

Data policy: нормативные и организационные барьеры для доступа к большим медицинским данным



В 2024 году инициативной группой в составе юридической фирмы VERBA Legal, НБМЗ и компании Цифровая Медицина при поддержке фармацевтической компании АО «Рош-Москва» было подготовлено аналитическое исследование с анализом российских и зарубежных подходов к доступу и обработке больших данных в здравоохранении. Были проанализированы локальные нормативные барьеры и международный опыт разных стран. Актуальность исследования была обусловлена несколькими причинами:

- Во-первых, большая часть медицинской информации переведена в цифровой вид в рамках первого этапа построения ЕГИСЗ и ЕЦК, сформирована базовая инфраструктура для цифровой трансформации системы здравоохранения. В связи с этим, вместе с переходом к «экономике данных» актуальным становится обеспечение возможности организации широкого, но регламентированного доступа к обезличенным медицинским данным.
- Во-вторых, сформированные механизмы доступа к большим медицинским данным могут позволить проводить глубинную аналитику потребностей населения и выводить управленческие и финансовые решения о лекарственном обеспечении на новый качественный уровень, обеспечив «прозрачность» проводимых процедур, а также их эффективность в условиях ограничений финансового и организационного характера.

Актуальность использования больших медицинских данных для фармацевтической индустрии

Будучи полноправными участниками системы здравоохранения, фармацевтические компании являются как источником (например, компании генерируют данные в рамках клинических исследований, проводят постоянный анализ потребностей врачей и пациентов, контроль товарных остатков лекарственных препаратов и т.д.), так и потребителем данных из системы здравоохранения. В связи с чем крайне важным для эффективной и корректной работы индустрии в части выполнения рыночных обязательств и создания востребованных здравоохранением инновационных продуктов является вопрос формирования прозрачной и понятной регуляторной среды обмена медицинскими данными.

Сегодня фармацевтическая индустрия в России и в мире выступает в качестве партнера системы здравоохранения – компании сектора готовы инвестировать в региональные и федеральные проекты по разработке и внедрению цифровых инструментов в здравоохранение, в том числе решений на основе искусственного интеллекта, компьютерного зрения, персональных цифровых помощников. Однако часть из таких инструментов может остаться невостребованной или использоваться ограниченно, поскольку отсутствует полноценный доступ к данным, на основе которых такие инструменты могут быть разработаны или использованы. Для эффективных инвестиций и успешного партнерства требуется уточнение стратегии доступа к данным и совместные усилия всех вовлеченных субъектов индустрии.

Потенциальные источники больших медицинских данных

Большая часть медицинских данных, которые потенциально могут быть использованы для вторичного анализа, формируется в рамках работы следующих элементов инфраструктуры данных:

Источник данных	«Держатель» источника данных
ЕГИСЗ	Министерство здравоохранения РФ
Архивы медицинских организаций	Соответствующие медицинские организации
Порталы региональных МИС	Региональные департаменты здравоохранения
Порталы ведомственных МИС	Соответствующие органы государственной власти
ВИМИС	НМИЦ
Дата-посредники (формальный статус не определен)	Владельцы датасетов
Технические интеграторы	Зависит от технического решения
Пациенты, получающие свои данные из различных источников	Пациент

Однако использование данных из указанных источников для целей, отличных от целей их сбора, на текущий момент ограничено в силу ряда технологических и нормативных барьеров.

НОРМАТИВНЫЕ БАРЬЕРЫ ДЛЯ ПОЛУЧЕНИЯ ДОСТУПА К ДАННЫМ О ЗДОРОВЬЕ

Данные о здоровье пациента относятся к персональным данным специальной категории, поэтому для их обработки, в том числе, получения доступа к таким данным, необходимо обеспечить соблюдение требований Федерального закона от 27 июля 2006 года № 152-ФЗ «О персональных данных» (далее – «152-ФЗ»). При этом многие требования, продиктованные 152-ФЗ и принятыми в его исполнение подзаконными актами, являются неоднозначными, либо в значительной степени строгими и ограничивающими возможность вторичной обработки данных о здоровье, в том числе обезличенных, что порождает нормативные барьеры для развития отрасли в направлении цифровизации и масштабирования работы с данными.

