ISSN 2079-3316 ПРОГРАММНЫЕ СИСТЕМЫ: ТЕОРИЯ И ПРИЛОЖЕНИЯ № 2(20), 2014, с.3–12

УДК 004.043

С. Г. Юрченко

Визуализация электронных клинических документов с учетом требований защиты персональных данных и данных с ограниченным допуском

Аннотация. В работе рассматривается проблема обезличивания персональных данных в электронных клинических документах, при условии соблюдения неизменности уже подписанного документа. Рассмотренные методы обезличивания основываются на включении в структуру документа специальной семантической разметки, обрабатываемой динамически при визуализации документа.

Ключевые слова и фразы: персональные данные, электронный документ, медицинская информационная система.

Введение

Согласно федеральному закону РФ №152-ФЗ [1] медицинские информационные системы (относящиеся к информационным системам персональных данных, ИСПД) должны обеспечивать конфиденциальность персональных данных пациентов, записи о которых имеются в МИС. Выбор средств информации для системы защиты персональных данных (согласно постановлению Правительства РФ №1119 [2]) «осуществляется оператором в соответствии с нормативными правовыми актами, принятыми Федеральной службой безопасности Российской Федерации и Федеральной службой по техническому и экспортному контролю».

В [1] оговаривается, что обеспечение конфиденциальности персональных данных не требуется в случае их обезличивания. Под обезличиванием понимаются «действия, в результате которых невозможно определить принадлежность персональных данных конкретному субъекту персональных данных». При этом до сих пор не

принято нормативных правовых актов, упомянутых в [2], которые регулировали бы процесс обезличивания персональных данных, по сути оставляя это на усмотрение разработчиков МИС.

В «Методических рекомендациях для организации защиты информации при обработке персональных данных в учреждениях здравоохранения, социальной сферы, труда и занятости» [3] для снижения класса ИСПД рекомендуется использовать сегментацию и обезличивание, то есть разделение данных, с тем, чтобы даже имея доступ к информации о процессе лечения пациента, нельзя было понять, к какому пациенту она относится, не имея доступа к более защищенной части МИС либо отдельной ИС, содержащей персональные данные.

Следует отметить, что в приказе ФСТЭК №21 [4] перечисляются организационные и технические меры по обеспечению безопасности персональных данных при их обработке в ИСПД, но обезличивание данных не упоминается.

Стандарт ГОСТ 52636-2006 [5] требует обеспечивать для электронной персональной медицинской записи (ЭПМЗ) «неизменность и достоверность на протяжении всего периода хранения». Что подразумевается под «неизменностью» в общем случае, не разъясняется, однако описывается в случае использования электронной цифровой подписи — «дайджест должен быть получен хэшированием всего содержимого ЭПМЗ, включая все прикрепленные файлы и все элементы формализованных данных, с тем чтобы ни один из этих элементов нельзя было изменить, не нарушив целостности ЭЦП». Стандарт не упоминает в рекомендуемой структуре ЭПМЗ в явном виде персональные данные, требуя указания лишь идентификатора пациента и необязательного номера истории болезни либо амбулаторной карты, которые сами по себе не позволяют идентифицировать пациента, не имея доступа к справочнику пациентов в МИС.

Следует отметить, что единых стандартов бланков медицинских документов не существует. Приказ Минздрава СССР от 04.10.1980 №1030 «Об утверждении форм первичной медицинской

документации учреждений здравоохранения» был отменен, и нового набора образцов бланков разработано не было, а использование старых образцов из данного приказа, письмом Департамента организации медицинской помощи и развития здравоохранения Минздравсоцразвития России от 30.11.2009 № 14-6/ 242888, разрешено, но не обязательно.

1. Обезличивание документов

Де факто, большинство медицинских документов, относящихся к пациенту, включают персональные данные, такие как ФИО и дату рождения, являясь, по сути, неотъемлемой частью ЭПМЗ.

Таким образом, с одной стороны, рекомендуется разделить персональные данные, позволяющие идентифицировать пациента, и медицинские данные о нём, созданные в рамках МИС. Но, с другой стороны, разделить их зачастую не представляется возможным без нарушения неизменности ЭПМЗ.

Пока не были приняты нормативные акты, регулирующие процесс обезличивания персональных данных, разработчикам МИС приходится разрабатывать собственные решения. Например, в «Руководстве по защите конфиденциальности персональной информации», разработанном Национальным институтом стандартов и технологии (США) [6], описываются следующие методы обезличивания:

- обобщение данных, делающее их менее точными;
- удаление данных всей записи, содержащей сведения о персоне, или ее частей;
- внесение «шума» в данные;
- замена данных аналогичными полями из других записей;
- замена данных усредненным значением.

Однако, даже выбрав надежный способ обезличивания данных, мы не решим этим проблему одновременного обеспечения возможности проверки аутентичности обезличенного документа первоначально созданному (возможно, подписанному электронной подписью).

1.1. Обезличивание различных форматов документов

Решением проблемы видится внесение специальной семантической разметки в документ на этапе подписания (когда документ приобретает свой окончательный вид) и использования для работы с документами в дальнейшем специального программного обеспечения, которое умеет работать с этой разметкой.

