ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 55036 — 2012 ISO/TS 25237:2008

Информатизация здоровья

ПСЕВДОНИМИЗАЦИЯ

ISO/TS 25237:2008
Health informatics — Pseudonymization
(IDT)

Издание официальное

Москва
Стандартинформ
2013

Предисловие

Цели и принципы стандартизации в Российской Федерации установлены Федеральным законом от 27 декабря 2002 г. № 184-ФЗ «О техническом регулировании», а правила применения национальных стандартов Российской Федерации — ГОСТ Р 1.0—2004 «Стандартизация в Российской Федерации. Основные положения»

Сведения о стандарте

1 ПОДГОТОВЛЕН Федеральным государственным учреждением «Центральный научно-исследовательский институт организации и информатизации здравоохранения Минздравсоцразвития» (ФГУ ЦНИИОИЗ Минздравсоцразвития) и Обществом с ограниченной ответственностью «Корпоративные электронные системы» на основе собственного аутентичного перевода на русский язык международного документа, указанного в пункте 4

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 468 «Информатизация здоровья» при ФГУ ЦНИИОИЗ Минздравсоцразвития — единоличным представителем ИСО ТК 215

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 29 октября 2012 г. № 585-ст

4 Настоящий стандарт идентичен международному документу ИСО/ТС 25237:2008 «Информатизация здоровья. Псевдонимизация» (ISO/TS 25237:2008 «Health informatics — Pseudonymization»)

При применении настоящего стандарта рекомендуется использовать вместо ссылочных международных стандартов соответствующие им национальные стандарты Российской Федерации, сведения о которых приведены в дополнительном приложении ДА

5 ВВЕДЕН ВПЕРВЫЕ

Информация об изменениях к настоящему стандарту публикуется в ежегодно издаваемом информационном указателе «Национальные стандарты», а текст изменений и поправок — в ежемесячно издаваемом информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ежемесячно издаваемом информационном указателе «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет

Содержание

Введение

Псевдонимизация — один из основных методов защиты конфиденциальности персональной медицинской информации.

Службы псевдонимизации могут использоваться как внутри страны, так и при трансграничной передаче данных.

Предметные области, в которых используются псевдонимы пациентов, включают в себя:

- вторичное использование медицинских данных (например, для научных исследований);
- клинические испытания и пострегистрационный мониторинг побочных действий лекарственных препаратов;
- анонимное лечение;
- системы идентификации пациентов;
- мониторинг и оценку состояния общественного здоровья;
- конфиденциальное информирование об угрозах безопасности пациентов (например, о побочных эффектах лекарственных средств);
- сопоставление показателей качества медицинской помощи;
- медицинскую экспертизу;
- защиту прав потребителей;
- обслуживание медицинских приборов.

В настоящем стандарте предлагаются: концептуальная модель областей применения псевдонимов; требования к надежной реализации псевдонимизации; указания, необходимые для обеспечения планирования и внедрения служб псевдонимизации.

Описание общего процесса применения псевдонимов в сочетании с политикой псевдонимизации может использоваться в качестве общего руководства для разработчиков систем, а также для оценки качества, необходимой для определения степени доверия к предлагаемым службам псевдонимизации.

Для обеспечения интероперабельности поставщиков и получателей псевдонимов, систем, предоставляющих службы псевдонимизации, систем управления учетными записями в настоящем стандарте определены интерфейсы служб псевдонимизации.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Информатизация здоровья

ПСЕВДОНИМИЗАЦИЯ

Health informatics. Pseudonymization

Дата введения — 2013—07—01

1 Область применения

В настоящем стандарте описаны принципы обеспечения конфиденциальности персональной информации с помощью служб псевдонимизации, предназначенных для защиты персональной медицинской информации, а также требования к этим службам. Он адресован организациям, гарантирующим доверенность операций, связанных с применением служб псевдонимизации.

В настоящем стандарте:

- определена базовая концепция псевдонимизации;
- приведен обзор различных сценариев обратимой и необратимой псевдонимизации;
- определена базовая методология служб псевдонимизации, охватывающая как технические, так и организационные аспекты;
- приведено руководство по оценке угрозы восстановления идентичности;
- определена общая схема политики и минимальные требования к доверенной практике применения службы псевдонимизации;
- определены интерфейсы, способствующие интероперабельности интерфейсов служб.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие международные стандарты:

ИСО 27799* Информатизация здоровья. Менеджмент безопасности информации по стандарту ИСО/МЭК 27002 (ISO 27799, Health informatics — Information security management in health using ISO/IEC 27002)

3 Термины и определения

В настоящем стандарте применены следующие термины с соответствующими определениями:

3.1

контроль доступа (access control): Средства, с помощью которых ресурсы системы обработки данных предоставляются только авторизованным субъектам в соответствии с установленными правилами.

[ИСО/МЭК 2382-8:1998, определение 08.04.01]

3.2 **обезличивание** (anonymization): Действия, в результате которых удаляется связь между совокупностью идентифицирующих данных и субъектом данных.

* Следует использовать последнее издание указанного документа, включая все поправки.

Издание официальное

3.3

обезличенные данные (anonymized data): Данные, по которым их получатель не может определить, к какому конкретному пациенту они относятся.

[Руководство по конфиденциальности организации General Medical Council]

3.4 **анонимный идентификатор** (anonymous identifier): Идентификатор лица, по которому невозможно однозначно установить, какое именно физическое лицо он обозначает.

3.5 **аутентификация** (authentication): Надежное установление подлинности объекта.

3.6

зашифрованный текст (ciphertext): Данные, полученные с помощью шифрования, семантическое содержание которых недоступно без применения криптографических средств.

[ИСО/МЭК 2382-8:1998, определение 08-03-8]

3.7

конфиденциальность (confidentiality): Состояние информации, при котором она недоступна неавторизованным лицам, субъектам или процессам.

[ИСО 7498-2:1989, определение 3.3.16]

3.8 **ключ шифрования контента** (content-encryption key): Криптографический ключ, используемый для шифрования содержания коммуникации.

3.9 **уполномоченный орган по защите прав субъектов персональных данных** (controller): Физическое или юридическое лицо, государственный орган, агентство или иная организация, которое самостоятельно или совместно с другими субъектами контролирует цели и способы обработки персональных данных.

3.10

криптография (cryptography): Дисциплина, определяющая принципы, методы и средства преобразования данных, предназначенная для защиты содержащейся в них информации от неустановленного изменения и несанкционированного использования.

[ИСО 7498-2:1989, определение 3.3.20]

3.11 **криптографический алгоритм** (cryptographic algorithm): Метод преобразования данных, обеспечивающий защиту содержащейся в них информации от неустановленного изменения и несанкционированного использования (шифрование).

3.12

управление ключами (key management, cryptographic key management): Генерация, хранение, распространение, удаление, архивирование или применение ключей в соответствии с политикой безопасности.

[ИСО 7498-2:1989, определение 3.3.33]

3.13

целостность данных (data integrity): Способность данных не подвергаться изменению или аннулированию в результате несанкционированного доступа.

[ИСО 7498-2:1989, определение 3.3.21]

3.14 **связывание данных** (data linking): Сопоставление и объединение данных, полученных из нескольких баз данных.

3.15 **защита данных** (data protection): Техническая и социальная система мероприятий по согласованию, управлению и обеспечению неприкосновенности, конфиденциальности и защиты информации.

3.16 **субъекты данных** (data-subjects): Лица, к которым относятся данные.

3.17

расшифрование (decipherment, decryption): Процесс восстановления соответствующих открытых данных из зашифрованных.

[ИСО/МЭК 2382-8:1998, определение 08-03-04]

Примечание — Если текст был зашифрован дважды, то однократное расшифрование не обеспечит восстановление исходных открытых данных.

3.18 **де-идентификация** (de-identification): Общее название любого процесса удаления связи между совокупностью идентифицирующих данных и субъектом данных.

3.19 **непосредственно идентифицирующие данные** (direct identifying data): Данные, непосредственно идентифицирующие конкретное лицо.

Примечание — Непосредственными идентификаторами являются те, которые могут использоваться для идентификации лица без привлечения дополнительной информации или данных из открытых источников.

3.20 **распространение** (disclosure): Разглашение данных или предоставление к ним доступа.

Примечание — Факт распространения не зависит от того, действительно ли получатель данных читал их, принял во внимание или оставил у себя.

3.21

шифрование (encipherment, encryption): Процесс криптографического преобразования открытых данных в зашифрованные.

[ИСО 7498-2:1989, определение 3.3.27]

Примечание — См. определение 3.10.

3.22 **идентификатор субъекта медицинской помощи** (subject of care identifier, healthcare identifier): Идентификатор лица, предназначенный для применения исключительно в сфере здравоохранения.

3.23

идентифицируемое лицо (identifiable person): Лицо, которое может быть прямо или косвенно идентифицировано, в частности, с помощью своего идентификатора или одного либо нескольких факторов, относящихся к его физической, физиологической, психологической, экономической, культурной или социальной идентичности.

[Директива 95/46/ЕС Европейского парламента и Совета Европейского союза от 24 октября 1995 года «О защите прав частных лиц при обработке персональных данных и свободном перемещении таких данных»]

3.24 **идентификация** (identification): Процесс сравнения предъявленных или наблюдаемых атрибутов субъекта с перечнем присвоенных идентификаторов в целях отделения одного субъекта от других субъектов.

Примечание — Идентификация субъекта в определенном контексте позволяет другому субъекту распознать, с каким именно субъектом он взаимодействует.

3.25

идентификатор (identifier): Информация, используемая для объявления идентичности перед тем, как получить потенциальное подтверждение от соответствующего опознавателя.

[ENV 13608-1]

3.26 **косвенно идентифицирующие данные** (indirectly identifying data): Данные, по которым можно идентифицировать конкретное лицо только в том случае, если они будут дополнены другими косвенно идентифицирующими данными.

Примечание — При совместном применении косвенные идентификаторы могут сузить популяцию, к которой принадлежит лицо, до одного человека.

Примеры — Почтовый индекс, пол, возраст, дата рождения.

3.27 **информация** (information): Совокупность данных с определенным смысловым содержанием.

3.28 **необратимость** (irreversibility): Ситуация, когда после преобразования идентификаторов в псевдонимы задача восстановления идентификатора по известному псевдониму является вычислительно неосуществимой.

3.29

ключ (key): Последовательность символов, управляющая операциями шифрования (0) и расшифрования (0).

[ИСО 7498-2:1989, определение 3.3.32]

3.30 **связывание информационных объектов** (linkage of information objects): Процесс, позволяющий установить логическую связь между различными информационными объектами.

3.31

другие фамилии, имена, отчества (other names): Фамилия, имя и отчество, под которыми данный пациент был некоторое время известен.

[ИСО 27931:2009]

3.32 **идентификация лица** (person identification): Процесс установления связи между информационным объектом и физическими лицом.

3.33 **идентификатор лица** (personal identifier): Информация, с помощью которой лицо может быть однозначно идентифицировано в определенном контексте.

3.34

персональные данные (personal data): Любая информация, относящаяся к идентифицированному или идентифицируемому лицу («субъекту данных»).

[Директива 95/46/ЕС Европейского парламента и Совета Европейского союза от 24 октября 1995 года «О защите прав частных лиц при обработке персональных данных и свободном перемещении таких данных»]

3.35 **первичное использование персональных данных** (primary use of personal data): Использование персональных данных при непосредственном оказании медицинской помощи.

3.36

конфиденциальность (privacy): Защита от вмешательства в личную жизнь или личные дела, выраженного в излишнем или неправомерном сборе и использовании персональных данных.

[ИСО/МЭК 2382-8:1998, определение 08-01-23]

3.37

обработка персональных данных (processing of personal data): Действие или совокупность действий с персональными данными, с помощью или без помощи средств вычислительной техники, включая сбор, накопление, систематизацию, хранение, уточнение или изменение, извлечение, консультирование, использование, распространение (в том числе передачу или иное предоставление доступа), сверку или комбинирование, блокирование, удаление или уничтожение.

[Директива 95/46/ЕС Европейского парламента и Совета Европейского союза от 24 октября 1995 года «О защите прав частных лиц при обработке персональных данных и свободном перемещении таких данных»]

3.38

оператор (processor): Физическое или юридическое лицо, государственный орган, агентство или иная организация, которое обрабатывает персональные данные с ведома уполномоченного органа по защите прав субъектов персональных данных.

[Директива 95/46/ЕС Европейского парламента и Совета Европейского союза от 24 октября 1995 года «О защите прав частных лиц при обработке персональных данных и свободном перемещении таких данных»]

3.39 **псевдонимизация** (pseudonymization): Особый случай обезличивания, при котором помимо удаления прямой связи с субъектом данных создается связь между конкретной совокупностью характеристик этого субъекта и одним или несколькими псевдонимами.

3.40 **псевдоним** (pseudonym): Идентификатор лица, отличающийся от того, который лицо обычно использует.

Примечания

1 Псевдоним может быть произведен из обычно используемого идентификатора с помощью обратимого или необратимого преобразования, но может и не иметь никакой связи с этим идентификатором.

2 Под псевдонимом в основном подразумевают идентификатор, по которому нельзя воспроизвести обычно используемый идентификатор лица. В этом случае информация, скрывающаяся за псевдонимом, функционально обезличена.

3.41 **лицо, получающее доступ к персональным данным** (recipient): Физическое или юридическое лицо, государственный орган, агентство или иная организация, которому предоставляются персональные данные.

3.42 **вторичное использование персональных данных** (secondary use of personal data): Любое использование персональных данных, отличающееся от первичного.

3.43

политика безопасности (security policy): Правила или руководящие принципы, установленные для обеспечения информационной безопасности.

[ИСО/МЭК 2382-8:1998, определение 08-01-06]

4 Обозначения и сокращения

HIPAA — Закон о преемственности страхования и отчетности в области здравоохранения (Health Insurance Portability and Accountability Act);

МИС — медицинская информационная система (Hospital Information System);

ВИЧ — вирус иммунодефицита человека (Human Immunodeficiency Virus);

IP — межсетевой протокол (Internet Protocol);

ЖНЗ — жертва нарушения закона (Victim of Violence).

5 Требования по обеспечению конфиденциальности идентичности лиц в сфере здравоохранения

5.1 Концептуальная модель псевдонимизации персональных данных

5.1.1 Общие сведения

Де-идентификация — общее название любого процесса удаления связи между совокупностью идентифицирующих данных и субъектом данных. Псевдонимизация представляет собой частный случай де-идентификации. Псевдоним представляет собой средство, с помощью которого можно связать между собой данные одного и того же лица, хранящиеся в разных записях или в разных информационных системах, не раскрывая его идентичность. Псевдонимизация может быть обратимой или необратимой, то есть может позволять или не позволять восстановление идентичности субъекта данных. В здравоохранении практикуются несколько сценариев псевдонимизации, рассчитанных на ускорение электронной обработки данных пациентов в условиях, когда пациенты предъявляют все более высокие требования к конфиденциальности. Некоторые примеры таких сценариев приведены в приложении А.

Примечание — Еще одной разновидностью де-идентификации является обезличивание. В отличие от псевдонимизации обезличивание не предоставляет возможность связывания информации об одном и том же лице, хранящейся в разных записях или в разных информационных системах. Следовательно, восстановление идентичности обезличенных данных невозможно.

5.1.2 Цель обеспечения конфиденциальности

Целью обеспечения конфиденциальности персональных данных, например с помощью псевдонимизации, является предотвращение несанкционированного или нежелательного распространения информации о лице, которое может повлечь за собой юридические, административные и экономические последствия. Обеспечение конфиденциальности персональных данных является частным случаем более общей задачи неразглашения информации, которая по определению включает в себя другие субъекты, например организации. Поскольку требования к конфиденциальности персональных данных хорошо проработаны, настоящая концептуальная модель сфокусирована на конфиденциальности. Решения по защите информации, рассчитанные на обеспечение конфиденциальности персональных данных, могут быть применены для защиты информации о других субъектах, например об организациях. Это может быть полезным для тех стран, где обеспечение коммерческой тайны регулируется законодательно.

В защите персональных данных выделяются две задачи: первая — защита оперативного доступа к персональным данным (например, в веб-приложениях), вторая — защита персональных данных, хранящихся в базах данных. Настоящий стандарт посвящен последней задаче.

Описанная ниже концептуальная модель предполагает, что данные могут извлекаться из баз данных, содержащих, например, информацию о лечении или диагнозах пациентов. При этом должно гарантироваться неразглашение идентичности субъектов данных. Научные работники изучают «случаи» заболеваний, то есть истории наблюдения пациентов, накопленные за длительное время и/или собранные из разных источников. Однако для агрегирования различных элементов данных в такие «случаи» необходимо использовать метод, позволяющий агрегировать данные, но при этом не нарушать конфиденциальности сведений о субъектах этих данных. В качестве такого метода может применяться псевдонимизация.

5.1.3 Неразглашение информации о субъектах

В качестве отправной точки концептуальная модель использует обеспечение конфиденциальности персональных данных, но понятие «субъект данных» не исчерпывается физическими лицами и может означать любую другую сущность, например, организацию, устройство или прикладную программу. Однако полезно сфокусировать модель на физических лицах, поскольку обеспечение конфиденциальности персональных данных регулируется законодательно и неразглашение информации прежде всего относится к персональным данным. В законодательстве о персональных данных содержится описание некоторых понятий, используемых в данной модели. В контексте здравоохранения обеспечение конфиденциальности персональных данных гораздо сложнее общих подходов к неразглашению информации, например об устройствах, поскольку для идентификации лица потенциально могут использоваться сведения о его фенотипе.

5.1.4 Сравнение персональных и де-идентифицированных данных

5.1.4.1 Определение персональных данных

В соответствии с Директивой о защите персональных данных Европейского парламента и Совета Европейского союза от 24 октября 1995 года (директива 95/46/EC) [7] под «персональными данными» должна пониматься любая информация, относящаяся к идентифицированному или идентифицируемому физическому лицу («субъекту данных»); при этом идентифицируемым считается лицо, которое может быть прямо или косвенно идентифицировано, в частности по номеру, идентифицирующему его, или по одному или нескольких факторам, специфичным для его физической, физиологической, психологической, экономической, культурной или социальной идентичности.

Это понятие используется и в законодательстве других стран, относящемся к тем же субъектам, что указаны в приведенном выше определении (например, в Законе HIPAA).

5.1.4.2 Идеализированная концепция идентификации и де-идентификации

В настоящем подпункте описана идеализированная концепция идентификации и де-идентификации. Она предполагает, что вне модели нет никаких данных, которые, к примеру, могли бы быть связаны с данными в составе модели, чтобы обеспечить (косвенную) идентификацию субъектов данных. В 5.1.5 приняты во внимание потенциальные источники информации, внешние по отношению к модели. Это необходимо для обсуждения угроз восстановления идентичности. При представлении функциональной архитектуры в проектах, описывающих информационные и коммуникационные технологии, никогда не изображаются данные, не используемые в модели. Но если моделируется идентификация субъектов, то критики модели апеллируют к информации, которая может быть добыта злоумышленником для идентификации субъектов данных или получения более точных сведений о них (например о принадлежности к определенной группе).

Как показано на рисунке 1, субъекты данных имеют ряд характеристик (например, фамилию, дату рождения, медицинские данные), которые хранятся в базе данных медицинской информационной системы (МИС) и являются персональными данными этих субъектов. Субъект данных идентифицируется в совокупности субъектов, если его можно однозначно выделить среди них. Это означает, что может быть найдена совокупность характеристик субъекта данных, по которой он может быть однозначно идентифицирован. В некоторых случаях для этого достаточно единственной характеристики (к примеру, уникального национального идентификатора субъекта). В других случаях для идентификации субъекта необходимо использовать несколько характеристик, например адрес, по которому он проживает с семьей, если такой адрес известен. Некоторые характеристики субъекта данных (например, дата и место рождения) более постоянны, чем другие (например адрес электронной почты).

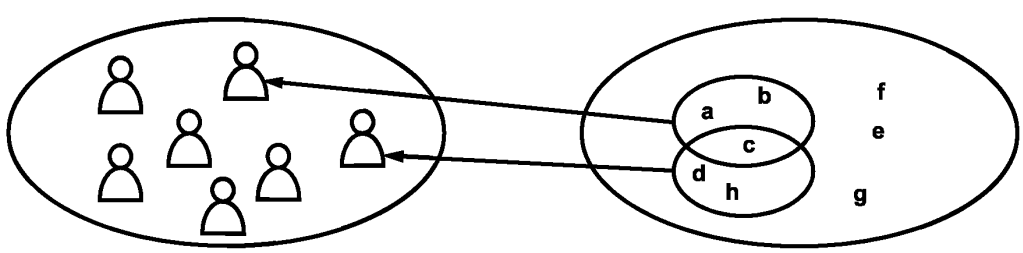


Рисунок 1 — Идентификация субъектов данных

Персональные данные можно разделить на две части в соответствии с критериями идентификации (см. рисунок 2):

- обрабатываемые данные: часть данных, содержащая характеристики, по которым субъект данных не может быть однозначно идентифицирован; концептуально эти данные являются обезличенными;
- идентифицирующие данные: часть данных, содержащая совокупность характеристик, по которым субъект данных может быть однозначно идентифицирован (например демографические данные).

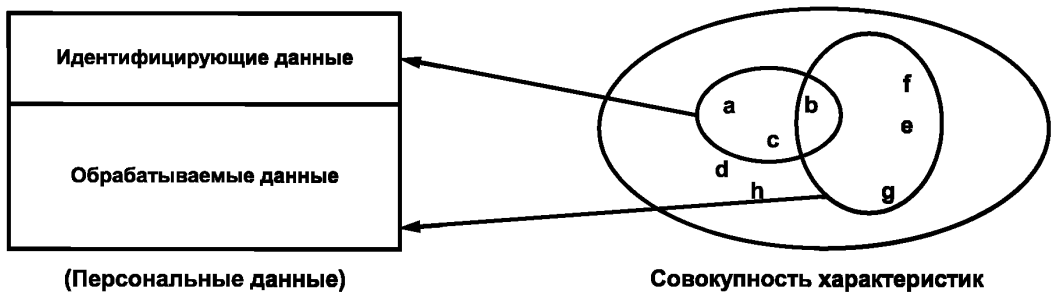


Рисунок 2 — Разделение персональных данных

Следует обратить внимание, что концептуальное разделение персональных данных на «идентифицирующие» и «обрабатываемые» может вести к противоречию, а именно в случае, когда непосредственно идентифицирующие данные в то же время являются обрабатываемыми. При разработке метода псевдонимизации необходимо стремиться к снижению уровня непосредственно идентифицирующих данных, например, агрегируя эти данные в группы. В отдельных случаях, когда это невозможно (например для даты рождения новорожденного), риск идентификации должен быть отражен в документе политики безопасности. В следующем разделе настоящего стандарта разделение данных на обрабатываемые и идентифицирующие обсуждается скорее с практической точки зрения, нежели концептуальной. С концептуальной точки зрения достаточно того, что такое разделение в принципе возможно. Важно отметить, что различие между обрабатываемыми и идентифицирующими данными не является абсолютным. Для научного исследования могут требоваться некоторые данные, которые принадлежат к числу идентифицирующих, например, год и месяц рождения.

