Подходы к деперсонализации информации в информационных системах медицинских и аптечных учреждений.

А.В. Курбесов, к.э.н., директор ООО «Лаборатория «Электронная медицина»

В связи с принятием закона 152-ФЗ «О персональных данных» перед многими предприятиями остро встала проблема хранения и обработки персональных данных. В лечебно-профилактических и аптечных учреждениях (ЛПУ и АУ) эксплуатируется большое количество программных комплексов в основе которых лежит персонифицированный учет медицинских услуг, поставленных диагнозов, отпущенных лекарственных средств и т.д. В настоящее время полное выполнение требований законодательства по защите персональных данных в указанных учреждениях крайне затруднено. Существующая нормативная база не дает однозначного ответа на вопрос проведены ли работы по защите персональных данных на предприятии в полном объеме. Это может привести к невозможности обработки персональных данных как в электронном виде, так и на бумажных носителях. Отсутствует полный спектр подзаконных актов регламентирующих правила защиты информации и устанавливающих однозначные критерии такой защиты. Невозможно однозначно рассчитать стоимость проведения работ по защите персональной информации. В настоящей работе выполнена попытка минимизировать объем, обрабатываемых вычислительными средствами, персональных данных путем их деперсонализации. Под деперсонализацией мы будем понимать организационно-технические мероприятия, призванные ликвидировать персонифицированную информацию в накапливаемых массивах данных. Это позволит выполнить требования соответствующих нормативных актов.

Цели обработки персональной информации в органах управления здравоохранением и системе ОМС территории.

В настоящее время в здравоохранении существует ряд задач, решение которых связано с обработкой персональных данных. Рассмотрим подробнее:
Первое: Обработка информации и получение статистической отчетности в интересах органов управления здравоохранения. Анализ показывает, что для качественной обработки указанной информации достаточно информации о возрасте, поле пациента и территории его проживания (с той или иной степенью детализации). Обработка информации в указанных целях не требует хранения и накапливания непосредственно персональных данных. Достаточно указания пола и возраста (даты рождения) пациента
Второе: Выставление счетов за пролеченных больных в системе обязательного медицинского страхования (ОМС). Данная задача объединяет четыре вида пациентов: застрахованных в системе ОМС территории и предъявивших полис; застрахованных в страховой медицинской организации (СМО), которые в силу объективных причин не смогли предъявить полис ОМС в момент обращения за медицинской помощью; иногородние пациенты; граждане других государств, не застрахованных в системе ОМС.
Для лиц, предъявивших полис территориальной системы ОМС, необходима система проверки полиса по принципу действующий/просроченный и ввода информации о поле и возрасте пациента. Этого достаточно для идентификации принадлежности пациента той или иной страховой компании. СМО, работающие на территории, должны предъявлять информацию о действительности/недействительности полиса страхования. В БД ЛПУ хранятся только реквизиты полиса, а персональная информация отсутствует. В электронном виде в СМО передаются только данные о реквизитах полиса, а конкретные фамилия, имя, отчество (ФИО) в электронных счетах за пролеченных больных отсутствуют.
К лицам, не сумевшим предъявить полис, относятся пациенты, которые поступили в медицинское учреждение для оказания экстренной помощи, но при этом у них отсутствовал

документ о страховании, или лица, проверка подлинности полиса которых дала отрицательный результат. Для идентификации таких пациентов может быть создан обобщенный сайт застрахованных территории, обеспечивающий хранение данных обо всех застрахованных в виде ФИО + дата рождения + реквизиты полиса. После отправки в такую систему запроса, содержащую информацию о ФИО пациента можно получить ответ о реквизитах страхового полиса гражданина. При этом информация, передаваемая таким образом, может распространяться по открытым каналам т.к. если запрос составлен в виде: ФИО + полная дата рождения, ответ должен прийти в виде «инициалы» (первая буква ФИО) + дата рождения + реквизиты полиса. Ответ, минимизированный указанным способом, не содержит никакой персональной информации требующей специальной защиты. Существенно большие трудности могут появиться если с помощью указанного метода не удается идентифицировать пациента. В настоящее время, в этом случае персоналом ЛПУ проводится поиск по косвенным признакам, например - по ФИО, ФИО + адрес или ФИО + адрес + место работы. Учитывая требования закона о персональных данных, такой поиск в автоматизированном виде, без разглашения персональных данных, выполняться не может. В этом случае необходимо привлекать представителей страховых компаний, имеющих договор с медицинским учреждением. Следует разработать форму официального запроса от ЛПУ к СМО. После заполнения запроса СМО должна самостоятельно произвести поиск данных о застрахованном в списках своей компании и идентифицировать его. При отрицательном ответе официально уведомить ЛПУ об отсутствии данных о данном пациенте. После получения указанного ответа от всех СМО необходимо предоставить ЛПУ право выставления случая лечения указанного пациента на территориальный фонд ОМС (ТФОМС). Следует указать, что, в настоящее время, счета за лечение пациентов без полиса не выставляются и оплата лечения таких пациентов не производится. Применение системы деперсонализации данных неизбежно приведет к увеличению таких неоплаченных случаев лечения, поэтому следует более точно сформулировать порядок оплаты именно этой категории пациентов.