Рассмотрим несколько стандартов, которые могут использоваться для создания и хранения медицинских документов: XML, HTML, PDF, текстовый файл, Microsoft Word.

1.1.1. Формат XML

В случае использования для хранения содержимого документа формата XML решение проблемы достаточно очевидно. Важно лишь выделить в структуре документа поля, для которых потребуется обезличивание, и добавить для них какой-нибудь специальный атрибут, а затем учитывать его при визуализации документа (с помощью языка XSLT или специализированного ПО).

1.1.2. Формат HTML

В случае использования HTML решение, аналогичное XML. Общепринятой практикой для внесения такого рода разметки являются «микроформаты» — способ семантической разметки сведений о разнообразных сущностях (событиях, организациях, людях, товарах и так далее) на веб-страницах с использованием стандартных элементов языка HTML (или XHTML). Пользователь-человек может воспринимать страницу с микроформатом как обычную веб-страницу (через браузер), тогда как программы-обработчики способны извлечь из такой страницы структурированную информацию, следуя определённым соглашениям [7]. Обычно для этого используется атрибут class, определенный для всех элементов HTML.

Например, используя микроформат hCard (позволяющий описывать информацию о людях, компаниях и организациях), документ может содержать следующую разметку:

```
<div class="vcard">
  <div>ФИО пациента:
    <span class="fn">Лукашин Евгений Михайлович</span>
  </div>
  <div>Дата рождения:
    <span class="bday">1939-01-01</span>
  </div>
  <div>Адрес проживания:
    <span class="adr">
      <span class="locality">Москва</span>,
      <span class="street-address">3-я улица Строителей, дом 25, кв.
12</span>
    </span>
  </div>
</div>
```

Близким аналогом микроформатов является стандарт HTML Microdata, разрабатываемый Веб-консорциумом W3C, однако, на момент написания статьи, он ещё не имел статуса «рекомендации», а лишь обсуждался в специальной рабочей группе [8].

Необходимо отметить, что документы в формате HTML могут быть открыты средствами Microsoft Office, а также, при соблюдении правильной структуры HTML, могут быть программным образом преобразованы в файлы PDF. Это позволяет обойти проблемы, присущие формату HTML, например, в вопросе гибкости настроек печати документов (нумерация страниц, задание размеров полей для четных и нечетных страниц и пр.).

1.1.3. Формат PDF

В случае использования формата PDF возможность вставлять невидимую разметку отсутствует, поэтому потребуется либо сгенерировать помимо обычной версии документа сразу же и обезличенную (которую можно рассматривать как прикрепленный файл к основному документу — п. 7.1.8 [5]), либо генерировать PDF «на лету», на основе какого-то иного формата, в котором присутствует возможность использования разметки, и который можно обезличить перед преобразованием в PDF.

1.1.4. Формат Microsoft Word

В случае использования формата документа Microsoft Word существует два варианта этого формата — бинарный (файлы с расширением .DOC) и Office Open XML (файлы с расширением .docx).

В первом случае использовать дополнительную разметку не получится. Поэтому возможно лишь использование поиска и замены конкретных слов или словосочетаний. Например, искать все вхождения словосочетания «Лукашин Евгений Михайлович» и заменить их на иной текст, например, «Фамилия1 Имя1 Отчество1». Либо найти первую по счету строку, начинающуюся на «ФИО пациента:» и заменить следующие за найденным текстом три слова или весь текст до конца строки (так как полное имя пациента может состоять из более чем 3 слов) на что-то другое. Очевидно, что такой вариант очень неудобен и не даёт полной гарантии того, что в документе не останется необезличенных данных о пациенте, по которым его удастся идентифицировать. Поэтому, например, в электронных системах обезличивания юридических документов, используемых в судах, после подобной машинной обработки обязательно требуется проверка документа вручную оператором (человеком) на предмет оставшихся необработанными или неправильно обработанных данных. Лишь после его подтверждения документ считается полностью обезличенным.

Во втором случае документ представляет собой ZIP-архив с набор файлов, в частности XML-документом, в котором и содержится набранный текст. Поэтому, в случае, если в бланк документа предварительно внести специальную разметку, появится возможность его машинной обработки с достаточной точностью. Естественно, при условии, что пользователь никак не будет влиять на эту разметку (заголовки полей), а будет лишь вводить текст в поля для ввода информации.

1.2. Этапы работы с документом

При создании документа в нём будет присутствовать невидимая для пользователя разметка данных, которые требуют обезличивания.

По окончании работы с документом (по прежнему включающим разметку обезличиваемых данных) он подписывается электронной подписью автора, что гарантирует его неизменность в дальнейшем. Защита персональных данных, содержащихся в необезличенном документе, осуществляется средствами МИС.

Обезличивание документа происходит внутри МИС перед выводом документа на просмотр или перед его выгрузкой в другие МИС. В обезличенном документе персональных данных, требующих защиты, не остается ни в явном, ни в неявном (недекларируемые возможности) виде. На обезличенный документ требование неизменности и достоверности не распространяется.

