

Алгоритм деперсонализации данных при информационном обмене в цифровой экономике*

А.В. СОЛОВЬЕВ^I, Н.Б. БАКАНОВА^{II}

^I Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия

^{II} Федеральный Исследовательский Центр Институт Прикладной Математики им. М.В. Келдыша Российской Академии Наук, г. Москва, Россия

Аннотация. В данной статье описаны проблема деперсонализации цифровых данных при информационном обмене в цифровой экономике, а также проблема моделирования состава информации цифровых данных. Приведен краткий обзор проблем и предложена формальная постановка задачи деперсонализации цифровых данных. Под деперсонализацией данных понимается представление персональных и/или конфиденциальных данных в виде, не позволяющем восстановить какую-либо информацию о субъекте персональных данных или о защищаемой конфиденциальной информации. В качестве решения поставленной задачи предложены алгоритм деперсонализации и математическая модель состава информации цифровых данных. Кратко описаны возможные области применения решения задачи деперсонализации.

Ключевые слова: цифровая экономика, долговременная сохранность, большие данные, распределенные реестры, деперсонализация.

DOI: 10.14357/20790279200202

Введение

В ряде статей автора исследования [1, 2] было показано, что в условиях цифровой экономики, когда «данные в цифровой форме» становятся «ключевым фактором производства во всех сферах социально-экономической деятельности» [3], задача обеспечения сохранности цифровых данных становится крайне важной.

Кроме того, при взаимодействии между хозяйствующими субъектами цифровой экономики, а также гражданами, вовлеченными в бизнес-процессы, возникает задача защиты персональных данных от утечки. Данная проблема была обозначена в работе [4] как проблема «деперсонализации» данных при информационном обмене.

В рамках данного исследования под деперсонализацией данных будем понимать представление персональных и/или конфиденциальных данных в виде, не позволяющем восстановить какую-либо информацию о субъекте персональных данных или о защищаемой конфиденциальной

информации¹. Остальные определения, такие как долговременная сохранность, цифровая экономика и другие, используемые в данной статье, даны в [1, 2, 4].

В данной статье представлен алгоритм решения задачи «деперсонализации» цифровых данных при передаче, представлена математическая модель цифровых данных.

1. Краткий обзор проблемы

В процессе информационного обмена цифровыми данными между хозяйствующими субъектами цифровой экономики, гражданами, вовлеченными в деловые процессы цифровой экономики, с высокой вероятностью возникает необходимость разграничения прав доступа на передаваемую информацию (цифровые данные). К этой информации могут быть отнесены как персональные данные, так и коммерческая или государственная тайна.

Например, при передаче данных цифрового профиля гражданина для различных ведомств, не

* Работа выполнена при частичной финансовой поддержке РФФИ в рамках научных проектов № 18-29-03070 и № 18-29-03085.

¹ Определение дано авторами по №152 ФЗ «О персональных данных» от 26.07.2006 г.

должна быть передана, в зависимости от конкретного бизнес-процесса, информация о его лицевых счетах, информация медицинской карты и др. При передаче информации между хозяйствующими субъектами не должна быть передана информация, составляющая банковскую, коммерческую и государственную тайну. В последнем случае, конечно, существуют различные средства, включая мандатный доступ, создание нескольких независимых контуров в информационной системе с разными правилами доступа [5, 6]. Однако, в условиях огромного роста цифровых данных увеличивается вероятность утечки данных вследствие человеческого фактора.

Тем самым, возникает задача деперсонализации цифровых данных при информационном обмене, т.е. необходимость реализации в рамках функционала информационных систем некоторой функции $f: d_i \rightarrow d'_i$, где $d_i \in D$ – исходные цифровые данные, созданные в рамках некоторой информационной системы, $d'_i \in D$ – «деперсонализированные» цифровые данные, D множество цифровых данных, существующих в рамках информационного обмена. Тогда должно существовать множество требований $T = \{NTD_i\}$, определяемых в нормативно-технической документации (НТД), которая составляет основу политик информационной безопасности, хранения и обработки персональных данных и т.д. Тем самым, правила выполнения f должны определяться в каждом конкретном случае множеством соответствующих T .

Еще одной проблемой является определить модель информационного состава d_p, d'_i , а также алгоритм работы функции f .

Попытки разработать математические модели состава информации для цифровых данных длительного хранения активно предпринимаются в настоящее время.

