УДК 61:681.518(075.8)

ОБЕЗЛИЧИВАНИЕ ПЕРСОНАЛЬНЫХ ДАННЫХ В ЗДРАВООХРАНЕНИИ

А.П. Столбов
ФГАОУ ВО Первый московский государственный медицинский университет им. И.М. Сеченова (Сеченовский университет), Москва

Рассмотрены методы обезличивания персональных данных – анонимизации и псевдонимизации в информационных системах в здравоохранении. Перечислены основные нормативные документы и стандарты, определяющие требования к защите персональных данных и их обезличиванию. Описаны варианты и примеры использования технологий псевдонимизации в здравоохранении. Сформулированы предложения по внедрению методов псевдонимизации в здравоохранение.

Ключевые слова: защита персональных данных, обезличивание персональных данных, информационные системы в здравоохранении

De-identification of Personal data in Health care

Andrey P. Stolbov
The First Sechenov Moscow State Medical University, Moscow

The methods of de-identification: anonymization and pseudonymization of personal data in information systems in health care. Lists the key normative documents and standards defining the requirements for the protection of personal data and their depersonalization. Described embodiments and examples of the use of technology pseudonymization in health care. Formulated proposals for the introduction of methods of pseudonymization in health care.

Keywords: protection of personal data, depersonalization of personal data, information systems in health care

Сегодня уже невозможно представить себе развитие здравоохранения без современных информационных технологий (ИТ) и медицинских информационных систем (МИС). Концепцией создания Единой государственной информационной системы в сфере здравоохранения (ЕГИСЗ) [1] и приоритетным проектом "Электронное здравоохранение" [2] предусмотрены ведение в лечебно-профилактических учреждениях электронных медицинских карт (ЭМК) и создание федеральной системы ведения интегрированной ЭМК гражданина (ИЭМК). В системе обязательного медицинского страхования (ОМС) обмен данными о застрахованных лицах и оказанной им медицинской помощи между лечебными учреждениями, страховыми медицинскими организациями (СМО) и фондами ОМС осуществляется по каналам связи [3]. Уже сейчас в медицинских организациях (МО), СМО, фондах ОМС и региональных МИС накоплены и обрабатываются огромные по объемам персонифицированные базы данных (БД), содержащие сведения, составляющие врачебную тайну, которые по закону должны быть надежно защищены от несанкционированного доступа, изменения и удаления.

Основные требования к организации обработки и защиты информации о физических лицах определены федеральным законом "О персональных данных" № 152-

ФЗ от 27.07.2006, постановлениями Правительства РФ № 687 от 15.09.2008 [4], № 1119 от 01.11.2012 [5], № 211 от 21.03.2012 [6], приказами ФСТЭК России № 17 от 11.02.2013 (в ред. от 15.02.2017) и № 21 от 18.02.2013, приказом ФСБ России № 378 от 10.07.2014. Реализация мероприятий по защите информации в ИС подразумевает применение целого комплекса организационных и технических мер и дорогостоящих специальных технических, программных, в том числе криптографических средств защиты информации.

Одним из способов обеспечения конфиденциальности персональных данных и снижения затрат на их защиту является обезличивание информации. В новой редакции федерального закона "Об основах охраны здоровья граждан в Российской Федерации" № 323-ФЗ от 21.11.2011 (ред. от 29.07.2017) сказано, что на федеральном уровне в ЕГИСЗ сбор и обработка данных персонифицированного учета лиц, которым оказывается медицинская помощь, а также лиц, в отношении которых проводятся медицинские экспертизы, медицинские осмотры и медицинские освидетельствования, осуществляется в обезличенном виде, в порядке, установленном Минздравом России по согласованию с Роскомнадзором[1].

Целью настоящей работы является анализ и обсуждение возможностей и особенностей применения методов обезличивания персональных данных в здравоохранении.

Основные понятия. Начнем с определения основных понятий и терминов.

Персональные данные – любая информация, относящаяся к прямо или косвенно определенному или определяемому физическому лицу – субъекту персональных данных (ст. 3 закона № 152-ФЗ). Состав персональных данных медицинских работников и пациентов, обработка которых допускается в системе здравоохранения и медицинского страхования, определен в ст.ст. 43, 44, 93, 94 федерального закона № 323-ФЗ от 21.11.2011 и ст. 44 закона "Об обязательном медицинском страховании в Российской Федерации" № 326-ФЗ от 29.11.2010. Далее, если это не указано особо, в качестве субъекта персональных данных будем рассматривать пациента.

Оператор персональных данных – юридическое или физическое лицо, которое: а) самостоятельно или совместно с другими лицами организует и/или осуществляет обработку персональных данных, и б) определяет цели их обработки, состав данных и выполняемые над ними действия (операции).

---

[1] Из этого, в частности, следует, что сведения в федеральном сегменте базы данных ИЭМК должны быть представлены в обезличенном виде. Заметим, что в техническом задании на создание федеральной подсистемы ведения ИЭМК в составе ЕГИСЗ (www.zakupki.gov.ru, 13.12.2011, заказ № 0173100005411000589) было указано, что в подсистеме должна быть реализована возможность использования технологий псевдонимизации, а также разработаны типовые сценарии формирования и доступа к обезличенным данным. Однако в документах по этой подсистеме, которые в настоящее время опубликованы на сайте www.portal.egisz.rosminzdrav.ru, о том, как это было реализовано ничего не сказано.

Идентификатор лица (personal identifier) – информация, с помощью которой лицо может быть однозначно определено в определенном контексте [20].

Служба идентификации субъектов персональных данных (пациентов) – специальные сотрудники, подразделение или организация, уполномоченные предоставлять авторизованным пользователям определенный, "стандартный" набор персональных данных конкретного пациента по запросу, содержащему его идентификатор (на основе [20]).

Обезличивание (де-идентификация, де-персонификация, de-identification) персональных данных – действия, в результате которых становится невозможным без использования дополнительной информации определить принадлежность персональных данных конкретному субъекту персональных данных. Основная цель обезличивания – обеспечение конфиденциальности персональных данных. Далее обезличенные данные будем назвать О-данными (записями, документами).

Деобезличивание (персонификация) – действия, в результате которых обезличенные данные принимают вид, позволяющий определить их принадлежность конкретному субъекту персональных данных – становятся персональными данными.

Анонимный идентификатор (anonymous identifier, AI) – идентификатор лица, по которому невозможно однозначно установить, какое именно физическое лицо он обозначает [20].

Анонимность данных – невозможность на основе этих данных однозначно установить их принадлежность определенному, конкретному лицу (персоне), без использования дополнительной информации.