Барьеры, связанные с необходимостью получения согласия на обработку данных

На текущий момент, правовой статус обезличенных медицинских данных в законодательстве не определен однозначно, понятие анонимизированных данных отсутствует, а обезличенные данные при буквальном толковании действующих норм сохраняют статус персональных данных специальной категории (данные о здоровье). Фактически единственным правовым основанием для обработки данных о здоровье, применимым к деятельности субъектов рынка, является получение согласия на обработку от субъекта данных, которое должно быть получено в письменной форме (бумажной, либо в форме электронного документа с использованием усиленной квалифицированной электронной подписи (УКЭП) или ПЭП при соблюдении ряда условий, опосредующих возможность ее применения). Получение такого согласия в строгом соответствии с требованиями 152-ФЗ затруднено или практически нереализуемо, что обусловлено, помимо соблюдения формы согласия, также следующими сложностями:

- Согласие должно быть оформлено отдельно для каждой цели обработки данных, при этом для случая обработки больших медицинских данных перечень соответствующих целей практически невозможно указать на этапе получения соответствующего согласия, а указание слишком широкой цели или вторичное использование данных, не связанных с первоначальной целью сбора, может рассматриваться как неправомерное.
- В согласии должны быть поименованы все субъекты, кому оператор вправе поручать обработку данных, что при вторичной обработке технически невозможно. При отсутствии поручения на обработку от первоначального оператора, другой самостоятельный оператор персональных данных должен иметь отдельно полученное согласие субъекта.

Требования к содержанию согласия обязывают указать в самой форме согласие большое количество сведений о субъекте персональных данных (ФИО, адрес, номер документа, удостоверяющего личность и т.д.). На текущий момент в информационных системах в сфере здравоохранения (ЕГИСЗ, РМИСы, МИСы) уже накоплено большое количество медицинских данных, которые хранятся в обезличенном виде. Возможность их использования на основании согласия субъекта этих данных означает, что такие данные должны быть сначала деобезличены для возможности идентификации субъекта таких данных, а затем у таких субъектов должно быть получено письменное согласие на обработку их данных, что только увеличивает уровень чувствительности и уязвимости таких данных и совершенно не способствует обеспечению их конфиденциальности и снижению ущерба в случае несанкционированного доступа.

- Согласие на обработку персональных данных может быть отозвано субъектом персональных данных в любой момент, что создаст для операторов обязанность уничтожить соответствующие данные, которые, в свою очередь, могут быть уже многократно обезличены и вторично обработаны. Аналогичным образом субъектом в адрес оператора могут быть направлены запросы на уточнение неактуальных данных.

Указанные ограничения делают модель вторичной обработки медицинских данных на основе согласия субъекта персональных данных нереализуемой с практической точки зрения.

Барьеры, связанные с необходимостью получения согласия на обработку данных:

На текущий момент, правовой статус обезличенных медицинских данных в законодательстве не определен однозначно, понятие анонимизированных данных отсутствует, а обезличенные данные при буквальном толковании действующих норм сохраняют статус персональных данных специальной категории (данные о здоровье). Фактически единственным доступным правовым основанием для обработки данных о здоровье, применимым к деятельности субъектов рынка, является получение согласия на обработку от субъекта данных, которое должно быть получено в письменной форме (бумажной, либо в форме электронного документа с использованием усиленной квалифицированной электронной подписи (УКЭП) или ПЭП при соблюдении ряда условий, опосредующих возможность ее применения). Получение такого согласия в строгом соответствии с требованиями 152-ФЗ затруднено или практически нереализуемо, что обусловлено, помимо соблюдения формы согласия, также следующими сложностями:

- Согласие должно быть оформлено отдельно для каждой цели обработки данных, при этом для случая обработки больших медицинских данных перечень соответствующих целей практически невозможно указать на этапе получения соответствующего согласия, а указание слишком широкой цели или вторичное использование данных, не связанных с первоначальной целью сбора, может рассматриваться как неправомерное.
- В согласии должны быть поименованы все субъекты, кому оператор вправе поручать обработку данных, что при вторичной обработке технически невозможно. При отсутствии поручения на обработку от первоначального оператора, другой самостоятельный оператор персональных данных для последующей обработки должен иметь отдельно полученное согласие субъекта.
- Требования к содержанию согласия обязывают указать в самой форме согласие большое количество сведений о субъекте персональных данных (ФИО, адрес, номер документа, удостоверяющего личность и т.д.). На текущий момент в информационных системах в сфере здравоохранения (ЕГИСЗ, РМИСы, МИСы) уже накоплено большое количество медицинских данных, которые хранятся в обезличенном виде. Возможность их использования на основании согласия субъекта этих данных означает, что такие данные должны быть сначала деобезличены для возможности идентификации субъекта таких данных, а затем у таких субъектов должно быть получено письменное согласие на обработку их данных, что только увеличивает уровень чувствительности и уязвимости таких данных и совершенно не способствует обеспечению их конфиденциальности и снижению ущерба в случае несанкционированного доступа.
- Согласие на обработку персональных данных может быть отозвано субъектом персональных данных в любой момент, что создаст для операторов обязанность уничтожить соответствующие данные, которые, в свою очередь, могут быть уже многократно обезличены и вторично обработаны. Аналогичным образом субъектом в адрес оператора могут быть направлены запросы на уточнение неактуальных данных.

