УДК 502.084
ГРНТИ 76.01.29

В. С. Блюм[1]
кандидат технических наук, доцент
С. В. Кулешов[2]
доктор технических наук
[1]*Санкт-Петербургский государственный университет аэрокосмического приборостроения*
[2]*Санкт-Петербургский институт информатики и автоматизации Российской академии наук*

ТЕХНОЛОГИЯ УПРАВЛЕНИЯ ПОТОКОМ ЦИФРОВЫХ ПЕРСОНАЛЬНЫХ МЕДИЦИНСКИХ ЗАПИСЕЙ

Предложен подход к формированию полной и достоверной базы интегрированных электронных медицинских карт на основе технологии распределенного реестра. Рассмотрены варианты реализации алгоритма консенсуса и реализации доступа, учитывающего интересы всех участников.

Ключевые слова*: распределенный реестр, блокчейн, цифровое здравоохранение.*

V. S. Blum[1]
Candidate of Technical Sciences, Associate Professor
S. V. Kuleshov[2]
Doctor of Technical Sciences
[1]*Saint-Petersburg State University of Aerospace Instrumentation*
[2]*Saint-Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences*

DIGITAL PERSONAL MEDICAL RECORD FLOW MANAGEMENT TECHNOLOGY

An approach to the formation of a complete and reliable database of integrated electronic medical records based on distributed registry technology has been proposed.The variants for the implementation of the consensus algorithm and the implementation of information access the interests of all participants are considered.

Keywords*: distributed ledger, blockchain, digital health.*

Под врачебной тайной (ст. 13 Федерального закона № 323) подразумеваются сведения: об обращении гражданина за медицинской помощью, о состоянии здоровья и поставленном диагнозе, а также врачебная информация, полученная в результате обследования или лечения [1]. Вся указанная информация составляет существо данных, которые содержит электронная медицинская карта (ЭМК). Поэтому не случайно при внедрении информационных технологий в системы ЭМК первостепенной является проблема сохранения врачебной тайны.

Рассмотрим информационную технологию, реализующую систему ЭМК на сети региональных и федерального центров обработки данных, которая обеспечивает не только надежное хранение достоверной первичной медицинской информации, но также ее защиту от несанкционированного доступа.

Первичная медицинская информация – система «больших данных»

Источниками информации о здоровье, точнее, о нездоровье нации, которая интегрируется государственной системой медицинской статистики, являются две категории объектов систе-

мы охраны здоровья – это множество действующих дипломированных врачей и множество лицензированных диагностических лабораторий. Назовем эти источники информации сертифицированными источниками медицинской информации (СИМИ). Именно эти, и только эти звенья системы охраны здоровья допущены государством к пациенту и вправе фиксировать сведения о состоянии его здоровья. Только они имеют право и обязаны генерировать информацию (писать тексты, строить графики, фиксировать изображения) о состоянии здоровья граждан своей страны, а также предлагать и протоколировать схемы их лечения.

Анализ информационной активности СИМИ показывает [2, 3], что ежедневно сертифицированные источники создают не менее миллиона персональных медицинских записей, которые на интервале в один год образуют «террикон» информации объемом 3–4 Тбайта.

Объективная общемировая тенденция к интегрированию первичной медицинской информации направлена на создание условий для оказания квалифицированной медицинской помощи пациенту, где бы он не находился, и минимизации вероятности врачебных ошибок, зависящих от полноты исходных данных. Эта тенденция воплощается в стремлении к переходу от разрозненных, не полных, слабо связанных медицинских данных об истории болезней пациента к его интегрированной электронной медицинской карте (ИЭМК).

Поскольку объектом диспансерного наблюдения являются все граждане страны, то объем актуальных данных ИЭМК об историях болезней граждан России составит не менее 500 Тбайт. То есть речь идет о «больших данных», к которым требуется ежедневный доступ.

Инфраструктура единой государственной информационной системы здравоохранения

Согласно постановлению Правительства РФ № 555 от 05.05.2018 в состав Единой государственной информационной системы здравоохранения (ЕГИСЗ) входят следующие компоненты и подсистемы:

– федеральный регистр медицинских работников (ФРМР);

– федеральный реестр медицинских организаций (ФРМО);

– федеральная электронная регистратура (ФЭР);

– федеральная интегрированная электронная медицинская карта (ФИЭМК);

– федеральный реестр электронных медицинских документов (ФРЭМД);

– специализированные регистры пациентов по отдельным нозологиям и категориям граждан;

– информационно-аналитическая подсистема мониторинга и контроля в сфере закупок лекарственных препаратов для обеспечения государственных и муниципальных нужд;

– подсистема автоматизированного сбора информации о показателях системы здравоохранения из различных источников и предоставления отчетности;

– федеральный реестр нормативно-справочной информации (ФНСИ);

– подсистема обезличивания персональных данных;

– геоинформационная подсистема;

– защищенная сеть передачи данных;

– интеграционные подсистемы.

