УДК 342.72

«BIG DATA»: НОВАЯ УГРОЗА ДЛЯ ПРАЙВЕСИ В УСЛОВИЯХ ИНФОРМАЦИОННОГО ОБЩЕСТВА

«BIG DATA»: A NEW THREAT TO PRIVACY IN THE INFORMATION SOCIETY

Серёгин В.А.,
доктор юридических наук, профессор,
профессор кафедры конституционного,
муниципального и международного права
Харьковского национального университета имени В.Н. Каразина

Статья посвящена определению проблемных аспектов обеспечения неприкосновенности частной жизни в условиях динамичного развития «Больших данных» как мощной формы интеллектуального анализа данных. В результате исследования получены выводы о необходимости стимулировать бизнес-структуры к внедрению новых бизнес-моделей, основанных на расширении прав потребителей. Предложено дополнить действующее информационное законодательство положениями, касающимися особенностей обработки персональных и неперсональных данных с использованием БД, принять соответствующие профессиональные кодексы поведения в сфере обработки персональных данных.

Ключевые слова: право на неприкосновенность частной жизни, прайвеси, информационное прайвеси, «Большие данные», персональные данные, защита персональных данных.

Стаття присвячена визначенню проблемних аспектів забезпечення недоторканності приватного життя в умовах динамічного розвитку «Великих даних» як потужної форми інтелектуального аналізу даних. У результаті досліджень отримані висновки про необхідність стимулювати бізнес-структури до впровадження нових бізнес-моделей, заснованих на розширенні прав споживачів. Запропоновано доповнити чинне інформаційне законодавство положеннями, що стосуються особливостей обробки персональних і неперсональних даних з використанням БД, прийняти відповідні фахові кодекси поведінки у сфері обробки персональних даних.

Ключові слова: право на недоторканність приватного життя, прайвесі, інформаційне прайвесі, «Великі дані», персональні дані, захист персональних даних.

The article is devoted to the definition of problematic aspects to ensure for inviolability of private life in conditions of «big data» dynamic development as a powerful form of data mining. As a result of research findings obtained on the need to stimulate business structure for the introduction of new business models based on the expansion of the consumers rights. It is proposed to complement existing information legislation provisions concerning the processing of personal and non-personal data using a BD to take appropriate professional codes of conduct in the field of processing of personal data.

Key words: right for inviolability of private life, privacy, informational privacy, „Big data", personal data, protection of personal data.

Постановка проблемы. Бурное развитие информационных технологий наряду с расширением горизонтов общественного развития содержит в себе и серьезные угрозы правам человека, в частности неприкосновенности частной жизни в информационной сфере или, как ее еще называют, информационному прайвеси. Это требует от правой науки адекватного и своевременного ответа на потенциальные угрозы информационному прайвеси, конкретных предложений по совершенствованию действующего законодательства и правоприменительной практики, направленных на формирование соответствующего правозащитного механизма.

Состояние исследования. Проблемы обеспечения информационного прайвеси традиционно находятся в поле зрения североамериканских и западноевропейских исследователей, став со второй половины XX века одним из трендов в развитии западного конституционализма. В этой связи достаточно вспомнить работы А. Аллен, Д. Брина,

Р. Гавизона, Дж. Инесс, Р. Познера, Д. Солова и др., ставшие доктринальной основой современного правового механизма защиты прайвеси во всем цивилизованном мире. Отдельно следует отметить работы П. Ома, Дж. Полонецки, И. Рубинштейн, О. Тене, посвященные непосредственно проблематике «Больших данных» в контексте обеспечения неприкосновенности частной жизни в информационной сфере. Вместе с тем для отечественной юриспруденции данная тематика является совершенно новой, поскольку «Большие данные» в силу серьезного отставания Украины от ведущих стран мира в области информационных технологий только в последние годы стали рассматриваться как реальная угроза неприкосновенности частной жизни.

Целью статьи является постановка проблемы «Больших данных» в конституционно-правовом и правозащитном аспекте, описание возможных угроз, порождаемых данной проблемой для неприкосновенности частной жизни, а также формулирование предложений по совершенствованию государственно-правовой политики Украины в этой сфере.

Изложение основного материала. Под понятием «Big Data» («Большие данные», или БД) принято понимать более мощную форму интеллектуального анализа данных, опирающуюся на огромные объемы информации, суперскоростные компьютеры и новейшие аналитические методы, способные обнаруживать скрытые, а порой даже неожиданные корреляции между фактами и явлениями окружающей действительности. Иными словами, БД представляет собой нетривиальное извлечение ранее неизвестных и потенциально полезных сведений из имеющихся баз данных. Такого рода интеллектуальный анализ опирается не на причинно-следственные связи, а на корреляции, которые обнаруживаются в результате применения соответствующих алгоритмов к большим базам данных. Следовательно, вновь открываемая таким образом информация является не только неинтуитивной и непредсказуемой, но также представляет собой результат достаточно неясного (или, как говорят на Западе, нетранспарентного) процесса.