Псевдоним (pseudonym) – идентификатор лица, по которому нельзя воспроизвести обычно используемый его идентификатор и установить его личность без использования дополнительной информации (на основе [20]).

Служба (сервисы) псевдонимизации – программные средства, предназначенные для выполнения функций формирования и присвоения псевдонимов пациентам, обезличивания и обратной персонификации (де-обезличивания) записей, а также специально уполномоченные сотрудники, подразделение или организация, осуществляющие администрирование и обеспечение применения этих программных средств пользователями.

В документе Европейского союза [7] и стандартах Международной организации стандартизации (ИСО, www.iso.org) сведения о человеке подразделяются на:

– сведения, позволяющие идентифицировать персону / личность (Personally Identifiable Information, PII); здесь и далее PII – это вполне определенный набор атрибутов (элементов структуры) данных – Ф.И.О., дата и место рождения, адрес места регистрации, жительства, место работы, должность, сведения о законных представителях, членах семьи, родственниках и т.д., различного рода идентификаторы,

сопоставленные с персоной: серия и номер паспорта, полиса ОМС, СНИЛС, ИНН, номер медицинской карты и т.п., а также контактные данные: номер телефона, адрес электронной почты[2] и т.д.; считаем, что всегда существует хотя бы одна совокупность PII-атрибутов, обеспечивающая однозначную идентификацию пациента; заметим, что с точки зрения возможности идентификации пациента, набор PII-атрибутов в записи может быть избыточным; некоторые рекомендации по составу атрибутов и алгоритмам идентификации пациентов приведены в ГОСТ ISO/TS 22220;

– сведения, соотносимые с конкретной личностью – персоной (the Information Correlation with the Person, ICP) – документированные сведения о человеке, в том числе о состоянии здоровья, оказанной медицинской помощи и т.д., – на основе которых невозможно однозначно определить их принадлежность конкретному лицу (персоне); то есть любой набор ICP-атрибутов является анонимным (anonymized data); в [20] такие данные названы "обрабатываемыми" или "деперсонифицированными";

– персонифицированные данные (personalized data, далее – исходные И-данные, записи) – данные о конкретном человеке (персоне), содержащие, в том числе сведения, позволяющие его идентифицировать; формально состав таких данных можно представить в виде пары (PII, ICP), где PII и ICP – наборы PII- и ICP-атрибутов.

В общем случае может осуществляться де-идентификация не только персональных данных физического лица, но и данных, позволяющих идентифицировать юридическое лицо – медицинскую или иную организацию. Это может быть необходимо для повышения "надежности" обезличивания данных, содержащих сведения о соответствующих организациях, если они при определенных условиях могут быть использованы для восстановления персональной принадлежности О-данных и установления личности пациента.

Далее, если специально не указано иное, словом "атрибут" будем обозначать и определенный элемент структуры данных (структуры записи), и его значение или величину (число, символьную строку, код и т.п.). Латинскими буквами в тексте и формулах будем обозначать как определенные атрибуты, так и их соответствующие наборы и/или категории (классы, виды).

Виды обезличивания персональных данных. Следует различать необратимое и обратимое обезличивание. В зависимости от того, какие действия выполняются и как реализуется процесс обезличивания, полученная деперсонифицированная О-запись может быть:

[2] Заметим, что в отличие от номера телефона, который по закону обязательно привязывается к "паспортным" данным абонента, личный, не служебный адрес электронной почты, если только он специально не указан самим субъектом персональных данных при регистрации у оператора, не позволяет идентифицировать персону, поскольку может присваиваться анонимно.

а) **анонимизированной**[3], когда все PII-атрибуты из записи удаляются либо изменяются таким образом, что выполнить её обратную персонификацию – сопоставить с PII какого-либо определенного пациента – невозможно (необратимое обезличивание); при этом О-запись помечается неким анонимным идентификатором AI; далее такие данные будем назвать А-данными;

б) **псевдонимизированной**, когда, как и при анонимизации, PII-атрибуты изменяются или удаляются из записи, но при этом обезличенные данные о пациенте помечаются его "секретным" псевдонимом Ps, присвоенным по определенным правилам, что позволяет при необходимости выполнить процедуру обратной персонификации – деобезличивание О-записи; далее такие данные будем назвать П-данными.

Как будет показано далее, псевдонимизация данных и организационно, и технологически существенно сложнее и затратнее, чем анонимизация. В тоже время надо иметь в виду, что:

– после выполнения процедуры анонимизации, А-данные о пациенте уже не могут быть корректно изменены и/или дополнены;

– собрать и объединить, интегрировать А-данные об одном и том же пациенте, полученные от разных источников и/или в разное время невозможно;

– изменение, актуализация А-данных может осуществляться только путем повторной анонимизации обновленных исходных персонифицированных данных о пациенте;

– в отличие от А-данных, П-данные о пациенте могут быть получены в разное время от одного или нескольких источников, объединены, дополнены – изменены по составу атрибутов и/или их значений без выполнения процедуры деобезличивания;

– изменение, актуализация и консолидация П-данных о пациенте могут осуществляться в инкрементном режиме с сохранением инкогнито пациента.

Конфиденциальность обезличенных данных. Полученные после обезличивания данные не содержат сведений, позволяющих идентифицировать личность, и поэтому они уже не являются персональными данными (с точки зрения закона № 152-ФЗ). В общем случае это позволяет существенно упростить организацию

[3] Сегодня иногда путают анонимизацию данных и оказание медицинской помощи на анонимной основе. После анонимизации медицинский документ уже невозможно сопоставить с конкретным человеком – пациентом, и поэтому его уже нельзя использовать по его прямому назначению – для принятия клинических решений, организации и учета выполнения врачебных назначений пациенту и т.д., а также для "для прокурора". При анонимном лечении пациент не сообщает свои реальные персональные данные и его личность в юридическом смысле не устанавливается, при этом медицинские документы могут быть сопоставлены с конкретным пациентом только если они содержат соответствующие биометрические данные, либо собственноручно сделанную запись и подпись пациента (правовые аспекты и особенности оказания медицинской помощи на анонимной основе, в том числе, возможно, с использованием процедур псевдонимизации, – это отдельная тема).

доступа к обезличенным данным и снизить совокупные затраты на создание системы защиты и обеспечение конфиденциальности информации.