Здесь наибольший интерес с точки зрения организации длительного хранения большого объема цифровых данных представляет опыт **National Archives and Records Administration (NARA)**. NARA в настоящее время является по факту крупнейшим в мире хранилищем данных в цифровой форме [7]. Если говорить об объемах данных, то можно привести следующую иллюстрацию: Пентагон и Госдепартамент США ежегодно передают в NARA десятки миллионы электронных документов для длительного хранения [8]. Речь о «деперсонализации» при вводе не идет, но используется в ручном режиме при рассекречивании документов.

Что касается модели цифровых данных, то, на основании анализа открытых источников, можно

утверждать, что разработчики NARA считают документом любую информацию в цифровом виде, в том числе включающую и программно-аппаратную реализацию [9]. Это приводит к тому, что в архив помещаются любые такие документы от файла до сервера (включая всю аппаратную часть, а также средства декодирования и отображения документа). При этом упор с самого начала сделан на создание распределенных средств хранения с многократным дублированием [10]. Данный подход приводит к тому, что информация надежно сохраняется, однако возникают проблемы с чтением документов (раскодированием форматов, интерпретацией) спустя многие годы [11]. В связи с чем архив работает в настоящее время в ограниченном режиме [9]. Из-за проблем с интерпретацией данных на длительном сроке хранения прослеживается четкая тенденция к сокращению набора принимаемых форматов данных [12]. Кроме того, можно отметить, что NARA не использует средства криптозащиты, делая упор на многократное резервирование данных. Модель документа включает полнотекстовый индекс, используемый для быстрого поиска [13], который можно отнести к метаданным документа, собственно метаданные [14]. Многие наработки по моделированию документов при долговременном хранении стандартизируются [7, 10]: 1) библиографический формат MARC, 2) формат хранения на оптических дисках ODISS, 3) формат архивных описаний GSA6710A, 4) формат хранения на магнитных лентах в кодировках ASCII и EBCDIC, 5) открытый стандарт кодированного архивного описания (EAD – Encoded Archival Description). Кроме того, разработаны: 1) бюллетень по допустимым файловым форматам передаваемых на архивное хранение [12], 2) требования для написания технических заданий для инструментов или сервисов управления цифровыми данными [15].

В математических моделях цифровых данных, описанных разработчиками электронных архивов Австралии и Новой Зеландии, в отличие от NARA, добавляются метаданные документа (сведения о создании, об авторстве, о системе классификации документа). Исключается из модели аппаратная среда. Однако, метаданные четко не описаны. Моделирование электронного документа в архивах Новой Зеландии включает в себя также необходимость миграции данных [16].

Разработчики архивов Великобритании определяют состав метаданных электронного документа, которые включаются в модель цифровых данных [17]. В модель также включаются: трехуровневая схема классификации документов

(класс-папка(дело)-том), сведения о пользователях (авторе и менявших документ), сведения о датах создания и изменения документов.

Разработчики архивов Германии и Республики Корея делают упор при моделировании документа на обеспечение неизменности и подтверждения авторства документа с использованием средств криптозащиты информации. В модель цифровых данных включается электронная подпись (ЭП), однако в модели отсутствуют метаданные и аппаратная составляющая [18, 19].

Разработчики Дании, Швейцарии при моделировании цифровых данных дополняют модель оригинала данных ее нормализованной цифровой копией (используются форматы PDF/A, TIFF, JPEG2000) [20]. В Дании модель дополняется еще и метаданными стандарта «дублинское ядро» (Dublin Core). Модель данных при долговременном хранении включает нормализованную копию (формат PDF) и метаданные (формат XML) [21].

Из приведенного краткого обзора можно сделать следующий вывод: проблема деперсонализации данных и моделирования состава информации цифровых данных являются актуальными, отсутствует решение данных проблем при одновременном обеспечении сохранности цифровых данных.

2. Постановка задачи деперсонализации цифровых данных

Математическая постановка задачи «деперсонализации» цифровых данных при информационном обмене данными в рамках цифровой экономики может быть сформулирована в следующем виде:

Дано:

- 1) Множество цифровых данных $D = \{d_i\}$.
- 2) Множество требований безопасности информации $T = \{NTD_i\}$, определяемых в НТД.

Найти:

- 1) Алгоритм реализации отображения функции $f: d_i \rightarrow d'_i, d'_i \in D$, при котором исполняются требования T .
- 2) Математическую модель состава передаваемой информации $D = \{d_i\}$.

Общая схема информационного обмена цифровыми данными в рамках цифровой экономики приведена в [4]. Для иллюстрации данной схемы приведем рисунок 1.

Как видно из рисунка, в процесс передачи заложена функция «деперсонализации» f .