ЕГИСЗ состоит из федеральной (центральной) информационной системы, на которую замкнуто 85 региональных информационных систем, на которые, в свою очередь, замкнуты более 70 тыс. медицинских информационных систем лечебно-профилактических учреждений.

В книге 3 системного проекта ЕГИСЗ «Решения по проектированию, построению и организации эксплуатации телекоммуникационных и вычислительных мощностей Центра обработки данных (ЦОД) ЕГИСЗ (основной, резервной и тестовой площадок)» в таблице «Характеристики ресурсов, требуемых для работы 1-й очереди компонентов ЕГИСЗ» приведен суммарный ресурс памяти федерального центра обработки данных, который составляет 30 Тбайт. Как следует из приведенных выше оценок потока первичных медицинских данных, такой объем памяти совершенно недостаточен для хранения и использования в ЦОД полной базы ИЭМК.

В настоящей статье будем полагать, что не только федеральный, но и все региональные ЦОД обладают ресурсом памяти, достаточным для хранения полной базы ИЭМК.

Аксиомы цифрового контура здравоохранения

С 1 января 2019 г. началась реализация государственного проекта «Создание единого цифрового контура в здравоохранении на основе единой государственной информационной системы здравоохранения (ЕГИСЗ)». Рассмотрим ряд аксиом, которым, на наш взгляд, должна удовлетворять подобная разработка для того, чтобы стать успешной.

Аксиома 1. Создание и внедрение медицинской информационной системы не должно нарушать или усложнять, а тем более, останавливать лечебно-диагностический процесс (ЛДП).

Аксиома 2. Любой контакт сертифицированного источника медицинской информации с пациентом в рамках ЛДП фиксируется в форме цифровой персональной медицинской записи (ЦПМЗ) и сохраняется как в локальной базе лечебно-профилактического учреждения (ЛПУ), так и в базе ИЭМК в ЕГИСЗ.

Аксиома 3. Существует лингвистическая модель языка ЦПМЗ, позволяющая выполнять автоматический анализ синтаксиса и семантики персональных медицинских записей.

Аксиома 4. Существует эффективная технология надежного хранения и защиты от несанкционированного доступа к базе ИЭМК.

Опираясь на приведенные утверждения, можно проводить анализ и оценку принимаемых решений по реализации цифрового контура здравоохранения.

В данной статье аксиомы используем для того, чтобы:

– показать, что формирование ЦПМЗ и ее трансляция в региональный ЦОД никаким образом не нарушают технологическую схему ЛДП (аксиома 1);

– обосновать необходимость введения особого правового статуса ЦПМЗ (аксиома 2);

– подтвердить уверенность в достижении цели стандартизации ЦПМЗ на основе разрабатываемой лингвистической модели;

– предложить конкретную схему реализации распределенной системы надежного, защищенного хранения данных ИЭМК на существующей инфраструктуре сети центров обработки данных ЕГИСЗ.

Технология распределенного реестра – способ сохранения врачебной тайны

Особенностью технологии распределенных реестров (РР) является то, что она ориентирована на работу с записями минимально возможной длины. Эта особенность привела к ряду неудач по внедрению технологий РР для областей применения, не связанных с криптовалютами и финансовой деятельностью вообще. Большой объем сохраняемых данных в блоках приводит к проблемам с производительностью при вычислении значений хеш-функций и масштабируемостью при недооценке скорости увеличения хранимых данных [4, 5].

Соответственно, при использовании технологии РР в области медицины целесообразно минимизировать объем данных, включаемых в транзакцию, а в дальнейшем, и в цепочку блоков, копии которых оказываются у всех участников РР. Так, «сырые данные» лабораторных исследований (компьютерная томография, магнитно-резонансная томография, сканированные в высоком разрешении РГ снимки и т. п.) могут размещаться в надежных облачных хранилищах (целесообразно создание специализированных сертифицированных облачных сервисов, направленных на хранение именно медицинских данных и сертифицированных по федеральному закону 152-ФЗ «О персональных данных» [5]), представляющих интерес только для самого пациента. Эти данные доступны по URI (universalresourceidentifier) из любой географической точки, где имеется доступ к Интернет, а URI включается в запись ЦПМЗ.

