

**Ю. П. Гаврильченко**

доктор юридических наук, профессор,  
профессор кафедры финансового права и правового  
регулирувания хозяйственной деятельности  
Белорусского государственного университета

**ПРОБЛЕМЫ ПРАВОВОГО РЕГУЛИРОВАНИЯ  
ПОНЯТИЯ БОЛЬШИХ ДАННЫХ (BIG DATA)**

*Аннотация.* Большие данные на современном этапе развития информационного общества характеризуются не просто как данные достаточно большого объема, но как информация, которая имеет размер настолько огромный, что возникают проблемы ее восприятия и обработки. В статье исследуются легальный, теоретический и практический подходы к понятию больших данных (big data). Автор приходит к выводу о необходимости закрепления соответствующего общего определения в белорусском законодательстве. За основу предлагается использовать понимание больших данных как массива данных, обладающего ценностью лишь в контексте соответствующей обработки. При этом подчеркивается, что ставшие уже классическими признаки больших данных (объем, скорость, разнообразие) в современных условиях дополняются новыми, что еще больше усложняет концепт больших данных и актуализирует вопросы правового регулирования данной сферы.

*Ключевые слова:* большие данные, big data, база данных, объемность, скорость, разнообразие, информация, информационное право.

**Введение.** В условиях становления информационного общества ежедневно генерируется так много данных, что раньше этот объем создавался десятилетиями. Еще в 1999 г. в мире формировалось около 1,5 миллиарда гигабайт информации, но уже в ходе исследования 2003 г. выяснилось, что объем удвоился [1] и темпы увеличения продолжают нарастать. При этом значительная часть данных проходит через облачные сервисы, то есть подлежит анализу. Таким образом, информация не только многократно увеличивается в размерах, но и обретает особую значимость. С учетом объема, скорости и изменчивости формируемых данных все очевиднее, что они имеют ценность уже не в качестве характеристики определенных объектов или субъектов, а как информация, с помощью которой можно видеть общие корреляции и закономерности. Большие данные обретают новые свойства, и это порождает множество вопросов: об их

соотношении с персональными данными, о больших данных в контексте конкурентного права, о правовом режиме больших данных и др. При этом базовым остается вопрос о понятии и сущности больших данных (big data). Несмотря на то что термин у всех на слуху и достаточно широко обсуждается научным сообществом, единообразного подхода к проблематике не существует. Целью настоящего исследования стало выявление ключевых подходов к понятию больших данных и раскрытие его содержания.

**Основная часть.** Вопрос о появлении проблематики больших данных лишен однозначности. Существует мнение, что, «по всей видимости, впервые термин „большие данные“ был использован в ежегодном отчете международной консалтинговой компании Gartner 2011 г. и отнесен тогда к новым технологиям, создающим „ажиотаж“» [2, с. 119]. Однако представляется обоснованной точка зрения о более раннем происхождении. Понятие в разном его значении использовалось в публикациях с конца XX в., а упоминалось еще раньше. При этом три ключевых признака больших данных были сформулированы уже в 2001 г. [3].

В современном белорусском законодательстве отсутствует общий подход к понятию больших данных, хотя само понятие в том или ином контексте встречается неоднократно. Определение приведено лишь применительно к образовательным стандартам высшего образования II ступени [4], где под большими данными понимаются структурированные и неструктурированные данные огромных объемов, которые эффективно обрабатываются с помощью масштабируемых программных инструментов (п. 3).

В источниках большие данные определяются в двух основных значениях:

1. Как массив данных. Этот подход считается общепринятым в англоязычной литературе. В частности, Оксфордский словарь английского языка определяет большие данные как «данные очень большого размера, обычно в той степени, в которой их манипулирование и управление представляют собой значительные логистические проблемы» [5]. Это определение в целом соответствует избранному белорусским законодателем и указывает на два фундаментальных признака, позволяющих отнести информацию к большим данным: «очень большой размер» («огромный объем») и, как следствие, сложность обработки. В связи с этим возникает вопрос о том,

на какой стадии существования информации можно говорить о больших данных: «Это сам объем информации (так называемые сырые данные, то есть абстрактные факты, необработанная информация)? Или это результат обработки такой информации?» [6, с. 27–28]. Белорусский законодатель не дает четкого ответа на эти вопросы, хотя упоминает структурированные и неструктурированные разновидности больших данных и таким образом косвенно соглашается, что информация может существовать в разной форме.

Практика показывает, что big data возникают уже в момент формирования сырых данных и остаются таковыми в процессе и после обработки, если продолжают отвечать всем ключевым характеристикам больших данных. При этом конечным или промежуточным результатом (результатами) обработки может стать достаточно краткий отчет, это никак не влияет на сами данные и возможность дальнейшей работы с ними.