Информационная система (ИС) конкретной организации – субъекта цифровой эконо-

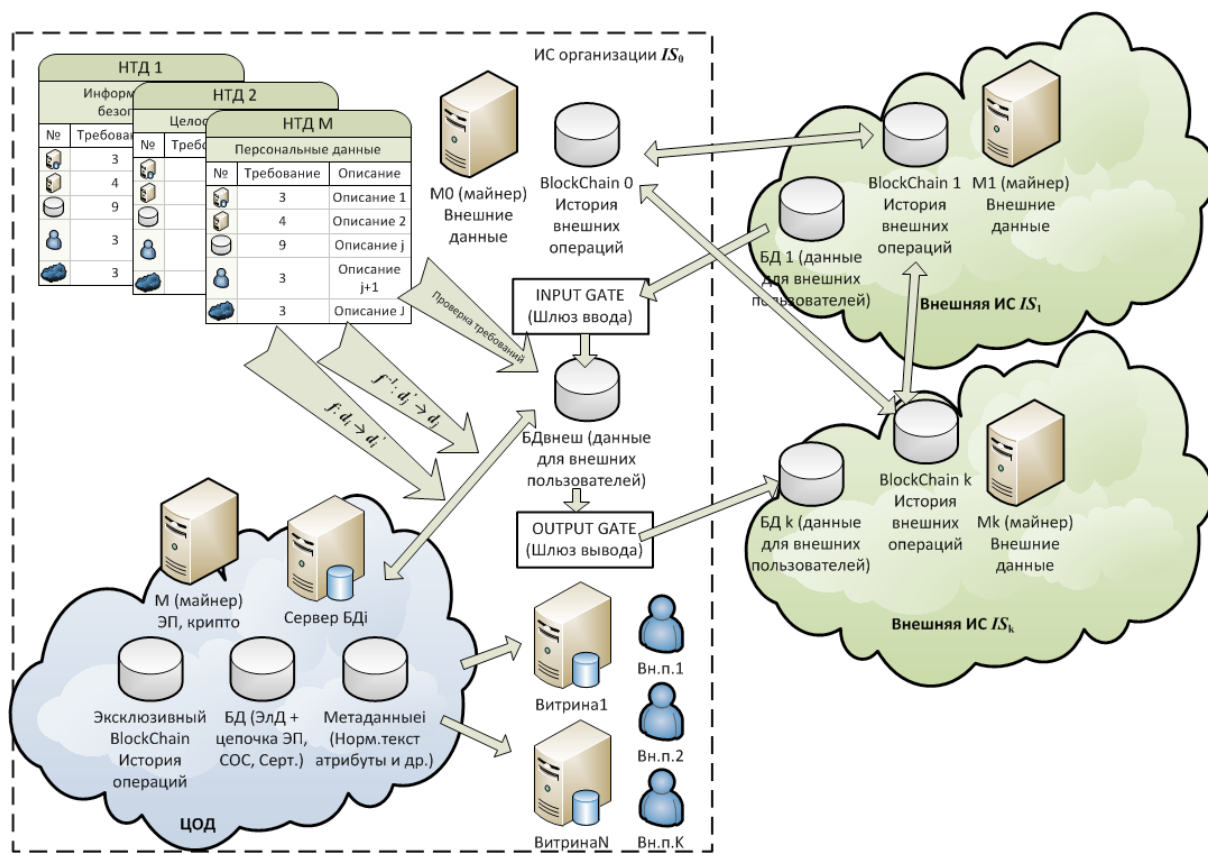


Рис. 1. Схема передачи цифровых данных между субъектами цифровой экономики

мики – должна реализовывать функцию «деперсонализации» данных $f: d_i \rightarrow d'_i$. Входами функции должны быть: цифровые данные d_i , требования защиты данных $T = \{NTD_i\}$, определяющие состав «закрытых» (персональных) данных в d_i . Для поиска персональных данных может применяться интеллектуальный анализ данных, анализ по ключевым словам, анализ состава данных и др. Кроме того, в структуре метаданных d_i должен быть предусмотрен атрибут «персональные данные», которым должны быть отмечены соответствующие данные. Данный атрибут должен устанавливаться как минимум при создании цифровых данных в исходной ИС, при изменении T , автоматически или автоматизированно.

Выходом функции «деперсонализации» данных f должны быть цифровые данные d'_i не содержащий персональных (закрытых) данных, возможно заверенные ЭП. Заверение ЭП необходимо для дальнейшего контроля аутентичности «деперсонализированных» данных d'_i . Хеш d'_i должен сохраняться во внешнем и приватном (эксклюзивном) блокчейнах (см. Рис. 1), возможно сохранение ЭП вместе со смарт-контрактом для проверки ЭП d'_i , для обеспечения возможности аудита в будущем. Также в блоках помещаются компоненты ЭП (сертификат открытого ключа, СОС, штамп времени).