Любая ЦПМЗ является объектом, содержащим врачебную тайну, поскольку связывает актуальную медицинскую информацию с конкретным пациентом. Синтаксическая категория ЦПМЗ формально может быть представлена как:

<ЦПМЗ> :: = <время сеанса>
<идентификатор СИМИ>
<идентификатор ПАЦИЕНТА>
<документ СИМИ>

либо

<ЦПМЗ>:: = <время сеанса>
<идентификатор СИМИ>
<идентификатор ПАЦИЕНТА>
<URI документа СИМИ>

Эти формулы указывают на обязательность временной привязки ЦПМЗ и наличия документа, который производит сертифицированный источник медицинской информациикак результат конкретной встречи с пациентом. Документ может содержать всего одно заветное слово – «здоров», которое также является объектом сохранения врачебной тайны.

Разделим проблему сохранения врачебной тайны на два уровня: уровень ЛПУ и уровень ЕГИСЗ.

Приобретая и внедряя медицинскую информационную систему в лечебно-диагностический процесс ЛПУ, руководство учреждения с необходимостью решает задачу информационной безопасности и сохранения врачебной тайны. С этой целью применяются организационные и технические (аппаратные и программные) методы.

Второй этап обеспечения информационной безопасности и сохранения врачебной тайны начинается с момента передачи ЦПМЗ из ЛПУ в региональный ЦОД с целью формирования и мониторинга интегральной электронной медицинской карты, а также для анализа общей интегрированной медицинской информации. Покажем, как на этом этапе может быть решена проблема защиты первичной медицинской информации путем внедрения технологии распре-

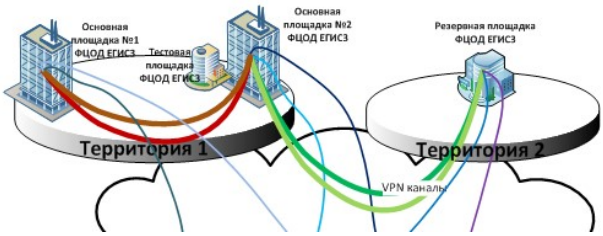

Рис. 1. Архитектура федеральных ЦОД ЕГИСЗ

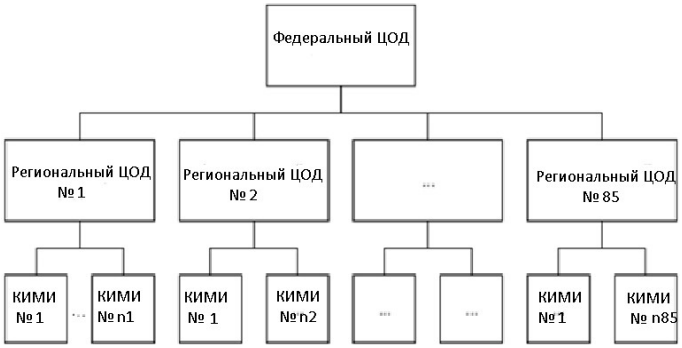

Рис. 2. Архитектура сети ЦОД ЕГИСЗ

деленного реестра и защиты сохраняемых блоков от несанкционированного доступа.

Сеть ЦОД ЕГИСЗ – технологическая основа распределенного хранения базы ИЭМК

Архитектура ЕГИСЗ предполагает развертывание федеральной компоненты, состоящей из основного и резервного центров обработки данных (рис. 1).

Для реализации инновационной технологии распределенного реестра для хранения базы ИЭМК в сети федеральных и региональных ЦОД необходимо обеспечить в каждом звене систему хранения данных (СДХ) с ресурсом памяти не менее 500 ТБ (рис. 2). Оценка потребной пропускной способности информационных каналов требует компьютерного моделирования и в данной статье обсуждаться не будет. На рис. 2 приведена архитектура сети ЦОД ЕГИСЗ ориентированная на формирование федеральной базы интегрированных электронных медицинских карт, интегрирующая информационные потоки от 85 региональных ЦОД.