2. Как технология обработки данных [7, с. 6], группа технологий и методов [8, с. 49], совокупность инструментов «обработки структурированных и неструктурированных данных огромных объемов из различных источников, подверженных постоянным обновлениям, в целях повышения качества принятия управленческих решений, создания новых продуктов и повышения конкурентоспособности» [9, с. 43]. Этот подход существует в русскоязычной литературе наряду с пониманием big data как массива информации. В данном контексте представляется более правильным все же говорить о технологии больших данных, а не больших данных как технологии, но анализируемое понимание остается достаточно распространенным, иллюстрируя суть big data: это не просто много информации, а настолько много, что для ее обработки нужны специалисты и особые технологии. Как отмечается в литературе, «данные – это новая „нефть“: они представляют собой ценность, однако в необработанном состоянии они не могут быть использованы» [10].

Таким образом, большие данные – это такой объем информации, который обретает значимость именно в контексте использования особого механизма обработки. Как отмечает М. С. Корнев, «здесь доминирующая характеристика в определении – это операционные трудности (важнее „как?“, а не „зачем?“)» [11, с. 81–82]. В этом контексте интересна дискуссия о признании больших данных в обработанном виде секретом производства (ноу-хау). Приводимая авторами аргументация

достаточно сомнительна и вряд ли применима к белорусскому праву [2, с. 166], но само ее существование демонстрирует ценность больших данных не как некоего объема информации, а как доступного для обработки (и практического использования) массива информации.

Традиционно выделяются три признака, характеризующие большие данные, так называемые 3 «V»: Volume (объем), Velocity (скорость) и Variety (разнообразие).

1. Объем. Размер данных сам по себе не является ключевым признаком больших данных. Более того, эксперты подчеркивают относительность этой характеристики: то, что является большим для одного, не будет иметь значительности для другого. Таким образом, любой подход к определению достаточности объема является слишком переменчивым и относительным, чтобы закреплять соответствующее определение в законодательстве.

В литературе отмечается, что «по состоянию на 2011 г. объем больших данных начинался с нескольких десятков терабайт (тысяч гигабайт) и заканчивался десятками петабайт (тысяч терабайт). На данный момент объемы больших данных измеряются уже в зеттабайтах (миллионах петабайт)» [2, с. 120]. Например, в базы данных Facebook ежедневно загружается 500 терабайт новых данных (в основном это фото и видео). Чтобы избежать двойственности, нередко предлагается использовать более простой критерий: большие – это данные, которые не помещаются на одном сервере. Как правило, они хранятся в кластере на нескольких серверах, поэтому работа с ними существенно затрудняется.

2. Скорость. Большие данные формируются с высокой скоростью, что порождает проблему хранения. Для работы с big data необходимы устройства, обладающие объемом не только достаточным, но и динамичным, способным к изменению, в связи с этим одним из принципов работы с большими данными является горизонтальная масштабируемость. Кроме того, возникает проблема скорости обработки информации. С возрастанием объема big data скорость обработки также должна увеличиваться, иначе со временем работа с информацией будет замедлена до критического уровня. Частично эти проблемы решаются в системах, где обработка предполагает использование распределенных вычислений, но вопросов все еще много, а критерий скорости остается достаточно

важным, чтобы закрепить его как характеристику больших данных на законодательном уровне.

3. Разнообразие. Например, один сервис велошеринга способен собирать информацию о погоде, загруженности улиц, разнообразные данные о пользователях и т. д. Эти сведения могут быть использованы для решения разных задач, но нуждаются в систематизации. Белорусский законодатель исходит из того, что большие данные классифицируются на структурированные и неструктурированные, однако на практике возможны промежуточные варианты. Большие данные могут быть структурированными (например, в форме таблиц), неструктурированными (например, результат поиска в «Гугл» – это массив данных, включающий картинки, видео, текстовые файлы) и полуструктурированными (смешанными).

Очевидно, что структурированные данные наиболее доступны для восприятия. Однако не любые массивные структурированные данные могут быть признаны big data в современном понимании. Например, И. В. Воробьева в своей статье исходит из того, что «примером структурированных данных являются многочисленные базы данных, ФМС, ГИБДД, портала Госуслуг и др.» [12, с. 197]. Но высказывание справедливо лишь в том случае, если эти базы данных соответствуют всем признакам big data, а не только объемны. Большие данные – это не просто некий значимый объем информации, они значительно сложнее этого понимания.