5.1.4.3 Концепция псевдонимизации

Для развития медицины важно, чтобы элементы данных конфиденциальных медицинских карт были доступны для ведения научной работы, контроля качества медицинской помощи, образования и других приложений. Согласно требованиям конфиденциальности и требованиям к научным исследованиям эти элементы должны быть модифицированы таким образом, чтобы идентичность субъекта была скрыта.

Никакая процедура де-идентификации не может удовлетворить различным требованиям применения персональных данных в медицине и в то же время гарантировать сокрытие идентичности. Для каждого процесса обработки медицинских карт должна быть построена модель угроз, учитывающая следующие факторы:

- цель обработки данных (например анализ);
- минимально необходимую информацию, которая должна быть предоставлена для достижения этой цели;
- угрозы распространения персональных данных (включая восстановление идентичности);
- существующие правила доступа к персональным данным.

Исходя из этих факторов, должны быть определены детали процесса доступа и модели угроз, а также стратегия сокрытия идентичности. Эту процедуру необходимо проводить для каждого нового процесса доступа к персональным данным, хотя для многих различных процессов доступа могут требоваться общая стратегия и общие детали. Во многих случаях при использовании персональных данных в образовании существуют общие цели и требования к минимально необходимой информации. Во многих клинических испытаниях лекарственных средств используется общая стратегия, а детали процесса доступа могут меняться. Де-идентификация может требоваться не только для обеспечения конфиденциальности, но и при других обстоятельствах, например при проведении простых и двойных слепых исследований, когда идентичность субъекта должна быть скрыта. Такие обстоятельства влияют на решение о предоставлении персональных данных.

В настоящем подпункте приведена терминология, используемая для описания сокрытия идентифицирующих данных.

Обезличивание (см. рисунок 3) представляет собой процесс удаления связи между идентифицирующей совокупностью характеристик и субъектом данных. Оно может быть осуществлено двумя разными способами:

- с помощью удаления или преобразования характеристик, при котором связь между характеристиками и субъектом данных либо прекращается, либо перестает быть уникальной и указывает на несколько субъектов данных;
- путем увеличения популяции субъектов данных, при котором связь между совокупностью характеристик и субъектом данных перестает быть уникальной.

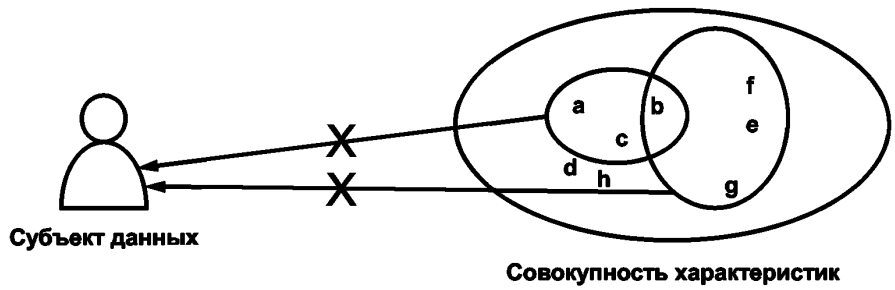


Рисунок 3 — Обезличивание

Псевдонимизация (см. рисунок 4) удаляет связь между совокупностью характеристик и субъектом данных и добавляет связь между этой совокупностью и одним или несколькими псевдонимами. Поскольку существуют связи между совокупностями характеристик и псевдонимами, то с функциональной точки зрения между несколькими псевдонимизированными совокупностями данных, относящимися к одному и тому же субъекту, можно установить связь, не раскрывая его идентичность. В результате, например, становится возможным проводить исследования реальных данных о пациенте, собранных в течение длительного времени, не нарушая требование конфиденциальности.

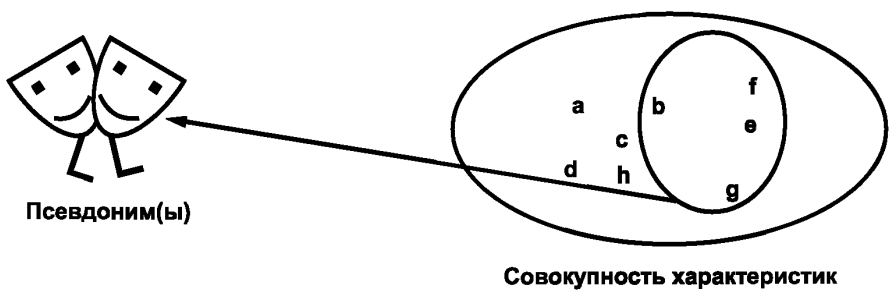


Рисунок 4 — Псевдонимизация

При необратимой псевдонимизации концептуальная модель не содержит метод определения субъекта данных по совокупности его характеристик и псевдониму.

При обратимой псевдонимизации (см. рисунок 5) концептуальная модель содержит способ восстановления связи между совокупностью характеристик и субъектом данных. Существуют два метода решения этой задачи:

- вычисление идентификации по обрабатываемым характеристикам; это можно осуществить, например с помощью расшифрования идентифицирующей информации, содержащейся в обрабатываемых характеристиках;
- вычисление идентификации по псевдониму, например с помощью таблицы соответствия.

Обратимая псевдонимизация может быть реализована несколькими способами, но при этом следует иметь в виду, что превращение псевдонима в идентифицирующую информацию должно выполняться авторизованным лицом или организацией в строго определенных обстоятельствах. Чтобы облегчить решение этой задачи, в разделе 9 описана рамочная политика процесса восстановления идентичности. Обратимая псевдонимизация требует большей защиты системы, выполняющей псевдонимизацию, нежели необратимая.

Обезличенные данные отличаются от псевдонимизированных тем, что для последних предусмотрен метод группировки по критериям, сформулированным в терминах характеристик персональных данных, из которых получены псевдонимизированные данные.

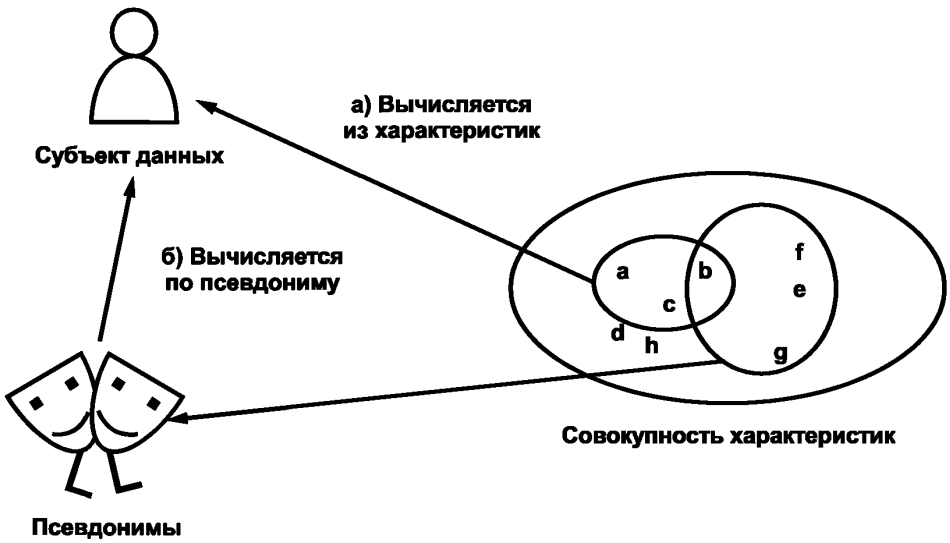


Рисунок 5 — Обратимая псевдонимизация

5.1.5 Псевдонимизация в реальном мире

5.1.5.1 Основания

В 5.1.4 описан концептуальный подход к псевдонимизации, при котором понятия «связь», «идентифицирующий», «псевдоним» и другие рассматриваются как абсолютные. На практике трудно оценить угрозу восстановления идентичности персональных данных. В настоящем подпункте уточняются понятия псевдонимизации и нежелательного/несанкционированного восстановления идентичности.

Согласно декларации 26 Европейской директивы о защите персональных данных «для определения того, является ли лицо идентифицируемым, необходимо учитывать все правдоподобные средства, которые оператор или любое другое лицо могут использовать для идентификации названного лица; при этом принципы защиты не должны примеряться к данным, сделанным обезличенными таким способом, что субъект данных более не идентифицируется; при этом нормы регулирования, указанные в содержании статьи 27, могут быть полезным правилом предоставления руководства по способам, с помощью которых данные делаются обезличенными и сохраняются в форме, при которой идентификация лица становится невозможной».

Как и само определение персональных данных, данная декларация делает упор на «идентификацию», то есть на связь между персональными данными и субъектом данных.

Смысл выражений «все правдоподобные средства» или «любое другое лицо» не является достаточно определенным. И поскольку смысл понятий «идентифицируемый» и «псевдоним» зависит от неопределенного способа действий («учитывать все правдоподобные средства») и неопределенных действующих лиц («любое другое лицо»), то концептуальная модель, описанная в настоящем стандарте, должна содержать «разумные» предположения обо «всех средствах», которые могут применяться «любым другим лицом» для установления связи характеристик с субъектом персональных данных.

В концептуальную модель будут внесены уточнения, отражающие различные возможности восстановления идентичности. Кроме того, в ней будут приняты во внимание такие понятия, как «базы данных наблюдений» и «злоумышленники».

5.1.5.2 Степени гарантий обеспечения конфиденциальности

В существующей терминологии некоторые определения страдали отсутствием точности, например определения терминов «псевдоним» или «идентифицируемый». Нереально предполагать, что можно устранить все неточности терминологии, поскольку псевдонимизация всегда является предметом статистики. Но можно оценить степень угрозы несанкционированного восстановления идентичности. Система классификации этой угрозы должна учитывать вероятность идентификации свойств данных, а также четкое представ-

ление о моделируемых сущностях и их связях. В одних случаях модель угроз может ограничиваться минимизацией угрозы случайного раскрытия информации или исключением смещений в двойных слепых исследованиях, в других она должна предусматривать возможность злоумышленных действий. Отсюда следует, что конкретные политики обеспечения конфиденциальности могут, например, сделать «границы неточности» более узкими и определить, что в конкретном контексте понимается под термином «идентифицируемость» в результате будет легче оценить ответственность.

Ниже приводится классификация степени угроз, которая, однако, требует дальнейшего уточнения, а именно в связи с тем, что для количественной оценки угрозы восстановления идентичности требуются математические модели. Угрозы восстановления идентичности существуют независимо от того, насколько хорош отдельный алгоритм преобразования одного файла персональных данных. Важным шагом процесса оценки угроз является анализ результирующей де-идентифицированной совокупности характеристик любых статических групп, которые могут быть использованы для восстановления идентичности. Это особенно важно в случаях, когда для целей обработки необходимы некоторые идентифицирующие характеристики. В настоящем стандарте такие математические модели не описаны, однако в библиографии приведены информационные ссылки на некоторые из них.

В отличие от идеализированной концептуальной модели, в которой не принимаются во внимание (известные или неизвестные) внешние источники данных, в методе оценки угроз восстановления идентичности должны быть сделаны определенные предположения о том, какие внешние источники данных могут быть использованы.

Модель реального мира должна принимать во внимание как непосредственно идентифицирующие, так и косвенно идентифицирующие данные. Необходимо анализировать каждый сценарий на предмет выделения требований к идентификаторам как к информационным объектам, а также определить, какие идентификаторы могут быть заменены пустыми значениями, какие могут быть искажены, какие требуется оставить в полной целостности, а какие должны быть заменены псевдонимами.

Ниже определены три уровня процедуры псевдонимизации, каждый из которых гарантирует определенную степень обеспечения конфиденциальности. Они характеризуют угрозы восстановления идентичности с учетом как непосредственно, так и косвенно идентифицирующих данных:

- уровень 1: угрозы, связанные с использованием элементов данных, идентифицирующих лицо;
- уровень 2: угрозы, связанные с использованием агрегатов данных;
- уровень 3: угрозы, связанные с использованием аномальной информации в заполненной базе данных.

На всех уровнях оценка степени угроз восстановления идентичности должна осуществляться как итерационный процесс с регулярным повторением оценки (согласно политике конфиденциальности). По мере накопления опыта способы обеспечения конфиденциальности и уровни угроз должны пересматриваться.

Наряду с регулярными повторениями процедура оценки может инициироваться определенными событиями, например, изменением состава обрабатываемых данных или включением в модель новых наблюдаемых данных.

При указании степени гарантий общее обозначение уровня 1, 2 или 3 может быть дополнено числом пересмотров (например, уровень 2+ для пересмотренного уровня 2; при этом целесообразно указывать последний пересмотр состава данных и поддерживать историю инцидентов и пересмотров в актуальном состоянии). По требуемой степени гарантий определяются технические и организационные меры, которые должны быть реализованы для обеспечения конфиденциальности персональных данных. Для более низкого уровня псевдонимизации требуются более серьезные организационные меры обеспечения конфиденциальности, чем для более высокого уровня псевдонимизации.

Обеспечение конфиденциальности уровня 1: удаление очевидных непосредственно идентифицирующих данных или легко получаемых косвенно идентифицирующих данных.

Первый, интуитивно очевидный, уровень обезличивания может быть достигнут с помощью применения простых практических методов. Эти методы обычно подразумеваются при обсуждении псевдонимизированных данных. Во многих случаях, в особенности если предполагается злоумышленник со скромными возможностями, этот уровень обезличивания может предоставлять достаточные гарантии. Данные считаются идентифицирующими, если в данном контексте для точного указания субъекта достаточно той информации, которая содержится в самих этих данных. Типичным примером служат фамилии, имена и отчества лиц. В 6.2.2 приведена спецификация элементов данных, которые могут рассматриваться как подлежащие удалению или агрегированию, чтобы результирующую совокупность характеристик можно было считать обезличенной.

Обеспечение конфиденциальности уровня 2: предполагается наличие злоумышленника, использующего внешние данные.

Второй уровень обеспечения конфиденциальности может быть достигнут, если будут приняты во внимание глобальная модель данных и потоки данных в этой модели. При определении процедур, необходимых для достижения этого уровня, необходимо провести статический анализ угроз восстановления идентичности различными действующими лицами. Кроме того, необходимо предполагать наличие злоумышленника, который для идентификации специфичных совокупностей характеристик использует внешние данные в дополнение к псевдонимизированным. Доступные внешние данные могут зависеть от законодательства конкретной страны и специфических знаний злоумышленника. Примером противодействующих процедур может служить удаление из данных абсолютных значений меток времени. Метка времени «Т» может быть привязана, например, к поступлению пациента в стационар, а время начала других событий, например выписки, может задаваться по отношению к этой метке. Злоумышленником считается организация или лицо, собирающее данные (санкционированно или несанкционированно) и задающееся целью несанкционированно привязать эти данные к субъектам данных, чтобы тем самым получить информацию, которая ему не предназначена. При анализе угроз данные, собираемые и используемые злоумышленником, называются «наблюдаемыми» данными. Следует иметь в виду, что неразрешенная или нежелательная деятельность злоумышленника состоит не в том, что он собирает данные, а в том, что он пытается привязать их к субъекту данных и тем самым получить несанкционированные сведения о субъекте данных.

Модель угроз может включать в себя определенные предположения о злоумышленнике и его действиях. Например, законодательство некоторых стран может разрешать сбор данных о выписанных пациентах организациям, которые напрямую не вовлечены в процесс лечения или связанный с ним процесс управления пациентами. Модель угроз может принимать во внимание вероятность доступности определенных совокупностей характеристик пациентов.

С концептуальной точки зрения злоумышленник может привносить элементы данных, которые в идеальном случае не существуют.

Документ политики конфиденциальности должен содержать оценку вероятности атак подобного рода.

Обеспечение конфиденциальности уровня 3: предполагается наличие злоумышленника, использующего аномальные данные.

Степень угрозы восстановления идентичности может серьезно зависеть от самих данных, например от наличия аномальных или редко встречаемых данных. Такие данные могут косвенным образом способствовать идентификации субъекта данных. Например, если в какой-то день клинику посетил только один пациент со специфичной патологией, то наблюдение за тем, кто посещал клинику в тот день, может косвенно способствовать идентификации этого пациента.

Статический анализ модели угроз, проводящийся для оценки процедуры псевдонимизации, сам по себе не может дать количественную оценку уязвимости содержания базы данных, поэтому для достижения более высокого уровня обезличивания необходимо проводить регулярный анализ накопленной базы данных.

На практике трудно доказать, что конфиденциальность может обеспечиваться на уровне 3.

5.2 Категории субъектов данных

5.2.1 Общие сведения

В настоящем стандарте основное внимание уделяется псевдонимизации данных о пациентах/получателях медицинской помощи, но изложенные в нем принципы могут применяться и к другим категориям субъектов данных, например к работникам здравоохранения или организациям.

В 5.2.2—5.2.4 перечислены специфичные категории субъектов данных и перечни особенностей этих категорий.

5.2.2 Пациент/получатель медицинской помощи

Решения о защите идентичности пациента могут быть обусловлены:

- требованиями законодательства по обеспечению конфиденциальности персональных данных;
- необходимостью соблюдения врачебной тайны для установления доверительных отношений между медицинским работником и пациентом;
- ответственностью за ведение регистров социально значимых заболеваний и других ресурсов информации об общественном здоровье;
- передачей минимально необходимой идентифицирующей информации третьим сторонам при оказании медицинской помощи (например для выполнения лабораторных анализов);

- обеспечением конфиденциальности при вторичном использовании медицинских данных для научной работы; необходимо иметь в виду, что в тех случаях, когда данные только псевдонимизированы, а не полностью обезличены, законодательство некоторых стран (например Германии) требует получения от пациента информированного согласия на использование его данных.

Для обеспечения преемственности медицинской помощи необходима сквозная идентификация пациентов, при которой информация, полученная в разных местах лечения одного и того же пациента, связана между собой. Если при оказании медицинской помощи данные псевдонимизируются в разных местах лечения, то существует угроза неправильной идентификации или отсутствия связи данных с пациентом. Если псевдонимизация осуществляется при непосредственном оказании медицинской помощи, то в тех случаях, когда пациент по соображениям безопасности здоровья отказывается от псевдонимизации, от него необходимо требовать информированное согласие.

5.2.3 Медицинские работники и организации

Псевдонимизация может также использоваться для защиты идентичности медицинских работников в разных ситуациях, включая:

- экспертизу качества медицинской помощи,
- информирование о врачебных ошибках или побочных действиях лекарств;
- анализ лечебно-диагностического процесса;
- экономический анализ медицинской помощи;
- статистику деятельности врача.

Требования к соответствующим мерам защиты являются основой местного законодательства и могут отличаться от требований к защите идентичности организаций.

5.2.4 Информация о медицинских устройствах

Обеспечение конфиденциальности в здравоохранении касается сведений о медицинских устройствах, например об имплантантах, так как их идентифицирующие данные напрямую связаны с пациентом. Другие медицинские и персональные устройства также могут быть связаны с пациентом (например устройства аппаратного дыхания). В этих случаях устройство может быть идентифицировано как субъект данных, и по характеристикам этого устройства можно различить пациентов. Необходимо также оценивать риск идентификации медицинских устройств, закрепленных за медицинскими работниками или другими служащими, поскольку по ним можно определить работника или организацию.

5.3 Категории данных

5.3.1 Обрабатываемые данные

Согласно настоящему стандарту данные можно разделить на те, которые приводят к идентификации субъекта данных, и те, которые представляют собой медицинскую информацию. Такое разделение полностью зависит от целевого уровня обеспечения конфиденциальности.

5.3.2 Наблюдаемые данные

Наблюдаемые данные отражают различные характеристики субъектов данных, регистрируемые в целях как можно более полного описания субъектов данных с намерением последующего восстановления идентичности субъектов данных или их принадлежности к определенным группам.

5.3.3 Псевдонимизированные данные

Возможны два типа псевдонимизированных данных:

при необратимой псевдонимизации данные не содержат информацию, позволяющую восстановить связь между псевдонимизированными данными и субъектом данных;

при обратимой псевдонимизации связь между псевдонимизированными данными и субъектом данных может быть восстановлена с помощью процедур, доступных только санкционированным пользователям.

Примечание — Обратимость является свойством, которое может быть обеспечено разными методами, например: 1) с помощью добавления к псевдонимизированным данным зашифрованных идентифицирующих данных; 2) ведения списка связей между псевдонимами и идентификаторами, хранящегося в безопасном месте.

5.3.4 Обезличенные данные

Обезличенными считаются данные, не содержащие информации, которая может быть использована для связи с субъектом, к которому они первоначально относились. Такая связь могла бы, например, обеспечиваться по фамилиям, именам, отчествам, дате рождения, регистрационным номерам или иной идентифицирующей информации.

5.3.5 Научные данные

5.3.5.1 Общие сведения

Использование медицинских данных для научной работы обычно представляет собой вторичное использование данных, помимо основного использования в процессе лечения пациента. Во многих законодательствах требуется, чтобы пациент давал информированное согласие на такое использование. Фундаментальный принцип защиты персональных данных состоит в том, что идентифицируемые персональные данные должны обрабатываться исключительно в декларируемых целях. Поэтому научно-исследовательские организации проявляют особый интерес к псевдонимизированным или даже обезличенным данным. Забота о конфиденциальности персональных данных, особенно в части медицинской информации, приводит к появлению новых нормативных требований к обеспечению права на конфиденциальность. Научные работники должны следовать этим требованиям, и во многих случаях им приходится менять традиционные подходы к обмену индивидуально идентифицируемой медицинской информацией.