Граждане, застрахованные на территории других субъектов РФ, отнесены в отдельную категорию т.к. информация об актуальности их полиса на конкретной территории отсутствует. Для осуществления межтерриториальных расчетов следует обратиться с предложением в Федеральный Фонд организовать единый сайт для проверки актуальности полиса по отрытым интернет-каналам. Также можно обязать все фонды субъектов федерации организовать сайты такого типа на каждой территории. Механизм функционирования проверки по таким сайтам аналогичен описанному выше. Если проверка не дает положительного результата, то оплатае такого пациента осуществляется как пациента без полиса. Передача в электронном виде счетов содержащих ФИО для проведения межтерриториальных расчетов может производиться либо в зашифрованном виде (при этом можно применять только сертифицированные ФСБ системы шифрования) либо передавать в электронном виде данные фактически деперсонализированные (с использованием ХЭШ-функции, см. ниже). При этом в бумажном виде передается информация об указанном пациенте в развернутом виде, что и позволяет проводить окончательную его идентификацию для последующей передаче информации в другие территориальные фонды. При этом можно вообще не производить шифрование данных об этой категории пациентов. Или производить ее минимально. В рамках отчетного периода (календарный месяц) количество лиц подпадающих под описанную категорию не превышает 500 человек (для поликлиники с численностью прикрепленного населения более 250000 человек). Указанная численность подпадает под третью, категорию секретности (в соответствии с приказом ФСТЭК и ФСБ № 55/86/20 от 13.02.2008г.), что значительно сокращает расходы на соблюдение режима секретности.

Лица, проживающие на территории иностранных государств. В настоящее время счета за таких пациентов ЛПУ практически не выставляются. Однако с рядом государств заключены договора о возможности оказания медицинской помощи гражданам. Информационный

обмен с такими государствами должен подчиняться оговоренным правилам. При их отсутствии можно применять к таким пациентам правила работы с иногородними гражданами.

Третье: Выставление счетов за пролеченных больных в системе добровольного медицинского страхования (ДМС) и платных больных. При лечении больных, застрахованных в системе ДМС, необходимо в качестве идентификатора пациента использовать реквизиты полиса. При этом автоматически осуществляется полная деперсонализация передаваемой информации. При учете платных больных целесообразно использовать внутренний код, присвоенный пациенту в лечебном учреждении и внесенные в те или иные бумажные документы (например - в «историю болезни»).

Четвертое. Учет граждан имеющих право на получение бесплатных медикаментов. Наиболее проблемная категория граждан. Существующая система прямо ориентирует задействованные в системе ЛПУ и АУ указывать персональные данные при осуществлении электронного документооборота в системе. При этом цепочка прохождения информации следующая Пенсионный фонд-Министерство здравоохранения территории – уполномоченная фарморганизация – ЦОД (центр обработки данных) – АУ – ЛПУ. Данные циркулируют в прямом и обратном направлении. Наличие централизованно распространяемых баз льготников существенно сокращает время ввода информации, и существенно снижает возможности при деперсонализации. Предлагается на уровне Министерства здравоохранения территории производить автоматическую деперсонализацию передаваемых баз по следующему алгоритму: персональные данные (ФИО пациента) заменить информацией полученной с использованием соответствующей ХЭШ – функции (см ниже). Дополнить данные полем контрольной суммой, заполняемой на основании вычисления ХЭШ – функции и значения поля СНИЛС. Справочник льготных категорий граждан должен передаваться в фарм-организацию с заполненными значениями ХЭШ – функции, СНИЛС и контрольной суммы, но без данных о ФИО. Вся обработка проводится при отсутствии информации о персональных данных пациента. Непосредственно персональные данные могут быть получены только на уровне Министерства здравоохранения при осуществлении обратной замены. Если при этом будет принято решения о допустимости выписки обезличенных рецептов, то можно полностью уйти от персонализации данных в этом случае. Если лекарственное средство отпускается гражданину, информация о котором отсутствует в БД, то работником АУ данные вводятся полностью, после чего производится их псевдо-шифрация и кодирования в виде ХЭШ – функции, что позволит в дальнейшем идентифицировать пациента.