3. Алгоритм деперсонализации данных

Исходя из общей постановки задачи «деперсонализации», алгоритм деперсонализации цифровых данных можно представить следующими шагами:

1. При создании шаблона для ввода цифровых данных d_i предусматриваются поля ввода, для которых устанавливается специальный атрибут «персональные (закрытые) данные».
2. При создании или вводе цифровых данных d_i в исходной ИС вся информация, вводимая в поля с установленным атрибутом «персональные (закрытые) данные» помечается атрибутом «закрытый».
3. При создании или вводе цифровых данных d_i в исходной ИС производится дополнительный автоматический анализ (может применяться интеллектуальный анализ данных, анализ по ключевым словам, анализ состава данных) вводимых данных на соответствие T .
4. Если в d_i выявлены данные, соответствующие требованиям T , то они отображаются для окончательного решения оператору ИС.
5. Оператор принимает решение по выявленным данным, после чего подтвержденные им данные помечаются атрибутом «закрытый».
6. Перед отправкой цифровых данных d_i во внешнюю ИС (другому субъекту цифровой экономики) производится дополнительный анализ данных d_i аналогично шагу 3 в случае, если множество T было изменено.
7. Из данных d_i формируются цифровые данные d'_i в которые не включаются все элементы данных, помеченные атрибутом «закрытый».
8. Непосредственно при передаче хеш d'_i сохраняется во внешнем (общем по отношению к принимающей и передающей ИС) блокчейне и в каждом приватном (эксклюзивном) блокчейнах принимающей и передающей цифровые данные ИС.

4. Математическая модель цифровых данных

Как было показано в [1], любые цифровые данные не существуют без определенной семантики. Это означает, что сами атомарные данные привязаны к некоторым метаданным, которые и определяют их смысл. Например, хранятся данные в виде «120». Что это такое? Расстояние? Время? Деньги? Когда же данные имеют вид «Скорость 120 километров в час», то не возникает сомнений что это такое. В этом случае ценность представляют также метаданные, которые и определяют семантику хранимых данных. Потеря метаданных критическим образом скажется на возможности использовать цифровых данных, даже надежно сохранив в течение десятилетий. Тогда цифровые данные представляют собой взаимосвязанные семантические блоки и взаимосвязи между ними в электронно-цифровой форме.

При этом семантические блоки можно определить как некоторые фрагменты цифровых данных, выделенные по смысловому содержанию.

Тогда в общем виде модель цифровых данных d_i можно представить графически в виде графа (или дерева, если до корня из любой листовой вершины имеется единственный путь), состоящего из взаимосвязанных семантических блоков B_j . Блоки в свою очередь представляют собой подграфы (поддерева), также состоящие из семантических блоков следующего уровня. Между d_i могут существовать различные отношения (связи), т.е. лес d_i может быть связан в единый граф $G(D)$.

В общем виде модель документа в ЭА в смысле представления необходимого состава информации определяется следующим образом:

$$d_i = \bigcup_{(j=1..N)} (B_j)$$

d_i разбивается на семантические блоки:

- если $B_k \cap B_j \neq \emptyset$, то $B_k \subseteq B_j$ или $B_j \subseteq B_k$;
- если $\exists B_j \subset B_k$, то $B_k = \bigcup B_j$ для всех $B_j \subset B_k$.

Теперь можно определить более конкретный состав B_j для d_i .

В общем случае модель цифровых данных можно представить следующим образом:

$$d_i = Md \cup Ord \cup ONd \cup FTIdx \cup CLIdx \cup LDIdx \cup Opd \cup DSign$$

где Md – множество семантических блоков метаданных d_i , в состав которых включается схема данных (расположение и подчиненность (иерархия) различных данных), возможно форма данных, сведения в формате дублированное ядро;

Ord – множество семантических блоков оригинальных данных d_i в исходном формате (электронный документ (оригинал) или оцифрованное изображение оригинального бумажного документа, которое далее также будем обозначать как оригинал), может включать ЭП, проставленные, например, в системе электронного документооборота исходной ИС;

ONd – множество семантических блоков нормализованной копии цифровых данных d_i (под нормализацией здесь понимается приведение к единому формату (набору форматов) долговременного хранения, например PDF/A, TIFF, PNG);

$FTIdx$ – множество семантических блоков, представляющих собой нормализованный текст или набор ключевых слов (тэгов) цифровых данных d_i (под нормализацией здесь понимается приведение всех слов в тексте к единственному числу именительного падежа и инфинитиву для глаголов), если подобная нормализация проводилась в исходной ИС;

$CLIdx$ – модель (множество семантических блоков) системы классификации цифровых данных d_i ($\langle CLIdx_1, \dots, CLIdx_k, \dots, CLIdx_K \rangle$ ($k=1, K$)). В простейшем случае представляет собой набор позиций классификаторов, с которыми связаны цифровые данные. В случае долговременного хранения данная часть документа является информацией о классифицировании и среде хранения (окружении) данных на момент их создания в исходной ИС;

$LDIdx$ – модель (множество семантических блоков) связей d_i с другими данными из множества D ;

Opd – модель (множество семантических блоков) журнала операций с цифровыми данными d_i .