Соблюдение врачебной тайны и автономии пациента весьма проблематично, поскольку многие традиционные подходы по защите данных с трудом адаптируются к возрастающей сложности данных, потоков информации и возможностям извлечения дополнительной пользы за счет объединения разрозненных данных. Получение классического информированного согласия пациента на обработку каждой совокупности данных может оказаться затруднительным или невозможным. Научный работник может свободно оперировать обезличенными данными, но не псевдонимизированными.

5.3.6 Идентификаторы пациентов

В системе здравоохранения приходится разрешать конфликтные ситуации в отношении требований к идентификации пациентов:

- если доступ к нескольким источникам медицинских данных, принадлежащих разным организациям, санкционирован, то хранящиеся в них сведения о заданном субъекте можно связать между собой. В зависимости от применения связанных данных, к связям могут предъявляться следующие требования:

правильность (связи с источниками данных не относятся к разным пациентам);

полнота (нет связей, которые не удалось установить из-за ошибки идентификации субъекта данных);

- если доступ к данным, идентифицирующим субъект, ограничен, то доверенный поставщик услуг может контролируемым образом обеспечить уполномоченным на то лицам связывание персональных данных с субъектом данных.

Некоторые законодательства могут ограничивать связывание информации из различных источников данных. Это обстоятельство также должно быть учтено. Когда субъект данных обращается к разным поставщикам медицинской помощи, то они нередко регистрируют его, используя внутреннюю систему нумерации. Административная и медицинская информация об этом субъекте нередко передается другим уполномоченным органам с этими местными номерами. Как следствие, уполномоченные органы, которым необходимо агрегировать полученную информацию, не могут быть уверены, что агрегированные данные являются полными.

Эту ситуацию можно избежать, используя структурированные подходы к управлению идентичностью. Подробное обсуждение управления идентичностью не является задачей настоящего стандарта, однако основой некоторых решений по управлению идентичностью может быть псевдонимизация.

5.3.7 Персональные данные жертв насилия и публичных лиц

Если идентификация жертв насилия может создавать угрозу их безопасности, то при их лечении или диагностике от медицинского персонала требуются дополнительные меры предосторожности. При непосредственном контакте с таким пациентом лечащие врачи могут его идентифицировать, а вспомогательные работники — нет.

Подобные моменты возникают при оказании медицинской помощи публичным лицам или тем, кто хорошо знаком медицинскому сообществу (нередко ошибочно называемым очень важными персонами), например, политикам, руководителям крупных предприятий и т. д.

5.3.8 Генетическая информация

В отношении генетической информации широко распространены две точки зрения: 1) генетическую информацию следует отличать от другой медицинской информации; 2) генетические данные должны защищаться таким же образом, как и любая другая медицинская информация. Точка зрения, согласно которой генетические данные должны быть обособлены от медицинских данных, получила название «исключительности генетических данных».

В 2004 году отдел С3 (этика и наука) Европейского директората С (наука и общество) опубликовал двадцать пять рекомендаций по этическим, юридическим и социальным аспектам генетического тестирования, например:

- следует избегать признания «исключительности генетических данных» в международном плане, на уровне Европейского союза и на уровне стран - членов Европейского союза. Однако публичное восприятие генетического тестирования как особой процедуры должно признаваться и приниматься во внимание;

- все медицинские данные, в том числе генетические, должны удовлетворять одинаково высоким стандартам качества и конфиденциальности.

В части конфиденциальности, неприкосновенности личной жизни и автономии эти рекомендации гласят:

- генетические данные, важные для клинического или семейного контекста, должны защищаться в той же мере, что и другие медицинские данные той же степени значимости;

- необходимо учитывать их релевантность для других членов семьи;

- необходимо учитывать право пациента знать или не знать свою генетическую информацию и обеспечивать это право соответствующими механизмами профессиональной практики;

- по отношению к генетическому тестированию должна быть выработана практика предоставления соответствующей информации, консультирования, информированного согласия, передачи результатов тестирования.

Существуют и другие мнения, отличающиеся от этих рекомендаций. Например, точное содержание генетической информации, собранной при исследовании популяций, неизвестно, и требования к защите генетической информации, хранящейся в базах данных и в банках человеческих тканей, могут быть со временем существенно повышены.

В сентябре 2005 года на 27-й Международной конференции по защите данных и обеспечению конфиденциальности была принята «Декларация Монтре», в которой было заявлено, что «...быстрый прогресс знаний в области генетики, может привести к тому, что информация о человеческой ДНК станет наиболее важной частью персональных данных; это ускорение знаний усиливает важность адекватной юридической защиты и обеспечения конфиденциальности...».

Тем не менее в тех случаях, когда генетическая информация является частью совокупности характеристик, которая должна быть псевдонимизирована, необходимо принимать во внимание существующее законодательство, методические руководства и публикации по этой тематике, которые охватывают более широкий спектр популяционных генетических исследований и функционирования банков человеческих тканей. В общих терминах этот ресурс данных может быть классифицирован как идентификация генов восприимчивости к заболеваниям или диагностических биомаркеров.

5.4 Доверенные службы

В случае, когда для синхронизации псевдонимов, присваиваемых разными учреждениями или организациями, необходима служба псевдонимизации, может понадобиться менеджер доверенных служб. Доверенные службы могут быть реализованы при большом числе провайдеров, включая коммерческие фирмы, общественные организации, государственные предприятия и учреждения. В различных законодательствах деятельность провайдеров может регулироваться нормативными документами или требованиями сертификации.

5.5 Необходимость восстановления идентичности псевдонимизированных данных

При псевдонимизации идентифицирующие данные отделяются от обрабатываемых данных и вместо них к обрабатываемым данным добавляется кодированное значение. Такой подход устраняет прямую связь между обрабатываемыми данными и персональными идентификаторами, но при этом в предписанных случаях при соблюдении мер безопасности можно восстановить идентичность субъекта данных. Эта возможность позволяет научным работникам исключать идентификаторы субъекта из обрабатываемых данных, но во многих (регулируемых) случаях восстанавливать исходные идентификаторы, если это необходимо. Некоторые такие случаи перечислены в приведенной системе кодирования.

Идентификация словаря данных: ИСО (1) стандарты (0) псевдонимизация (25237) цели восстановления идентичности (1).

Кодированные значения:

1) верификация/проверка целостности данных (data integrity verification/validation);

2) верификация/проверка дублирования записей данных (data duplicate record verification/validation);

3) запрос дополнительных данных (request for additional data);

4) установление связи с источниками дополнительной информации (link to supplemental information variables);

5) нормоконтроль (compliance audit);

6) передача значимых результатов (communicate significant findings);

7) дополнительное исследование (follow-up research).

Для контроля эти значения должны быть указаны при санкционированном восстановлении идентичности. Методы такого восстановления должны быть хорошо защищены, что можно обеспечить с помощью использования доверенной службы генерации ключей расшифровки идентификаторов и управления этими ключами. Доверенные службы могут безопасно определять критерии восстановления, автоматизировать восстановление в соответствии с этими критериями и управлять ими.

5.6 Характеристики службы псевдонимизации

Существуют два основных сценария применения служб псевдонимизации:

1) псевдонимы присваиваются одной организацией (или для одной организации) либо для одной цели — в этом случае обычно служба управляет идентификаторами, которые присвоены одной организацией или известны этой организации;

2) псевдонимы предоставляются службами псевдонимизации — в этом случае обычно служба представляет псевдоидентификаторы нескольким независимым организациям, чтобы можно было связать между собой медицинскую информацию об одном и том же пациенте, собранную в этих организациях, и при этом защитить идентичность пациента.

В обоих случаях служба должна функционировать таким образом, чтобы угроза несанкционированного восстановления идентичности субъектов этой службы была сведена к минимуму.

Служба, предназначенная для защиты идентичности пациентов, должна удовлетворять следующим минимальным требованиям доверенной практики:

- необходимо обеспечить доверие получателей медицинской помощи в том, что медицинская информационная система способна управлять конфиденциальностью их информации;
- необходимо, чтобы была обеспечена физическая защита службы;
- необходимо, чтобы была обеспечена безопасность функционирования службы;
- ключи восстановления идентичности, таблицы преобразования и меры защиты должны подвергаться независимому контролю нескольких лиц или нескольких организаций на предмет соответствия заявленным гарантиям безопасности службы;
- служба должна контролироваться (например, по договору или в силу служебных обязанностей) лицом, ответственным за безопасность исходных идентификаторов;
- в подтверждение заявленных уровней обеспечения конфиденциальности должны предоставляться юридические и технические регламенты применения ключей восстановления идентичности и мер защиты;
- качество службы и ее доступность должны быть описаны и соответствующим образом обеспечены при предоставлении информации и доступа;
- если некоторые идентификаторы не нужны для обработки персональных данных, то они могут быть заменены пустыми значениями;
- некоторые идентификаторы могут быть искажены способом, соответствующим цели обработки персональных данных.

6 Процесс псевдонимизации (методы и реализация)

6.1 Критерии конструирования

При псевдонимизации данных необходимо отделить идентифицирующие данные от обрабатываемых.

Разделение идентифицирующих и обрабатываемых данных в соответствии с заявленными уровнями обеспечения конфиденциальности и построенной моделью угроз является ключевым шагом процесса псевдонимизации данных. В дальнейшем идентифицирующая часть подвергнется преобразованию, а обрабатываемая останется без изменения. Процесс псевдонимизации преобразует заданные идентификаторы в псевдоним. С точки зрения наблюдателя, результирующие псевдонимы не содержат идентифицирующую информацию (что является основой криптографических преобразований). Такое преобразование может быть осуществлено различными способами в зависимости от требований проекта. Псевдонимизация может:

- всегда преобразовывать данный идентификатор в один и тот же псевдоним.Поскольку сочетание сохранения связи между совокупностями характеристик одного и того же лица и обеспечения конфиденциальности субъектов данных является основной причиной применения псевдонимизации, этот вариант используется наиболее часто;

- преобразовывать данный идентификатор в разные псевдонимы:

в зависимости от контекста (контекстно-зависимый псевдоним);

в зависимости от времени (например, каждый раз другой или меняющийся через определенные интервалы времени);

в зависимости от места (например, меняющийся, если данные передаются из разных мест).

6.2 Моделируемые сущности

Модель псевдонимизации включает в себя четыре сущности, а именно (см. рисунок 6):

- источник данных;
- провайдер службы идентификации лиц;
- провайдер службы псевдонимизации;
- получатель данных.

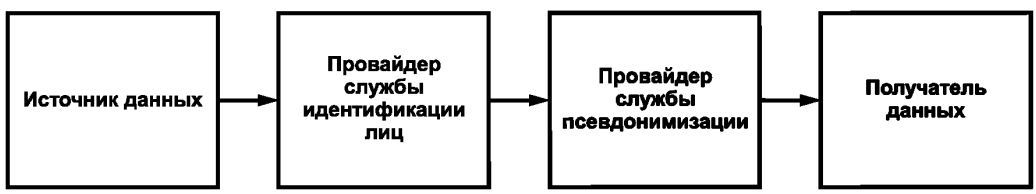


Рисунок 6 — Моделируемые сущности

Источник данных представляет собой сущность, выполняющую следующие функции:

- подготовку и структурирование данных для служб идентификации и псевдонимизации. Служба псевдонимизации должна определять, что требуется делать с элементом данных. Это можно задать либо с помощью разметки элементов данных, либо с помощью размещения элементов данных в определенные места, в каждом из которых осуществляется заранее заданная обработка;

- передачу данных службе идентификации лиц, а затем службе псевдонимизации. Это можно сделать с помощью вызова клиента службы идентификации, а затем клиента службы псевдонимизации;

- считывание и обработку результирующего кода, возвращаемого службой псевдонимизации. Этот шаг может состоять в простой регистрации результата, если вызов службы был успешен, либо, если при вызове возникла ошибка, в повторном вызове или в отправке предупреждения в зависимости от возвращенной информации.

Получатель данных представляет собой сущность, которая принимает псевдонимизированные данные от службы псевдонимизации и обеспечивает дальнейшую обработку данных. В зависимости от местного законодательства и уровня обеспечения конфиденциальности даже псевдонимизированные данные могут не удовлетворять требованиям закона о персональных данных. Получатель данных выполняет следующие функции:

- расшифровку данных, полученных от службы псевдонимизации;

- запись полученных данных в целевые хранилища информации в соответствии с системными правилами (проверка дубликатов, обновление ранее записанных данных и т. д.);

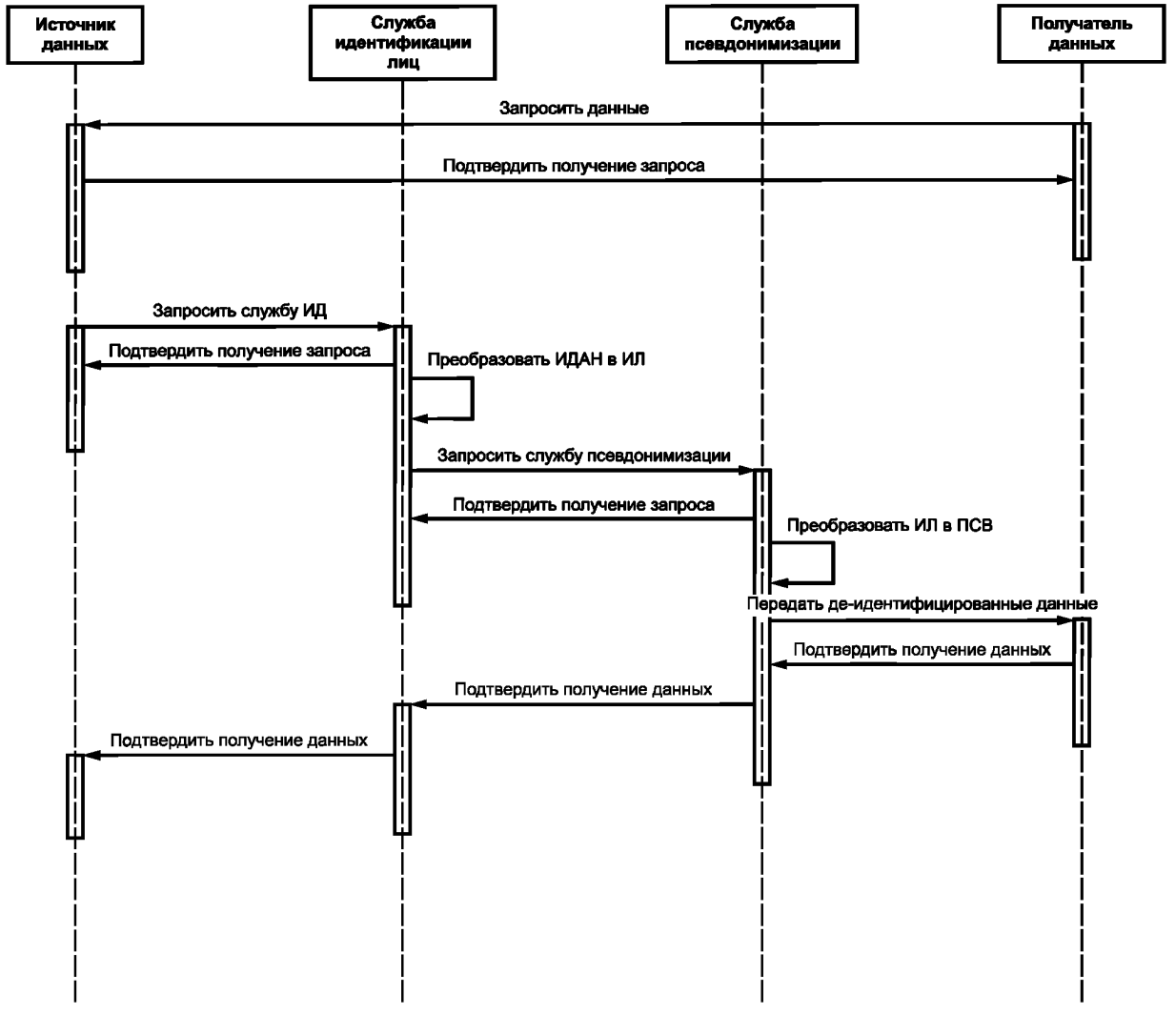
- статистический анализ результирующего набора характеристик.

Службы идентификации лиц и псевдонимизации представляют собой сущности, выполняющие соответственно установление идентичности пациента и процессы псевдонимизации. Вся информация, которая требуется политикой безопасности во время сеанса для принятия решения о разрешении обработки, должна быть предоставлена во время сеанса. Если псевдонимизация требуется для объединения данных, собираемых несколькими независимыми организациями, или для снижения угрозы несанкционированного восстановления идентичности, то такие службы идентификации должны обеспечиваться службой псевдонимизации.

Служба идентификации лиц управляет идентификаторами, передаваемыми службе псевдонимизации. Она взаимодействует с источником данных либо непосредственно, либо с помощью определенного отношения. В зависимости от архитектурного решения идентификаторы пациента у источника данных и у службы идентификации лиц могут отличаться.

6.3 Модель потоков работ

На рисунке 7 показаны типичные потоки данных между сущностями модели потоков работ. Они также идентифицируют типы сообщений, требуемых для передачи данных или уведомления о статусе операций с помощью сообщений подтверждения. Извлечение данных и интеграция данных выходят за рамки настоящей модели и осуществляются соответственно приложениями источника данных и получателя данных.



Служба ИД — служба идентификации; ИЛ — идентификатор лица; ИДАН — идентифицирующие данные; ПСВ — псевдоним

Рисунок 7 — Потоки данных

События обмена данными включают в себя:

- запрос получателем данных той информации, которая должна быть включена в запросы идентификации и псевдонимизации и соответствующие подтверждения;
- запрос идентификаторов пациента для передачи службе псевдонимизации и соответствующее подтверждение приема;
- передачу запроса на псевдонимизацию;
- передачу службой псевдонимизации подтверждения приема запроса на псевдонимизацию;
- передачу псевдонимизированных данных получателю данных; по завершении процесса псевдонимизации данные передаются на требуемую обработку;

- подтверждение получения: получатель данных подтверждает получение псевдонимизированных данных (подтверждение может также содержать результат проверки данных, например правильность их формата);

- если требуется сквозное подтверждение, то служба псевдонимизации транслирует подтверждение получения службе идентификации лиц, которая в свою очередь транслирует его источнику данных.

Если в процессе исполнения службы псевдонимизации обнаруживаются ошибки, то источнику данных возвращаются коды ошибок (например, неправильный формат данных, ошибка аутентификации источника данных).

6.4 Подготовка данных

Прежде чем передать данные службе псевдонимизации, источник данных должен их подготовить. Такая подготовка необходима для выполнения требований обеспечения конфиденциальности, определенных политикой конфиденциальности.

Согласно концептуальной модели применения служб псевдонимизации, данные должны быть разделены на две части: первая содержит идентифицирующие данные, а вторая — исключительно обезличенные данные.

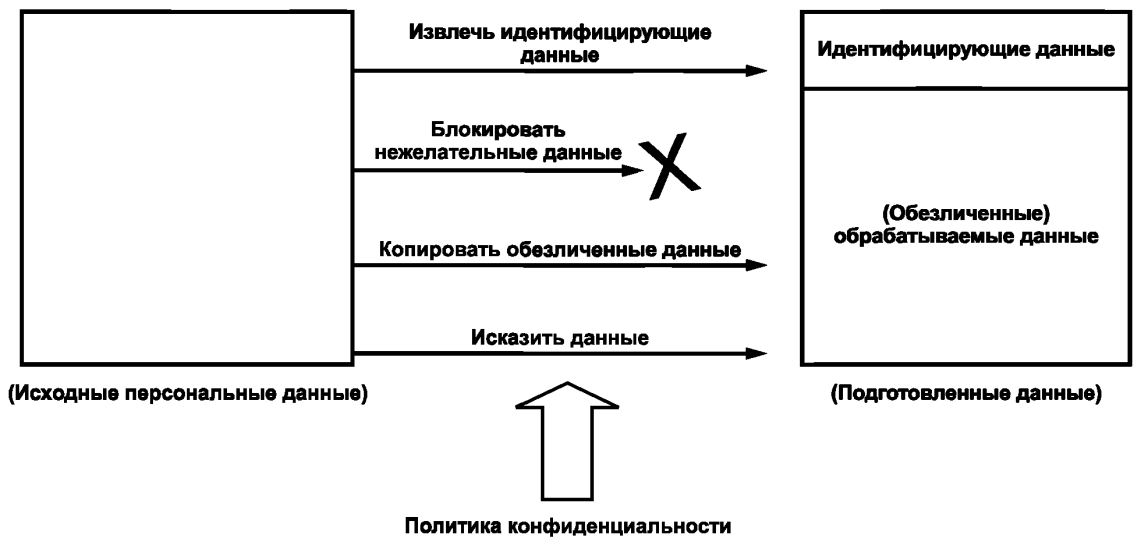


Рисунок 8 — Подготовка данных

Структурирование данных может быть выполнено с помощью разметки элементов данных, создания таблицы ссылок на данные или с помощью помещения элементов данных в заранее определенное место.

В соответствии с концептуальной моделью и требуемыми уровнями обезличивания различаются следующие способы подготовки:

- элементы данных, которые используются для связывания, группировки, поиска обезличенной информации, поиска совпадений и т. д., должны быть выделены и размечены таким образом, чтобы служба псевдонимизации знала, где найти их и как их обработать;

- в зависимости от политики конфиденциальности необходимо соответствующим образом разметить те элементы, которые должны быть преобразованы специальным образом, например, абсолютные значения даты и времени должны быть преобразованы в относительные, даты рождения в возрастные группы;

- идентифицирующие элементы, которые в соответствии с политикой конфиденциальности не могут участвовать в дальнейшей обработке приложением получателя данных, должны быть исключены;

- обезличенная часть исходных персональных данных помещается в обрабатываемую часть подготовленных данных.

6.5 Шаги обработки в потоке действий

Основные шаги псевдонимизации показаны на рисунке 9.