Кроме названных трех характерных черт больших данных, ставших уже классическими, в современных исследованиях называются также другие «V»:

4. Value (ценность). Большие данные могут быть использованы для самых разнообразных целей, а потому их ценность очевидна. «Если взглянуть на большие данные с экономической точки зрения, то они, по сути, являются информацией, способной к обороту» [13, с. 93]. И уже это порождает правовые сложности, поскольку в науке до сих пор ведутся дискуссии о способности информации в целом (а не только отдельных ее видов) выступать в качестве объекта гражданско-правовых отношений [14].

Возникает вопрос: если большие данные в силу своей ценности могут быть использованы в качестве товара, то каков правовой режим этого объекта? В литературе big data относят к объектам интеллектуальной собственности [15], к циф-

ровым активам [16], называют конструкцией «имущественного комплекса, которая представляет собой совокупность разнородных вещей, имущественных прав (обязательственных, исключительных прав)» [13], относят к новым правовым режимам, особым информационным услугам, оказываемым на основе оригинальной технологии [17], ноу-хау [2, с. 167].

Считая данный вопрос предметом, достойным отдельной дискуссии, отметим, что на этом фоне выделяется тезис о распространении на большие данные правового режима баз данных: «Наиболее близким объектом гражданских прав, который может воплощать в себе ценность больших данных, является такой результат интеллектуальной деятельности, как база данных» [2, с. 167]. И это мнение вызывает сомнения уже в контексте определения понятия больших данных. Во-первых, при таком подходе неструктурированные большие данные не попадают под правовое регулирование. Во-вторых, по структурированным тоже возникают вопросы практического применения.

В литературе справедливо подчеркивается, что «применительно к хранилищам big data прежде всего надо отметить проблематичность их отнесения к базам данных. Это, в частности, связано с тем, что множество лиц может одновременно собирать и хранить данные, полученные из одних и тех же открытых источников, – в этих условиях не представляется возможным определить, права какого лица будут подлежать защите» [18]. Кроме того, правовым режимом баз данных охраняется ее целостность (весь массив данных), отдельные элементы остаются незащищенными, хотя могут иметь самостоятельную значимость. Существующие проблемы соотношения больших данных и баз данных подтверждаются судебной практикой в Российской Федерации: иск ООО «В Контакте» к ООО «ДАБЛ», иск ООО «Хэдхантер» к ООО «Стафори» и компании «КлаудФлэр Инк.» о защите исключительных прав на базу данных резюме HeadHunter (история данных дел может быть отслежена по новостным публикациям).

5. Veracity (достоверность). Как самостоятельная характеристика, она вытекает из ценности big data. Значимы только истинные данные, поэтому достоверность становится обязательным свойством больших данных.

6. Variability (изменчивость) характеризует не только динамичный характер тех или иных сведений и больших данных в целом, но и такую важную составляющую работы с big

data, как учет контекста. Одна и та же информация в разном контексте может нести различную смысловую нагрузку, и эта вариативность определяет ценность системы целиком и ее взаимосвязь с другими системами, поскольку данные, вырванные из контекста, могут утратить свойство достоверности.

7. Visualization (визуализация) – достаточно неоднозначная характеристика больших данных, которая, возможно, в большей степени связана с восприятием, нежели с big data как таковыми. Предполагается, что именно визуализация позволяет человеку полноценно воспринимать (и анализировать) большие данные. Следовательно, обработка больших данных предполагает обязательность той или иной формы визуализации как важный этап.

**Заключение.** Таким образом, большие данные на современном этапе развития информационного общества характеризуются не просто как данные достаточно большого объема, но как информация, которая имеет размер настолько огромный, что возникают проблемы ее восприятия и обработки. Учитывая этот факт, русскоязычная литература оперирует двумя ключевыми подходами к понятию больших данных – это массив информации и также технология обработки такого массива. С учетом усложнения больших данных традиционные их признаки (объем, скорость и разнообразие) дополняются новыми (ценность, достоверность, изменчивость, визуализация), однако происходящие процессы пока не нашли полноценного отражения в нормативном правовом поле. При этом существующее в белорусском законодательстве определение хоть и применимо только к образовательным стандартам высшего образования II ступени, в целом достаточно точно отражает суть больших данных и может быть использовано в качестве общего подхода, который, однако, следует дополнить некоторыми характеристиками.

### *Список использованных источников*

1. Lyman, P. How much information 2003? [Электронный ресурс] / P. Lyman, H. R. Varian. – Режим доступа: <https://www2.sims.berkeley.edu/research/projects/how-much-info-2003>. – Дата доступа: 28.08.2021.

2. Цифровое право : учебник / под общ. ред. В. В. Блажеева, М. А. Егоровой. – М. : Проспект, 2020. – 640 с.