Международный институт МакКинси (MGI) недавно определил БД как «набор данных, размер которых выходит за пределы возможностей типичного программного обеспечения для сбора, хранения, управления и анализа баз данных» [1]. Все крупнейшие интернет-компании – Google, Facebook, Amazon, eBay, Microsoft и Yahoo! – занимаются Большими Данными в той или иной форме и обращаются с данными в качестве основного актива и источника получения прибыли. В целом, отчет MGI наглядно демонстрирует, что БД могут составлять существенную ценность для мировой экономики, повышения производительности труда, конкурентоспособности компаний и государственного сектора, а также создания значительных экономических выгод для потребителей. Действительно, в результате тематических исследований в докладе MGI утверждается, что БД будут генерировать 300 млрд долл. прибыли в год в отрасли здравоохранения США, 250 млрд евро прибыли в год в европейской системе публичного управления, более 100 млрд долл. дополнительного дохода для поставщиков услуг (данных) о местоположении, увеличат на 60% чистую прибыль в сфере розничной торговли, и до 50% – товарные объемы в производстве [1].

Вместе с тем БД бросают вызов еще недавно казавшейся устойчивой и вполне современной системе международной и национальной защиты персональных данных, а в более широком плане – и сложившемуся к началу XXI века организационно-правовому механизму защиты информационного аспекта права на неприкосновенность частной жизни – так называемого информационного прайвеси. Это происходит двумя путями: с одной стороны, БД ставят по сомнение различия между персональными и неперсональными данными, а с другой – вступают в противоречие с принципом минимизации данных и подрывают принцип осознанного выбора.

Дело в том, что БД присущи три характерные черты. Во-первых, это доступность данных в массовом масштабе, собранных не только в Интернете или за счет использования мобильных устройств (с возможностью отслеживания местонахождения клиента и тысячей «приложений», разделяющих данные с нескольких сторон, взаимодействия со смарт-средой, системой мониторинга в физической среде), но также и за счет организма самого человека, который в настоящее время используется и как пучок данных для генетического тестирования, и как объект для аутентификации с помощью биометрических данных. Кроме того, Web 2.0 сервисы позволяют пользователям создавать и добровольно делиться с другими огромным количеством персональных данных о себе, своих друзьях и родственниках. Хотя лица разглашают эти данные в основном добровольно, для определенных социальных целей, организации рады собирать и получать прибыль от их анализа. Второй особенностью БД является использование компьютеров с высокой скоростью передачи данных в сочетании с петабайтами (т.е. миллионами гигабайт) емкости памяти, в результате чего обработка данных становится дешевой и эффективной. Еще более повышает этот эффект использование «облачных» технологий. Третей и последней особенностью БД является использование новых вычислительных структур (как-то Apache Hadoop) для хранения и анализа этого огромного объема данных.

В свете этих трех характеристик невозможно переоценить то огромное изобилие цифровых данных, которые теперь доступны организациям, а также новых способов, с помощью которых БД объединяют эти различные наборы данных. Поэтому не удивительно, что феномен БД существенно актуализует и усугубляет существующие проблемы юридической защиты частной жизни от дальнейшего отслеживания и профилирования. С приходом БД, куки и веб-маяки больше не являются главны-

ми виновниками проблем. Скорее, технологии профайлинга, распространяемые сейчас на все аспекты и фазы индивидуальной и социальной жизни, с БД приобретают значительно большую мощь для того, чтобы найти скрытые корреляции и сделать интересные предсказания, некоторые из которых могут принести пользу отдельным лицам или даже обществу в целом, в то время как другие могут быть более чем проблемными.

БД закономерно вызывают у правозащитников целый ряд сомнений и порождают попытки обратить внимание общественности на два основных вопроса: о неприкосновенности частной жизни (прайвеси) и дискриминации. С учетом ограниченности объема данной работы обратимся исключительно к неприкосновенности частной жизни, оставляя менее важные вопросы, касающиеся дискриминации, в качестве темы для дальнейших исследований.

Наиболее совершенным и эффективным механизмом защиты информационного прайвеси в современном мире считается тот, который существует в рамках ЕС. Нормативную основу этого механизма составляет Директива о защите данных (Data Protection Directive) [2]. В настоящее время обсуждаются Общие правила защиты данных (General Data Protection Regulation) [3], призванные заменить существующую Директиву. Эти Правила предполагают признание новых прав человека и в то же время возложение новых мер ответственности на организации, занимающиеся сбором и обработкой персональных данных. Однако похоже, что БД, подобно цунами, способны сокрушить все эти реформаторские усилия.