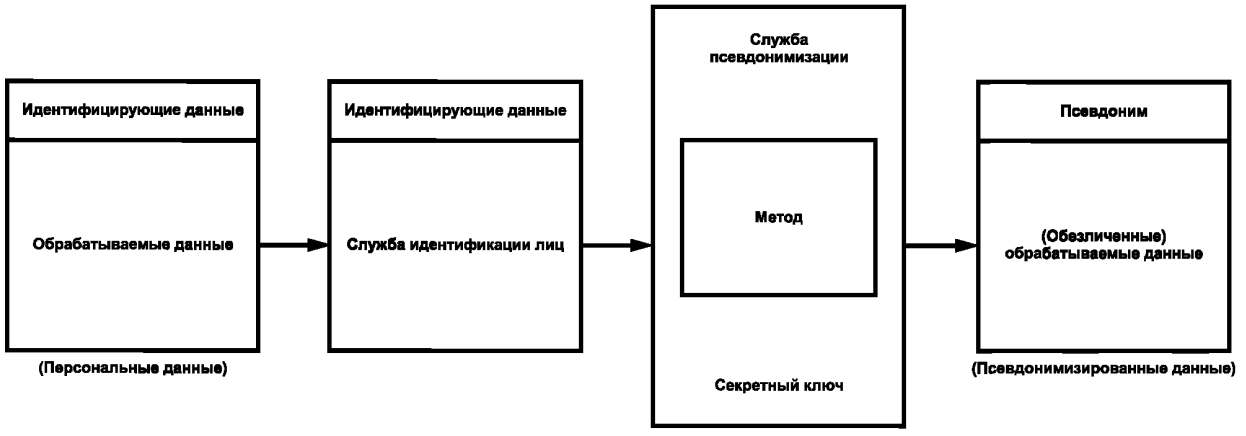


Рисунок 9 — Процесс псевдонимизации

Они состоят в следующем:

1) в источнике данных, который предоставляет данные службе псевдонимизации, данные расщепляются на идентифицирующие данные и обрабатываемые обезличенные данные. Какие именно считаются идентифицирующими данными, а какие обезличенными, зависит от требуемой степени обезличивания в политике безопасности системы сбора персональных данных;

2) служба псевдонимизации разбирает заголовок и выполняет действия, предписанные ее политикой. К ним относятся псевдонимизация элементов данных, вычисление относительных дат, удаление элементов данных и, возможно, шифрование отдельных элементов данных, если это требуется политикой обеспечения конфиденциальности. При этом для службы псевдонимизации содержание обрабатываемых данных остается невидимым. Это предпочтительный способ действий этой службы. Однако политика может предписывать иное, например, что содержание обрабатываемых данных также подлежит разбору. Обычно разбор осуществляется для проверки наличия нежелательного содержания (например, идентификаторов в элементах данных). Все проверки осуществляются «без сохранения состояния». Служба псевдонимизации не сохраняет ранее обработанные данные и поэтому не может сравнить с ними или выполнить проверку с участием этих данных. Учитываются только те данные, которые поступили в данном сеансе.

Примечание — Обработка псевдонимов и обезличивание обрабатываемых данных могут выполняться отдельными компонентами и службами;

3) по завершении обработки служба псевдонимизации отправляет псевдонимизированные данные в хранилище по защищенному каналу;

4) когда приложение хранилища получает данные, оно применяет соответствующие бизнес-правила для помещения данных в хранилище. Они могут включать в себя проверку дублирования данных, проверку отсутствующих элементов, требуемые процедуры подтверждения и т. д.

Для доверенной реализации метода псевдонимизации в случае, когда данные собираются несколькими независимыми организациями, служба псевдонимизации или доверенная третья сторона, осуществляющая псевдонимизацию, действительно является необходимой. Это обусловлено следующими тремя основными причинами:

- поскольку одна взаимодействующая сторона не всегда доверяет другой, то доверие может быть установлено опосредовано, если каждая из двух сторон доверяет независимой третьей стороне. Обе стороны связаны нормами регулирования, указанными в соглашении о политиках безопасности и обеспечении конфиденциальности, заключенном со службой псевдонимизации;
- использование службы псевдонимизации предлагает единственно надежную защиту против некоторых типов атак на процесс псевдонимизации;
- облегчается реализация дополнительной технологии усиления безопасности (ТУБ) и дополнительных функций обработки данных.

Примеры — Контролируемая обратимость, не ослабляющая конфиденциальность; интерактивные базы данных.

6.6 Обеспечение конфиденциальности с помощью псевдонимизации

6.6.1 Концептуальная модель проблемных областей

В настоящем стандарте основное внимание уделяется собираемой или хранящейся информации и в меньшей степени – диалоговому использованию систем пациентами. (Информация пациентов, введенная или измененная, может рассматриваться как хранящаяся.)

Существует много причин для обеспечения конфиденциальности с помощью сокрытия идентичности. В любом случае политика конфиденциальности должна задавать предмет защиты конфиденциальности с помощью псевдонимизации и определять, что считается идентифицирующей информацией, а что считается неидентифицирующей информацией.

С функциональной точки зрения важно определить, требуется ли обратимость псевдонимизации и каковы будут последствия обратимости, чтобы процедурно и технически способствовать санкционированному применению обратимости и препятствовать другим применениям.

В зависимости от схемы управления идентификацией системам управления идентификацией могут требоваться сложные функции псевдонимизации, включающие в себя преобразование псевдонимов при переходе от одного домена к другому.

Двумя важными элементами концепции псевдонимизации являются:

- домен, в котором будет использоваться псевдноним;
- защита ключа псевдонимизации или функции рассеивания.

6.6.2 Непосредственная и косвенная идентификация персональной информации

6.6.2.1 Общие сведения

Персональные данные могут идентифицироваться непосредственно или косвенно. Данные считаются непосредственно идентифицируемыми, если лицо может быть идентифицировано с помощью атрибута данных либо по его значению, либо по ссылке на общедоступный ресурс (или ресурс, доступ к которому регулируется альтернативной политикой), содержащий идентификаторы этого лица. В качестве таких атрибутов могут выступать хорошо известные идентификаторы (например, номер телефона или адрес) или числовые идентификаторы (например, номер направления на исследование, номер результата диагностического исследования, объектный идентификатор документа, номер лабораторного анализа). Косвенным идентификатором служит атрибут, который в сочетании с другими косвенными идентификаторами может обеспечить уникальную идентификацию лица (например, почтовый индекс, пол, дата рождения). К таким атрибутам также относятся защищенные косвенные идентификаторы (например, дата процедуры или дата медицинского изображения). Доступ к ним ограничен, но тем не менее они могут использоваться для идентификации пациента.

6.6.2.2 Характеристики, идентифицирующие лицо

К характеристикам, идентифицирующим лицо, относятся:

- имена лица (включая предпочтительные, юридические, другие, под которыми известно это лицо); под именами понимается фамилия и все другие элементы именования лица, описанные в документе ИСО/ТС 22220;
- идентификаторы лица (такие, как номер лицевого счета пациента, номер медицинской карты, номера лицензий/сертификатов, номер карточки социального страхования, номер страховых полисов, идентификаторы и регистрационные номера транспортных средств, в том числе номера государственной регистрации; эти номера могут включать в себя дополнительные атрибуты, например идентификацию организаций, присвоивших идентификаторы, типы идентификаторов и обозначения);
- биометрические данные (образцы речи, отпечатки пальцев, фотографии и т.д.);
- цифровые сертификаты, идентифицирующие лицо;
- девичья фамилия матери и другие подобные характеристики отношений с другими лицами (например, идентификация семьи);
- адрес регистрации по месту жительства;
- телекоммуникационные адреса (номер телефона, мобильного телефона, факса, пейджера, адрес электронной почты, адрес в сети Интернет, IP-адреса, идентификаторы и серийные номера устройств);
- связи с субъектами медицинской помощи (мать, отец, сестра или брат, ребенок);
- описания татуировок и других особых примет.

В зависимости от используемого стандарта формата данных могут быть дополнительные спецификации, которым надо следовать (например, дополнение 55 к стандарту DICOM).

6.6.2.3 Агрегируемые характеристики

Если данные используются для целей статистики, то следует избегать указания абсолютных значений:

- даты рождения (они являются очень существенными идентификаторами). Указание возраста безопаснее, но все же тоже может представлять собой угрозу в сочетании с наблюдаемыми данными. Поэтому лучше использовать возрастные группы или категории. Чтобы определить безопасные диапазоны возраста, надо провести анализ угроз восстановления идентичности, описание которого выходит за рамки настоящего стандарта;

- даты госпитализации и выписки и прочие, они также могут заменяться на категории периодов, но при этом события лечения должны регистрироваться по отношению к основному событию (например х месяцев после лечения);

- коды административно-территориальных образований; если они слишком точны, то такие географические указания следует агрегировать. Если эти коды имеют иерархическую структуру, то можно отбросить уточняющую часть кода, например, если вторая половина почтового индекса идентифицирует образование с 20 000 жителей или менее, то ее надо заменить на 000[1].

Демографические данные являются косвенными идентификаторами и по возможности должны либо удаляться, либо агрегироваться до уровня, определенного предметной областью или законодательством. Если надо оставлять эти данные, то необходимо провести анализ угрозы восстановления идентичности и соответствующим образом противодействовать обнаруженным угрозам. Демографические данные включают в себя:

- язык домашнего общения;
- язык общения с другими людьми;
- религию;
- этническую принадлежность;
- пол лица;
- страну рождения;
- профессию;
- криминальную историю;
- судебные решения;
- другие адреса (например, адрес места работы, адреса временной регистрации, почтовые адреса);
- порядок рождения (второй или следующий близнец при кратных родах).

Должен быть составлен документ политики, содержащий оценки возможности атак в данном контексте в качестве оценки угрозы для обеспечения конфиденциальности на уровне 2. К идентифицированным угрозам должны быть добавлены меры противодействия.

6.6.2.4 Аномальные характеристики

В зависимости от модели угроз аномальные характеристики должны удаляться. К таким характеристикам относятся:

- редко встречаемые диагнозы;
- необычные процедуры;
- некоторые профессии (например профессиональный теннисист);
- определенные нехарактерные особенности популяции в информационном ресурсе;
- различные уродства.

Должен быть составлен документ политики, содержащий оценки возможности атак в данном контексте в качестве оценки угрозы для обеспечения конфиденциальности на уровне 3. К идентифицированным угрозам должны быть добавлены меры противодействия.

Долговременные ресурсы данных, в которых используется псевдонимизация, должны быть предметом регулярного анализа угроз, связанных с наличием потенциально идентифицирующих аномальных характеристик. Такой анализ должен проводиться не реже одного раза в год. К идентифицированным угрозам должны быть добавлены меры противодействия.

6.6.2.5 Структурированные компоненты данных

Наличие у данных структуры определяет, какую информацию в них можно обнаружить и где ее можно обнаружить. При анализе угроз восстановления идентичности необходимо сделать предположения о том, что может привести к (нежелательной) идентификации, начиная от простых очевидных оценок до анализа заполненных баз данных и выводов зависимостей. В отличие от «структурированного» текста гарантированный результат автоматического анализа «свободного» текста невозможен.

---

[1] Требование, указанное в разделе 164.514 Закона HIPAA.

6.6.2.6 Неструктурированные компоненты данных

При наличии неструктурированных данных вопрос об их разделении на идентифицирующие и обрабатываемые данные становится основным. Свободный текст должен считаться уязвимым и рассматриваться как подлежащий удалению. Неструктурированные компоненты данных должны быть предметом следующих действий:

- определить, что в этих компонентах в соответствии с политикой обеспечения конфиденциальности (и желательным уровнем защиты конфиденциальности) может считаться идентифицирующей информацией;
- удалить данные, не являющиеся необходимыми;
- включить в политики требование, согласно которому неструктурированные данные не должны содержать непосредственно идентифицирующую информацию.

Те неструктурированные компоненты данных, которые в соответствии с политикой считаются неидентифицирующими, должны объединяться в качестве обрабатываемых данных.

Свободный текст

При использовании существующих подходов к псевдонимизации свободный текст нельзя рассматривать как гарантированно обезличенный. Весь свободный текст должен быть предметом анализа угрозы восстановления идентичности и соответствующей стратегии противодействия обнаруженным угрозам.

Противодействовать угрозам восстановления идентичности из оставленного свободного текста можно следующим образом:

- реализовывать политику, требующую, чтобы в свободном тексте не содержалась непосредственно идентифицирующая информация (например, фамилии, имена, отчества пациентов, номера пациентов);
- проверять свободный текст на наличие в нем идентифицирующей информации (что можно сделать, например в случае, когда свободный текст генерируется из структурированного текста);
- пересматривать, переписывать или превращать данные в кодированную форму.

По мере развития процедур разбора и обработки текста на естественном языке для «очистки» данных и алгоритмов псевдонимизации угрозы восстановления идентичности из свободного теста могут сделать приведенные требования менее строгими.

Свободный текст должен быть пересмотрен, переписан или каким-либо способом превращен в кодированную форму.

Текстовые/голосовые данные, содержание которых не может быть разобрано

Как и свободный текст, текстовые/голосовые данные, содержание которых не может быть разобрано, например речевые данные, должны быть удалены.

Медицинские изображения

Некоторые медицинские изображения содержат внутри себя идентифицирующую информацию (например, наложение на рентгеновском снимке содержит идентификаторы пациента). При возникновении проблемы с такими идентифицирующими данными, включенными в структурированный и кодированный заголовок сообщения в формате DICOM, необходимо применять указания, изложенные в дополнении 55 к стандарту DICOM (Attribute Level Confidentiality — конфиденциальность на уровне атрибутов). Необходимо также оценить угрозу восстановления идентичности по идентифицирующим характеристикам изображения или по обозначениям, являющимся частью изображения.

6.6.2.7 Оценка угроз предположений

Следует понимать, что псевдонимизация не может полностью защитить персональные данные, поскольку она не в полной мере противодействует атакам предположений. Применение служб псевдонимизации и обезличивания необходимо дополнить оценкой угроз, стратегиями противодействия угрозам, политиками информированного согласия и другими средствами анализа данных, предварительной обработки, последующей обработки. Администратор псевдонимизированных хранилищ должен нести ответственность за анализ содержания хранилищ данных на предмет угрозы предположений и за предотвращение раскрытия содержания отдельных записей. Источник информации должен нести ответственность за предварительный анализ/предварительную обработку передаваемых данных, защищающие эти данные от предположений, основанных на наличии аномальных данных, вложенных идентифицирующих данных и других ненамеренных раскрытий информации.

6.6.2.8 Конфиденциальность и безопасность

Всегда существует угроза, что псевдонимизированные данные могут быть связаны с субъектом данных. В свете этой угрозы собранные данные должны рассматриваться как «персональные данные» и должны использоваться только для заявленных целей сбора. Во многих странах законодательство требует, чтобы псевдонимизированные данные защищались наравне с идентифицирующими данными.

7 Процесс восстановления идентичности (методы и реализация)

Рассмотрим два разных контекста восстановления идентичности по псевдонимизированной информации:

- восстановление как часть обычной обработки;
- восстановление как исключительная обработка.

Восстановление как часть обычной обработки

Если восстановление идентичности является частью обычной обработки, то условия и процедуры восстановления должны быть элементом общей конструкции процессов. Пусть, например, псевдонимизированные запросы передаются из системы ведения электронной медицинской карты в патологоанатомическое отделение с использованием де-идентификации. Система ведения электронной медицинской карты получает результаты в псевдонимизированной форме, выполняет восстановление их идентификации и автоматически вставляет в медицинскую карту.

Восстановление идентичности в процедурах восстановления характеризуется тем, что оно осуществляется автоматически, прозрачным способом и не требует санкционирования каждого случая.

Когда восстановление идентичности является частью обычной обработки, то необходимо заботиться об обеспечении целостности данных (полноты и отсутствии преднамеренной или случайной модификации). В большинстве случаев при обработке требуется и обеспечивается та же степень целостности, что и для персональных данных. Это не обязательно имеет место для научной обработки данных, которая по этой причине относится к категории «исключительная обработка».

Исключительная обработка

Когда восстановление идентичности является исключением из стандартного способа обработки данных, то процесс восстановления должен требовать:

- специальных процедур аутентификации;
- нестандартного вмешательства со стороны провайдера службы псевдонимизации.

Если восстановление идентичности или де-идентификация рассматриваются как исключение из общего правила, то политика безопасности должна описать обстоятельства, которые могут приводить к восстановлению идентичности.

Документ политики безопасной обработки данных должен определять случаи, которые можно предвидеть, и его содержание должно отвечать следующим требованиям:

- каждый исключительный случай должен быть описан, и для него должны быть указаны один или несколько сценариев восстановления идентичности;
- должна быть обеспечена идентификация лица, инициирующего запрос на восстановление идентичности;
- в соответствии с процедурами авторизации должна быть проведена верификация прав лица на восстановление идентичности. В таких случаях все организации должны быть информированы о событии восстановления идентичности. Описанный процесс восстановления должен быть запущен только после надлежащего санкционирования (электронного или иного) и должен следовать сценарию, описанному в политике;
- исключительное восстановление идентичности должно выполняться только доверенным провайдером службы (в случае, когда необходим провайдер службы псевдонимизации, способный обеспечить восстановление идентичности);
- оператор данных, чья идентификация была восстановлена, должен проводить интенсивное тестирование целостности (правильности, полноты данных). В особенности это касается случая, когда меняется конечная цель обработки данных, например, псевдонимизированные научные данные преобразуются в данные, предназначенные для диагноза или лечения;
- в политике должно быть указано, кто будет оператором персональных данных, полученных после восстановления идентичности, и какова конечная цель обработки этих данных. Необходимо указать источник восстановленных данных (в качестве меры предосторожности, поскольку де-идентифицированные данные могут быть не настолько полны или надежны, как исходные персональные данные, которые были преобразованы для помещения в научные базы данных или в хранилища клинических данных).

В непредвиденных исключительных случаях применяются те же правила, что и в случаях, которые можно предвидеть. Отличие в том, что нет априорных сценариев восстановления идентичности. При этом необходимо оценить, насколько обоснована потребность в восстановлении идентичности. Ответственность за такие случаи возлагается на оператора.

Исключением из этого правила может быть восстановление идентичности в связи с осуществлением правосудия. Оно не описано в настоящем стандарте, однако предполагается, что органы, затребовавшие восстановление идентичности, реализуют должные меры обеспечения конфиденциальности переданных им персональных данных.

Техническая осуществимость

В случае, когда восстановление идентичности является частью обычной обработки или предусматривается для ряда заранее описанных сценариев, должна существовать техническая возможность восстановления.

Существует несколько методов выполнения восстановления идентичности:

- прямо или косвенно идентифицирующие данные (например, список местных идентификаторов) могут быть зашифрованы и сохранены вместе с псевдонимизированными данными. Только назначенный доверенный провайдер службы может расшифровать эти данные и связать косвенно идентифицирующие данные с субъектом данных;
- доверенный провайдер службы (провайдер службы псевдонимизации или провайдер службы депозитария) может хранить у себя список связей между псевдонимами и идентификаторами (прямо или косвенно идентифицирующими субъект данных).

8 Спецификация интероперабельности интерфейсов (методы и реализация)

Интероперабельность служб псевдонимизации должна быть определена на нескольких уровнях. Реализация псевдонимизации может, например, использовать промежуточную службу псевдонимизации или может представлять собой модуль местной разработки, добавленный к программному обеспечению сбора персональных данных.

Службы псевдонимизации могут быть ориентированы на пакетную обработку данных или могут встраиваться в оперативные базы данных для обеспечения псевдонимизированного доступа.

В настоящем стандарте основное внимание уделяется базовым механизмам, использующим службу псевдонимизации.

Интероперабельность может обеспечиваться с помощью следующих средств:

- один или несколько механизмов обмена данными между моделируемыми сущностями (источник, служба псевдонимизации, потребитель) и контроля выполняемых действий. Это не является особой проблемой, поскольку для обмена могут использоваться стандартные протоколы, например html. При необходимости могут быть разработаны конвертеры форматов обмена в качестве компонентов средств предварительной или последующей обработки;
- криптографические алгоритмы. Псевдонимизация на основе криптографических алгоритмов будет использовать цепочку базовых криптографических и сопутствующих алгоритмов: хеширование, генерацию случайных чисел, генерацию ключей, шифрование и базовые функции логического преобразования битовых строк;
- обмен ключами.

Поскольку псевдонимизация может использоваться в самом различном контексте, то важно ограничить контекст, для которого будет определена интероперабельность.

Общий подход к реализации большинства служб псевдонимизации состоит в использовании криптографических преобразований идентифицируемых данных.

Чтобы два независимых провайдера служб псевдонимизации были интероперабельными, необходимо одно из двух:

- интегрировать данные друг с другом: данные субъекта, обработанные одним провайдером, могут быть связаны с данными того же субъекта, обработанные другим провайдером, без непосредственного восстановления идентичности этого субъекта данных;
- контролируемым способом преобразовывать результаты псевдонимизации, полученные одним или несколькими провайдерами, без непосредственного восстановления идентичности этого субъекта данных.

9 Определение политики функционирования служб псевдонимизации (методы и реализация)

9.1 Общие сведения

Важно дополнять технические меры соответствующими нетехническими мерами. Нетехнические меры обычно устанавливаются политиками, соглашениями и нормами.

9.2 Политика обеспечения конфиденциальности

Для каждого процесса обработки персональных данных или проекта сбора таких данных, использующих псевдонимизацию, должна быть разработана политика обеспечения конфиденциальности, описывающая требования к псевдонимизации. Она может быть самостоятельной или являться частью общей политики безопасности.

Политика обеспечения конфиденциальности должна содержать:

- описание обработки, в которой участвует псевдонимизация;
- идентификацию оператора персональных данных;
- идентификацию оператора псевдонимизированных данных;
- описание метода псевдонимизации;
- идентификацию организации, в которой выполняется псевдонимизация;
- защиту, хранение и манипулирование «секретами» псевдонимизации (обычно ключом шифрования или таблицей связей); описание того, что случится, если организация перестанет существовать (или как минимум прекратит деятельность по псевдонимизации), описание того, для каких доменов и приложений будет использоваться секрет и как долго он будет оставаться действующим (для случая смены секрета должны быть описаны возможности и процедуры связи с унаследованными данными);
- детальное описание того, является ли псевдонимизация обратимой, какая и чья санкция требуется;
- описание ограничений, предъявляемых к получателю псевдонимизированных данных (например, действия с этими данными, пересылка третьей стороне, политики хранения данных):

он не может сделать их общедоступными;

он должен защищать их от несанкционированного доступа и использовать их внутри своей организации для преобразования их в де-идентифицированные данные, которые агрегированы и в таком виде могут быть сделаны общедоступными или проданы покупателям;

он должен удалить их, если они не используются, либо он больше не собирается защищать эти данные.