$DSign$ – модель (множество семантических блоков) ЭП цифровых данных. Модель может содержать в себе сертификаты ЭП, цепочки сертификатов, сертификаты удостоверяющих центров (УЦ), списки отзыва сертификатов (СОС).

В контексте наличия в d_i закрытых (персональных) данных, пересылка блоков $LDIdx$, Opd , $DSign$ не имеет смысла, т.к. вряд ли привязана к цифровым данным ИС получателя. Вместо $DSign$ в исходной (передающей) ИС формируется новое множество $DSign1$, т.к. из-за того, что в общем случае $d'_i \neq d_i$ проверка $DSign$ в ИС принимающей цифровые данные будет невозможна.

Тогда математическая модель для d'_i будет выглядеть следующим образом:

$$d'_i = Md' \cup Ord' \cup ONd' \cup FTIdx' \cup CLIdx' \cup DSign1$$

где $Md' = Md - MdP$, MdP – семантические блоки Md , содержащие персональные данные;

$Ord' = Ord - OrdP$, $OrdP$ – семантические блоки Ord , содержащие персональные данные и т.д.

5. Возможное практическое применение

В результате решения задачи «деперсонализации» цифровых данных получен алгоритм и математическая модель общего вида исходных и передаваемых цифровых данных.

Рассмотренное решение задачи подходит для решения следующих проблем практической организации информационного обмена данными в цифровой экономике:

- Организация обмена данными между внутренними пользователями системы и внешними по отношению к ней пользователями и ИС.
- Подтверждение неизменности данных при обмене.
- Контроль процесса информационного обмена, отслеживание полной цепочки изменений данных.

Авторы считают, что показанное в статье решение подойдет для создания распределённых сервисов подтверждения квалификации, проверки качества дипломов о высшем образовании, лицензий на осуществление профессиональной деятельности и сроков их действия, HR-систем при проведении собеседований, учётных системы ВУЗ-ов и центров профессиональной подготовки, учётных систем государственных институтов и ведомств, занимающихся сертификацией и лицензированием разного рода деятельности

Другими областями применения является создание распределённых систем управления знаниями, применяемых в судебной практике, страховании и медицине.

В условиях развивающейся цифровой экономики могут быть и другие применения: передача юридически значимых документов, истории взаимодействия граждан с органами власти, наблюдательные и страховые дела и др.

Заключение

В данной статье описана постановка задачи «деперсонализации» цифровых данных при информационном обмене в цифровой экономике. Предложен алгоритм деперсонализации, а также математическая модель цифровых данных. Рассмотрены возможные перспективы применения.

В дальнейшем авторы планируют серию статей, посвященных практическому применению решений поставленной задачи, а также разработку детальных постановок для частных проблем и задач, возникающих при передаче цифровых данных.

Разработка комплексного подхода к обеспечению сохранности и защите персональных (конфиденциальных) цифровых данных при информационном обмене цифровой экономике позволит создавать распределённые программные решения, которые с одной стороны будут обеспечивать надёжность и неизменность данных, а с другой защиту цифровых данных при передаче.