Указанные ограничения делают модель вторичной обработки медицинских данных на основе согласия субъекта персональных данных нереализуемой с практической точки зрения.

Барьеры, связанные с отсутствием проработанной нормативной базы в части обезличивания персональных данных и их дальнейшего использования:

На текущий момент в законодательстве о персональных данных отсутствует достаточное регулирование вопросов, связанных с проведением процедуры обезличивания персональных данных и дальнейшего их использования. В частности, среди основных проблем можно выделить:

- Отсутствие легального определения понятия обезличенных данных и, в связи с этим, неоднозначность статуса данных, полученных в результате обезличивания.
- Не закреплены понятия и критерии разграничения псевдонимизированных и анонимизированных данных, разграничение понятий «обезличивание» и «анонимизация».
- Отсутствие установленных правил обезличивания персональных данных коммерческими субъектами. На текущий момент процедура обезличивания персональных данных регулируется Приказом № 341н и Приказом № 996, которые имеют ограниченную сферу действия и не распространяются на обезличивание персональных данных коммерческими субъектами.
- Установлен узкий перечень законных оснований для обработки обезличенных данных о здоровье, отличных от согласия субъекта персональных данных, в частности, обработка данных в статистических или исследовательских целях возможна только для общей категории персональных данных, но не для медицинских данных как персональных данных специальной категории.

ПОТЕНЦИАЛЬНЫЕ СПОСОБЫ ПРЕОДОЛЕНИЯ ВЫЯВЛЕННЫХ НОРМАТИВНЫХ ОГРАНИЧЕНИЙ

По результатам проведенного анализа можно выделить следующие направления работы, имеющие целью совершенствование действующего нормативного регулирования в области обработки обезличенных медицинских данных:

Совершенствование нормативной базы в части процедуры обезличивания персональных данных:

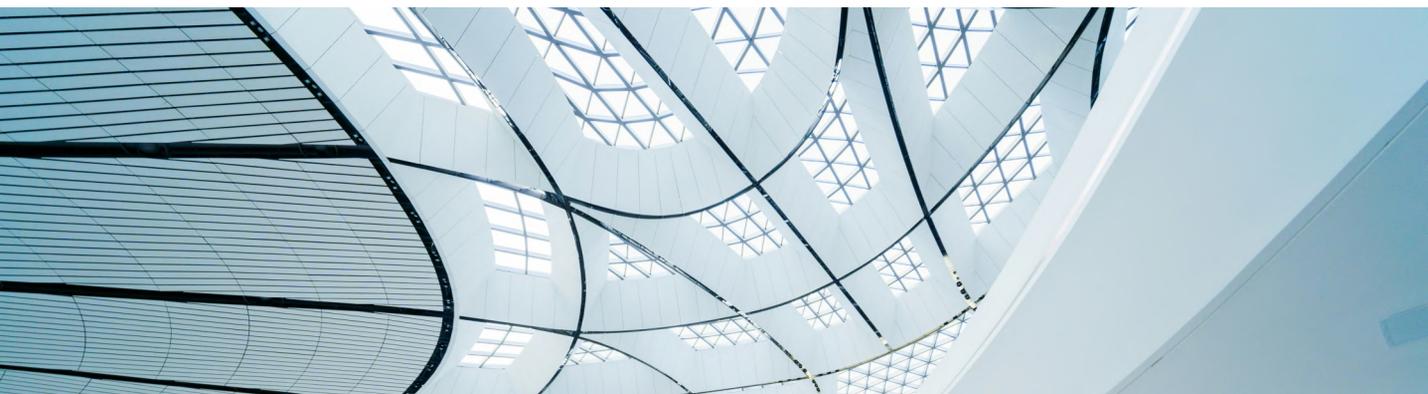
- Уточнение статуса обезличенных данных путем введения понятий псевдонимизированных и анонимизированных данных и условий их обработки;
- Разработка правил и методов обезличивания персональных данных для коммерческих субъектов путем принятия нового правового акта или распространения действующего регулирования на более широкий круг субъектов;
- Разработка системы оценки рисков деобезличивания датасетов и введение обязанности оператора осуществлять такую оценку при формировании датасета, содержащего псевдонимизированные или анонимизированные данные (в зависимости от степени и глубины обезличивания) с составлением файла оценки риска.