Дело в том, что БД бросают вызов самим основам европейской Директивы о защите данных (и всех подобных законов о защите информационного прайвеси, в частности и Закона Украины «О защите персональных данных» [4]), позволяя реидентифицировать субъектов данных, используя неличные (обезличенные) данные, что крайне ослабляет возможности анонимизации как эффективной стратегии, поскольку ставит под сомнение фундаментальное различие между личными и неличными данными. БД также значительно усугубляют моральный ущерб, связанный с накоплением информации о человеке, – то, что профессор Д. Солов называет аггрегацией (aggregation) [5, с. 477, с. 506]. В своем массовом масштабе постоянный мониторинг с использованием различных источников и сложных аналитических возможностей БД делают аггрегацию более детальной, более показательной и в то же время более агрессивной.

Безусловно, реидентификация только усиливает вред, связанный с аггрегацией, позволяя контроллерам данных «привязать» еще больше сведений к имеющейся информации об индивидууме, приводя в итоге к тому, что Дж. Ом называет «базой данных разорения» [6]. БД также актуализируют вопрос об автоматизированном принятии решений, касающихся жизнедеятельности человека, – таких, как кредитные рейтинги, перспективы трудоустройства и право на страховое покрытие или социальное пособие, – автоматизированных процессов, основанных на алгоритмах и искусственном интеллекте.

О влиянии БД на действующее законодательство о защите информационного прайвеси говорилось выше. Однако БД имеют еще более широкое воздействие на законодательство о защите данных, в частности на Директиву ЕС о защите данных и ориентированное на нее национальное законодательство. Напомним, что упомянутая Директива в основном опирается на принципы прозрачности и согласия, чтобы пользователи делали осознанный выбор в отношении обмена личными данными с организациями. Директива содержит многие другие требования по обработке данных (ограниченное целеполагание, качество данных, безопасность, доступ и т.д.), однако все они, кроме требования безопасности, имеют достаточно ограниченное применение, поскольку непосредственно зависят от осознания индивидуумом того, какие данные о нем были обработаны.

Интеллектуальный анализ данных и БД серьезно усугубляют эту проблему, поскольку подрывают основы «информированного выбора» в трех направлениях.

Во-первых, фирмы, которые полагаются на поиск полезных данных, могут оказаться не в состоянии обеспечить надлежащее уведомление субъекта данных по той простой причине, что они не знают (и не могут знать) заранее, что они могут обнаружить. Во-вторых, поскольку пользователи не обладают информацией о потенциальных корреляциях, они не могут сознательно дать согласие на использование своих данных для интеллектуального анализа. В-третьих, законы о неприкосновенности частной жизни применяются только к персональным данным, то есть данным, относящимся к идентифицированным или идентифицируемым лицам. В то же время вовсе не ясно, будут ли принципы, составляющие ядро современной защиты информационного прайвеси, – прозрачность, согласие, минимизация данных, доступ и т.д. – забыты по отношению к переносимым данным, либо они будут применяться и к вновь открывшимся знаниям, полученным от персональных данных, особенно тогда, когда эти данные, которые были анонимными или обобщенными, превращаются в групповые профили, то есть профили, применяемые к лицам в качестве членов контрольной группы, хотя данное лицо может на самом деле и не проявлять данное свойство.

БД также ставят под сомнение три устоявшиеся нормативные положения законодательства о прайвеси, в т.ч. и Директивы о защите данных.

Во-первых, остается ли юридически состоятельным разделение данных на персональные и неперсональные. Как выше было отмечено, интеллектуальный анализ данных извлекает новые знания как из персональных, так и из неперсональных данных, порождая, таким образом, нормативную дилемму: следует ли Директивой о защите данных охватывать не только персональные данные, но также и любые

неперсональные данные, которые формирует основу для экстракции данных нового знания, и (будучи единожды созданными) будут ли они регулироваться как персональные данные? Если это так, то для рамок Директивы о защите данных не существует потенциально никаких ограничений; если нет, то интеллектуальный анализ данных в значительной степени может избежать регулирования и контроля, хотя он позволяет сделать выводы о ранее частной информации и/или использовать групповые профили, способные причинить столько же или даже больше вреда как регулируемым базам, так и использованию персональных данных.

Во-вторых, анонимизация – процесс удаления идентификаторов для создания анонимных наборов остаточных данных – еще недавно казался вполне эффективным для защиты пользователей от отслеживания и профилирования. Однако за последние несколько лет было несколько печально известных случаев реидентификации лиц по кросс-ссылкам анонимных наборов данных с соответствующим набором данных, который включал идентификаторы. Как уже отмечалось, БД усиливает проблему, опираясь на дополнительные данные, высокоскоростные компьютеры, а также улучшение аналитических методов.