Обоснование ограничений по формированию блока данных базы ИЭМК

Охрану здоровья граждан России реализует интегральный лечебно-диагностический процесс, каждое событие которого фиксируется (должно фиксироваться) в форме персональной медицинской записи. В государственном проекте создания цифрового контура здравоохранения поставлена задача формирования записей медицинских карт в цифровой форме, т. е. в форме ЦПМЗ, к 2024 г.

В работах [8–10] разработана векторно-событийная модель (ВС-модель) полной базы ИЭМК, для которой предложен метод задания отношения строгого порядка на множестве всех ЦПМЗ.

На рис. 3 представлена временная плоскость, на которой в ограниченном временном про-

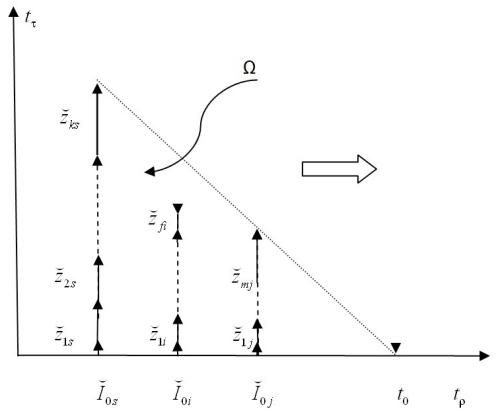

Рис. 3. Модель потока событий в базе ИЭМК

странстве размещены и упорядочены все события, отраженные в историях болезни всех ныне живущих пациентов.

На рис. 3 задана система координат, в которой осью абсцисс – время жизни популяции и осью ординат – время жизни пациентов. Пациенты и их ИЭМК упорядочены по дате рождения и порядку учета внутри каждой даты.

Гипотенуза равностороннего треугольника (рис. 3) – это отрезок прямой, который задает текущий момент времени для всех живых пациентов. Этот отрезок определяет интерфейс системы здравоохранения. В данной модели системы охраны здоровья в текущий момент времени именно на этом ограниченном отрезке, и только на нем, могут возникнуть новые события (ЦПМЗ).

Интервал времени в BC-модели отображается в форме сектора, ограниченного гипотенузами соответствующих треугольников, внутри которого будут сосредоточены все события лечебно-диагностического процесса на заданном интервале h (рис. 4).

Естественным ограничением при формировании блока данных в схеме распределенного реестра может выступать фиксированный временной интервал. В интересах клинической практики таким интервалом может быть выбран 1 час.

Однако каким бы ни был выбран интервал для формирования блока, с учетом 11 часовых поясов в России, новый блок полной базы ИЭМК будет формироваться с задержкой в 12 часов.

Учитывая указанную специфику сдвигов времени на территории России, целесообразно новый блок формировать, как сумму 11 фрагментов. Каждый фрагмент должен доставляться во все узлы сети и подвергаться подтверждению с помощью алгоритма консенсуса. Заключительный этап сборки блока должен происходить после завершения сборки данных в самом западном регионе. На этом этапе выполняется сложение 11 подтвержденных фрагментов данных и формирование итогового хеш-кода.

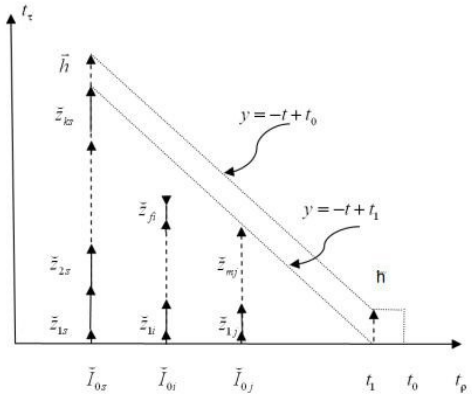

Рис. 4. Динамика потока событий в базе ИЭМК

Таким образом, технологическая схема в каждом звене распределенной сети ЦОД ЕГИСЗ включает следующую последовательность операций:

1) получение вновь сформированной ЦПМЗ в региональном ЦОД;

2) контроль легитимности ЦПМЗ по реестру СИМИ;

3) аутентификация пациента;

4) рассылка ЦПМЗ в федеральный и региональные ЦОД;

5) контроль признака завершения формирования блока;

6) проверка условия консенсуса блока;

7) расчет хеш-кода нового блока.