Литература

1. Соловьев А.В., Баканова Н.Б. Проблемы долговременной сохранности больших данных // Информационные технологии и вычислительные системы, №2, М.:2019. С.44-53. DOI 10.14357/2071863219020.
2. Соловьев А.В. Обеспечение сохранности электронных документов при обмене данными в цифровой экономике // Труды XXV Международной научно-практической конференции «Документация в информационном обществе: задачи архивоведения и документоведения в условиях цифровой экономики» (Москва, Московский финансово-юридический университет (МФЮА), 7-8 ноября 2018 г.) – Росархив, ВНИИДАД, – М.:2019. С.90-98. ISBN 978-5-89789-147-4.
3. ПРОГРАММА «Цифровая экономика Российской Федерации». УТВЕРЖДЕНА распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р. М.:2017 – 88 с.
4. Соловьев А.В. Электронные документы и задача обеспечения сохранности при обмене данными в цифровой экономике / А.В. Соловьев, И.А. Тарханов // Труды Института системного анализа РАН (ИСА РАН), Том 68, выпуск 1, М.: 2018. С.42-53. DOI 10.14357/20790279180104.
5. Акимов Г.П., Даниленко А.Ю., Пашкин М.А., Папкина Е.В., Подрабинович А.А. Особенности использования электронной подписи в защищенных информационных системах // Проблемы информационной безопасности. Компьютерные системы. 2017. № 4. С. 95–101.
6. Даниленко А.Ю., Папкина Е.В., Пашкин М.А., Соловьев А.В. Применение технологии блокчейн в информационных системах. Часть 2. Подтверждение авторства и обеспечение целостности. // Системы высокой доступности. 2018. Т. 14. № 1. С. 9–11.
7. Рысков О.И. Об основных направлениях деятельности зарубежных архивных органов в области исследования и нормативного регулирования работы с электронной документацией / О.И. Рысков // Секретарское дело. – 2005. - № 3. – С.76.
8. Афанасьева Л.П. Автоматизированные архивные технологии / Л.П. Афанасьева // Федеральное агентство по образованию. Государственное Образовательное учреждение высшего профессионального образования Российский Государственный Гуманитарный университет – М.: 2005. – С. 114.
9. Miller J. NARA to suspend development of ERA starting in 2012 [Электронный ресурс] / J. Miller – Режим доступа: FederalNewsRadio.com <http://www.federalnewsradio.com/?sid=2204570&nid=35>
10. Рысков О.И. Основные направления деятельности национальных архивов США и Соединенного Королевства Великобритании и Северной Ирландии в области управления электронными документами правительственных учреждений / О.И. Рысков // Отечественные архивы. – 2004. - № 3.
11. Carlstrom G. Is DoD's new pay system fair? [Электронный ресурс] G. Carlstrom // FederalTimes.com – Режим доступа: <http://federaltimes.com/index.php?S=3502888>
12. Блог Национальных Архивов США: <http://blogs.archives.gov/records-express/2013/11/01/opportunity-for-comment-transfer-guidance-bulletin/>
13. Lipowicz A. NARA officials defend searchability of electronic archive [Электронный ресурс] / A. Lipowicz // Federal Computer Week.– Режим доступа: <http://fcw.com/articles/2011/11/01/nara-officials-defending-searchability-of-electronic-archive.aspx>
14. National Archives Announces a New Model for the Preservation and Accessibility of Presidential Records. U.S. National Archives and Records Administration [Электронный ресурс] – 2017 – Режим доступа: <https://www.archives.gov/press/press-releases/2017/nr17-54>
15. Universal Electronic Records Management (ERM) Requirements. U.S. National Archives and Records Administration, 2017: <https://>

- www.archives.gov/records-mgmt/policy/universalermrequirements
16. Hutar J. Rendering Matters [Электронный ресурс]. Режим доступа: <http://archives.govt.nz/rendering-matters-report-results-research-digital-object-rendering>.
17. Типовые требования к автоматизированным системам электронного документооборота. Спецификация MoReq // Office for Official Publications of the European Communities as INSAR Supplement VI, ISBN 9289412909.
18. Preservation of Evidence of Cryptographically Signed Documents // BSI Technical Guideline TR-03125 – Version 1.2 – Federal Office for Information Security – 2015 – 183 p.
19. Храмцовская Н.А. Южная Корея: Стандарт функциональных требований к системе управления архивными документами – и не только [Электронный ресурс] – Режим доступа: http://rusrim.blogspot.ru/2010/11/blog-post_12.html.
20. Das Archiv muss zu den Leuten gehen [Электронный ресурс] // Das Zürcher Staatsarchiv zügelt ins Internet. Neue Zürcher Zeitung Online. Режим доступа: (http://www.nzz.ch/nachrichten/medien/das_archiv_muss_zu_den_leuten_gehen_1.740545.html?printview=true)
21. Statens Arkiver. Format- og Strukturkonverteringsprojektet. [Электронный ресурс]. Режим доступа: http://www.sa.dk/content/dk/forskning_og_udvikling/udviklingsprojekter/format_og_strukturkonverteringsprojektet.
19. Храмцовская Н.А. Южная Корея: Стандарт функциональных требований к системе управ-

Соловьев Александр Владимирович. Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия. Главный научный сотрудник. Доктор технических наук. Количество печатных работ: 105. Область научных интересов: системный анализ, системы управления базами данных, теория надежности, математическое моделирование, долговременное хранение электронных документов. E-mail: soloviev@isa.ru