Совершенствование нормативной базы в части обеспечения наличия правовых оснований для вторичной обработки обезличенных медицинских данных:

- Введение дополнительного основания для обработки персональных данных специальной категории, касающихся состояния здоровья – обработка в исследовательских целях, а также конкретизация понятия «исследовательские цели».
- Введение исключения из принципа ограничения обработки данных только конкретными и заранее определёнными целями, а также допущение изменения целей обработки при обработке данных в исследовательских и статистических целях.
- Дополнение Положения о ЕГИСЗ регламентацией предоставления доступа к обезличенным данным для «внешних» пользователей, организация процесса создания и систематизации баз данных по нозологиям.

Механизмы преодоления технологических барьеров для организации доступа к данным.

- Разработка стандартов обмена медицинскими данными:
 - Широкое внедрение специализированных стандартов на основе FHIR (в виде ГОСТ или ПНСТ).
 - Разработка единой модели медицинских данных и новых СЭМД (на ее основе) для конкретных групп потребителей данных;
 - Создание постоянной общественной рабочей группы для взаимодействия регулятора с потребителями данных.
- Разработка регламента взаимодействия с техническими операторами, обеспечивающими интероперабельность данных и систем («шины данных»), внедрение в систему обмена данных провайдеров, имеющих статус «иной информационной системы».
- Изучение релевантных зарубежных практик по моделям организации доступа к данным на предмет возможностей имплементации схожих подходов:
 - Внедрение модели доверенного третьего лица (дата-посредника): использование защищенных внешних платформ для доступа к данным. Платформы могут быть реализованы на базе государственных или квазигосударственных структур: FinData (Финляндия), Health Data Hub (Франция), FDZ Gesundheit (Германия).
 - Введение механизма предварительной оценки и одобрения запроса на доступ к данным в качестве обязательного этапа получения доступа. Обеспечение безопасности обрабатываемых данных в рамках защищенных сред.
 - Внедрение дополнительных обязательств для получения данных с целью проведения научных исследований: в некоторых странах введено требование обязательной публикации научных работ по итогам использования данных, к которым запрашивался доступ (Финляндия, Франция, Австралия).



ОДИН ИЗ ПРИМЕРОВ ЗАРУБЕЖНОГО ОПЫТА ОРГАНИЗАЦИИ ДОСТУПА К ДАННЫМ ДЛЯ ВТОРИЧНОГО ИСПОЛЬЗОВАНИЯ: БРАЗИЛИЯ

Основным нормативным правовым актом, регулирующим защиту данных в Бразилии, является Закон о защите персональных данных от 14 августа 2018 года - General Personal Data Protection Law (LGPD). LGPD представляет собой наиболее полный нормативный документ о защите данных, который охватывает деятельность контролеров и обработчиков данных и устанавливает требования к обработке данных.

В законодательстве Бразилии разделены понятия псевдонимизированных и анонимизированных данных. Под псевдонимизацией понимается обработка данных, посредством которой данные больше не могут быть прямо или косвенно связаны с физическим лицом, за исключением использования дополнительной информации, хранящейся отдельно контролером в контролируемой и безопасной среде. Псевдонимизированные данные остаются персональными данными. При этом анонимизированные данные (под которыми понимаются данные, относящиеся к субъекту, на основании которых он не может быть идентифицирован с учетом использования разумных и доступных на момент обработки технологий) уже не считаются персональными данными для целей регулирования LGPD, в связи с чем на них не распространяются требования, применимые к обработке персональных данных.

В отношении осуществления процесса обезличивания уполномоченным органом в Бразилии было издано разъяснение о порядке и способах анонимизации данных, в котором предусмотрена необходимость выполнения следующих действий:

- **Документирование процесса анонимизации.** В ходе процедуры оператор должен фиксировать совершаемые с данными манипуляции, в том числе с целью недопущения искажения изначального набора данных;
- **Оценка рисков.** Оператору необходимо оценивать риски повторной идентификации данных с учетом их характера и уникальности, а также учитывать риски при выборе конкретного метода обезличивания;
- **Учет ограничений методов обезличивания.** Оператор обязан учитывать особенности методов обезличивания и особенности категорий обезличиваемых данных для того, чтобы снижать риск идентификации субъектов и поддерживать уровень их полезности для целей обработки.