9.3 Доверенная практика деятельности

Чтобы обеспечить эффективную защиту конфиденциальности, служба псевдонимизации в здравоохранении должна преследовать следующие цели:

- надежную и защищенную привязку уникальных псевдонимов к лицам или организациям, являющимися субъектами псевдонимизированной медицинской информации;
- защиту псевдонимов от несанкционированного восстановления идентичности;
- обеспечение санкционированного восстановления исходных идентификаторов лица в соответствии с условиями политики восстановления, согласованными между провайдером и подписчиком служб.

Для достижения указанных выше целей необходимо использовать средства, вызывающие доверие у всех, кому необходима конфиденциальность персональной медицинской информации, защищаемой с помощью службы псевдонимизации. Поскольку псевдонимизация особо пригодна при проведении первичных и вторичных исследований и анализа медицинских данных, то пациенты, чья информация исследуется, равно как и общественность, должны испытывать доверие к информационным ресурсам, использующим эти службы для защиты персональной медицинской информации. Вряд ли медицинские работники или пациенты будут сотрудничать в сборе персональной медицинской информации для анализа, если у них сложится представление, что такие службы защиты идентификации не надежны.

Чтобы обеспечить достижение указанных выше целей, должны быть выполнены определенные условия.

Служба псевдонимизации должна:

- быть строго независимой от организаций-поставщиков исходных данных;
- гарантировать безопасность и доверенность своих методов действий и предоставлять своим подписчикам информацию о том, какими средствами это достигается;
- гарантировать безопасность и доверенность своих программных модулей;
- предусматривать поддержку исходных кодов, процессов и целостности своих программных модулей;
- обеспечивать проверку целостности программного кода с помощью его цифровой подписи;
- гарантировать безопасность и доверенность своей операционной среды, платформы и инфраструктуры;
- избавлять сетевой трафик от обменов данными, не являющимися необходимыми;
- обеспечивать остановку всех процессов операционной системы, не являющихся необходимыми;
- обеспечивать технические, физические, процедурные меры защиты и меры защиты, связанные с персоналом, в соответствии со стандартом ИСО 27799;

- обеспечивать мониторинг и оценку качества служб и программ;
- поддерживать необходимое качество службы;
- вести мониторинг несанкционированного проникновения в сеть и атак вредоносных программ.

Управление криптографическими ключами:
- должно контролироваться несколькими лицами;
- идентификаторы должны шифроваться двумя ключами: одним, принадлежащим источнику данных, и другим, принадлежащим службе псевдонимизации.

Установленный экземпляр службы псевдонимизации:
- должен быть задокументирован;
- должен регистрировать выполнение действий и обеспечивать возможность их аудита в целях демонстрации целостности службы.

Доступность службы псевдонимизации:
- должна обеспечиваться процедурами резервного копирования и восстановления;
- должна дополняться планом восстановления операций при возникновении чрезвычайной ситуации.

Процедуры внутреннего аудита:
- должны быть документированы;
- должны выполняться не реже одного раза в месяц.

Процедуры внешнего аудита:
- должны проводиться таким образом, чтобы подписчики службы и другие участвующие стороны могли убедиться, что деятельность службы полностью соответствует опубликованным требованиям;
- должны проводиться аудитором, полностью независимым от контролируемой стороны и принадлежащим к организации, не связанной с провайдером службы псевдонимизации;
- должны проводиться аудитором, не имеющим финансового интереса в контролируемой стороне;
- должны проводиться аудитором, имеющим знания в области информационных систем, достаточные для членства в соответствующем профессиональном сообществе;
- должны обеспечивать немедленное информирование подписчиков службы обо всех службах псевдонимизации, которые признаны аудитором не соответствующими заявленным требованиям.

Участники:
- должны управлять целостностью ключей организации, используемых для псевдонимизации;
- обеспечивать в смежных информационных системах технические, физические, процедурные меры защиты и меры защиты, связанные с персоналом, в соответствии со стандартом ИСО 27799;
- нести ответственность за обезличивание обрабатываемых данных и защиту конфиденциальности всех псевдонимизированных ресурсов, используемых своей организацией.

Должна проводиться оценка угроз, связанных с доступом источника данных к результирующим псевдонимам, и требования к мерам противодействия должны быть включены в политику операционных процедур.

9.4 Реализация доверенной практики восстановления идентичности

Служба псевдонимизации должна поддерживать контролируемое восстановление идентичности. При обеспечении такой поддержки служба псевдонимизации должна предоставлять подписчикам и субъектам данных политику восстановления, в которой должны быть указаны критерии, предъявляемые к санкционированному событию восстановления идентичности:
- восстановление идентичности должно контролироваться несколькими лицами и несколькими организациями;
- процедура срочного восстановления идентичности должна быть описана в определенной политике и реализована для передачи информации тем, кто нуждается в таком восстановлении (например службе санитарно-эпидемиологического контроля).

Регистрация событий восстановления должна:
- проводиться для всех событий восстановления идентичности в соответствии с документом RFC 3881;
- как минимум включать в себя следующую информацию о:
  - стороне, которой раскрывается идентичность;
  - дате и времени восстановления идентичности;
  - причине восстановления из числа определенных в 5.5.

Процедура восстановления идентичности, выполняемая службой псевдонимизации, должна восстанавливать идентичность только по местному псевдониму, предоставленному организации—источнику информации.

Оператор персональных данных несет ответственность за восстановление идентичности пациента и может в соответствии с местным законодательством выполнять дополнительную проверку запроса на восстановление идентичности.

Приложение А
(справочное)

Сценарии псевдонимизации в здравоохранении

А.1 Введение

В настоящем приложении описан ряд высокоуровневых ситуаций или «сценариев», иллюстрирующих базовые функциональные и технические требования к службам псевдонимизации, предназначенным для обеспечения взаимодействия различных организаций сферы здравоохранения.

Вначале приводятся общие требования, касающиеся базовых принципов конфиденциальности и безопасности, а также показана необходимость применения этих принципов для сферы здравоохранения. Затем приведены описания каждого сценария по следующей форме:

1) описание сценария или ситуации при оказании медицинской помощи, в которых требуются службы псевдонимизации;

2) соответствующие функциональные и технические требования к службе псевдонимизации.

А.2 Объяснение сценариев

Сценарии, описанные в А.3.1—А.3.5, показывают, как службы псевдонимизации могут применяться в здравоохранении. Каждый сценарий описывает возможные применения службы псевдонимизации в здравоохранении.

Таблица с описанием сценариев содержит следующие названия граф:

Тип ИД

Указывает кому принадлежит идентификатор: пациенту или, например, медицинскому работнику.

Уникальность

Чтобы оценить, каким образом можно использовать псевдонимизированную базу данных, необходимо знать, однозначно ли входной параметр идентифицирует лицо в данном контексте. Это особенно важно, если надо однозначно связать между собой данные об одном и том же лице, собираемые за длительный промежуток времени разными организациями. Важно также установить, можно ли связать между собой все данные о лице, собираемые одной организацией, или существует риск наличия синонимов в целевых базах данных.

Чувствительность данных

При проектировании службы псевдонимизации полезно иметь характеристику чувствительности данных. Степень чувствительности должна определяться на основе требований законодательства или коммерческой ценности. Например, с юридической точки зрения информация о наличии у физических лиц вируса иммунодефицита человека (ВИЧ) может иметь гораздо большую степень чувствительности и потребовать соразмерного анализа угроз. А информация о частоте выздоровления при применении конкретного метода лечения заболевания с юридической точки зрения не является чувствительной, но может быть критичной с позиции коммерциализации.

Источники данных: кратность источников данных и их взаимосвязь

Требования, которые могут предъявляться законодательством к различным реализациям псевдонимизации, существенно зависят от этой характеристики. Важно также знать, получают ли данные непосредственно от субъекта данных.

Контекст/назначение: коммерческое, медицинские исследования, лечение пациентов

В этой графе дается краткое описание контекста.

Наличие поиска/связывания

Наличие поиска является очень важным элементом общей конструкции решения по применению псевдонимизации. Должна быть определена степень точности поиска, а также возможность поиска псевдонимизированных элементов путем задания непсевдонимизированных элементов (например, территории). Реализация функции поиска требует применения службы псевдонимизации и может ограничиваться.

Обратимость/восстановление

В этой графе указано, является ли восстановление идентичности желательным, запрещенным, желательным в контролируемых условиях или оно должно осуществляться при пока неизвестных, но в будущем желательных условиях.

Должна также приводиться информация о том, какой объем восстановления идентичности приемлем.

Связывание во времени

Угроза восстановления идентичности тем выше, чем больше объем псевдонимизированной информации, которая может быть связана воедино. Ограничивая связывание псевдонимизированной информации во времени, можно уменьшить объем связываемой информации. Конечно, это может войти в противоречие с требованием проведения научных исследований, для которых требуется информация о пациенте за длительное время.

Связь между областями применения

Использование конкретного ключа шифрования или метода псевдонимизации должно ограничиваться как можно более узкой областью применения. Поэтому в сценариях важно описать область, в которой будет применяться псевдонимизация, как долго она будет применяться и какая требуется связь с другими областями применения. В свою очередь, от этого зависит необходимость привлечения посреднической организации.

Этот аспект должен также учитывать кооперацию различных посреднических организаций.

А.3 Сценарии применения псевдонимизации в здравоохранении

Характеристики сценариев применения псевдонимизации в здравоохранении показаны в таблице А.1.

Таблица А.1 — Характеристики сценариев

Сценарий		Субъект данных			Источники данных		Требования к функциям и их выполнению			
		Тип ИД	Уникальность	Чувствительность (перс. данных согласно законам)	Кратность источников данных	Первичный/вторичный	Контекст/назначение	Наличие поиска	Восстановление	Связывание во времени
1	Псевдон. медпомощь	ИД пациента	Уникален в присвоившей системе (МИС)	Высокая	Единственный	Первичный	Лечение	Нет	Да	Да
2	Клинические испытания	ИД пациента	Уникальность не гарантируется	Высокая	Мультицентровый	Первичный	Научная работа	Да	Нет (политика исключения)	Да/нет
3	Клинические результаты	ИД пациента / ИД медработника	Уникальность не гарантируется	Высокая	Мультицентровый	Вторичный	Научная работа	Да	Нет (политика исключения)	Да
4	Мониторинг обществ. здоровья	ИД пациента / ИД медработника	Уникальность не гарантируется	Высокая	Мультицентровый	Первичный/вторичный	Управление общественным здоровьем	Да	Да, в строго контролируемых случаях	Да
5	Отчет о побочных действиях лекарств	ИД пациента / ИД медработника	Уникален	Низкая (врач) Высокая (диагноз)	Мультицентровый	Первичный	Научная работа	Да	Да	Да
6	Исследование вне сферы здравоохранения	ИД пациента, другие ИД лица	Очень гетерогенная, уникальности нет	Высокая для медицинских данных	Мультицентровый	Вторичный	Немедицинское исследование	Да	Нет	Да

Окончание таблицы А.1

Сценарий		Субъект данных			Источники данных		Требования к функциям и их выполнению			
		Тип ИД	Уникальность	Чувствительность (перс. данных согласно законам)	Кратность источников данных	Первичный/вторичный	Контекст/назначение	Наличие поиска	Восстановление	Связывание во времени
7	Маркетинговое исследование в здравоохранении	ИД врача/ ИД пациента	Уникален	Низкая (врач) Высокая (диагноз)	Мультицентровый	Вторичный	Немедицинское исследование	Да	Нет	Да

1) Псевдонимизированная медицинская помощь (Псевдон. медпомощь).
2) Клинические испытания и постмаркетинговые исследования (Клинические испытания).
3) Вторичное использование клинических данных, например, научное исследование (Клинические результаты).
4) Мониторинг и исследование общественного здоровья (Мониторинг обществ. здоровья).
5) Конфиденциальный отчет о побочных действиях лекарств (Отчет о побочных действиях лекарств).
6) Исследование вне сферы здравоохранения (Исследование вне сферы здравоохранения, ранее — группы потребителей).
7) Маркетинговое исследование в здравоохранении (Маркетинговое исследование в здравоохранении. Включает сравнительный отчет показателей качества, экспертизу, контрольную деятельность, клиническую квалификацию/правильность счетов на оплату лечения, выставление счетов на оплату лечения).

А.3.1 Направление на лабораторный анализ (псевдонимизированная медицинская помощь)

Сценарий приведен в качестве примера (в данной группе).

В данном сценарии служба псевдонимизации используется для защиты идентификации пациентов и согласованного отслеживания информации об одном и том же пациенте при обмене данными между разными системами.

Лечащему врачу требуется послать биоматериал на исследование (лабораторные анализы). Согласно политике конфиденциальности информация, идентифицирующая пациента, не должна передаваться в направлении на анализы. Однако врачу важно, чтобы результаты анализов соответствовали направлению, а лабораторной службе важно иметь возможность сравнить текущие и предыдущие результаты анализов биоматериала пациента. Прежде чем послать направление в лабораторию, с помощью доверенной службы псевдонимизации генерируется псевдоним, затем он включается в направление, и результаты анализов возвращаются с этим псевдонимом. По псевдониму восстанавливается идентичность пациента, и результаты включаются в запись его медицинской карты.

Действующие лица: заказчик исследования (например, в условиях стационара — лечащий врач), исполнитель заказа (например, клиническая лаборатория), служба псевдонимизации, медицинская информационная система (МИС).

Предусловия: заказчик исследования выбирает список анализов, которые должен провести исполнитель заказа. Список анализов привязывается к субъекту данных с помощью уникального идентификатора пациента, присвоенного ему при поступлении в стационар.

Постусловия: заказчик исследования получает от исполнителя результаты анализов и включает их в медицинскую карту субъекта данных, используя уникальный идентификатор, ранее включенный в направление на анализы.

Поток работ / события/ действия

Ввод заказа в медицинскую информационную систему (МИС):

- МИС аутентифицирует лечащего врача;
- лечащий врач вводит направление на лабораторные анализы, указывая в нем уникальный номер субъекта данных, присвоенный ему при поступлении в стационар;
- лечащий врач проверяет соответствие направления политикам (например, получателю не разрешается получать идентифицирующую информацию, пациент является очень важным лицом) и принимает решение о применении мер защиты.

Псевдонимизация:

- МИС вызывает службу псевдонимизации, передавая ей в качестве параметра уникальный номер пациента;
- служба псевдонимизации обрабатывает этот номер;
- служба псевдонимизации возвращает МИС псевдоним;
- МИС передает исполнителю направление, в котором уникальный номер пациента заменен на псевдоним:

соединяется с лабораторной информационной системой;
передает сообщение;
получает подтверждение приема.

Исполнитель выполняет лабораторные анализы в соответствии с направлением, в котором указан псевдоним пациента (при выполнении анализов специалист проводит сравнение с результатами, ранее сделанными для этого пациента).

Исполнитель направления отправляет в МИС результаты, в которых указан псевдоним пациента:
- соединяется с МИС;
- передает сообщение;
- получает подтверждение приема.

Восстановление идентичности результатов:
- МИС передает псевдоним службе псевдонимизации;
- служба псевдонимизации проверяет в соответствии с политикой восстановления идентичности, разрешено ли оно для данного пользователя МИС;
- служба псевдонимизации обрабатывает псевдоним;
- служба псевдонимизации возвращает МИС истинный идентификатор пациента;
- МИС находит по этому идентификатору медицинскую карту пациента и включает в нее результаты.

Другие примеры / замечания

Он-лайновые консультации через Интернет — оказание медицинской помощи лицу (тому же самому) в разное время.

Публично известное лицо обращается к медицинскому работнику за помощью. Желая обеспечить конфиденциальность этого эпизода и последующего лечения, пациент требует, чтобы во всех этих случаях использовались его псевдонимизированные идентификаторы.

А.3.2 Клинические испытания

А.3.2.1 Общие сведения

Существует большое количество разновидностей клинических испытаний. Сбор данных при клинических испытаниях лекарств для представления в Управление по санитарному надзору за качеством пищевых продуктов и медикаментов регулируется рядом процедурных норм. Испытываются новые диагностические устройства, например, с помощью ROC-анализа, а также новые процедуры, при этом в соответствии с действующим законодательством требуется псевдонимизация не только для обеспечения конфиденциальности. При применении такого научного метода, как двойное слепое исследование, псевдонимизировать надо даже внутренние данные.

На рисунке А.1 указаны различные места, в которых данные могут быть изменены для добавления идентифицирующих атрибутов клинических испытаний (ИКИ) и/или удаления атрибутов в целях псевдонимизации.

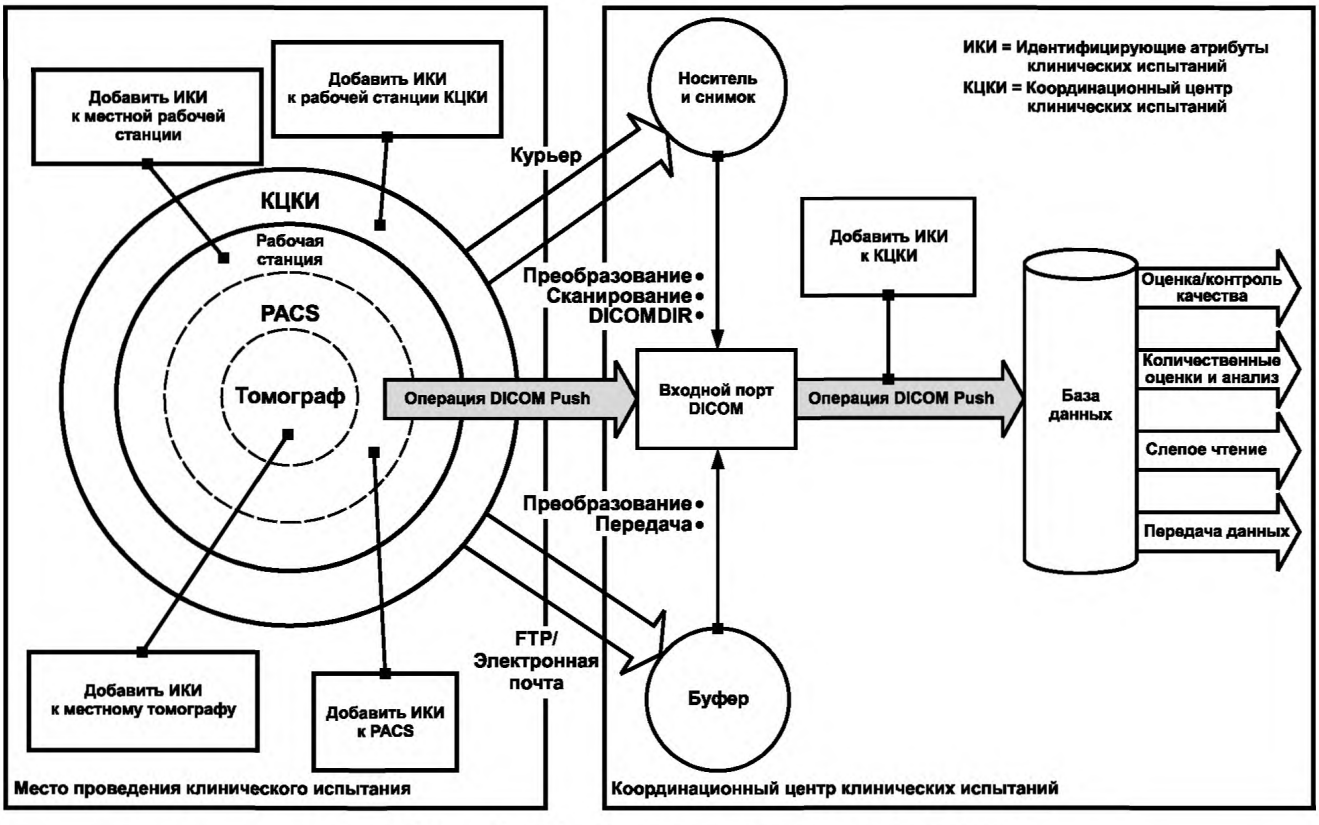


Рисунок А.1 — Модификация данных клинических испытаний

В отличие от преподавания с использованием медицинских данных пациентов, в процессе проведения клинических испытаний обычно участвуют несколько сторон:

Спонсор клинических испытаний, который задает научные требования к испытанию. Обычно в этих требованиях указаны данные, которые в целях научного анализа должны быть сохранены, данные, которые должны быть сделаны слепыми, и данные, которые должны быть удалены.

Координационный центр клинических испытаний, который координирует, собирает и готовит данные. Этот центр может также обеспечивать псевдонимизацию данных, зависящую от выбранных процедур и соглашений о месте проведения клинических испытаний.

Использование нескольких *мест проведения клинических испытаний*, в которых осуществляется фактическая клиническая деятельность. В них псевдонимизируют данные при взаимодействии со спонсором клинических испытаний и в соответствии со своими политиками обеспечения конфиденциальности и требованиями данного спонсора.

Привлечение других *контролеров*, например, Управление по санитарному надзору за качеством пищевых продуктов и медикаментов, которые рассматривают результаты клинического испытания.

Если необходимо, чтобы пациенты, участвующие в клиническом испытании, получили уведомления о результатах, важных для их лечения, то в таком испытании может потребоваться обратимость псевдонимизации. Ее можно обеспечить несколькими способами. Один из них: контролер, который получает результаты испытания, должен иметь возможность сообщить кому-либо (например, агенту клинического испытания), что «пациент X, участвовавший в клиническом испытании Y, должен быть уведомлен о результатах ...».

А.3.2.2 Где используется псевдонимизация

Заранее очень трудно принять какое-либо конкретное решение о том, какие данные должны быть сделаны слепыми и каким именно образом. Диапазон исследуемых характеристик может быть очень широким, и нередко информация о них не может быть сделана слепой. В каждом клиническом испытании должны быть определены собственные правила «слепоты» и псевдонимизации, хотя эту работу можно упростить, если за основу взять правила, предложенные в предшествующих аналогичных испытаниях.