Баканова Нина Борисовна. Институт Прикладной Математики им. М.В. Келдыша Российской Академии Наук», 125047, г. Москва, Миусская площадь, дом 4. Зав. сектором. Доктор технических наук. Количество печатных работ: более 50. Область научных интересов: системный анализ, управление и обработка информации, проектирование информационных систем, поддержка принятия решений, проблемно-ориентированные системы, экспертные системы. E-mail: nina@keldysh.ru

Data Depersonalization Algorithm for Information Exchange in the Digital Economy

A.V. Solovyev^I, N.B. Bakanova^{II}

^I Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow, Russia

^{II} M.V. Keldysh, Institute of Applied Mathematics of the Russian Academy of Sciences, Moscow, Russia

Abstract. This article describes the problem of digital data depersonalization in information exchange in the digital economy, as well as the problem of modeling the composition of digital data information. A brief overview of the problems is given and a formal statement of the problem of the depersonalization of digital data is proposed. Depersonalization of data means the presentation of personal and / or confidential data in a form that does not allow the recovery of any information about the subject of personal data or about protected confidential information. As a solution to the problem, a depersonalization algorithm and a mathematical model of the composition of digital data information are proposed. Possible areas of application for solving the depersonalization problem are briefly described.

Keywords: *digital economy, long-term preservation, big data, distributed registries, depersonalization*

DOI: 10.14357/20790279200202

References

1. *Solovyev A.V., Bakanova N.B.* Problemy dolgovremennoy sokhrannosti bol'shikh dannykh [Long-Term Preservation Issues for Big Data] // *Informatsionnyye tekhnologii i vychislitel'nyye sistemy* [Information Technology and Computing Systems], Vol.2, 2019. P.44-53. DOI 10.14357/2071863219020.
2. *Solovyev A.V.* Obespecheniye sokhrannosti elektronnykh dokumentov pri obmene dannymi v tsifrovoy ekonomike [Securing Electronic Documents in Data Exchange in the Digital Economy] // *Trudy XXV Mezhdunarodnoy nauchno-prakticheskoy konferentsii «Dokumentatsiya v informatsionnom obshchestve: zadachi arkhivovedeniya i dokumentovedeniya v usloviyakh tsifrovoy ekonomiki»* (Moskva, Moskovskiy finansovo-yuridicheskiy universitet (MFYUA), 7-8 noyabrya 2018 g.) [Proceedings of the XXV International Scientific and Practical Conference "Documentation in the Information Society: Problems of Archival and Document Management in the Digital Economy" (Moscow, Moscow University of Finance and Law (IFLA), November 7-8, 2018)] – Rosarchiv, VNIIDAD, – 2019. P.90-98. ISBN 978-5-89789-147-4.
3. PROGRAM "Digital Economy of the Russian Federation". APPROVED by the order of the Government of the Russian Federation at July 28, 2017. № 1632-р. М.:2017 – 88 p.
4. *Solovyev A.V., Tarkhanov I.A.* Elektronnyye dokumenty i zadacha obespecheniya sokhrannosti pri obmene dannymi v tsifrovoy ekonomike [Electronic documents and the task of ensuring safety during data exchange in the digital economy] // *Trudy Instituta sistemnogo analiza RAN* [Proceedings of the Institute for System Analysis of the Russian Academy of Sciences], Vol.68(1), 2018. P.42-53. DOI 10.14357/20790279180104.
5. *Akimova G.P., Danilenko A.Yu., Pashkina E.V., Pashkin M.A., Podrabinovich A.A.* Osobennosti ispol'zovaniya elektronnoy podpisi v zashchishchennykh informatsionnykh sistemakh [Features of the use of electronic signatures in secure information systems] // *Problemy informatsionnoy bezopasnosti. Komp'yuternyye sistemy.* [Information Security Issues. Computer systems.] 2017. Vol.4. P.95–101.
6. *Danilenko A.Yu., Pashkina E.V., Pashkin M.A., Solovyev A.V.* Primeneniye tekhnologii blockchain v informacionnih sistemakh. Chast 2. Podtverzhdeniye avtorstva I obespecheniye celostnosti [The use of blockchain technology in information systems. Part 2. Confirmation of authorship and integrity.] // *Sistemy vysokoy dostupnosti* [High Availability Systems]. 2018. T. 14. № 1. P. 9–11.
7. *Ryskov O.I.* Ob osnovnih napravleniyah deyatel'nosti zarubezhnykh arhivnykh organov v oblasti issledovaniya I normativnogo regulirovaniya raboti s elektronnoi dokumentatsiey [On the main activities of foreign archival bodies in the field of research and regulatory work with electronic documentation] // *Sekretarskoye delo* [Secretarial business]. – 2005. - № 3. – P.76.
8. *Afanasyeva L.P.* Avtomatizirovannye arhivnyye tekhnologii [Automated Archive Technologies] // *Federal'noye agentstvo po obrazovaniyu. Gosudarstvennoye Obrazovatel'noye uchrezhdeniye vysshego professional'nogo obrazovaniya Rossiyskiy Gosudarstvenniy Gumanitarniy universitet* [Federal Agency for Education. State Educational Institution of Higher Professional Education Russian State University for the Humanities] – М.: 2005. – P.114.
9. *Miller J.* NARA to suspend development of ERA starting in 2012 [Electronic resource] / J. Miller – Access mode: <http://www.federalnewsradio.com/?sid=2204570&nid=35>
10. *Ryskov O.I.* Osnovniye napravleniya deyatel'nosti nacional'nykh arhivov USA I Soedinennogo Korolevstva Velikobritanii I Severnoy Irlandii v oblasti upravleniya elektronnyimi dokumentami pravitel'stvennykh uchrezhdeniy [The main activities of the national archives of the United States and the United Kingdom of Great Britain and Northern Ireland in the field of management of electronic documents of government agencies] // *Otechestvenniye Arhivy* [Russian archives]. – 2004. - № 3.
11. *Carlstrom G.* Is DoD's new pay system fair? [Electronic resource] G. Carlstrom // *FederalTimes.com* – Access mode: <http://federaltimes.com/index.php?S=3502888>
12. US National Archives Blog [Electronic resource] – Access mode: <http://blogs.archives.gov/records-express/2013/11/01/opportunity-for-comment-transfer-guidance-bulletin/>– 2019/03/22.
13. *Lipowicz A.* NARA officials defend searchability of electronic archive [Electronic resource] / A. Lipowicz // *Federal Computer Week.*– Access mode: <http://fcw.com/articles/2011/11/01/nara-officials-defending-searchability-of-electronic-archive.aspx>
14. National Archives Announces a New Model for the Preservation and Accessibility of Presidential Records. U.S. National Archives and Records Administration [Electronic resource] – 2017 –