При необходимости отображения документа в рамках МИС проверяется наличие прав у пользователя, и, в зависимости от этого, ему показывается оригинал документа, либо его обезличенная версия.

При выгрузке документов в другие МИС осуществляется предварительное обезличивание документа.

1.3. Дополнительные возможности при использовании разметки в HTML

Помимо использования разметки для обезличивания HTML документов, ее можно использовать и для иных целей:

- извлечение фрагментов текста для последующего использования в МИС;
- показ частей документов лишь при наличии соответствующих прав у пользователя;
- различные варианты визуального представления одного и того же документа.

1.3.1. Извлечение фрагментов текста

Использование специальной разметки в HTML-документах позволяет легко извлекать содержимое отдельных полей либо целых разделов документов. Для чего это может понадобиться? Например, в некоторых видах осмотров пациента врачом-специалистом присутствует поле «Жалобы на момент поступления», в которое, очевидно, должны попадать данные из осмотра в приемном отделении. Для того, чтобы пользователю не приходилось копировать данные вручную из одного документа в другой, содержимое поля «жалобы» можно выгрузить при подписании осмотра в ПО и затем вставить его во вновь созданные документ.

1.3.2. Права на просмотр содержимого документа

Определенные поля или разделы документа могут содержать информацию, которую должны иметь возможность видеть лишь пользователи, обладающими соответствующими правами (привилегиями), скажем, лишь лечащий врач и заведующий отделением. При использовании специальной разметки в HTML-документе соответствующие фрагменты документа могут не отображаться при работе с ним пользователя.

1.3.3. Варианты визуального представления документа

В зависимости от назначения документа или иных условий может понадобиться скрывать или показывать некоторые поля или разделы документа. Например:

- некоторая информация может показываться лишь при редактировании документа, но не будет попадать в подписанный документ;
- поля или разделы документа, которые не видны на экране, но видны при печати (в бумажной копии), например, на основе выписного эпикриза может печататься выписка, которая выдается на руки пациенту, где будет присутствовать эмблема и контактная информация медицинского учреждения;
- в зависимости от некоторых условий поля или разделы могут вообще не показываться, например, очевидно, что гинекологи-

ческий раздел при осмотре пациента-мужчины не имеет никакого смысла, однако создавать два разных варианта осмотра для мужчин и женщин видится излишним.

Заключение

В работе описаны методы решения задачи обезличивания электронных клинических документов, при условии соблюдения ФЗ №152[1] и соответствия информационной системы стандарту [5]. Предложен легкореализуемый подход к обезличиванию персональных данных, не требующий ручной работы оператора, при использовании XML и HTML-форматов документов. Выбранный для хранения документов формат HTML позволяет программными средствами преобразовывать его в широко распространенные форматы PDF и Microsoft Word. Также рассмотрены преимущества данного подхода для решения иных задач, связанных с электронными клиническими документами.

Список литературы

[1] Федеральный закон Российской Федерации от 27 июля 2006 г. N 152-ФЗ «О персональных данных»

[2] Постановление Правительства Российской Федерации от 1 ноября 2012 г. N 1119 г. Москва «Об утверждении требований к защите персональных данных при их обработке в информационных системах персональных данных»

[3] «Методические рекомендации для организации защиты информации при обработке персональных данных в учреждениях здравоохранения, социальной сферы, труда и занятости» (Министерство здравоохранения и социального развития Российской Федерации, 2009 г.)

[4] Приказ Федеральной службы по техническому и экспортному контролю (ФСТЭК России) от 18 февраля 2013 г. N 21 г. Москва «Об утверждении Состава и содержания организационных и технических мер по обеспечению безопасности персональных данных при их обработке в информационных системах персональных данных»

[5] ГОСТ 52636-2006 «ЭЛЕКТРОННАЯ ИСТОРИЯ БОЛЕЗНИ. Общие положения»National Institute of Standards and Technology. SP 800-122. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). URL: http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf
[6] Микроформат — Википедия. URL: https://ru.wikipedia.org/wiki/Микроформат
[7] HTML Microdata — World Wide Web Consortium. URL: http://www.w3.org/TR/microdata/

Рекомендовал к публикации *к.т.н. Я.И. Гулиев*

Об авторе:

Сергей Геннадьевич Юрченко
Руководитель группы разработки в Интерин Технологии, м.н.с. Института Программных Систем им. А.К. Айламазяна РАН.
e-mail: yurch@interin.ru

Образец ссылки на публикацию:
С. Г. Юрченко. Визуализация электронных клинических документов с учетом требований защиты персональных данных и данных с ограниченным допуском // Программные системы: теория и приложения: электрон. научн. журн. 2013. Т. 4, № 3(17), с. 3–12. URL: http://psta.psiras.ru/read/psta2014_2_3-12.pdf

S. G. Yurchenko. Visualization of electronic clinical documents, adjusted for protection of personal data and other restricted access information.

Abstract. The work considers the problem of depersonification of electronic clinical documents, while keeping in mind the constancy of signed document contents. The author considers the method of depersonification by including special semantic markup in document structure, which is processed during the visualization.

Key Words and Phrases: personal data, electronic document, medical information system.