А.3.2.3 Требования к псевдонимизации

Проведение некоторых клинических испытаний может регулироваться особыми нормами сбора данных. Например, эти нормы могут включать в себя ведение полного регистрационного журнала и документирование всех изменений данных, в том числе предназначенных для целей де-идентификации. Эти нормы существенным образом влияют на выбор методов де-идентификации.

Сценарий приведен в качестве примера (в данной группе).

Предоставление данных для клинических испытаний. Этот сценарий описывает сбор данных из одного источника медицинской помощи и ресурсы данных, используемых для клинического испытания.

Действующие лица: пользователь системы (например, исследователь, медицинский сотрудник, участвующий в лечении, информационный ресурс клинического испытания, информационный ресурс поставщика медицинской помощи (МИС)).

Предусловия: пациент участвует в клиническом испытании; у исследователя есть система сбора данных, которая равным образом удовлетворяет требованиям клинического испытания и внешнего информационного ресурса; доступна местная информационная система; получено информированное согласие пациента; выполнен этап клинического испытания.

Постусловия: информационный ресурс клинического испытания имеет все необходимые данные о каждом эпизоде лечения пациента, участвующего в испытании. Внешний информационный ресурс (например, МИС, система ведения электронной медицинской карты) имеет все необходимые данные о каждом эпизоде лечения пациента.

Поток работ / события / действия

Участники предоставления данных для клинических испытаний показаны на рисунке А.2.

Течение процесса:

- пользователь из числа медицинских работников, обеспечивающих лечение, аутентифицируется в МИС;
- МИС инициирует регистрацию доступа в журнале, фиксируя время регистрации;
- медицинский работник вводит данные в систему сбора данных;
- система обезличивает или псевдонимизирует данные;
- система передает релевантные данные информационному ресурсу клинического испытания;
- информационный ресурс клинического испытания получает данные;
- МИС сбора данных отправляет информацию о лечении внешнему информационному ресурсу, и исследователь клинического испытания просматривает и проверяет (с помощью службы eSignature или иного механизма проверки), точно ли эти данные отражают собранные данные, требуемые для испытания.

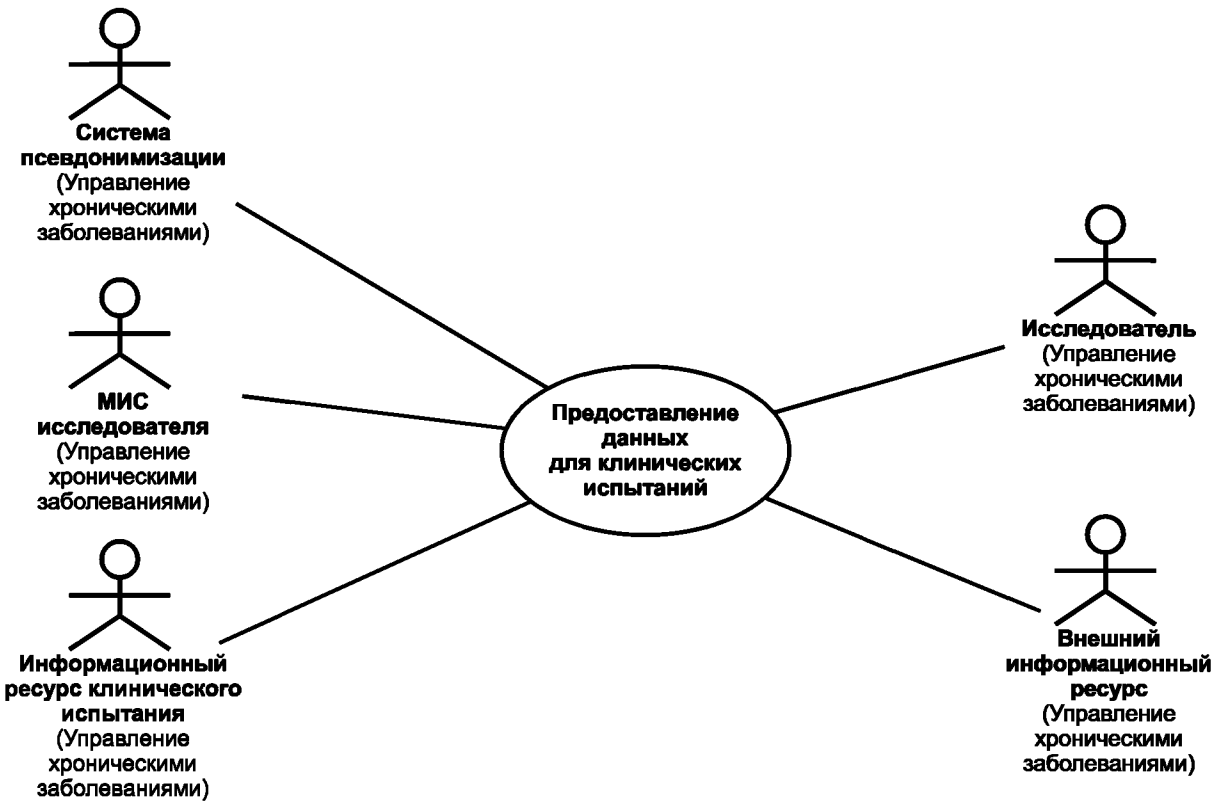


Рисунок А.2 — Участники предоставления данных для клинических испытаний

А.3.3 Клиническое исследование

Сценарий приведен в качестве примера (в данной группе).

Вторичное использование клинических данных для научных целей.

Отдел нефрологии собирает в стационарах и медицинских центрах данные о профильном лечении пациентов, страдающих диабетом. Для лечения диабета использовались разные схемы лекарственной терапии. Данные о лечении хранятся в нескольких системах, в качестве идентификатора пациента используется национальный страховой номер индивидуального лицевого счета. Какое-то время спустя, соискатели ученой степени решили проанализировать результаты успешного лечения. Это использование данных является вторичным, поскольку оно не предназначено для продолжения лечения. Соискатели должны собрать сведения из разных баз данных и сгруппировать их таким образом, чтобы они относились к одному и тому же пациенту. Однако медицинские организации — владельцы информации — не собираются раскрывать идентифицирующие персональные данные. Поэтому все данные должны быть псевдонимизированы с помощью службы псевдонимизации, которая удалит непосредственно идентифицирующие данные и заменит их на псевдонимы в соотношении один к одному. Исследователи не будут знать идентификацию пациентов, но тем не менее смогут группировать данные, относящиеся к одному и тому же пациенту.

При научном анализе результатов лечения обнаружилось, что для выполнения важнейшего исследования, помимо собранных данных, необходима связанная с ними дополнительная информация. Соответствующая экспертная комиссия санкционирует восстановление идентичности пациентов, чтобы запросить разрешение и узнать о заинтересованности участия в дальнейшем научном исследовании у лиц, которые могли бы составить исследуемую группу.

С одним и тем же набором данных можно проводить разные программы исследования.

Действующие лица: пользователь системы (например, исследователь, медицинский сотрудник, участвующий в лечении, информационный ресурс научного исследования, информационный ресурс поставщика медицинской помощи (МИС)).

Предусловия: медицинская карта, представляющая интерес для исследователя (данные обо всех эпизодах лечения или только о тех, которые удовлетворяют критериям темы исследования популяционной группы); у исследователя есть система сбора данных, которая удовлетворяет потребности как научного исследования, так и МИС; местная информационная система; информированное согласие пациента, полученное в соответствии с местным законодательством; завершенная часть эпизода лечения.

Постусловия: информационный ресурс научного исследования содержит все релевантные псевдонимизированные данные всех эпизодов лечения пациентов популяционной группы. Обеспечена защита конфиденциальности этих данных. Внешний информационный ресурс (например, МИС, система ведения электронной медицинской карты) имеет все релевантные данные о каждом эпизоде лечения пациента.

Поток работ / события / действия

Течение процесса:

- пользователь из числа медицинских работников, обеспечивающих лечение, аутентифицируется в МИС;
- МИС инициирует регистрацию доступа в журнале, фиксируя время регистрации;
- медицинский работник вводит данные в систему сбора данных;
- в целях обеспечения конфиденциальности система генерирует агрегированные данные;
- система проверяет наличие уникально идентифицирующих характеристик данных (например, редких диагнозов) или сочетаний характеристик;
- система обезличивает или псевдонимизирует данные;
- система передает релевантные данные информационному ресурсу научного исследования;
- информационный ресурс научного исследования получает данные;
- МИС сбора данных отправляет информацию о лечении местной МИС.

Другие примеры / замечания

Подготовка данных для образования:

Отчет о сравнительных показателях качества: данные об эпизодах лечения и выписной эпикриз передаются поставщиком медицинской помощи в научно-исследовательскую базу данных. Идентификаторы пациента псевдонимизируются с помощью службы псевдонимизации, поскольку имеются идентифицирующие группировки и данные, необходимые для выравнивания рисков. Чтобы еще лучше защитить данные от атак предположения, проводится агрегирование данных, например, сведений о продолжительности лечения. Идентичность поставщика медицинской помощи также псевдонимизируется в целях защиты идентичности врачей и медицинских организаций.

Экспертиза: разработан новый метод хирургической операции. Врачи используют службу псевдонимизации для предоставления регистру общего пользования информации о проведенном лечении и побочных эффектах. Этот регистр предназначен для проведения экспертных оценок трендов и сравнения результатов для разных диагностически связанных групп и сопутствующих заболеваний. Конфиденциальность пациентов и врачей обеспечивается с помощью псевдонимов, получаемых от службы псевдонимизации. Это позволяет собрать вместе данные пациента, предоставленные разными врачами, чтобы оценить полную картину медицинской помощи.

При экспертной оценке случаев лечения выявлено, что у пациента, получавшего лечение у нескольких поставщиков медицинской помощи, повышен риск послеоперационного осложнения. Для этого случая производится восстановление идентичности, чтобы пригласить пациента на дополнительное обследование и лечение.

А.3.4 Мониторинг общественного здоровья

Сценарий приведен в качестве примера (в данной группе).

Способность быстро выявить случаи угроз общественному здоровью и соответствующим образом мобилизовать ресурсы для проведения санитарных мероприятий может спасать жизни. Информация, собранная в больницах, в других медицинских учреждениях и вспомогательных организациях, может в электронной форме отправляться в санитарно-эпидемиологическую службу в целях социально-гигиенического мониторинга, который практически в реальном времени дает представление о состоянии общественного здоровья и информирует ответственные службы о необходимости реагирования на случаи угроз общественному здоровью. Идентификация пациентов при мониторинге не обязательна. Собранные данные могут передаваться органам государственного и местного управления, а также медицинским организациям для обеспечения скоординированных действий.

Действующие лица: пользователь системы (например, санитарный врач, медицинский сотрудник, участвующий в лечении), информационная система санитарно-эпидемиологической службы, клинический информационный ресурс, информационный ресурс санитарно-эпидемиологической службы.

Предусловия: заданы критерии механизма фильтрации передаваемых данных; определены алгоритмы выявления событий; пациент правильно идентифицирован; поставщик медицинской помощи правильно идентифицирован; доступны службы псевдонимизации, де-идентификации и восстановления идентичности.

Постусловия: данные переданы несколькими клиническими информационными ресурсами информационной системе санитарно-эпидемиологической службы, в которой реализованы функции, необходимые для выявления угроз общественному здоровью, то есть эта система обеспечивает социально-гигиенический мониторинг и анализ общественного здоровья, выявление и исследование угроз, передает уведомления о состоянии общественного здоровья и предупреждения об угрозах и другие данные, связанные с угрозами общественному здоровью.

Поток работ / события/ действия

Сбор данных информационной системой санитарно-эпидемиологической службы:

- клинический информационный ресурс обеспечивает ввод данных о случаях медицинской помощи в систему ведения электронной медицинской карты (ЭМК);
- система ведения ЭМК обеспечивает хранение данных, необходимых информационной системе санитарно-эпидемиологической службы;

- клинический информационный ресурс инициирует ведение регистрационного журнала событий, учитывающего их время появления;

- клинический информационный ресурс просматривает и проверяет (с помощью службы eSignature или иного механизма проверки), точно ли данные в системе ведения ЭМК отражают исходные данные;

- клинический информационный ресурс на основе критериев фильтрации извлекает информацию для предоставления (передачи) информационной системе санитарно-эпидемиологической службы;

- клинический информационный ресурс вызывает соответствующую службу для псевдонимизации данных;

- клинический информационный ресурс предоставляет (передает) релевантные данные информационной системе санитарно-эпидемиологической службы, используя защищенные средства передачи сообщений;

- клинический информационный ресурс получает подтверждение приема данных от информационной системы санитарно-эпидемиологической службы.

Выявление угроз общественному здоровью:

- поставщик медицинской помощи получает от информационной системы санитарно-эпидемиологической службы телефонное или защищенное электронное уведомление о подозрении на угрозу общественному здоровью;

- клинический информационный ресурс по мере необходимости предоставляет информационной системе санитарно-эпидемиологической службы дополнительные данные;

- поставщик медицинской помощи получает от информационной системы санитарно-эпидемиологической службы телефонное или защищенное электронное предупреждение о выявленной угрозе общественному здоровью;

- клинический информационный ресурс получает от информационной системы санитарно-эпидемиологической службы уведомление о необходимом продолжении лечения пациента, у которого выявлено заболевание, угрожающее общественному здоровью.

Последующий мониторинг выявленного события:

- клинический информационный ресурс собирает дополнительные данные, необходимые для ликвидации вспышки заболевания, особенно сведения о новых и рано выявленных случаях заболевания или подозрений на них, и предоставляет их информационной системе санитарно-эпидемиологической службы;

- аутентифицированный клинический информационный ресурс вызывает службу псевдонимизации для восстановления идентичности пациента в целях продолжения лечения пациента и скрининга семьи и контактов пациента в соответствии с санитарно-эпидемиологическими правилами;

- клинический информационный ресурс ежедневно передает информационной системе санитарно-эпидемиологической службы данные о проводимых мероприятиях;

- клинический информационный ресурс получает от информационной системы санитарно-эпидемиологической службы защищенные электронные уведомления об изменениях санитарно-эпидемиологической обстановки.

Быстрое реагирование на событие:

- клинический информационный ресурс получает от информационной системы санитарно-эпидемиологической службы защищенные электронные предписания по проведению противоэпидемических (профилактических) мероприятий в соответствии с санитарно-эпидемиологическими правилами;

- клинический информационный ресурс отправляет подтверждение получения от службы социально-гигиенического мониторинга защищенных электронных предписаний по проведению противоэпидемических (профилактических) мероприятий в соответствии с санитарно-эпидемиологическими правилами.

Другие примеры / замечания

Еженедельно информационные системы врачей общей практики передают сведения о случаях аллергических заболеваний и гриппа в центральное национальное хранилище. Перед отправкой идентификаторы пациента и врача псевдонимизируются с помощью службы псевдонимизации, и информация о точном месте жительства пациента заменяется на более общую. Центральное национальное хранилище используется для формирования предупреждения о санитарно-эпидемиологической обстановке по гриппу и аллергиям, для которых идентифицирующие данные не требуются.

А.3.5 Информирование о побочных действиях (лекарственных средств)

Описание сценария: мониторинг безопасности терапии. Этот сценарий описывает деятельность, связанную с мониторингом безопасности терапии. Он включает в себя постмаркетинговый мониторинг и информирование о побочных действиях лекарственных средств.

Действующие лица:

- пользователь системы (например, медицинский сотрудник, участвующий в лечении);
- служба обезличивания/псевдонимизации;
- медицинская система;
- информационный ресурс сбора данных о побочных действиях.

Предусловия:

- пациент получает обычную медицинскую помощь;
- на пациента оказано воздействие (лекарственным средством, медицинским прибором, окружающей средой, например, укореняющимся сумахом);

- медицинский сотрудник, участвующий в лечении, имеет доступ к местной информационной системе, в которой предусмотрены функции по информированию о побочных действиях;
- эта информационная система находится в состоянии готовности;
- пациент предоставил информированное согласие;
- местная информационная система имеет доступ к службе псевдонимизации.

Постусловия:

- информационный ресурс сбора данных о побочных действиях имеет все необходимые сведения;
- при необходимости обеспечивается дополнительное обследование пациента.

Поток работ / события / действия

Участники мониторинга безопасности терапии показаны на рисунке А.3.

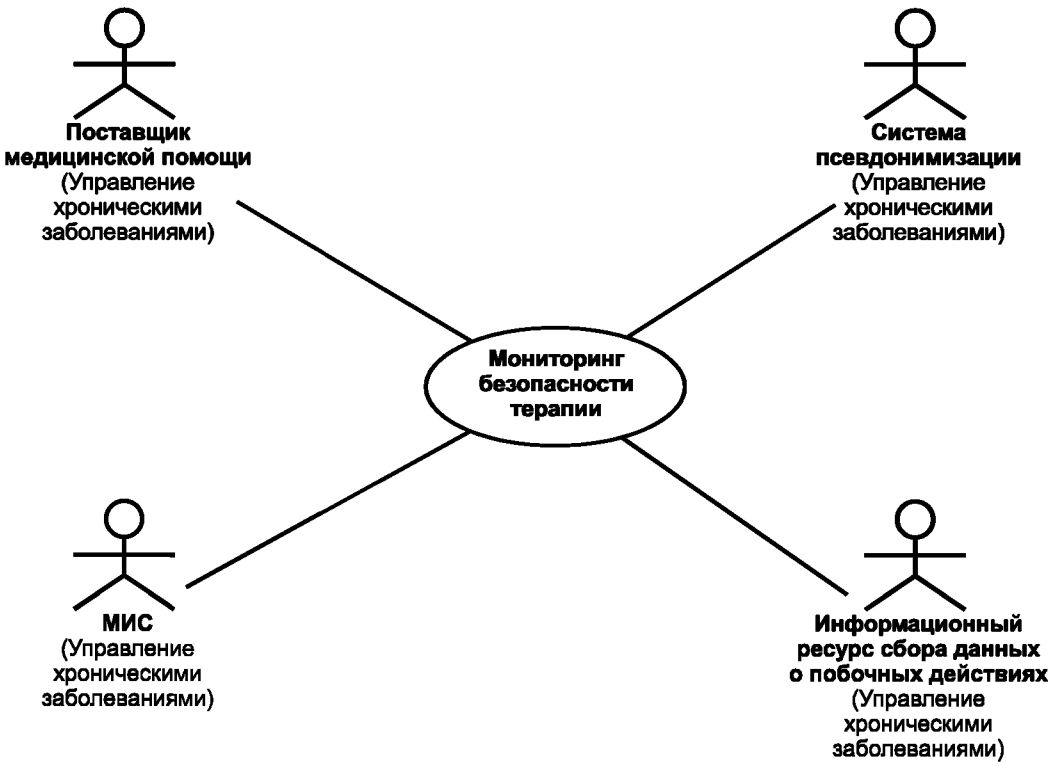


Рисунок А.3 — Участники мониторинга безопасности терапии

Течение процесса:

- пользователь из числа медицинских работников, обеспечивающих лечение, аутентифицируется в МИС;
- МИС инициирует регистрацию доступа в журнале, фиксируя время регистрации;
- медицинский работник присваивает пациенту уникальный идентификатор;
- медицинский работник документирует симптомы пациента, жизненно важные показатели, диагнозы, а также указывает, вызваны ли они воздействием или нет;
- медицинский работник принимает решение о составлении специального отчета (например, отчета о побочном действии, отчета о выявлении аллергии);
- медицинский работник отправляет эту информацию/отчет в соответствующую организацию (службу контроля побочных действий, профессиональную ассоциацию медицинских и фармацевтических работников, Управление по санитарному надзору за качеством пищевых продуктов и медикаментов, Центр контроля и профилактики заболеваний, санитарно-эпидемиологическую службу);
- медицинский работник отправляет эту информацию в электронную медицинскую карту пациента, после чего вызывается служба обезличивания или псевдонимизации;
- выявляется необходимость дополнительного лечения пациента;
- аутентифицированный поставщик медицинской помощи производит восстановление идентичности пациента.

Другие примеры/замечания

Добровольная система отчетов о побочных действиях используется для создания базы данных, предназначенной для обеспечения безопасности пациентов. Чтобы защитить идентификацию как пациента, так и поставщика медицинской помощи, используется псевдонимизация, осуществляемая с помощью соответствующей служ-

бы. При обменах данными о дальнейшем лечении пациентов и запросах дополнительных деталей об уже зарегистрированных случаях побочных действий используются псевдонимы, что устраняет угрозу идентификации пациента или поставщика медицинской помощи.

А.3.6 Исследование немедицинских аспектов с использованием медицинских данных

Описание сценария: нормы регулирования требуют проведения оценки долгосрочного финансового эффекта использования ремней безопасности. С этой целью инициируется исследование, в котором собираются данные из протоколов аварий, отчетов бригад скорой помощи, сведений о регистрации автотранспортных средств, медицинских карт стационарных больных, медицинских данных о восстановительном лечении и медицинских карт муниципальных медицинских учреждений. Идентификация субъектов и соответствующие сведения о распределении рисков, извлекаемые из этих данных, псевдонимизируются с помощью службы псевдонимизации и собираются в научной базе данных для последующей обработки.

А.3.7 Маркетинговое исследование

Описание сценария: группа поставщиков медицинской помощи договорилась об обмене данными о предоставлении услуг. Представление этих данных включает в себя и коммерческую информацию, и в целях ее защиты идентичность поставщиков защищается с помощью методов псевдонимизации.

А.3.8 Сбор данных в учебных целях

А.3.8.1 Общие сведения

Для преподавания используются данные об избранных случаях лечения реальных пациентов, из которых удаляется идентифицирующая и посторонняя информация. Это можно сделать с помощью обезличивания, но когда данные относятся к живым пациентам, то через какое-то время может потребоваться обновление этих данных. Чтобы сохранить связь между реальным пациентом и его данными, собираемыми в учебных целях, можно псевдонимизировать эти данные и накапливать их по мере поступления новой информации.

Учебные данные могут предоставляться только студентам по месту сбора данных или публиковаться, чтобы ими могли воспользоваться студенты всего мира. В первом случае правила псевдонимизации могут разрешать детализацию клинических случаев, во втором — медицинские данные должны быть сокращены до минимума, необходимого в дидактических целях.

