшей в 2006 году и приписываемой британскому математику Клайву Хамби<sup>2</sup>. Примечательно, что ее смысл не в признании равноценности данных и нефти, а в констатации того, что, как и сырая нефть, необработанные (сырые) данные не представляют собой особой ценности: для того чтобы данные создавали действительную прибыль, они должны быть использованы – подвергнуты анализу, визуализированы, интегрированы и т.д. для целей принятия того или иного решения. Иными словами, огромные объемы собранной и накопленной информации, требующие существенных затрат на хранение, имеют только потенциальную коммерческую ценность. Реальную же коммерческую ценность данные, полученные из потоков *big data*, могут приобрести лишь тогда, когда они используются для решения той или иной научной/социальной/коммерческой задачи.

Под использованием *big data* понимают, прежде всего, аналитику больших данных (англ. *big data analytics*), которая в настоящее время признается «новой формой производства знаний» и

---

<sup>2</sup> [https://ana.blogs.com/maestros/2006/11/data\\_is\\_the\\_new.html](https://ana.blogs.com/maestros/2006/11/data_is_the_new.html)

предполагает действия по структуризации данных, созданию алгоритмов анализа данных, агрегации и анализу данных, выявлению связей между данными, установлению закономерностей и скрытых тенденций, построению прогнозов и т.п.

Применительно к *big data analytics*, по моему мнению, следует обратить внимание на следующие черты *big data*.

Во-первых, значимым является то, что *big data* не создаются специально для проведения какого-либо анализа, а возникают естественным образом: упомянутый в первой части статьи поток данных – это результат непрерывной генерации данных, которые появляются в процессе осуществления различной деятельности и/или при решении разных задач. Иными словами, под *big data*, как правило, понимают вторичные данные, которые первоначально предназначались для других целей, – их рассматривают как побочный результат различных процессов.

Так, пользователи соцсетей размещают в своих аккаунтах личную информацию и фото, преследуя цель поведать о себе миру; автоматическое сообщение контроллером сведений о состоянии и работоспособности технического узла предназначено для предотвращения нарушений его работы; онлайн-отзывы о ресторане/гостинице/прокате авто обычно направлены на улучшение качества услуг соответствующих компаний; передача с метеостанции данных о температуре воздуха, осадках, снежном покрове нацелена на генерацию необходимых сведений о метеорологических условиях и т.д.

Такого рода информация вливается в бесконечный непрерывный поток *big data* и, будучи собранной и (или) накопленной компанией, может использоваться при проведении *big data analytics* – для принятия взвешенных, обоснованных решений и построения прогнозов.

Во-вторых, надо специально обратить внимание на то обстоятельство, что *big data analytics* осуществляется не вручную, а автоматизированно. Причем системы анализа данных и поддержки принятия решений предполагают создание специальных алгоритмов анализа данных и иных инструментов. Иными словами,

применительно к процессу *big data analytics* особую значимость приобретают специально разрабатываемые технологические решения, которые позволяют эффективно и оперативно решать поставленные задачи. Поэтому неудивительно, что здесь под термином “*big data*” обычно подразумевают уже не совокупность данных, а именно техническое решение, алгоритм.

В качестве примера можно привести разработку российского стартапа Synqera<sup>3</sup>, созданного на основе исследования, согласно которому покупатели тратят миллионы на импульсные покупки. Разработанная стартапом вычислительная платформа за 40 секунд – время ожидания, которые покупатель в среднем проводит на кассе в ожидании оплаты товаров, анализирует информацию о каждом покупателе: историю его покупок, покупательские предпочтения, возраст, пол и даже настроение – на кассах магазинов сенсорные экраны с датчиками распознают эмоции покупателей. Полученный результат обогащается бизнес-информацией об акциях или скидках магазина, а также данными из открытых источников (из соцсетей или о погоде). По результатам каждому покупателю отправляются таргетированные сообщения, предоставляются персональные скидки и специальные предложения и т.п., способствующие совершению этим покупателем упомянутых импульсных покупок. Таким образом, вычислительная платформа позволяет использовать время, проведенное покупателем на кассе, для стимуляции таких покупок.

В-третьих, следует отметить, что применительно к *big data analytics* приоритетное значение приобретает не количественная, а качественная сторона данных.