В настоящее время в большинстве развитых стран возможность сбора А-информации без согласия субъекта персональных данных независимо от целей обработки нормативно регулируется только в части обязательных требований к процедурам гарантированной анонимизации данных[4]. Что касается П-данных, то возможность их сбора и обработки с той или иной целью, а также допустимость открытого, публичного доступа к таким данным, во многих странах строго регламентируется национальным законодательством. Например, в Германии для обработки псевдонимизированных данных требуется специальное согласие субъекта персональных данных. Сегодня, в связи с интенсивной "интернетизацией" экономики, социальной сферы и здравоохранения, появлением социальных сетей и развитием технологий, называемых Big Data, вопросы правового регулирования и регламентации процессов сбора и использования псевдонимизированных данных находятся в зоне особого внимания как государственных органов, так и общественных организаций. В российском законодательстве каких-либо явных ограничений и требований относительно сбора и обработки обезличенных данных пока еще нет, хотя уже обсуждаются законопроекты, регламентирующие использование Big Data, в том числе в сфере здравоохранения и медицинской науки.

Таким образом очевидно, что организация практического применения технологий анонимизации и псевдонимизации в здравоохранении, когда медицинская, клиническая и административная информация о пациенте должна быть документирована и юридически значима, требует соответствующего правового, организационного, методического и технического обеспечения.

Применение методов псевдонимизации в здравоохранении. Общемировая практика показывает, что псевдонимизацию данных, содержащих сведения, относящиеся к врачебной тайне, целесообразно применять:

1) в тех случаях, когда врач, медработник, участвующий в процессе оказания медицинской помощи пациенту, непосредственно работает только:

– с медицинскими документами, например, при подготовке экспертного заключения ("второе мнение"), при описании и интерпретации медицинских изображений и диаграмм: рентгеновских снимков, томограмм, ЭКГ и т.д., в том числе полученных по каналам связи (телерадиология, телекардиология, телеморфология и т.д.); следует заметить, что международным стандартом представления и передачи медицинских

[4] Практически всегда существует определенный риск восстановления персональной принадлежности обезличенных, псевдонимизированных данных и компрометации псевдонима. Выбор методов и разработка процедур обезличивания должны осуществляться с учетом анализа и оценки указанных рисков на основе сответствующей модели угроз. Описание подходов к построению такой модели приведено в ГОСТ Р 55036.

изображений DICOM (см. ГОСТ Р ИСО 12052 и ГОСТ Р ИСО 17432) предусмотрена возможность псевдонимизации данных;

– с биоматериалами при проведении лабораторных исследований; сегодня такая практика все чаще используется медицинскими организациями при взаимодействии с внешними, централизованными лабораториями, что существенно упрощает их работу и позволяет снизить затраты на защиту информации.

В описанных выше случаях псевдонимизация может осуществляться как локально, так и централизованно (см. далее). При этом после получения П-документа с результатами исследования или заключением перед их записью (включением) в медицинскую карту пациента предварительно должна осуществляться обратная персонификация документа[5];

2) при "вторичном" использовании массивов медицинских данных, когда: а) необходимо обеспечить доступ к ним большого количества различных пользователей для решения управленческих, научных и иных задач, и при этом б) может потребоваться идентифицировать пациента (субъекта данных) – осуществить обратную персонификацию, например, для получения дополнительных сведений о пациенте и т.д.; типичными примерами такого рода массивов данных могут быть:

– территориально-популяционные нозологические регистры;

– геномные (ДНК-) медицинские регистры; заметим, что в законодательстве большинства европейских стран и у нас в РФ действует принцип "одинаковых требований к уровню конфиденциальности" геномной, генетической и иной медицинской информации;

– регистры потенциальных и фактических доноров и реципиентов органов и тканей человека; в качестве примера можно назвать всемирную поисковую систему доноров костного мозга (www.bmdw.org); подобная система сейчас создается и у нас в России [10];

– регистры, формируемые при проведении научных исследований; особенно эффективно это при сборе, унификации и консолидации данных многоцентровых исследований [18];

– регистры пациентов, используемые при проведении клинических испытаний (см. ГОСТ Р 52379, ГОСТ Р 56044 и ГОСТ Р ИСО 14155); заметим, что применение методов псевдонимизации позволяет при этом обеспечить также и высокий уровень "ослепления" пользователей – экспертов, работающих с этими данными;

– регистры лиц, с имплантированными медицинскими изделиями, которые используются при сборе катамнестической информации и оценке безопасности и эффективности применения медицинских изделий (см. [11]);

---

[5] Заметим, что в обоих приведенных выше случаях специальное согласие пациента на передачу его П-данных из МО, в которую он обратился, в другую МО не требуется (см. п. 8 части 4 ст. 13 закона № 323-ФЗ).

– регистры, формируемые при сборе извещений о побочном действии лекарственных препаратов, медицинских изделий и биомедицинских клеточных продуктов;

– базы данных персонифицированного учета объемов и результатов медицинской помощи, оказанной по программам ОМС, которые формируются и ведутся в СМО и территориальных фондах ОМС.

Очевидно, что при работе с перечисленными выше регистрами и базами данных псевдонимизация должна, как правило, осуществляться централизованно. В тоже время, у автора есть положительный опыт работы с обезличенной базой данных об оказанной медицинской помощи для выявления клинически связанных случаев обращения пациента в различные лечебно-профилактические учреждения (так называемых эпизодов), псевдонимизация которой осуществлялась локально в территориальном фонде ОМС.

3) при централизованном ведении баз данных, в которых собираются и накапливаются различные медицинские и иные сведения о персоне – конкретном человеке, поступающие из множества учреждений, в которых он проходил обследование, лечение или реабилитацию [12]; такая персоно-центрированная модель сбора данных о состоянии здоровья и оказанной медицинской помощи за рубежом получила название Long Life Personal Health History, а у нас – интегрированная ЭМК – ИЭМК [13, 14] (правильнее – "интегральная"); в ГОСТ Р ИСО/ТО 20514 этому соответствует термин Electronic Health Record for Integrated Care (ICEHR); как правило, такие БД содержат записи об анамнезе жизни, сигнальные (витальные) данные, выписные эпикризы по каждому случаю оказания медицинской помощи пациенту и др. Основная цель создания подобных БД – обеспечение преемственности, безопасности, качества и эффективности медицинской помощи. К ним организован удаленный доступ авторизованных пользователей – прежде всего, врачей различных учреждений, которым при этом "видны" реальные PII-данные пациента. Иным категориям пользователей доступ к таким БД предоставляется только в режиме чтения А- или П-записей – реальные PII-данные пациентов для них закрыты. В отличие от медицинских регистров, предназначенных для решения узко специальных задач, такие БД, очевидно, обладают гораздо большим "аналитическим потенциалом" и могут быть использованы при решении самых разных клинических, управленческих, научных и учебных задач, в том числе для формирования и ведения медицинских регистров. Псевдонимизация записей в таких БД должна осуществляться централизованно. Примером подобной БД является единая база данных "выписных" эпикризов в национальной системе здравоохранения Великобритании, ведение которой осуществляется с использованием псевдонимов, для чего создана специальная служба Secondary Uses Service (SUS), обеспечивающая

удаленный авторизованный доступ врачей и пациентов к этой базе данных через систему Spine [15].