Нередко студенты собирают персональные учебные данные о случаях медицинской помощи, которые им представляются интересными, что является предметом более строгих ограничений. Законодательство о персональных данных запрещает студентам копировать медицинские карты соответствующих пациентов. Если информация из этих медицинских карт надлежащим образом псевдонимизируется, то студенты могут получить разрешение на копирование псевдонимизированных данных в целях собственного обучения.

А.3.8.2 Использование псевдонимизации

Ключевым моментом псевдонимизации является задание имени и идентификатора пациента, когда они упоминаются в учебных данных. При разборе клинического случая обычно используются обороты наподобие «Мужчина 50 лет обратился...». При составлении описания клинического случая потребуется удалить из собранных источников фамилии, имена и отчества или заменить их на псевдонимы, указать возраст вместо даты рождения (или релевантный диапазон возраста), а также удалить всю иную идентифицирующую информацию, не имеющую отношения к цели разбора случая. Полученное таким путем описание клинического случая может быть опубликовано.

А.3.8.3 Требования к псевдонимизации

При формировании учебных данных часто требуется создать защищенную базу данных, в которой будет храниться связь между фактической идентификацией пациента и присвоенными ему псевдонимами. Нередко источниками учебных данных являются медицинские информационные системы нескольких учреждений, поэтому надо либо физически перемещать базу данных в каждое из этих учреждений, либо обеспечить к ней удаленный доступ из этих учреждений. Потребуется также преобразовывать данные в формат, который может восприниматься различными системами.

Генерация псевдонимов должна соответствовать как местным правилам, например, регламентирующим способ генерации, так и общим правилам, регламентирующим, например, генерацию слепых дат.

Преподаватель должен будет задать правила сохранения, псевдонимизации или удаления характеристик субъекта данных. Эти правила должны одинаково применяться разными системами в разное время при создании псевдонимизированных учебных данных.

А.3.9 Служба сопровождения

Чаще всего обезличивание медицинских карт производится тогда, когда доступ к этим картам должны получить сотрудники службы сопровождения медицинской информационной системы. Например, если сбой компьютера привел к ошибке в данных пациента, то службе сопровождения может понадобиться копия этих данных, чтобы установить причину ошибки. Обычно для целей анализа достаточно одной медицинской карты, которая обезличивается и сокращается. Служба сопровождения получает только те характеристики и аномальные данные, которые необходимы для этого анализа. Идентификация пациента и предыдущие данные могут быть удалены.

Иногда для анализа приходится копировать не одну медицинскую карту, а связанную группу карт, но и в этом случае копируемые данные могут быть существенно сокращены. В такой ситуации данные должны быть псевдонимизированы в целях сохранения связи между анализируемыми картами, но при этом можно использовать необратимую псевдонимизацию.

Приложение В
(справочное)

Построение модели угроз конфиденциальности

В.1 Введение

Описание метода оценки угроз конфиденциальности не входит в область применения настоящего стандарта. Однако ниже приводится ряд положений, которые могут помочь тем, кому приходится строить модель угроз.

В настоящем стандарте описана модель, рассматривающая три уровня обеспечения защиты конфиденциальности. Эти уровни выбраны в зависимости от степени сложности восстановления идентичности субъекта данных.

Настоящий стандарт может также помочь сформулировать ряд требований, которые должны быть приняты во внимание при разработке метода оценки угроз.

ИСО/МЭК 15408-2 содержит справочное приложение по конфиденциальности. Его можно использовать в качестве отправной точки, но оно в большей степени фокусируется на использовании ресурсов.

Функции безопасности объекта оценки (ОО) содержат спецификации, которые могут быть полезными, но они не описывают понятие уровня обезличивания.

В настоящем приложении приводится адаптированный обзор факторов оценки угроз, взятый из отчета D2.1 «Inventory Report on Privacy Enhancing Techniques» (Обзорный отчет по методам усиления конфиденциальности), датированного 27.02.2004 г. и представленного в рамках проекта PRIDEH-GEN (IST-2001-38719). Разрешение авторов отчета имеется.

Ключевым элементом модели угроз конфиденциальности является оценка влияния наблюдаемых данных, которые могут быть получены злоумышленником. Наблюдаемые данные могут включать в себя данные о событиях, которые злоумышленник может получить только незаконно, но к ним также относится информация, которую злоумышленник может получить законно. Может оказаться, что злоумышленником является обычный пользователь системы, который случайно или путем несанкционированных действий получил дополнительные данные, с которыми он не должен соприкасаться.

Важно иметь в виду, что эта информация обычно не входит в рамки модели информационных объектов прикладной системы. Тем не менее при оценке конфиденциальности данных, обрабатываемых системой, необходимо сделать определенные допущения о наблюдаемых данных.

Для создания методологии оценки угроз конфиденциальности необходим формализованный способ описания этих угроз и вероятности восстановления идентичности субъекта данных.

Базовая модель атак восстановления идентичности, показанная на самом высоком уровне абстракции на рисунке В.1, включает в себя три основные сущности. Хотя в таком виде она может показаться простой и прямолинейной, ее можно детализировать до такой степени сложности, которая будет охватывать все реальные аспекты восстановления идентичности и обеспечения конфиденциальности.

Рисунок В.1 — Атаки восстановления идентичности

В данной модели три основные сущности:

a) обезличенная база данных[2]

Эта база данных содержит обезличенные записи. В ней хранятся данные о неизвестных субъектах. Она является источником, содержащим потенциально чувствительную информацию, которая не должна быть раскрыта;

b) злоумышленник

Злоумышленником считается лицо, которое намеревается злоупотребить информацией, содержащейся в обезличенной базе данных. Для этого ему нужно связать обезличенную информацию с реальными лицами, то есть восстановить идентичность субъектов данных, хранящихся в этой базе;

c) база данных наблюдений

База данных, собираемых злоумышленником, содержащая идентифицирующую информацию.

Предположения о характере умысла очень важны для анализа угроз. Злоумышленниками могут быть люди, пользующиеся случаем ради забавы и любопытства, или тренированные профессионалы, принадлежащие к группе, поставившей себе определенную цель в получении конфиденциальной информации и располагающей значительными финансовыми и техническими ресурсами.

Злоумышленник наполняет свою базу данных «наблюдениями», релевантными для атаки. Эти наблюдения могут иметь различные источники, например:

- информация, хранящаяся в существующих базах данных, содержащих идентифицирующую информацию;
- социальные контакты: сбор информации, на которую в обычных условиях злоумышленник не имеет права, путем общения с другими людьми;
- фактические данные, то есть данные, собранные с помощью наблюдений в прямом смысле этого слова.

В настоящем контексте «релевантные» (для атаки) наблюдения означает, что собираемая информация должна иметь отношение к содержанию обезличенной базы данных, то есть эта информация или имеет прямое отношение к этому содержанию, или тесно связана с ним.

B.2 Модель угроз, цели и средства злоумышленника

Одно из достоинств понятия «база данных наблюдений» состоит в том, что оно позволяет определить категории методов, которые мог бы использовать реальный злоумышленник. Действительно, анализ безопасности и защиты обычно (неявно) основан на оценке угрозы воздействия на систему. Например, модель безопасности современной криптографии открытых ключей основана на том факте, что злоумышленник может располагать только ограниченными финансами (вычислительными ресурсами).

В основном уровень угрозы злоумышленника определяется двумя характеристиками: 1) цель атаки (что будет делать злоумышление после атаки?) и 2) средства воздействия, которыми может располагать злоумышленник (рисунок B.2). Последние связаны с «ценностью», которую информация, восстановленная из обезличенной базы данных, имеет для злоумышленника. Если чувствительная информация, скрытая в обезличенных данных, может обеспечить злоумышленнику большой выигрыш (например, медицинские карты для страховой компании), он будет готов вложить в процесс восстановления идентичности большие средства.

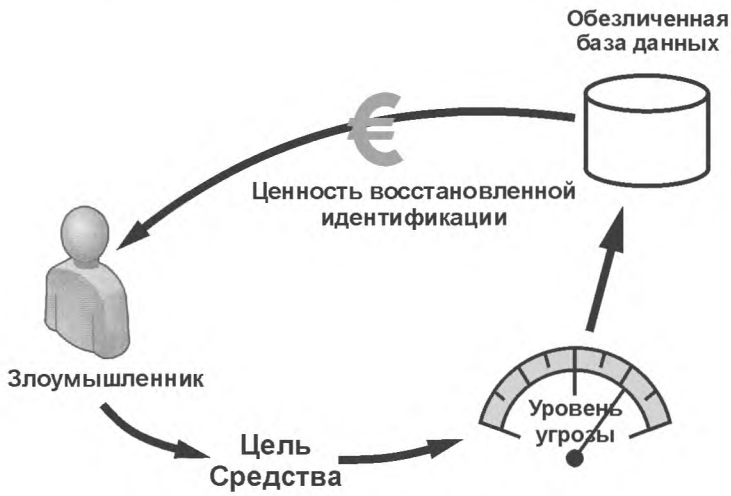


Рисунок B.2 — Цель и средства злоумышленника

[2] Под обезличенностью здесь имеется в виду, что база данных не содержит непосредственно идентифицирующую информацию (например, фамилию субъекта данных), то есть обезличенная база данных упоминается в широком смысле. Достаточно очевидно, что модели угроз строятся только для таких баз данных.

После определения уровня злоумышленника важно включить в модель угроз его «цели». Обеспечение конфиденциальности касается защиты персональных данных, а не только идентифицирующих данных, связанных с конкретной записью в базе данных. Это тонкое различие отражается в трех различных целях злоумышленника, описанных в модели, а именно:

a) восстановление идентичности (полное):
   определение, к кому именно относится конкретная обезличенная запись;
   определение, какие обезличенные записи относятся к определенному лицу;
b) восстановление информации (частичное восстановление идентичности);
c) определение принадлежности к базе данных:
   содержит ли база данных записи, относящиеся к некоторому лицу;
   отсутствуют ли в базе данных записи, относящиеся к некоторому лицу.

Понятие полного восстановления идентичности хорошо знакомо. Оно состоит в попытке (частичного) преобразования обезличенной базы данных к своему идентифицирующему эквиваленту. В наиболее общем случае злоумышленник пытается восстановить идентичность субъектов данных по всей обезличенной базе данных. Однако на практике это редко имеет место, и злоумышленник пытается или определить, к кому относится обезличенная запись, представляющая особый интерес (например, к кому относится запись о высоком доходе), или извлечь всю информацию о конкретном лице (например, страховой агент пытается установить, не страдает ли определенное лицо заболеванием сердца).

Две другие цели (частичное восстановление идентичности и определение принадлежности к базе данных) обсуждаются не часто, поскольку теория соответствующего анализа достаточно сложна. В этом случае для получения необходимой информации злоумышленнику не обязательно восстанавливать идентичность всей информации о лице. Иногда достаточно извлечь из обезличенной базы данных только отдельную характеристику лица, даже не зная, какие записи связаны с этим лицом.

Наконец, в некоторых ситуациях интерес может состоять лишь в определении того, включена ли информация о лице в базу данных. Такое включение само по себе может быть конфиденциальной (чувствительной) информацией. Примером могут служить базы данных, содержащие информацию о пациентах, больных ВИЧ. В этом случае цель злоумышленника может состоять лишь в том, чтобы определить, включена ли информация о людях из интересующего его списка в обезличенную базу данных, и ему нет необходимости восстанавливать идентификацию каждой записи.

Понятно, что методы, применяемые злоумышленником, зависят от целей, которые он пытается достичь. Стратегии атаки тесно связаны с целями. Хотя эти стратегии и укладываются в общую модель, они достаточно сильно отличаются и некоторые их аспекты должны обсуждаться отдельно.

Очевидно, что полное или частичное восстановление идентичности в том смысле, как они определены в настоящем стандарте, тесно связаны. Частичное восстановление идентичности представляет собой промежуточное состояние между отсутствием восстановления и полным восстановлением идентичности (конкретного субъекта данных) из обезличенной базы данных (рисунок В.3). Другими словами, оно соответствует ситуации, когда полное восстановление идентичности отсутствует, но использованные процессы (алгоритмы) восстановления привели к получению некоторой информации из обезличенной базы данных.

Рисунок В.3 — Полное и частичное восстановление идентичности

При всех формах восстановления идентичности злоумышленник следует в основном одной и той же процедуре. Используя свои наблюдения и содержание обезличенной базы данных, он составляет для каждого идентификатора лица в базе данных наблюдений (Ном-ИД) список обезличенных идентификаторов (Обезл-ИД), которые могут ему соответствовать.

Связь между данными наблюдений и обезличенными данными может быть установлена большим числом способов, выбор которых зависит от конкретной ситуации. Однако важно добавить к общей модели некоторые подходы классификации, чтобы лучше понять механизмы связывания.

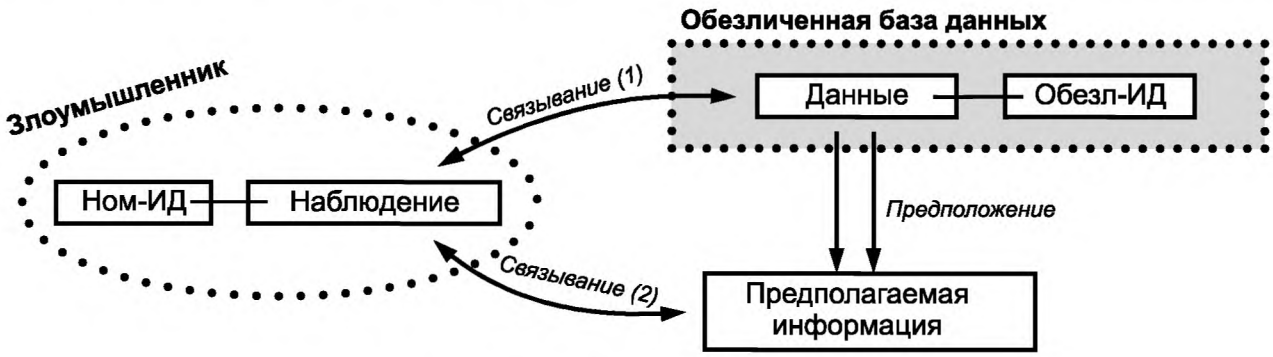


Рисунок В.4 — Механизмы связывания

На рисунке В.4 показано, что связь между Ном-ИД и Обезл-ИД может быть установлена непосредственно, используя характеристики, хранящиеся в соответствующих базах данных (связывание (1)), или с помощью промежуточного шага (связывание (2)). В первом случае данные, хранящиеся в обезличенной базе данных, непосредственно отображаются на данные наблюдений. Это означает, что злоумышленник с помощью связывания баз данных может прочитать некоторые характеристики записей обезличенной базы данных. Используя эти характеристики, злоумышленник может определить, соответствует ли обезличенная запись идентифицирующей записи базы данных наблюдений.

Во втором случае для получения возможности связывания двух источников информации необходимо сделать промежуточный шаг. Записи наблюдений не имеют прямой связи с записями обезличенной базы данных, но можно предположить наличие такой связи, используя характеристики, хранящиеся в обезличенных записях. Согласно настоящему приложению эта ситуация эквивалентна предположению связи обезличенных записей с записями наблюдений, используя характеристики, хранящиеся в записях наблюдений.

Применяемые алгоритмы связывания и предположений о связях обычно специфичны для конкретных данных и прикладных программ. Однако некоторые алгоритмы рассчитаны на общие типы данных. На высшем уровне абстракции можно оперировать важным понятием «уверенности» в сконструированной связи.

Как алгоритмы связывания, так и алгоритмы предположений о связях не обязательно опираются на достоверные факты. Связь, сконструированная злоумышленником, не обязательно правильная. В зависимости от предположений, сделанных злоумышленником, а также полноты и достоверности его наблюдений, сложности и неопределенности обезличенных данных, некоторые связи могут быть правдоподобнее других. Поэтому злоумышленник должен присвоить связи между идентификаторами некоторую вероятность (конечно, точная связь получит значение вероятности, равное 1). Например, в обезличенной базе данных содержатся суммы зарплат, и злоумышленник не может узнать из нее, чьи это зарплаты. Однако он может сделать предположения на основе таких характеристик, как выполняемая работа, площадь дома, модель автомобиля. Злоумышленник никогда не может быть уверен в правильности своей догадки, его предположения верны только с определенной вероятностью.

В.3 Полное или частичное восстановление идентичности

Если в конце процедуры связывания злоумышленник может сопоставить единственный идентификатор записи наблюдений с идентификатором обезличенной записи, то это означает вероятное восстановление идентичности соответствующей обезличенной записи. Правдоподобность этого восстановления зависит от вероятностных характеристик используемых правил связывания и предположений, а также от третьего фактора — отношений между субъектами, чьи данные хранятся в базе данных наблюдений и в обезличенной базе данных. На рисунке В.5 показаны различные возможные зависимости между базой данных наблюдений (обозначенной буквой «Н») и обезличенной базой данных (обозначенной буквой «О»). Крестики обозначают субъекты данных, а не записи базы данных (элемент множества, обозначенный крестиком, означает, что данные этого субъекта хранятся в базе данных).

Пока Н является подмножеством О или наоборот, восстановление идентичности является истинным, если выявлена уникальная связь между Ном-ИД и Обезл-ИД. Если такого включения нет, то обнаруженная уникальная связь не гарантирует, что существует истинное соответствие между наблюдаемым и обезличенным идентификатором.

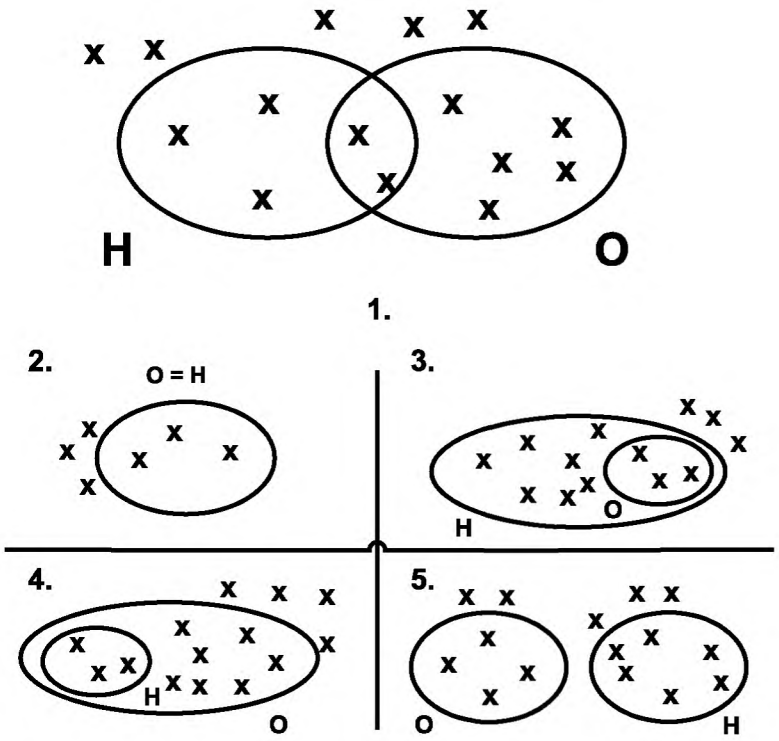


Рисунок В.5 — Зависимости между базой данных наблюдений и обезличенной базой данных

Если уникальная связь между идентификатором записи базы данных наблюдений и идентификатором обезличенной записи не установлена, то это не означает, что никакая информация не раскрыта. Если совокупность идентификаторов записей базы данных наблюдений может быть связана с совокупностью идентификаторов обезличенных записей, то информация, общая для всех этих обезличенных записей, может быть связана с совокупностью идентификаторов записей базы данных наблюдений (например, можно удостовериться, что обо всех наблюдаемых субъектах есть записи в обезличенной базе данных, см. рисунок В.5). Важно отдавать себе отчет в полном объеме утечки информации. Если злоумышленника только такой факт и интересует, то он достиг своей цели и атака оказалась успешной.

При применении описанного метода восстановления идентичности или выявления информации таблица возможных связей между идентификаторами записей базы данных наблюдений и идентификаторами обезличенных записей должна обновляться и вычисляться заново при каждом обнаружении новой частичной информации или при полном восстановлении идентификации субъекта.

В.4 Пример восстановления идентичности

Описанные выше понятия можно пояснить на простом примере. На рисунке В.6 показано содержание базы данных наблюдений и соответствующей обезличенной базы данных. В этом примере обезличенная база данных содержит три записи, каждая из которых содержит четыре статические характеристики, которые могут принимать значения А или Б (вопросительный знак означает, что точное значение характеристики неизвестно). Злоумышленник может непосредственно и правильно наблюдать только две из этих характеристик, и знать идентификацию всех людей, чьи записи хранятся в обезличенной базе данных. В этом случае правила связывания чрезвычайно просты — значение или то же самое, или нет.

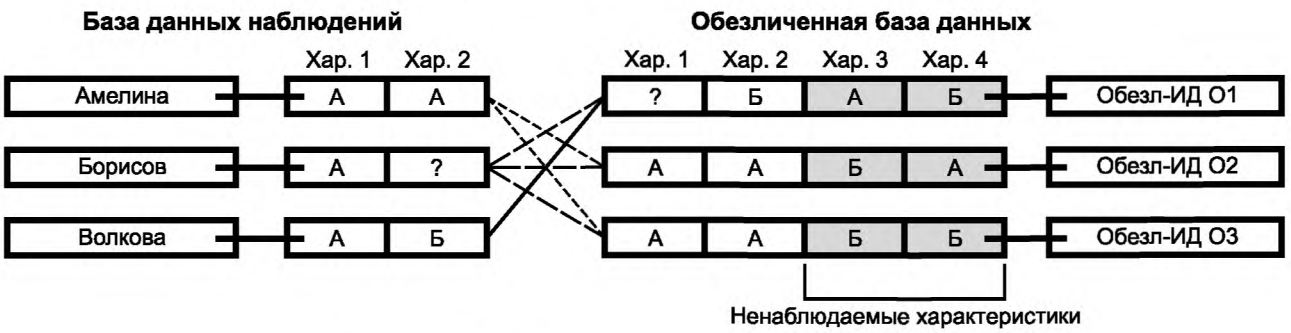


Рисунок В.6 — Пример восстановления идентичности

Применяя простой алгоритм связывания записей, который непосредственно сравнивает соответствующие характеристики записей обезличенной базы данных и базы данных наблюдений (рисунок В.4, *связывание (1)*), можно составить следующие таблицы соответствия (проиллюстрированные на рисунке В.6 различными линиями между двумя наборами записей).