3. Data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges; (also) the branch of computing involving such data [Electronic resource]. – Mode of access: <http://www.oed.com/view/Entry/18833?eid=301162177>. – Date of access: 07.08.2018.

4. Об утверждении образовательных стандартов высшего образования II ступени [Электронный ресурс] : постановление М-ва образования Респ. Беларусь от 26 июня 2019 г. № 81 : в ред. постановления М-ва образования Респ. Беларусь от 23.06.2020 г. // ЭТАЛОН. Законодательство Республики Беларусь / Нац. центр правовой информ. Респ. Беларусь. – Минск, 2021.

5. Оксфордский словарь английского языка [Электронный ресурс]. – Режим доступа: [https://en.oxforddictionaries.com/definition/big\\_data](https://en.oxforddictionaries.com/definition/big_data). – Дата доступа: 20.09.2021.

6. Санникова, Л. В. Цифровые активы: правовой анализ : монография / Л. В. Санникова, Ю. С. Харитоновна. – М. : Изд-во «4 Принт», 2020. – 304 с.

7. Амбарцумов, Р. А. Правовое регулирование технологии big data в Российской Федерации / Р. А. Амбарцумов // Вопр. рос. юстиции. – 2019. – № 4. – С. 6–11.

8. Борисов, А. В. Современные решения и подходы к обработке массивов неструктурированной текстовой информации в области больших данных / А. В. Борисов // Проблемы науки и образования. – 2017. – № 1 (83). – С. 49–52.

9. Савельев, А. И. Проблемы применения законодательства о персональных данных в эпоху «больших данных» (Big Data) / А. И. Савельев // Право. Журн. Высш. шк. экономики. – 2015. – № 1. – С. 43–66.

10. Arthur, C. Tech giants may be huge, but nothing matches big data [Electronic resource] / C. Arthur. – Mode of access: <http://www.theguardian.com/technology/2013/aug/23/tech-giants-data>. – Date of access: 18.09.2021.

11. Корнев, М. С. История понятия «большие данные» (Big Data): словари, научная и деловая периодика / М. С. Корнев // Вестн. РГГУ. – 2018. – № 1. – С. 81–85. – (Серия «История, филология, культурология, востоковедение»).

12. Воробьева, И. Б. Применение больших данных (big data) при прогнозировании и расследовании преступлений / И. Б. Воробьева // Вестн. СГЮА. – 2021. – № 3. – С. 195–202.

13. Лаптева, А. М. Правовой режим цифровых активов (на примере Big Data) / А. М. Лаптева // Журн. рос. права. – 2019. – № 4. – С. 93–104.

14. Городов, О. А. О нетрадиционных объектах гражданских прав / О. А. Городов // Правоведение. – 2013. – № 6. – С. 96–107.



15. Егорова, М. А. Особенности нормативного регулирования цифровой экономики и проблемы антимонопольного регулирования на цифровых рынках как средство защиты национальных интересов / М. А. Егорова // Юрист. – 2018. – № 11. – С. 7–10.

16. Санникова, Л. В. Цифровые активы как объекты предпринимательского оборота / Л. В. Санникова, Ю. С. Харитонова // Право и экономика. – 2018. – № 4. – С. 27–34.

17. Сергеев, А. П. Большие данные: в поисках места в системе гражданского права / А. П. Сергеев, Т. А. Терещенко // Закон. – 2018. – № 11. – С. 106–123.

18. Рожкова, М. А. Что такое большие данные (big data), чем они отличаются от обычных данных и в чем состоит проблема правового регулирования big data [Электронный ресурс] / М. А. Рожкова // Закон.ру. – Режим доступа: <https://zakon.ru/blog/2019/4/22>. – Дата доступа: 14.09.2021.

29.09.2021

**Yu. P. Gavrilchenko**

Doctor of Law, Professor, Professor of the Department  
of Financial Law and legal regulation of economic activities  
of the Belarusian State University

## **PROBLEMS OF LEGAL REGULATION OF THE CONCEPT OF BIG DATA**

*Abstract.* Big data at the present stage of the development of the information society is characterized not just as data of a sufficiently large volume, but as information that has a size so huge that problems arise with its perception and processing. The article explores legal, theoretical and practical approaches to the concept of big data. The author comes to the conclusion that it is necessary to consolidate the corresponding general definition in the Belarusian legislation. As a basis, it is proposed to use the understanding of big data as an array of data that has value only in the context of appropriate processing. At the same time, it is emphasized that the classic features of big data (volume, speed, diversity) are supplemented with new ones in modern conditions, which further complicates the concept of big data and actualizes the issues of legal regulation of this sphere.

*Key words:* big data, database, volume, speed, diversity, information, information law.