В-третьих, является ли минимизация данных – идея, что обработка персональных данных должна быть ограничена до минимально необходимой суммы, – способной противостоять натиску БД. Проще говоря, минимизация данных враждебна по отношению к основной направленности БД, которая открывает новые корреляции с применением изощренных аналитических методов к массовому сбору данных, и стремится сделать это без каких-либо ограничений. Поскольку требования к минимизации данных будут вредить БД и связанным с ними экономическим и социальным выгодам, регуляторы должны ожидать увидеть, что это требование в целом соблюдаться не будет.

На наш взгляд, разделение данных на персональные и неперсональные продолжает оставаться юридически состоятельным и в условиях существования «Больших данных». Более того, указанное разделение является научно-методологической основой для дальнейшего совершенствования правозащитного механизма, направленного на обеспечение информационного прайвеси. В этой связи европейской Директивой о защите персональных данных и национальным законодательством в этой сфере должны охватываться не только персональные данные, но также и любые неперсональные данные, которые формирует основу для экстракции новых данных, способных идентифицировать личность. Однако соответствующие ограничительные меры должны касаться неперсональных данных только в части их обработки и распространения полученной в результате их интеллектуального анализа идентифицирующей информации.

В то же время «Большие данные» не дают достаточных оснований для того, чтобы отказываться от анонимизации. Прежде всего, следует учитывать, что использование БД-технологий – крайне затратный процесс, позволить себе который могут лишь немногие, финансово состоятельные компании. К тому же нужно четко осознавать, что анонимизация не является панацеей от раскрытия персональных данных, а лишь входит в общий организационно-правовой механизм обеспечения информационного прайвеси в качестве важной его составляющей. Ведь проблема на самом деле – не в самой возможности преодоления барьера анонимизации, а в отсутствии должных рычагов воздействия на распростанителей итоговой, полученной в результате многопланового анализа идентифицирующей информации. При распространении принципов защиты информационного прайвеси и на обработку неперсональных данных, прежде всего в части распространения итоговых данных, – многие потенциальные угрозы, исходящие сегодня от БД, будут в значительной степени снивелированы.

В свою очередь, на первый план в борьбе с угрозами для информационного прайвеси, исходящими от БД, выходит минимизация данных. Естественно, что заинтересованные субъекты будут стараться от такой минимизации уклониться, обосновывая необходимость в получении все новых и новых данных, в том числе – и персонального характера. Однако в этом случае должны вступать в действие государственные регуляторы, общественные правозащитные организации и, естественно, механизмы судебной защиты.

Выводы. Проведенное исследование свидетельствует о необходимости разработки и принятия комплексных мер по защите информационного прайвеси от потенциальных угроз, исходящих от «Больших данных». В частности, действующее информационное законодательство следует дополнить положениями, касающимися особенностей обработки персональных и неперсональных данных с использованием БД, принять соответствующие профессиональные кодексы поведения в сфере обработки персональных данных. В свою очередь, органы, осуществляющие регулирование и контроль в данной сфере, должны стимулировать бизнес-структуры к внедрению новых бизнес-моделей, основанных на расширении прав потребителей, используя при этом такие меры, как гибкое регулирование и снижение штрафных санкций. Особое внимание должно быть уделено правозащитной практике, направленной на неуклонное соблюдение всеми участниками информационных отношений принципа минимизации данных, если это касается вопросов обеспечения информационного прайвеси.

Выработка конкретных предложений по совершенствованию информационного законодательства в аспекте противодействия угрозам информационному прайвеси, исходящим от БД, с учетом передового зарубежного опыта, является перспективным направлением дальнейших исследований в данной сфере.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ:

1. Big Data: The Next Frontier for Innovation, Competition, and Productivity ; McKinsey Global Institute, May 2011 [Электронный ресурс]. – Режим доступа : http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

2. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data // Official Journal of the European Communities. – 1995. – L 281. – P. 31-50.

3. Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), Brussels, 25.1.2012 [Электронный ресурс]. – Режим доступа : http://ec.europa.eu/justice/ data-protection/document/review2012/com_2012_11_en.pdf.

4. Закон України «Про захист персональних даних» від 1 червня 2010 р. № 2297-VI [Электронный ресурс]. – Режим доступа : http://zakon4.rada.gov.ua/ laws/show/2297-17.

5. Solove D.J. A Taxonomy of Privacy / D.J. Solove // Pennsylvania Law Review. – 2006. – Vol. 154. – № 3. – P. 477-560.

6. Ohm P. Don't Build a Database of Ruin / P. Ohm // Harvard Business Review. – 2012. – August 23.