Доступ к перечисленным выше базам данных может быть организован так же как и к анонимизированным данным – без возможности персонификации записей.

Методы обезличивания. В настоящее время разработано много различных методов обезличивания [16-19]. Приказом уполномоченного органа по защите прав субъектов персональных данных – Роскомнадзора – в 2013 г. были утверждены требования по обезличиванию персональных данных [8], опубликованы методические рекомендации по применению этого приказа [9]. С июля 2013 г. введен в действие ГОСТ Р 55036 [20], который был разработан путем перевода технических спецификаций ISO/TS 25237:2008 (в январе 2017 г. ИСО был издан "настоящий стандарт" ISO 25237:2017 Health informatics. Pseudonymization), с июля 2017 г. – ГОСТ Р ИСО/МЭК 27038 [21]. Однако, все эти документы имеют "рамочный" характер – для практического использования методов анонимизации и псевдонимизации в российском здравоохранении необходимо определить организационные процедуры, разработать алгоритмы и специальные программные средства, исходя из потребностей различных прикладных задач и условий их применения, и издать соответствующие нормативно-методические документы.

Далее будем рассматривать процессы обезличивания только структурированных данных – записей и документов. При этом полагаем, что структура записи (документа) соответствует 2-му (section-level templates) или 3-му (entry-level templates) уровню формализации по ГОСТ Р ИСО/HL7 27932.

Процедуры и алгоритмы обезличивания основаны на том, что любая структурированная персонифицированная запись (документ) может быть представлена в виде двух наборов атрибутов – (PII, ICP) и при этом PII-атрибуты, в свою очередь, в общем случае могут быть разделены на пять наборов атрибутов:

PII = (ID, pV, pR, pN, pF), где

ID – идентификаторы пациента; следует различать: а) внутренние ID, присваиваемые самим оператором (например, номер медицинской карты), и б) внешние ID, которые присвоены пациенту другим оператором (третьим лицом), например, номер полиса ОМС или СНИЛС; как правило, один из внешних идентификаторов пациента принимается в качестве "стандартного", единого для всех организаций – источников данных о пациентах (далее – ID_S; считается, что такой идентификатор всегда существует); при обезличивании все идентификаторы либо удаляются, либо заменяются другими идентификаторами пациента:

– при анонимизации – на анонимный идентификатор пациента AI – некий чисто условный код, который не связан с PII какого-либо пациента;

– при псевдонимизации – на псевдоним пациента Ps (см. далее);

pV – допускающие (предусматривающие) замену на их обобщенное значение pG; например, вместо полного адреса указывается только название или код населенного пункта или района, вместо возраста – код возрастной группы (к pV-атрибутам относятся также конкретные даты: рождения, смерти, обращения в МО, госпитализации, выписки и т.п., которые должны либо заменяться на обобщенные периоды: номер недели, месяца, квартала и т.д., либо удаляться или заменяться "пустыми" значениями); в исходной И-записи pV-атрибуты могут отсутствовать; далее замену pV-атрибутов на их обобщенные значения будем называть генерализацией и обозначать pV -> pG;

pR – допускающие (предусматривающие) замену на вычисляемый атрибут pC, значение (величина) которого рассчитывается по определенному алгоритму; например, вместо даты рождения пациента вычисляется и указывается его возраст или код возрастной группы, вместо роста и веса – величина индекса массы тела (в данной работе рост, вес и другие антропометрические данные пациента будем относить к PII-атрибутам); в исходной записи pR-атрибуты могут отсутствовать; далее замену pR-атрибутов на вычисляемые атрибуты будем обозначать pR -> pC; преобразования pV -> pG и pR -> pC в [8, 9] обобщенно называются методом изменения семантики или состава данных;

pN – номинальные, не подлежащие генерализации или какому-либо иному преобразованию (Ф.И.О., контактные реквизиты, текстовые поля в "свободном формате", в которых могут содержаться персональные данные и др.); при обезличивании они удаляются из записи (не включаются в О-запись) либо заменяются "пустыми" значениями;

pF – идентифицирующие других физических лиц (не пациентов), которые в определенном контексте и/или в сочетании с другими атрибутами могут быть использованы для определения личности пациента (поэтому они отнесены к PII-атрибутам); могут отсутствовать; при обезличивании они либо удаляются из записи, либо при необходимости могут заменяться:

– при анонимизации – на анонимные идентификаторы этих физических лиц pA, не позволяющие определить их истинные, реальные реквизиты (Ф.И.О., адрес, СНИЛС и т.д.), что можно представить в виде преобразования pF -> pA;

– при псевдонимизации – на соответствующие "секретные" персональные псевдонимы pP, которые присваиваются по определенным правилам, что позволяет осуществить обратную персонификацию pP -> pF – сопоставление псевдонимов с реальными реквизитами указанных лиц.

Наличие в записях pF-атрибутов и их замена на pA и pP позволяет при обезличивании сохранить информацию о различных видах взаимосвязей (отношений) между субъектами персональных данных, представляемых с помощью указанных атрибутов. В [8, 9] это свойство названо структурированностью обезличенных данных.

Далее "обезличенные" идентификаторы субъектов персональных данных обобщенно будем обозначать pU = {AI, Ps} – для пациента и pX = {pA, pP} – для иных физических лиц.

Все PII-атрибуты при обезличивании могут также представляться в O-записи "пустыми" значениями. Возможно также шифрование PII-атрибутов на "секретных" ключах службы псевдонимизации, однако на практике этот метод используется крайне редко и здесь не рассматривается.

Что касается обезличивания мультимедийных файлов – фото-, видео- и аудио-записей, прилагаемых к медицинским документам, а также сканов медицинских документов, содержащих в том или ином виде информацию, позволяющую идентифицировать личность пациента, то это особая, отдельная проблема. Некоторые общие рекомендации на эту тему приведены в [21].

Анонимизация данных реализуется путем удаления из записи, генерализации и/или замены PII-атрибутов – выполнения следующих преобразований:

(ID, pN, pV, pR, pF, ICP) -> (AI, (pG, pC, pA, ICP));

где (pG, pC, pA, ICP) – обрабатываемые данные о пациенте (как видим, ICP могут быть дополнены pG-, pC- и pA-атрибутами); AI – анонимный идентификатор пациента, которым помечены обрабатываемые данные. Напомним, что pV-, pR- и pF-атрибуты могут отсутствовать. Атрибуты с "пустыми" значениями здесь и далее не показаны.