Введение уникального идентификатора пациента.

Кардинальным решением вопроса о деперсонализации данных в системе ОМС и органах управления здравоохранением может быть ввод уникального идентификатора человека на территории РФ или соответствующего субъекта РФ. Прообразы такого идентификатора уже имеются: СНИЛС, ИНН и т.д. Однако сложность ведения БД силами стороннего ведомства (пенсионный фонд, налоговые органы) не позволяют признать указанные идентификаторы в качестве надежной альтернативы персональным данным. Целесообразно разработать собственную систему кодирования в рамках ведения реестра застрахованных. Реквизиты страхового полиса при этом использоваться не могут т.к. полис у человека может меняться, а идентификатор должен обеспечить гарантированную уникальность на протяжении всей жизни пациента. Поэтому в качестве уникального идентификатора целесообразно применение ХЭШ – функции над следующими данными (ФИО + дата рождения + место рождения) что практически гарантирует уникальность кода (на 1000000 человек не более 1 повторения и им можно пренебречь). После внедрения указанного метода, прочие методы станут не актуальными.

Решение собственных проблем лечебного учреждения

Существуют ряд задач, связанных с персональными данными, которые решаются силами лечебно-профилактического учреждения в собственных интересах: отслеживание процесса лечения пациента, расчет экономической эффективности его лечения, ведение истории болезни и т.д. Этих проблемы связаны с наличием инструментария, позволяющего накапливать данные, о том или ином пациенте. Именно они называются персональными и именно для защиты такой информации разработана соответствующая законодательная база. Однако существует метод позволяющий решить указанную проблему. При этом в регистратуре поликлиник и приемниках стационаров необходимо присваивать пациенту идентификатор позволяющий обеспечивать уникальность пациента. Данный идентификатор необходимо заносить в историю болезни пациента и (или) в специализированные бумажные документы с обеспечением хранения указанных документов в соответствии с действующим законодательством. Одновременно в историю болезни вносится (распечатывается) результат выполнения ХЭШ-функции над следующими данными: ФИО + дата рождения + место рождения (см ниже). Это обеспечит возможность последующего поиска и сбора персональной информации на конкретного пациента без указания персональных данных. Однако данная система должна предусматривать высокую ответственность оператора при первоначальном вводе информации т.к. ошибка даже в одном символе в поле ФИО изменит результат выполнения ХЭШ-функции и сделает невозможным интеграцию данных из различных историй болезни (например при многократной госпитализации пациента в один и тот же стационар в различные отделения или со значительным временным интервалом) .

Использование псевдо-шифрации пациента

Еще одним методом минимизации количества хранения персональных данных и их деперсонализации является метод, который мы назвали методом псевдо-шифрации пациента. Основа этого метода обрезание нескольких символов в реквизитах ФИО, что приводит к фактическому исчезновению персональных данных, позволяя достаточно точно идентифицировать человека. Следует разделить шифрацию на три различных уровня:

1. Низкий. При этом из фамилии имени и отчества пациента, в процессе ввода данных в информационную систему, удаляется по одному символу (байту) и дальнейшая обработка ведется в обычном режиме. Такой метод шифрации практически не препятствует функционированию информационной системы в полном объеме, однако по формальным признакам можно утверждать, что работа ведется с деперсонализированными данными, т.к. людей указанных в таких электронных документах фактически не существует.
2. Средний. При этом из фамилии имени и отчества пациента в процессе ввода данных в информационную систему, удаляется от 40-50% символов и дальнейшая обработка предполагает определенные интеллектуальные усилия пользователей информационной системы. Однако это позволяет вести обработку большинства персональных данных и обеспечивает достаточно полную идентификацию человека.
3. Высокий. При этом от фамилии имени и отчества пациента, в процессе ввода данных в информационную систему, остаются только инициалы. Данные при этом являются фактически полностью деперсонализированными, но можно осуществлять достаточно значительный контроль принадлежности тех или иных данных конкретному пациенту.

Этот метод можно отнести к методу по использованию некоторой специальной разновидности ХЭШ-функции.