- Access mode: <https://www.archives.gov/press/press-releases/2017/nr17-54>
15. Universal Electronic Records Management (ERM) Requirements. U.S. National Archives and Records Administration [Electronic resource], 2017, Access mode: <https://www.archives.gov/records-mgmt/policy/universalemrequirements>
 16. *Hutar J.* Rendering Matters [Electronic resource]. Access mode: <http://archives.govt.nz/rendering-matters-report-results-research-digital-object-rendering>.
 17. Typical requirements for automated electronic document management systems. Specification MoReq // Office for Official Publications of the European Communities as INSAR Supplement VI, ISBN 9289412909.
 18. Preservation of Evidence of Cryptographically Signed Documents // BSI Technical Guideline TR-03125 – Version 1.2 – Federal Office for Information Security – 2015 – 183 p.
 19. *Chramzovskaya N.A.* South Korea: Standard functional requirements to the control system archival documents – and not only [Electronic resource]. Access mode: http://rusrim.blogspot.ru/2010/11/blog-post_12.html.
 20. Das Archiv muss zu den Leuten gehen [Electronic resource] // Das Zürcher Staatsarchiv zügelt ins Internet. Neue Zürcher Zeitung Online. Access mode: (http://www.nzz.ch/nachrichten/medien/das_archiv_muss_zu_den_leuten_gehen_1.740545.html?printview=true)
 21. Statens Arkiver. Format- og Strukturkonverteringsprojektet. [Electronic resource]. Access mode: http://www.sa.dk/content/dk/forskning_og_udvikling/udviklingsprojekter/format-_og_strukturkonverteringsprojektet.

A.V. Solovyev. Chief Researcher, Department 94 ISA FRC CSC RAS. Moscow, prosp. 60-let Oktyabrya, 9. Doctor of Technical Sciences. Number of publications: 105. Area of scientific interests: system analysis, database management systems, reliability theory, mathematical modeling, electronic document management, electronic archive, long-term storage of electronic documents. E-mail: soloviev@isa.ru

N.B. Bakanova. Head of sector, M.V. Keldysh, Institute of Applied Mathematics of the Russian Academy of Sciences, 125047, Moscow, Miusskaya Square, Building 4. Doctor of Technical Sciences. Number of publications: 50. Area of scientific interests: system analysis, management and information processing, information systems design, decision-making support, problem-oriented systems, expert systems, E-mail: nina@keldysh.ru