Присвоение псевдонимов. Начальным этапом процесса псевдонимизации является присвоение пациенту псевдонима Ps, которое осуществляется специальной службой (сервисом) псевдонимизации. При этом для разных целей и задач могут использоваться разные псевдонимы одного и того же лица. Для одних задач псевдоним пациента может быть как разовым, так и постоянным (например, при обмене медицинскими документами между МО и клинической лабораторией), для других – только постоянным, в частности, при ведении БД ИЭМК. Принципиально важно, что псевдоним:

– должен быть уникальным в системе идентификации и учета пациентов, для использования в которой он предназначен;

– не должен совпадать ни с одним из идентификаторов, сопоставленных с пациентом и используемых в других системах учета и идентификации физическим лиц или относящихся к ним документов, записей и т.д., например, с номером медицинской карты, полиса ОМС, серией и номером паспорта, СНИЛС и т.д.;

– никогда не указывается вместе с персональными данными пациента в первичных медицинских, и иных документах, доступ к которым может получить пациент или его законный представитель;

– во всех случаях не известен ни врачу, ни пациенту и поэтому не может быть ими передан кому-либо или раскрыт, случайно или намеренно, в отличие от других идентификаторов, например, номера медкарты, полиса ОМС и СНИЛС;

– может быть сопоставлен с персональными данными пациента, раскрыт или передан кому-либо только в строго определенных специальных случаях, предусмотренных законодательством либо соглашением между пользователями ИС, в которой обрабатываются П-данные, по жестко контролируемым процедурам с обеспечением установленных требований по защите персональных данных и сохранению врачебной тайны.

В зависимости от метода формирования псевдонимы могут быть:

а) назначаемыми – с помощью "секретных" таблиц соответствия (PII : Ps); уникальный Ps при этом может формироваться на основе некоторого порядкового или составного номера, с помощью датчика случайных чисел либо с использованием хэш-функции, вычисляемой по значению определенной уникальной совокупности PII-атрибутов $PII_P \subseteq PII$ или, чаще всего – "стандартного" ID_S пациента, например, СНИЛС; в случаях, когда алгоритм формирования псевдонима опубликован, в качестве Ps целесообразно использовать зашифрованное значение кода-идентификатора, сформированного указанным выше способом, иначе возможна компрометация псевдонима путем простого последовательного подбора; важно, что во всех перечисленных случаях обратное вычисление PII_P на основе Ps невозможно;

б) вычисляемыми – путем шифрования на "секретных" ключах службы псевдонимизации определенного набора атрибутов PII_P, обеспечивающего однозначность идентификации пациента и уникальность псевдонима; в этом случае "секретная" таблица соответствия (PII_P : Ps) может не формироваться и не храниться – прямое PII_P -> Ps и обратное Ps -> PII_P криптопреобразование выполняются "на лету"; при этом могут применяться как симметричные, так и асимметричные методы шифрования.

Присвоение псевдонима пациенту и иным физическим лицам может выполняться как заблаговременно, так и непосредственно в процессе формирования псевдонимизированной О-записи о пациенте.

При необходимости по аналогичным правилам псевдонимы могут присваиваться также иным физическим лицам, сведения о которых содержатся в исходных записях о пациенте. При этом также используется соответствующий "стандартный" идентификатор субъекта персональных данных. Из-за ограниченного объема статьи описание указанных процессов здесь не приводится.

Псевдонимизация, как и анонимизация данных, может быть реализована путем выполнения следующих преобразований:

(ID, pN, pV, pR, pF, ICP) -> (ID_S, (pG, pC, pF, ICP)) -> (Ps, (pG, pC, pP, ICP));

где используется таблица соответствия (ID_S : Ps) либо псевдоним Ps вычисляется по идентификатору ID_S пациента; (pG, pC, pP, ICP) – обрабатываемые данные о пациенте, помеченные его псевдонимом Ps; если PII включают данные о законных представителях и/или родственниках пациента pF, и они необходимы для дальнейшей обработки, то их также надо заменить их псевдонимами pP. В ряде случаев при обезличивании может потребоваться также удаление из записей, замена на псевдонимы или генерализация данных о врачах и/или медицинских организациях. Например, индивидуальный код врача может быть заменен на код должности или специальности.

Сформированная П-запись может быть направлена в виде сообщения (электронного документа) адресату-получателю и/или сохранена в базе данных DB для дальнейшего использования (см. рис. 2). При этом в базе данных может осуществляться сбор, консолидация и накопление П-данных о пациенте, полученных от множества разных источников.

Указанные выше преобразования данных при обезличивании могут осуществляться в автоматическом режиме с помощью соответствующих программных средств (генераторов таблиц соответствия, специальных шлюзов или серверов псевдонимизации и обратной персонификации и т.д.[6]), под контролем уполномоченных сотрудников службы псевдонимизации. При этом, как правило, они не имеют доступа к персонифицированным данным о пациентах, содержащим сведения, составляющие врачебную тайну.

Формирование обезличенной записи о пациенте может осуществляться с использованием заранее определенного шаблона (template) О-записи, в котором перечислены все атрибуты исходных И-записей, и при этом каждый атрибут имеет специальную пометку, обозначающую, какое действие с ним выполняется – элиминация или замена:

– удаление – для pN-, ID- и pF-атрибутов (не включаются в О-запись);

– замена на "пустое" значение pZ – для любых PII-атрибутов;

– замена на анонимный идентификатор – для pN, ID – на AI, для pF – на pA;

– замена на псевдоним – для pN, ID – на Ps, для pF – на pP;

– замена на обобщенное значение pG (генерализация) – для pV-атрибутов;

– замена на вычисляемый атрибут pC – для pR-атрибутов;

– включение в О-запись без изменений – для ICP-атрибутов.

При формировании О-записи операции по замене одного и того же PII-атрибута на "пустые", обобщенные или вычисляемые значения могут применяться одновременно.

[6] Для этого, например, могут использоваться специальные USB-устройства, "на борту" которых выполняются все криптопреобразования, хранятся ключи шифрования и аутентификации пользователей и т.д., подобно тому, как это сделано сегодня в USB-токенах.

Все PII-атрибуты, не помеченные в шаблоне как заменяемые, безусловно удаляются (не включаются в O-запись).

Очевидно, что все перечисленные выше операции удаления или замены атрибутов, кроме замены на вычисляемый псевдоним, необратимы – восстановить исходные значения атрибутов при обратной персонификации невозможно. В связи с этим при обезличивании возможна частичная потеря информации о пациенте в O-записи.

Генерализация и замена атрибутов при обезличивании записей могут осуществляться и для ICP-атрибутов.

Должны быть приняты единые правила генерализации и расчета значений вычисляемых атрибутов.