Использование ХЭШ-функций

Хэш-функция (функция свёртки) - это функция, отображающая аргумент произвольной конечной длины в образ фиксированной длины. Результат работы данной функции называют хэшем, хэш-кодом или дайджестом сообщения (англ. message digest). Если хэш-функция зависит от секретного ключа, то она называется ключевой, в противном случае бесключевой. Простым примером хеширования может служить нахождение контрольной суммы сообщения: сумма кодов всех входящих в него символов, от которой берётся несколько последних цифр. Полученное число является примером хэш-кода исходного сообщения. Применение хеширования:

1. Проверка на наличие ошибок. Например, контрольная сумма может быть передана по каналу связи вместе с основным текстом. На приёмном конце, контрольная сумма может быть рассчитана заново и её можно сравнить с переданным значением. Если будет обнаружено расхождение, то это значит, что при передаче возникли искажения.
2. Проверка пароля. В большинстве случаев пароли фразы не хранятся на целевых объектах, хранятся лишь их хэш-значения. В нашем случае пароль и есть персональные данные. Хранить пароли в открытом виде нецелесообразно, так как в случае несанкционированного доступа злоумышленник узнает все пароли и сразу сможет ими воспользоваться, а при хранении хэш-значений он узнает лишь хэш-значения. В ходе процедуры аутентификации вычисляется хэш-значение введённой парольной фразы, и сравнивается с сохранённым.
3. Ускорение поиска данных. Например, при записи текстовых полей в базе данных может рассчитываться их хэш-код и данные могут помещаться в раздел, соответствующий этому хэш-коду. Тогда при поиске данных надо будет сперва вычислить хеш-код текста и сразу станет известно, в каком разделе их надо искать, то есть, искать надо будет не по всей базе, а только по одному её разделу.

В настоящее время существует три способа построения хэш-функции:
на основе трудно решаемой математической задачи;
на основе алгоритмов блочного шифрования;
разработанные с нуля.
Каждый из перечисленных выше методов обладает своими достоинствами и недостатками, однако наиболее распространенными на сегодняшний день оказались два последних метода. Это связано с тем, что при построении хэш-функции с нуля появляется возможность учета эффективности программной реализации.В свою очередь, широкое использование кэш-функций, построенных на основе алгоритмов блочного шифрования, является результатом тщательной проработки вопроса стойкости многих из существующих блочных алгоритмов.

Юридические проблемы обработки персональных данных

Необходимо доработать нормативную базу, чтобы каждый гражданин, обратившийся в СМО, ЛПУ или АУ при начале обслуживания давал согласие на передачу своих персональных данных всем субъектам систему ОМС, ДЛО и здравоохранения. Формулировки должны подтверждать согласие гражданина на максимально возможную обработку и передачу информации в рамках действующего законодательства. Согласие должно обеспечивать возможность использование персональных данных в объемах обеспечивающих бесперебойный информационный обмен между всеми субъектами в объеме предусмотренными существующей нормативной базой.

Проблемы обработки персональных данных в СМО

Одна из наиболее сложных в автоматизации проблем. Полная деперсонализации информации невозможна т.к. необходимо проводить учет застрахованного контингента и передача этой информации в ТФОМС. Целесообразным представляется выделить и защитить компьютеры, на которых производится учет застрахованных, а в другие информационные системы (контроль качества лечения, углубленная экспертиза и т.д.) предоставлять информацию только о реквизитах полиса без персональной информации. Таким образом, объем циркулирующей в системе персональной информации будет существенно сокращен.

Проблемы обработки персональных данных в органах управления здравоохранением. В данном случае может вообще отсутствовать потребность передачи персональных данных т.к. эти органы заинтересованы в передаче либо статистической информации (которая вполне может быть деперсонализированна) или информации о гражданах, которым отпущены льготные (бесплатные) рецепты. В этом случае в системе могут использоваться данные документа, на основании которого отпуск бесплатного (льготного) ЛС был осуществлен.

Проблемы обработки персональных данных в ТФОМС. Аналогично можно производить обработку данных в системе ОМС. Практически возможно производить деперсонализированную обработку информации для всех категории застрахованных на конкретной территории. Исключение необходимо предусмотреть только для иногороднего контингента и базы данных застрахованного контингента. При обработке сведений об этой категории граждан следует предусмотреть защитные меры. Хотя система ХЭШ-функций и псевдо-шифрации позволяет минимизировать обрабатываемые персональные данные.

Полное или частичное применение описанных методов позволит провести деперсонализацию медицинской информации и значительно сократит расходы ЛПУ и АУ на соблюдение условий конфиденциальности.