Также в данном разъяснении приводятся примеры конкретных методов, которые могут быть использованы для обезличивания данных, а также анализ ограничений таких методов. Для текстовых данных предпочтительными являются метод замены данных случайными значениями, метод перестановки атрибутов данных и метод обобщения данных. При обезличивании изображений возможно использование методов размытия, пикселизации или, например, маскировки глаз. Описываемые методы приводятся в качестве примеров и не являются исчерпывающим перечнем, ключевым фактором при выборе конкретного способа обезличивания данных является контроль рисков повторной идентификации субъектов с учетом категории обезличиваемых данных.

Проект CIDACS

Министерством здравоохранения Бразилии основан Центр CIDACS. Его целью является интеграция государственных баз данных, созданных различными государственными органами, а также расширение возможностей для их использования в научных исследованиях в области здравоохранения и медицины.

CIDACS разработал платформу, способную обеспечить условия для управления данными в рамках реализации исследовательских проектов, в том числе предоставлять доступ к данным для третьих лиц. Платформа позволяет собирать, хранить и изучать большие объемы данных, а также каталогизировать данные. Деятельность CIDACS основана на использовании данных, поступающих от различных государственных ведомств и агентств, которые были собраны ими в процессе предоставления государственных услуг и реализации государственных программ в социальной сфере (медицинские услуги, социальное обеспечение и т.п.).

Использование баз данных на платформе в настоящий момент возможно только при заключении соглашения между CIDACS и научной некоммерческой или государственной организацией для реализации конкретного проекта, прошедшего этическую экспертизу, а также методологическую оценку. Платформа позволяет использовать сведения из хранилища данных национальной информационной системы здравоохранения (DATASUS), которая объединяет все население страны.

В рамках платформы авторизованным пользователям по согласованным проектам может предоставляться доступ, в частности, к обезличенным данным следующих категорий:

- Смертность (причина смерти, дата смерти, дата рождения, пол, образование, наличие детей и др.).
- Рождаемость (дата рождения матери, образование, семейное положение, количество детей и аборт, количество родовых консультаций, срок беременности, тип родов, дата рождения ребенка, пол, раса, вес при рождении, наличие врожденных заболеваний и пр.).
- Проведенные госпитализации (причина госпитализации, описание болезни, диагноза, назначенное лечение, собранные анализы).
- Информация о медицинских учреждениях (число больничных коек, оборудование, профиль учреждения, количество персонала и их специализация, виды оказываемых медицинских услуг).

Также в рамках CIDACS может быть предоставлен доступ к базе социально-демографических сведений (возраст, пол, раса, сведения о зарплате и т.п.) 100 миллионах граждан Бразилии, которые когда-либо обращались за любой государственной социальной помощью (база CadUnico).

Перед началом проведения исследования специалисты CIDACS проводят процедуру псевдонимизации или анонимизации данных (зависит от потребности конкретного исследования) в безопасной среде путем замены прямых и косвенных идентификаторов конкретного субъекта хэш-кодами, что обеспечивает невозможность процесса идентификации субъектов данных. Исследователи могут работать только с псевдонимизированными и анонимизированными данными, а также обязуются соблюдать требования LGPD при получении доступа к данным CIDACS.

Деятельность по предоставлению доступа и по дальнейшему использованию медицинских данных подпадает под специально установленное правовое основание для обработки данных: «допускается обработка обезличенных данных о здоровье строго в целях проведения научных исследований» (ст. 7, 11, 13 LGPD).

На портале CIDACS приводится список из более чем 200 исследований, которые были проведены совместно с центром CIDACS с 2016 года университетами, научными институтами, медицинскими учреждениями и государственными органами, например в 2023 году были проведены: исследование по вопросу постнатального развития у 2 млн. новорожденных; международное исследование фенотипов воспаления при астме у пациентов, которые проживают на четырех континентах.

Контакты:



Александр Панов

Старший партнер юридической фирмы
VERBA Legal

alexander.panov@verba.legal



Карина Колобова

Старший юрист практики фармацевтики
и здравоохранения юридической фирмы
VERBA Legal

karina.kolobova@verba.legal

info@verba.legal

+7 (495) 374 74 03

**Гоголевский бульвар.д 11,
Москва, 119019**



t.me/verbalegal



t.me/IAR_VERBA



t.me/ma_nagogolevskom