В зависимости от того, как реализован процесс формирования обезличенной записи, следует различать:

а) цензурирование исходной И-записи:

$$(PII, ICP) \rightarrow (pU, (pZ, ICP));$$

где pU – анонимный идентификатор AI или псевдоним Ps пациента; при цензурировании всегда создается новый документ, в котором все PII-атрибуты заменены на "пустые" значения pZ; при этом, как правило, в O-запись включаются все ICP-атрибуты; общие требования и рекомендации по выполнению цензурирования документов приведены в [21];

б) сборку (компиляцию) O-записи из атрибутов одной или нескольких И-записей о пациенте:

$$\{(ID_j, pN_j, pV_j, pR_j, pF_j, ICP_j) \mid j = 1, \ldots, N\} \rightarrow (pU, (pG, pC, pX, ICP));$$

при этом O-запись может включать консолидированные наборы атрибутов данных, полученных в результате преобразований pV -> pG, pR -> pC и pF -> pX, а также экстракции и/или объединения атрибутов из ICP-данных множества И-записей:

$$pG = pG_1 \cup \ldots \cup pG_N; \quad pC = pC_1 \cup \ldots \cup pC_N; \quad pX = pX_1 \cup \ldots \cup pX_N;$$

$$ICP = ICP'_1 \cup \ldots \cup ICP'_N; \quad \text{для } j = 1, \ldots, N;$$

где $ICP'_j \subseteq ICP_j$ – наборы ICP-атрибутов исходных И-записей ICP_j, включаемые в O-запись в соответствии с шаблоном. Как правило, O-запись не содержит атрибутов с "пустыми" значениями. Исходные записи о пациенте могут быть разного типа и находится в разных хранилищах (базах) данных. В шаблоне указываются типы исходных записей и при необходимости – места их хранения, а также ссылки на программные модули для выполнения генерализации и расчета значений вычисляемых атрибутов.

Как видим, цензурирование документа и сборка обезличенного документа из множества исходных И-записей – это разные процессы, технологии и возможные области их применения. Сложность и техническая осуществимость процедур

обезличивания зависят от уровня формализации контента в записи – состава атрибутов, формы и способов представления данных.

Деобезличивание. Персонификация П-данных – восстановление их принадлежности определенному пациенту – осуществляется с помощью службы (сервисов) псевдонимизации по запросам авторизованных пользователей при наличии у них соответствующих полномочий. При этом, как уже было отмечено выше, возможна потеря части исходной информации о пациенте, поскольку не представляется возможным восстановить:

– удаленные / не включенные в О-запись PII-атрибуты;

– исходное значение pG- и pC-атрибутов;

– исходное значение атрибутов с "пустыми" значениями.

Идентификация, аутентификация, проверка полномочий пользователей и предоставление им по запросу, содержащему псевдоним Ps пациента, его стандартного идентификатора ID_S, осуществляются службой псевдонимизации. Через службу (сервисы) идентификации пациентов по запросу, содержащему ID_S пациента, авторизованным пользователям может быть предоставлен определенный "стандартный" набор $PII_S \subseteq PII$ его персональных данных. Проверка полномочий, обработка запросов и предоставление пользователям указанных данных о пациентах обеими службами могут осуществляться в автоматическом режиме с помощью соответствующих программных средств (сервисов).

Обезличивание и обратная персонификация записей могут осуществляться как в пакетном, так в транзакционном режимах. Описание возможных вариантов и особенностей реализации указанных режимов, а также процедур и способов идентификации, аутентификации и проверки полномочий пользователей в виду ограниченного объема статьи здесь не приводится[7].

Организация процессов псевдонимизации и обратной персонификации. Возможности, процедуры и эффективность использования методов псевдонимизации при решении различных медицинских, управленческих, научных, учебных и иных задач во многом зависят от того, как организованы присвоение псевдонимов, де-персонификация и обратная персонификация П-данных – локально или централизованно.

---

[7] Заметим, что в настоящее время нет утвержденных Минздравом России нормативных документов, регламентирующих права доступа различных категорий работников к тем или иным видам медицинских документов или их разделам. Правила (политика) доступа к медицинской документации определяются локальными нормативными актами, изданными в учреждениях. При использовании бумажных носителей отсутствие четких, формализованных правил доступа к документам не создавало особых проблем с точки зрения обеспечения конфиденциальности персональных данных пациентов и сохранения врачебной тайны. При работе с электронным документами и базами данных разделение полномочий и регламентация доступа пользователей к информации является обязательным требованием. Некоторые общие рекомендации по организации доступа в медицинских ИС приведены в ГОСТ Р ИСО/ТС 22600, ГОСТ Р 54472, ГОСТ Р ИСО 21091.

В первом случае присвоение псевдонимов пациентам, де-персонификация и обратная персонификация данных осуществляются в самой МО собственной, локальной службой (сервисами) псевдонимизации LPS. Модель потоков данных о пациенте для этого случая в виде направленного линейного графа показана на рис.1, где U_M – множество пользователей – врачей, медсестер и т.д., непосредственно работающих с первичной медицинской документацией (ЭМК) и персональными данными пациента, при этом Ps пациента им не известен; U_E – внешние пользователи, например, сотрудники внешней клинической лаборатории, которые работают только с биоматериалами и псевдонимизированными электронными медицинскими П-документами; направленными стрелками обозначены соответствующие потоки данных (документов), в надписях указан состав данных в потоках: pG, pC, pP и ICP_M – данные о пациенте, которые передаются внешним пользователям, ICP_E – получаются от внешних пользователей. Атрибуты с "пустыми" значениями в О-запись, как правило, не включаются и здесь не показаны.

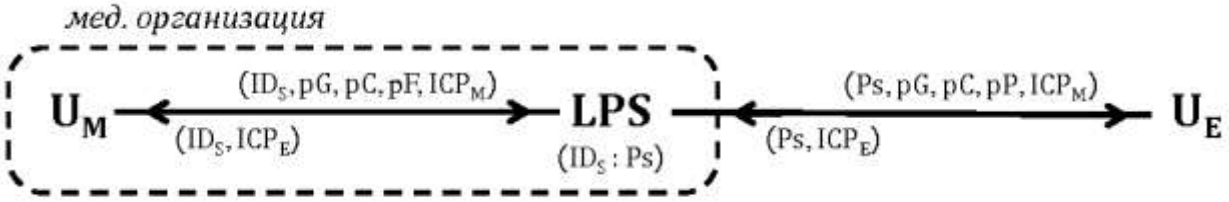

Рис. 1 – Модель локальной псевдонимизации

Очевидно, что область применения технологий локальной псевдонимизации весьма ограничена.