Это проявляется, в частности, в том, что для решения различных задач подразумевается задействование разных доступных для аналитики данных – собранных в различном контексте и полученных из многих источников (например, современные автомобили сегодня способны накапливать данные о водителе, погоде и окружающей среде, самом авто и допущенных системой ошибках,

---

<sup>3</sup> <https://www.retail-loyalty.org/news/synqera-predstavila-vysokotekhnologichnoe-reshenie-dlya-rosta-pokupatelskoy-loyalnosti/>



подключенных устройствах и др.). При этом *big data analytics* допускает использование одних и тех же доступных для аналитики данных для достижения различных целей: для этого они соответствующим образом трансформируются и агрегируются – с добавлением новых наборов данных или без таковых. То есть анализу могут быть подвергнуты все доступные компании данные либо выборки из собранных и накопленных данных – в зависимости от поставленной задачи. Это позволяет акцентировать внимание на том, что упомянутые данные могут использоваться бесконечное количество раз и в целях, которые заранее сложно предвидеть.

С учетом изложенного применительно к аналитике больших данных термин “*big data*” употребляется обычно для обозначения именно алгоритмов и иных IT-решений в сфере анализа данных, тогда как объемы данных уходят на второй план.

Таким образом, как и в случаях со всеми иными IT-решениями, здесь затрагивается сфера интеллектуальной собственности: в частности, алгоритмы относят к патентуемым техническим решениям (изобретениям), а компьютерные программы (программы для ЭВМ) – к числу объектов авторских прав. Вследствие этого отношения по поводу *big data*, если понимать их как алгоритмы и иные IT-решения в сфере анализа данных, предполагают субъектный состав, аналогичный тому, что был указан в первой части настоящей статьи: исключительные (имущественные) права на соответствующие объекты возникают у первоначальных правообладателей (как правило, разработчиков) и могут быть переданы заинтересованным лицам на условиях лицензии (ст. 1235 ГК РФ) или отчуждены в полном объеме (ст. 1234 ГК РФ).

В то же время здесь находится место и для договоров оказания услуг – не только с системными интеграторами, внедряющими такого рода IT-решения, но и с разработчиками алгоритмов (датамайнерами), когда они не передают права на разработанные ими алгоритмы, а с их помощью по заданию заказчика осуществляют анализ *big data*, выявляя взаимосвязи, делая прогнозы и вырабатывая решения. В этом ключе можно вспомнить и разработчиков сервисов, которые

предоставляют заинтересованным лицам возможность использования своих технических решений в сфере аналитики данных.

Примечательно, что результат применения алгоритма нередко рассматривается как самостоятельный объект интеллектуальной собственности, имеющий при этом значительную коммерческую ценность. Однако с такой позицией сложно согласиться. Как уже указывалось, анализ *big data* осуществляется автоматически, поэтому, не отрицая участия физических лиц в управлении программным обеспечением, сложно признать, что полученный продукт есть результат интеллектуальной деятельности именно человека.

\* \* \*

В отечественных публикациях вопросы правового регулирования *big data* обычно анализируются авторами в контексте законодательства о персональных данных. Это обусловлено тем, что, как уже указывалось, персональные данные, поступающие из социальных источников, становятся составляющими *big data*, что влечет необходимость решения вопросов надлежащего информирования субъекта персональных данных о целях использования его данных, получения согласия на использование таких данных и т.п.

Вместе с тем не менее частым является рассмотрение *big data* в контексте интеллектуальной собственности. Причем высказываются как предложения признать их новым объектом интеллектуальной собственности, включив в перечень ст. 1225 ГК РФ «Охраняемые результаты интеллектуальной деятельности и средства индивидуализации», так и идеи о некотором противостоянии или конкуренции *big data* и интеллектуальной собственности.

На мой взгляд, для таких умозаключений нет серьезных оснований. Подобные выводы могут стать следствием слишком упрощенного, одностороннего понимания *big data*, тогда как это многоаспектное явление. Вариативность трактовки *big data* становится помехой для определения единого для всех случаев понятия (сформулировать всеобъемлющую дефиницию очевидно не

получается, но попытки с завидным упорством предпринимаются многими авторами), однако она не должна препятствовать определению природы прав, возникающих по поводу *big data*, и правовых норм, подлежащих применению в том или ином случае.

***М. А. РОЖКОВА,***

*член Экспертного совета Комитета Государственной Думы  
по информационной политике,  
информационным технологиям и связи, эксперт РАН,  
эксперт Института развития интернета,  
президент IP CLUB, профессор-исследователь БФУ им.  
Иммануила Канта, доктор юридических наук*

*Библиографический список*

1. Рожкова М. А. Персональные и неперсональные данные как объекты гражданских прав // Хозяйство и право. 2019. № 5. С. 15-21.