Идентификатор лица	Может соответствовать
Амелина	О2, О3
Борисов	О1, О2, О3
Волкова	О1

Идентификатор обезличенной записи	Может соответствовать
О1	Борисов, Волкова
О2	Амелина, Борисов
О3	Амелина, Борисов

В этих таблицах показано соответствие между известной фамилией и идентификатором обезличенной записи и наоборот, построенное с помощью данного правила связывания. Если используются несколько алгоритмов связывания, то они должны выполняться совместно, и построение соответствующих таблиц может стать достаточно сложным.

Из таблиц видно, что обезличенная запись с идентификатором О1 может соответствовать только Волковой, то есть субъект данных этой записи полностью установлен и злоумышленник теперь знает, что у Волковой две не наблюдаемые характеристики имеют значения А, Б. Используя это знание, злоумышленник может обновить ранее построенные таблицы и получить следующий результат.

Идентификатор лица	Может соответствовать
Амелина	О2, О3
Борисов	О2, О3

Идентификатор обезличенной записи	Может соответствовать
О2	Амелина, Борисов
О3	Амелина, Борисов

По оставшимся записям обезличенной базы данных злоумышленник не может однозначно определить, к кому именно они относятся. Однако он может предположить, что с вероятностью 50 % запись с идентификатором О2 относится либо к Амелиной, либо к Борисову. В реальной (большой) базе данных подобная информация мало что дает для восстановления идентичности субъекта данных, однако даже и в этом случае злоумышленник может получить полезные ему сведения.

Хотя в описанном выше примере полное восстановление идентичности не имеет места, тем не менее определенная утечка информации существует, поскольку в записях обезличенной базы данных с идентификаторами О2 и О3 характеристика 3 имеет одно и то же значение. Отсюда злоумышленник может заключить, что эта характеристика и у Амелиной, и у Борисова имеет значение Б. Но какие именно значения у них имеет оставшаяся характеристика 4, у него нет точной информации.

Если все, что злоумышленник хотел знать об Амелиной и Борисове, состоит в значении характеристики 3, то его попытка раскрыть информацию оказалась полностью успешной. Если характеристика 4 содержит нужную ему информацию, то его успех оказался частичным. Однако учитывая тот факт, что теперь о Борисове известно значение характеристики 2, хотя она и не была наблюдаемой, раскрытие информации оказалось еще большим.

Показанная модель и процедуры применимы не только к таким простым структурам данных, которые были использованы в этом примере. Записи баз данных содержат многочисленные характеристики, зависящую от времени информацию или сочетания различных типов данных, которые также укладываются в представленную модель. Однако реализация соответствующих правил связывания может оказаться более сложной.

В.5 Получение новой информации

Целью восстановления идентичности является получение конфиденциальной информации о каком-либо лице (рисунок В.7). Хотя это достаточно очевидно, важно не забывать этот факт. Если злоумышленник может наблюдать все характеристики записей обезличенной или псевдонимизированной базы данных, то его база данных наблюдений есть не что иное, как полностью идентифицируемая версия защищенной базы данных. Из обезличенной базы данных нельзя извлечь никакой дополнительной информации, которая еще не доступна злоумышленнику, последний не может извлечь из обезличенной базы данных никаких новых знаний, а, значит, с обезличенной базой данных не связаны угрозы обеспечению конфиденциальности. Вся информация и так доступна.

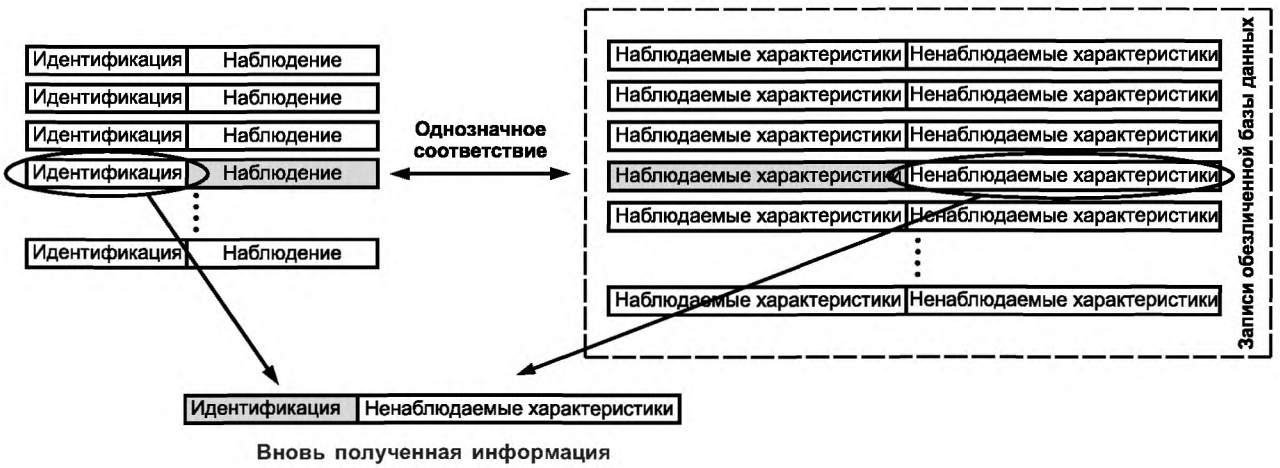


Рисунок В.7 — Извлечение новой информации из обезличенной или псевдонимизированной базы данных с помощью восстановления идентичности

Существуют пограничные случаи, например, наблюдения могут быть существенно неопределенными. В этом случае обезличенная база данных используется как средство верификации наблюдений.

В.6 Принадлежность к базе данных

Помимо восстановления идентичности субъекта данных, цель злоумышленника может состоять исключительно в определении принадлежности субъекта к базе данных. «Принадлежность» к базе данных сама по себе может быть той информацией, что нужна злоумышленнику, примером может служить база данных ВИЧ-инфицированных пациентов.

Отсутствие принадлежности определяется относительно легко. Если вероятность связи между искомым субъектом и каждой записью обезличенной базы данных равна нулю, то, несомненно, этот субъект не принадлежит этой базе данных (конечно, при условии, что база данных наблюдений и обезличенная база данных не содержат ошибок).

К сожалению, нельзя воспользоваться рассуждением от обратного. Если злоумышленник не может доказать, что субъект не принадлежит обезличенной базе данных, из этого не обязательно следует, что об этом субъекте нет ни одной записи в этой базе. Как показано на рисунке В.8, если данные записей базы данных наблюдений и обезличенной базы данных совпадают, то эти записи могут принадлежать единственному субъекту, входящему в пересечение множеств $O \cap H$, или двум разным субъектам (один из разности множеств $O \setminus H$, другой из разности множеств $H \setminus O$). Этой связи должна быть присвоена некоторая вероятность, по которой потом будет вычислена вероятность принадлежности к базе данных.

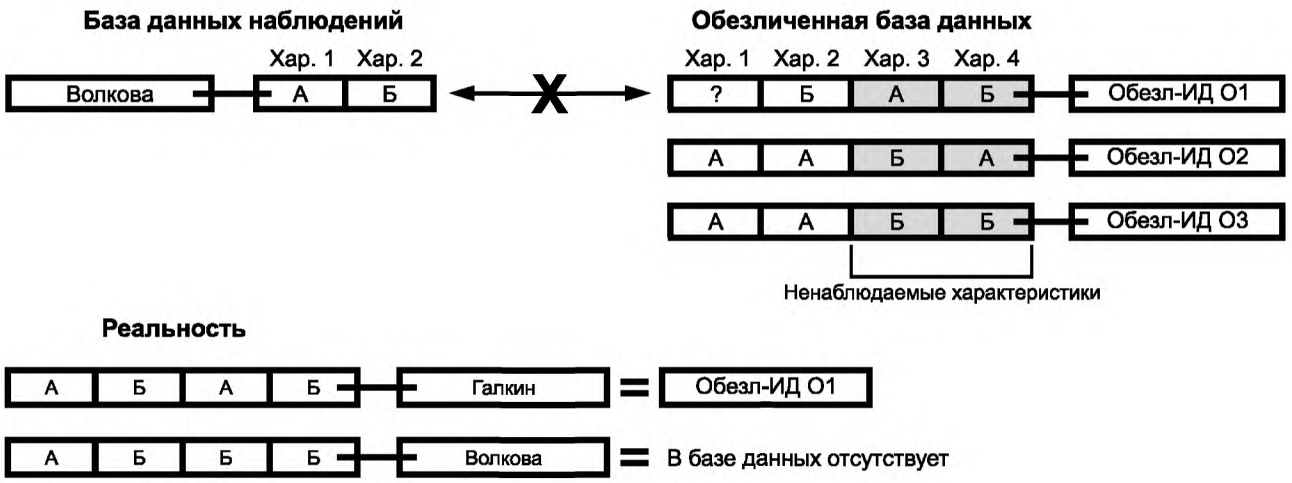


Рисунок В.8 — Пример определения принадлежности субъекта к базе данных

Это утверждение иллюстрируется примером, показанным на рисунке В.8. Рассмотрим обезличенную базу данных с тремя записями. Злоумышленнику требуется узнать, принадлежит ли Волкова к этой базе данных. Используя наблюдения и правило простого совпадения, он обнаруживает, что единственная запись обезличенной базы данных, которая может соответствовать Волковой, имеет идентификатор «Обезл-ИД О1». Отсюда, однако, не следует, что эта запись действительно соответствует Волковой. Как показано на рисунке В.8, она принадлежит Галкину. Если злоумышленник присвоит высокую вероятность единственной связи, установленной с помощью его правил связывания (в данном случае сравнение пар двух характеристик), то он может прийти к ошибочному заключению, что запись с идентификатором «Обезл-ИД О1» содержит сведения о Волковой.

На рисунке В.9 показан пример применения псевдонимизации в клинических испытаниях.

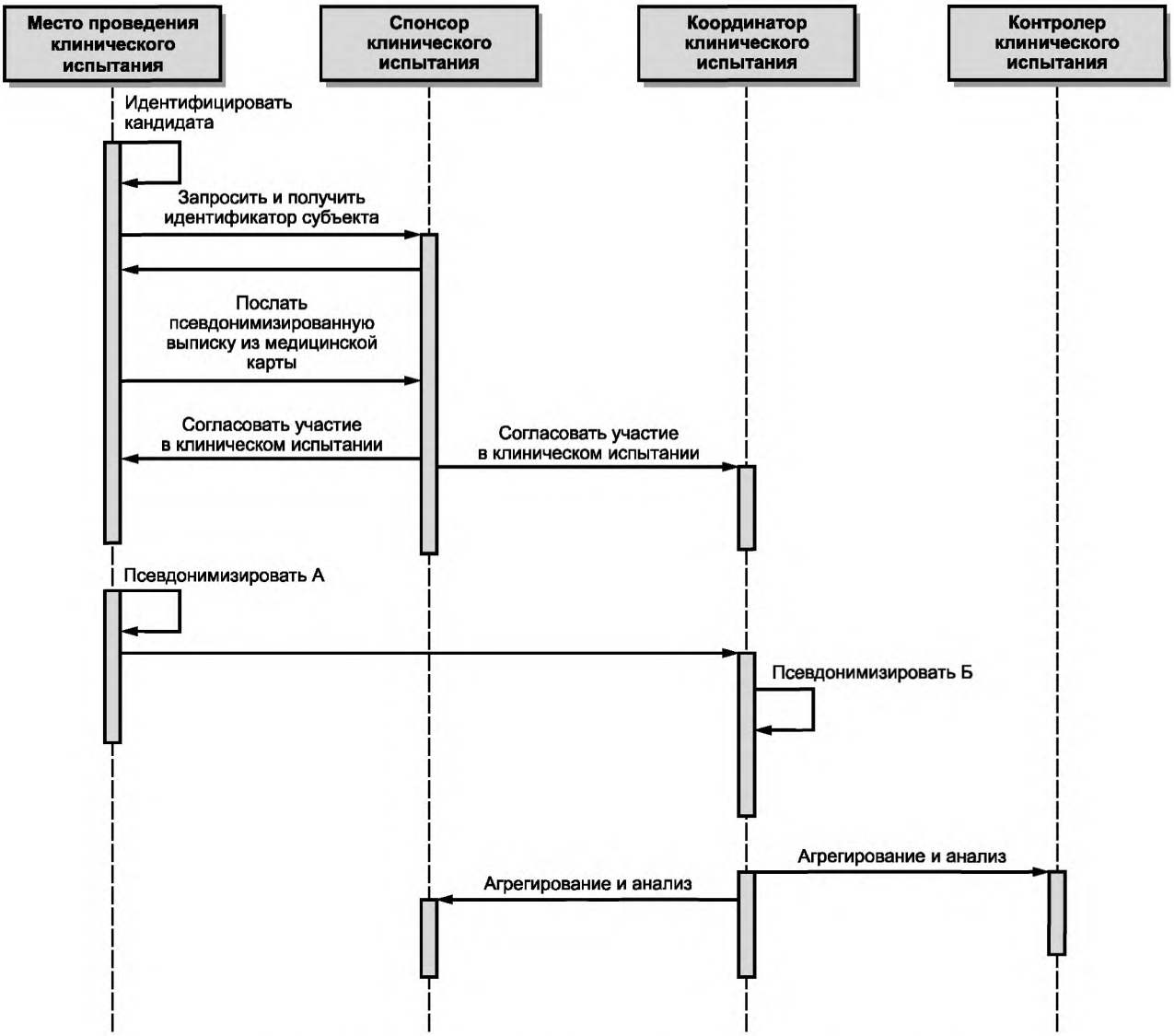


Рисунок В.9 — Псевдонимизация субъектов клинических испытаний

Псевдонимизация субъектов клинического испытания осуществляется следующим образом:

a) спонсор клинического испытания инициировал испытание с идентификатором ИД. Детали испытания обычно не сообщаются кандидатам на участие в испытании или медицинской организации. Это противоречит принципу двойного слепого исследования. Раскрываются только информация, необходимая для оценки кандидата и назначения испытания, а также риски участия;

b) организация — место проведения клинического испытания использует эту информацию для подбора кандидата. Спонсор или координатор клинического испытания присваивают этому кандидату некоторый номер. Нередко это порядковый номер запроса на согласование кандидата. В этот момент месту проведения клинического испытания не сообщается ничего, кроме интереса к этому кандидату;

c) место проведения клинического испытания готовит краткую выписку из медицинской карты кандидата, указывая в ней идентификатор клинического испытания и номер кандидата вместо его истинной идентификации. Выписка содержит только ту информацию, что нужна для оценки пригодности кандидата;

d) спонсор или координатор клинического испытания определяет пригодность кандидата и согласует его участие в клиническом испытании;

e) время от времени место проведения клинического испытания посылает данные координатору испытания. Эти данные псевдонимизируются с помощью удаления идентификаторов, не являющихся необходимыми, некоторой другой информации. В них указаны идентификатор клинического испытания и ранее присвоенный номер субъекта. Все изменения, подстановки и удаления данных, а также передача данных отражаются в детальном журнале регистрации транзакций;

f) эти данные защищенным способом передаются координатору клинического испытания. Получение данных также регистрируется в журнале;

g) эти данные еще раз псевдонимизируются и сокращаются. (Координатор клинического испытания может беспокоиться о наличие излишней информации и исключить некоторые данные из тех, что были получены.) Кроме того, удаляются данные, по которым можно определить место проведения испытания. Эти действия записываются в журнале таким образом, чтобы необходимые данные можно было восстановить;

h) для оценки результатов испытания повторно псевдонимизированные данные анализируются и агрегируются;

i) агрегированные данные и сопутствующие псевдонимизированные данные защищенным способом передаются спонсору клинического испытания, контролерам и т. д. Эта передача полностью регистрируется в журнале.

Примечания

1 Следует обратить внимание, что в этом примере не используется служба псевдонимизации. Системы, участвующие в информационном взаимодействии, выполняют псевдонимизацию собственными средствами.

2 Следует обратить внимание, что в этом примере для генерации псевдонимов не используются криптографические средства. Использование случайной последовательности чисел, назначаемой спонсорами клинических испытаний, более надежно. (Псевдонимы, полученные с помощью не очень корректно примененных криптографических средств, чрезвычайно подвержены атаке по словарю, поскольку список фамилий достаточно ограничен.)

3 Следует обратить внимание, что восстановление данных требует несколько шагов. При восстановлении используются журналы регистрации изменения данных и различные записи в базе данных места проведения испытания, а не вложенные зашифрованные исходные данные. Чтобы проследить связь с изначальным лицом, необходимо сначала получить данные от координатора клинического испытания, затем извлечь данные из журналов регистрации транзакций, а потом вернуться к месту проведения клинического испытания для окончательной идентификации лица.

Приложение ДА
(справочное)

Сведения о соответствии ссылочных международных стандартов ссылочным национальным стандартам Российской Федерации

Таблица ДА.1

Обозначение ссылочного международного стандарта	Степень соответствия	Обозначение и наименование соответствующего национального стандарта
ИСО 27799	—	*

* Соответствующий национальный стандарт отсутствует. До его утверждения рекомендуется использовать перевод на русский язык данного международного стандарта. Перевод данного международного стандарта находится в Федеральном информационном фонде технических регламентов и стандартов.

Библиография

[1] A function to hide person-identifiable information. Computing for Health Intelligence, Bulletin 1 (version 3) 15 July 2002
[2] ANSI X9.52-1998, Triple Data Encryption Algorithm Modes of Operation
[3] Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER). RFC-2313 PKCS #1: RSA Encryption, Version 1.5, March 1998
[4] BERMAN, J.J., Confidentiality for Medical Data Miners, Artificial Intelligence in Medicine, November 2002
[5] DE MOOR, G. and CLAERHOUT, B., PPeP Privacy Protection in e-Pharma: Leading the Way for Privacy Protection in e-Pharma, White Paper, 2003
[6] DE MOOR, G., CLAERHOUT, B. and DE MEYER, F., Privacy Enhancing Techniques: the Key to Secure Communication and Management of Clinical and Genomic Data, Methods of Information in Medicine, 42, pp79-88, 2003
[7] 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
[8] EL KALAM, A.A., DESWARTE, Y., TROUESSIN, G. and CORDONNIE, E., A new method to generate and manage anonymized healthcare information
[9] HARA, K., OHE, K., KADOWAKI, T., KATO, N., IMAI, Y., TOKUNAGA, K., NAGAI, R. and OMATA, M., Establishment of a method of anonymisation of DNA samples in genetic research, J. Hum. Genet., 48(6), pp 327-30, 2003
[10] HES, R. and BORKING, J., Privacy-Enhancing Technologies: The path to anonymity, Revised Edition, Registratiekamer, The Hague, August 2000
[11] ROBERTS, I., Pseudonymization in CLEF, PT001 (draft)
[12] IHLE, P., KRAPPWEIS, J. and SCHUBERT, I. Confidentiality within the scope of secondary data research — approaches to a solution of the problem of data concentration, Gesundheitswesen, 63 Suppl. 1: pp S6-12, 2001
[13] INFOSEC/TTP Project, Code of Practice and Management Guidelines for Trusted Third Party Services, CASTELL, S., (Ed.), Ver. 1.0, Castell, Spain, October 1993
[14] INFOSEC/TTP Project, Trusted Third Party Services: Functional Model, MULLER, P. (Ed.), Ver. 1.1, Bull. Ingenierie, France, December 1993
[15] INFOSEC/TTP Project, Trusted Third Party Services: Requirements for TTP Services
[16] ISO 7498-2 Information processing systems — Open Systems Interconnection — Basic Reference Model — Part 2: Security Architecture
[17] ISO/IEC 8825-1 Information technology — ASN.1 encoding rules: Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER) — Part 1
[18] LANGHEINRICH, M., Privacy by Design — Principles of Privacy-Aware
[19] Latanya Sweeney, Phd, http://privacy.cs.cmu.edu/people/sweeney/
[20] LOWRANCE, W., Learning from experience: privacy and the secondary use of data in health research, J. Health Serv. Res. Policy, Suppl 1: pp S1:2-7, July 8, 2003
[21] PRIDEH Privacy Enhancement in Data Managemtn in E-Health. Final Report, Deliverable D4.4 Report Version 2.0. Technical Recommendations, Guidelines and Business Scenarios, July 2003
[22] NIH Publication Number 003-5388. Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule, http://privacyruleandresearch.nih.gov
[23] RFC-2630 Cryptographic Message Syntax, June 1999
[24] SCHNEIER, B., Applied Cryptography: Protocols, Algorithms, and Source Code in C, 2nd edition, John Wiley, 1996
[25] Ubiquitous Systems, http://www.vs.inf.ethz.ch/publ/papers/privacy-principles.pdf
[26] WESTIN, A., Privacy and freedom, 1967
[27] ISO/IEC 2382-8 Information technology — Vocabulary — Part 8: Security
[28] ISO/IEC 15408-2 Information technology — Security techniques — Evaluation criteria for IT security — Part 2: Security functional components
[29] ISO/TS 22220 Health Informatics — Identification of subjects of health care
[30] ENV 13608-1 Health informatics — Security for healthcare communication — Part 1: Concepts and terminology
[31] RFC 3881, Security Audit and Access Accountability Message XML Data Definitions for Healthcare Applications

УДК 004:61:006.354 ОКС 35.240.80 П85 ОКСТУ 4002

Ключевые слова: здравоохранение, информатизация здоровья, псевдоним, обезличивание, персональная информация

Редактор *Н. Н. Кузьмина*
Технический редактор *В. Н. Прусакова*
Корректор *С. И. Фирсова*
Компьютерная верстка *В. Н. Романовой*

Сдано в набор 24.06.2013. Подписано в печать 18.07.2013. Формат $60\times84^{1}/_{8}$. Бумага офсетная. Гарнитура Ариал.
Печать офсетная. Усл. печ. л. 6,05. Уч.-изд. л. 5,40. Тираж 60 экз. Зак. 892.

ФГУП «СТАНДАРТИНФОРМ», 123995 Москва, Гранатный пер., 4.
www.gostinfo.ru info@gostinfo.ru
Набрано и отпечатано в Калужской типографии стандартов, 248021 Калуга, ул. Московская, 256.