Во втором случае организуется единая централизованная служба (сервисы) псевдонимизации CPS, что позволяет: а) осуществлять сбор, хранение и обработку П-данных об одном и том же пациенте из множества различных источников – медицинских учреждений, СМО и др.; и б) обеспечить доступ большого количества различных пользователей к массивам П-данных (документов).

Соответствующая этому варианту обобщенная модель доступа к центральной базе П-данных DB, содержащей записи о пациентах вида (Ps, ICP_{DB}), показана на рис. 2, где $\{U_{Mj}\}$ – это, как и в предыдущем случае, множество пользователей, работающих с персонифицированными медицинскими данными (МО, СМО и др.); авторизованный доступ этих пользователей к базе данных DB по чтению и/или записи осуществляется через службу (сервисы) псевдонимизации CPS; при этом они "видят" только реальные ID_S и "свои" PII_j пациента, Ps пациента им не известен; U_A – пользователи, которые во всех случаях работают только с П-данными – при решении профессиональных задач им не нужны PII пациентов, для них эти данные являются анонимизированными; PIS – служба идентификации пациентов, предоставляющая авторизованным пользователям

определенный "стандартный" набор PII_S персональных данных конкретного пациента по запросу, содержащему его стандартный идентификатор ID_S; U_C – пользователи, которые в строго определенных случаях при решении своих прикладных задач имеют право по псевдониму пациента Ps получить через службу псевдонимизации CPS авторизованный доступ к его стандартному идентификатору ID_S и затем обратиться в службу идентификации пациентов PIS для получения "стандартного" набора PII_S его персональных данных.

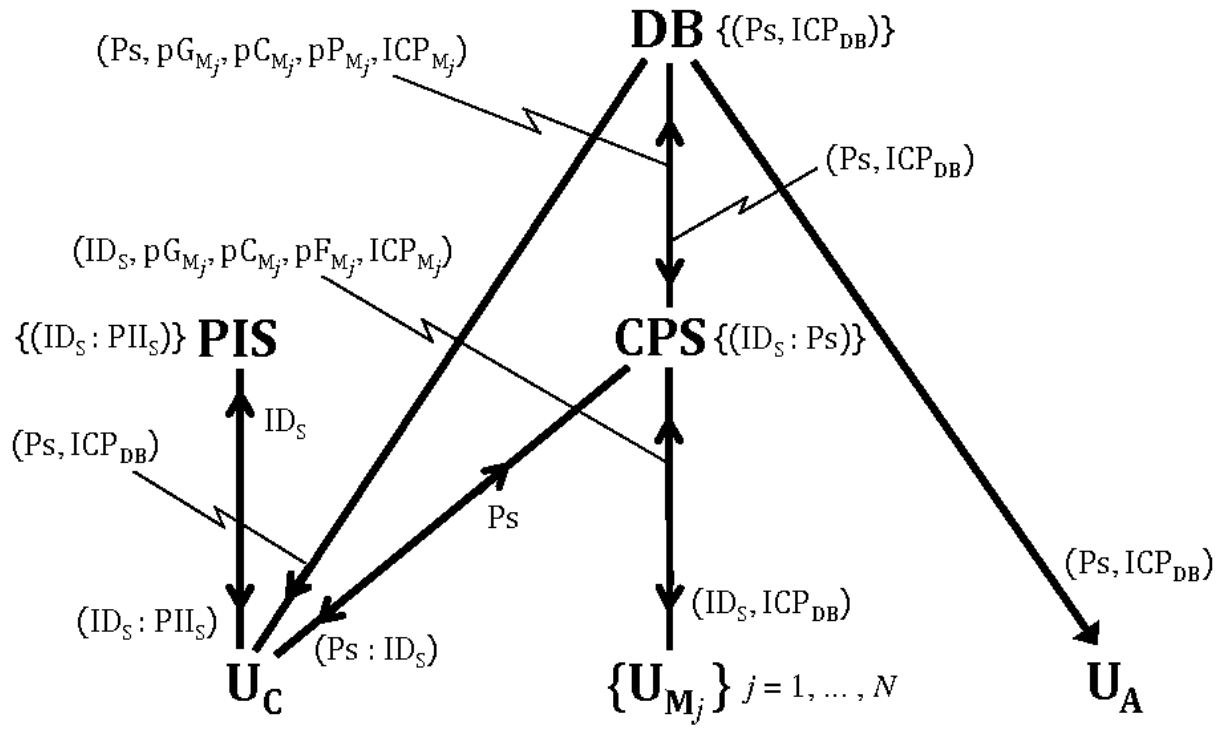

Рис. 2 – Модель доступа к данным через централизованную службу псевдонимизации

Формирование записей о пациентах в DB осуществляется на основе данных, поступающих от U_M-пользователей через службу (сервисы) псевдонимизации CPS. "Интегральная" запись о пациенте ICP_{DB} получается в результате объединения всех П-записей, полученных в разное время от разных медицинских организаций:

$$(Ps, ICP_{DB}) = (Ps, ((pG_{M_1}, pC_{M_1}, pP_{M_1}, ICP_{M_1}) \cup ... \cup (pG_{M_N}, pC_{M_N}, pP_{M_N}, ICP_{M_N})));$$

Как правило, удаление PII-атрибутов и преобразования pV -> pG и pR -> pC выполняются на стороне источника, замена стандартных идентификаторов на соответствующие псевдонимы пациента и иных физических лиц – через службу (сервисы) псевдонимизации.

Примером описанной централизованной модели, как уже было отмечено, может быть псевдонимизированная БД ИЭМК, с которой могут работать все три категории пользователей, показанные на рис. 2. Например, лечащий врач (U_M) может запросить и получить из БД анамнестическую информацию, сигнальные данные и эпикриз по

последнему случаю оказания медицинской помощи пациенту; врач-эксперт страховой компании (U_C) – запросить и получить персональные данные пациента в случае необходимости проведения дополнительной экспертизы по первичной медицинской документации; ученые при выполнении обсервационного ретроспективного эпидемиологического исследования имеют доступ только к П-данным (U_A).

Возможна также смешанная модель псевдонимизации (multi-centric pseudonymisation [18]), когда псевдонимы присваиваются в МО, а централизованной службой осуществляется координация их присвоения (синхронизация псевдонимов, см. пп. 5.4, 5.6 в [20]). Смешанная модель организационно и технологически заметно сложнее централизованной и поэтому наименее предпочтительна, и здесь не рассматривается.

Правила обезличивания, в том числе алгоритмы преобразования pV- и pR-атрибутов, шаблоны для формирования О-записей, процедуры анонимизации, присвоения псевдонимов и обратной персонификации данных должны быть утверждены соответствующими нормативными документами или определены соглашениями между участниками – пользователями ИС, в которой обрабатываются обезличенные данные.