Алгоритм достижения консенсуса

Большинство известных алгоритмов консенсуса [10] изначально предназначались для криптовалют, и содержат в себе майнинг – получение дохода за поддержание функционирования распределенного реестра (РР) на конкурентной основе между участниками этого процесса. Но в этом нет необходимости для построения распределенного реестра в области здравоохранения.

Наиболее известный алгоритм консенсуса Proof-of-Work (PoW) вряд ли будет эффективен, в первую очередь, в связи с бессмысленными энергетическими затратами – большое количество узлов производят вычисления, но в реальности только один (первый) проводит успешную работу и получает вознаграждение.

В качестве возможных алгоритмов консенсуса для построения РР ЦПМЗ могут быть выбраны PoS, DPoS, PoAuthority, либо их модификации.

В алгоритмах Proof-of-Stake (PoS) или Delegated Proof-of-Stake (DPoS) создателем следующего блока в цепочке блоков выбирается узел, который обладает бóльшим балансом – количеством ресурсов (либо выбранный представитель в алгоритме DPoS).

В алгоритме консенсуса Proof-of-Authority (PoAuthority) для работы не требуется иметь вообще какого-либо майнинга (а также расходов на его обслуживание), как в случае с PoW или PoS. В блокчейн сети, базирующейся на PoAuthority, все транзакции и блоки проверяются посредством одобренных аккаунтов (валидаторов). Проведение транзакций и создание блоков, проходит в автоматическом режиме при помощи вычислительных мощностей валидато-

ра. Негативным моментом является повышение централизации за счет участия валидаторов.

Предоставление доступа к ЦПМЗ может реализовываться на базе смарт-контрактов по следующей схеме:

1) пациент, прошедший авторизацию и аутентификацию, предоставляет ЛПУ разрешение на временное предоставление доступа к своим данным ЦПМЗ;

2) смарт-контракт временно разблокирует ЭМК пациента для внесения изменений. Пациент уведомляется о любых фактах доступа и изменения данных ЦПМЗ;

3) ЛПУ использует локальный идентификатор медицинского работника (МР) или диагностической лаборатории (ДЛ) для доступа к нужным данным ЦПМЗ. После завершения медицинских процедур ЛПУ обновляет записи ЭМК пациента, добавляя новые транзакции в РР;

4) смарт-контракт автоматически информирует об изменениях пациента и врача: пациент информируется об изменениях в ЭМК, смарт-контракт информирует врача о получении пациентом набора услуг в рамках посещения ЛПУ.

Каждая ЦПМЗ должна подписываться электронной подписью, подтверждающей валидность и значимость записи обоими заинтересованными сторонами – МР (ДЛ) и пациентом. Электронная подпись важна при предоставлении данных из РР страховым компаниям для оплаты действий МР и ДЛ и исключения случаев «приписки» консультаций или лечебных процедур, которые фактически не были оказаны.

Алгоритм хеш-функции

Алгоритм SHA был разработан Национальным институтом стандартов и технологии США (NIST) и опубликован в виде федерального стандарта обработки информации в 1993 г. Пересмотренная версия вышла в 1995 г.

Алгоритм SHA обладает тем свойством, что каждый бит хеш-кода зависит от всех битов хешируемых данных. Сложное многократное использование базовых функций в результате дает хорошее перемешивание, это означает, что практически невероятно, чтобы два набора входных данных породили один и тот же хеш-код, несмотря на то, что они оказываются подобными по структуре.

SHA-256 представляет собой однонаправленную функцию для создания цифровых отпечатков фиксированной длины (256 бит, 32 байт) из входных данных размером до 2,31 эксабайт и является частным случаем алгоритма из семейства криптографических алгоритмов SHA-2 (SecureHashAlgorithmVersion 2), опубликованным АНБ США в 2002 г.

Существенными свойствами алгоритма хеш-функции, позволяющими использовать ее в технологии РР являются:

– невозможность определения аргумента функции (подписываемого сообщения) по значению хеш-функции более простым способом, чем полный перебор;

– возникновение больших изменений значения функции («лавинный эффект» хеш-функции, рис. 5) при небольшом изменении аргумента (сообщения);

– невозможность нахождения различных аргументов (сообщения) с одинаковыми хеш-значениями.

На рис. 5 проиллюстрированы свойств хеш-функции.

Актуальная цепочка данных базы ИЭМК

Применение технологии распределенного реестра к потоку ЦПМЗ приведет к возникновению в каждом узле сети идентичных цепочек блоков, идентифицированных временными интервалами.