Выше были описаны только самые общие принципы и способы обезличивания персональных данных о состоянии здоровья и примеры их применения в здравоохранении. Остались не рассмотренными многие важные вопросы, связанные, в частности, с оценкой рисков восстановления персональной принадлежности обезличенных данных и построением модели угроз конфиденциальности персональных данных (см. в [20]), с обработкой биометрических данных, аномальными и особыми случаями идентификации и псевдонимизации данных о пациентах, и многие другие.

В заключение хотелось бы еще раз подчеркнуть, что применение методов псевдонимизации при ведении медицинских регистров и иных полицевых баз данных предоставляет качественно новые возможности сбора, обработки и использования содержащейся в них ценнейшей информации, как при решении задач практического здравоохранения, так и для решения научных и образовательных задач.

ВЫВОДЫ

1. Псевдонимизация данных в территориально-популяционных и иных медицинских регистрах и хранилищах данных позволит значительно сократить совокупные расходы на их создание и эксплуатацию, снизить затраты на сбор, обработку и предоставление доступа к информации, необходимой для контроля, планирования и принятия решений органами управления здравоохранением, проведения научных исследований, клинических испытаний и в учебных целях, что, в свою очередь, будет способствовать повышению их результативности и эффективности.

2. Представляется целесообразным на законодательном уровне определить понятие псевдонимизации медицинских данных, определить статус псевдонимизированных данных как не конфиденциальных, сбор и обработка которых не требует получения специального согласия пациента.

3. Для практического использования методов анонимизации и псевдонимизации необходимо активизировать работу по созданию нормативно-технических и методических документов, регламентирующих процессы псевдонимизации и обратной персонификации медицинских данных, определить требования к соответствующим программным и техническим средствам, разработать типовые сценарии и регламенты формирования и использования обезличенных данных при решении различных практических задач в здравоохранении, медицинской науке и образовании.

Автор будет признателен всем, кто пришлет свои замечания и предложения по рассмотренным вопросам по электронной почте на адрес ap100Lbov@mail.ru.

ЛИТЕРАТУРА

1. Концепция создания единой государственной информационной системы в сфере здравоохранения. Приказ Минздравсоцразвития России от 28.04.2011 г. № 364.

2. Приоритетный проект "Совершенствование процессов организации медицинской помощи на основе внедрения информационных технологий до 2025 г.("Электронное здравоохранение")". Утвержден решением Совета по стратегическому развитию при Президенте РФ, протокол № 9 от 25.10.2016.

3. Общие принципы построения и функционирования информационных систем и порядок информационного взаимодействия в сфере ОМС (АИС ОМС). Приказ Федерального фонда ОМС № 79 от 07.04.2011 (в ред. приказа № 169 от 09.09.2016).

4. Положение об особенностях обработки персональных данных, осуществляемой без использования средств автоматизации. Постановление Правительства РФ от 15.09.2008 № 687.

5. Требования к защите персональных данных при их обработке в информационных системах персональных данных. Постановление Правительства РФ от 01.11.2012 г. № 1119.

6. Перечень мер, направленных на обеспечение выполнения обязанностей, предусмотренных Федеральным законом "О персональных данных" и принятыми в соответствии с ним нормативными правовыми актами, операторами, являющимися государственными или муниципальными органами. Постановление Правительства РФ от 21.03.2012 № 211 (ред. от 06.09.2014).

7. General Data Protection Regulation (GDPR), Regulation (EU) 2016/679, 27 April 2016. Доступно по: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679. Ссылка активна на 01.07.2017.

8. Требования и методы по обезличиванию персональных данных. Приказ Роскомнадзора от 05.09.2013 г. № 996.

9. Методические рекомендации по применению приказа Роскомнадзора от 5 сентября 2013 г. № 996 "Об утверждении требований и методов по обезличиванию персональных данных". Утверждены руководителем Роскомнадзора 13.12.2013.

10. Макаренко О.А., Алянский А.Л., Иванова Н.Е. и др. Эффективность поиска неродственного донора гемопоэтических стволовых клеток с помощью российской поисковой системы Bone Marrow Donor Search: опыт НИИ детской онкологии, гематологии и трансплантологии им. Р.М. Горбачевой // Клиническая онкогематология, 2017, № 10, сс. 39-44.

11. IMDRF/Registry WG/N33:2016 Principles of International System of Registries Linked to Other Data Sources and Tools, 30.09.2016.

12. Столбов А.П., Кузнецов П.П. Современные модели организации использования персональных данных о состоянии здоровья // Проблемы стандартизации в здравоохранении, 2010, № 1-2, сс. 19-24.

13. Тавровский В.М. Структура, содержание и ведение интегральной электронной медицинской карты (ИЭМК). Опубликовано 25.04.2015. Доступно по: http://gosbook.ru/node/89840. Ссылка активна на 01.07.2017.

14. Зарубина Т.В., Швырев С.Л., Соловьев В.Г., Раузина С.Е., Родионов В.С., Пензин О.В., Сурин М.Ю. Интегрированная электронная медицинская карта: состояние дел и перспективы // Врач и информационные технологии, 2016, № 2, сс. 35-44.

15. HSCIC Data Pseudonymisation Review – Interim Report, 31-07-2014 – http://content.digital.nhs.uk/media/14828/HSCIC-Data-Pseudonymisation-Review-Interim-Report/pdf/HSCIC_Data_Pseudonymisation_Review_Interim_Report.pdf. Ссылка активна на 01.07.2017.

16. Рябко С.Д. Об обезличивании персональных данных // Информационная безопасность, 2009, № 5. – www.itsec.ru/articles2/bypub/insec-5-2009. Ссылка активна на 01.07.2017.

17. Саксонов Е.А., Шередин Р.В. Процедура обезличивания персональных данных // Наука и образование, 2011, № 3, март 2011, электронный журнал. – http://technomag.edu.ru/doc/173146.html. Ссылка активна на 01.07.2017.

18. Lo Iacono Luigi. Multi-centric Universal Pseudonymisation for Secondary Use of the EHR, 2007. – http://geneva2007.healthgrid.org/proceedings/proceedings/pdf/25.pdf. Ссылка активна на 01.07.2017.

19. Sweeney L. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002. - P. 557-570.

20. ГОСТ Р 55036-2012 / ISO/TS 25237:2008 Информатизация здоровья. Псевдонимизация.

21. ГОСТ Р ИСО/ МЭК 27038-2016 / ISO/IEC 27038:2014 Информационные технологии. Методы обеспечения безопасности. Требования и методы электронного цензурирования.

Статья опубликована в журнале "Врач и информационные технологии", 2017, № 3 сс. 76-91