Потенциально бесконечная цепочка блоков данных, в которую воплощается база ИЭМК, на самом деле имеет естественное ограничение, что непосредственно следует из анализа выше приведенной ВС-модели информационного потока. Таким ограничением является следующее утверждение: «В данном блоке и во всех ранее соз-

Рис. 5. Иллюстрация свойств хеш-функции: изменение одного символа в аргументе дает полное изменение значения хеш-функции

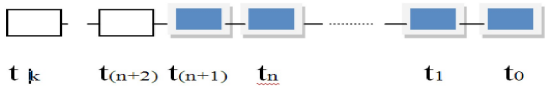

Рис. 6. Конечная цепочка актуальных блоков

данных блоках нет ЦПМЗ о состоянии живого пациента». Очевидно, что в каждый текущий момент времени крайним блоком будет блок, в котором хранится ЦПМЗ о рождении самого старшего из ныне живущих пациентов (рис. 6).

Можно утверждать, что технология распределенного реестра хорошо согласуется с задачей сохранения записей о событиях интегрального лечебно-диагностического процесса.

Для повышения эффективности управления записями целесообразно в заголовке каждого блока указывать перечень идентификаторов пациентов, ЦПМЗ которых попали в данный блок. Такое решение позволит оперативно формировать однозначное отображение ИЭМК пациента в последовательность блоков актуальной цепочки.

Заключение

Рассмотрена технологическая схема сбора и хранения полной, достоверной и доступной для автоматизированной обработки базы ИЭМК всех граждан России.

Создание единой базы данных ЭМК на основе технологии распределенных реестров позволит:

– видеть полную историю по каждому пациенту;

– организовать преемственность медицинской помощи;

– повысить достоверность собираемой статистики;

– упорядочить деятельность страховых компаний.

Предложенная технология исключает какие-либо несанкционированные изменения медицинской информации и гарантирует равные информационные возможности для любого гражданина страны, в каком бы регионе он не оказался.

Следует отметить, что перед повсеместным внедрением РР для медицинских данных потенциальные проблемы (возможность потери ключей, реализация разделения прав доступа и т. д.), требуют дальнейшей проработки.

Реализация схемы распределенного реестра для хранения базы ИЭМК создает все необходимые информационные условия для мониторинга состояния здоровья нации как на региональном, так и на федеральном уровне, а также позволяет оперативно оценивать эффективность использования ресурсов системы здравоохранения и маневрировать этими ресурсами.

Библиографический список

1. Федеральный закон 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации».
2. *Блюм В. С., Заболотский В. П.* Мысленный эксперимент по организации учета и обработки информационных медицинских услуг //Врач и информационные технологии.2009. № 4.С. 27–35.
3. *Блюм В. С., Виноградов В. М., Карташев А. В.* Информатизация здравоохранения и иммунокомпьютинг. // Врач и информационные технологии. 2009. № 3.С. 17–27.
4. *Цыганов С. Н.* Применение технологии блокчейн для хранения данных электронных медицинских карт пациентов // Фундаментальные исследования.2017. №11–12. С. 338–343.
5. *Кулешов С. В., Зайцева А. А.* Перспективы использования технологии распределенных реестров для произведений цифрового искусства // в сб. Технологическая перспектива: новые рынки и точки экономического роста. Материалы 4-ой Международной научной конференции. Под. ред. проф. Кораблевой О. Н. и др. СПб.: Издательство «Астерион», 2018. С. 20–22.
6. Федеральный закон 152-ФЗ «О персональных данных».
7. *Блюм В. С.* Дискретно-событийная модель здравоохранения и федеральный сервис «Интегрированная электронная медицинская карта» // Математическая морфология. Электронный математический и медико-биологический журнал. 2012. Т. 11. Вып. 4. С. 4–15. URL: http://sgma.alphadesign.ru/MMORPH/TITL.HTM (дата обращения 24.09.2019).
8. *Блюм В. С.* Инновационная государственная система медицинской статистики // Актуальные проблемы экономики и управления. 2015. № 2. С. 80–88.
9. *Блюм В. С., Инкин В. А.* Метод визуализации математической модели базы интегрированных электронных медицинских карт // Актуальные проблемы экономики и управления. 2016. № 2(10). С. 88–94.
10. Обзор 9 алгоритмов блокчейн консенсуса. URL: https://digiforest.io/blog/blockchain-consensus-algorithms (дата обращения 23.08.2019).