

Алгоритм инвентаризации цифровых носителей при долговременном хранении*

А. В. Соловьев, А. Ю. Даниленко, Г. П. Акимова, М. А. Пашкин,
А. А. Подрабинович, И. В. Туманова

Федеральное государственное учреждение Федеральный исследовательский центр
"Информатика и управление" Российской академии наук, г. Москва, Россия

Аннотация. В статье описаны проблемы надежности и технологического старения цифровых носителей информации, а также изменения аппаратно-программной среды хранения цифровых данных, возникающие на длительном сроке хранения (десятилетия). Сделан краткий обзор различных типов цифровых носителей информации с оценкой сроков надежного гарантированного хранения данных. Описана проблема вынужденного изменения аппаратно-программной среды хранения цифровых данных, возникающая вследствие ограниченного срока эксплуатации операционных систем и программного обеспечения систем хранения, а также вследствие ограниченного срока эксплуатации технических средств информационных систем. Для решения представленных проблем предложен алгоритм инвентаризации цифровых носителей при организации долговременного хранения. Оговорены основные аспекты и правила его применения. Описаны аспекты практического применения разработанного алгоритма. Сделан вывод, что предложенный алгоритм, прошедший проверку практикой, позволят решить поставленную в статье проблему. В заключении рассмотрены перспективы развития и использования предложенного алгоритма.

Ключевые слова: долговременное хранение, надежность, метаданные, алгоритм инвентаризации, цифровизация, миграция данных, технологическое старение.

DOI 10.14357/20718632200304

Введение

В условиях стремительной цифровизации, ведущие экономические державы пытаются решить проблему долговременной сохранности цифровых данных [1, 2]. Проблема сохранности является не только актуальной [3-5], но и крайне сложной, так как к ней относятся проблемы аутентичности, надежности, независимости данных от носителей информации и программно-аппаратной среды, интерпретируемости, потери семантики, безопасности, устойчивости к внешним воздействиям [6, 7].

В Российской Федерации, где в условиях цифровизации, согласно [8], цифровые данные стали «ключевым фактором производства», решение задачи долговременной сохранности также крайне актуально. Одной из важнейших проблем при долговременном хранении цифровых данных является проблема выхода из строя носителей информации (дисков, лент, оптических носителей, SSD – Solid state disk и др.). Ни один производитель подобной техники не гарантирует сохранность ее в течение десятилетий (тем более столетий), а, следовательно, встает про-

* Работа выполнена при частичной финансовой поддержке РФФИ в рамках научного проекта № 18-29-03070.

блема своевременной диагностики носителей цифровых данных и своевременного переноса их на другие носители.

Решению проблемы контроля цифровых носителей информации при долговременном хранении данных посвящена данная статья. В работе предложен алгоритм инвентаризации носителей, с помощью которого может быть решена поставленная проблема в рамках программного обеспечения аппаратно-программных систем хранения цифровых данных.

1. Краткий обзор технических средств хранения цифровых данных

Гарантийные сроки надежного хранения большинства жестких дисков – 5 лет. Производители оптических дисков однократной записи (носители типа WORM – write once read many), называли изначально сроки в 50-100 лет, но затем и они были существенно уменьшены (кроме того, для них нужны идеальные условия хранения) до 20-25 лет максимум, после чего данные должны быть перезаписаны. На основе опыта авторов работы по созданию электронных архивов (ЭА) с использованием DVD-R ведущих производителей, автор может утверждать, что на практике срок хранения DVD-R еще ниже, проверки и перезаписи нужно осуществлять не реже 1 раза в 5 лет.

Даже для специально предназначенных для ЭА накопителей на базе технологии UDO (Ultra Density Optical, разработка компании Plasmon) на основе ультраплотной записи [9] не подтверждена возможность их работы в течение многих десятилетий. Накопители UDO используются, например, для хранения медицинских изображений, медицинских документов, медицинских карт пациентов. Причем гарантированный срок хранения не превышает 5 лет. UDO представляет собой картридж 5.25” с оптическим диском внутри. Объем диска на данный момент составляет от 60 Гб до 120 Гб. Для записи может использоваться как красный лазер (650нм), так и сине-фиолетовый (405нм), причём во втором случае максимальный объем диска может достигать 500 Гб. Оптический диск не подвержен размагничиванию, как магнитные носители.

Используемые для резервного копирования магнитные ленты являются крайне неустойчивым к внешним воздействиям носителям, поскольку требуется их перемотка 1 раз в полгода и тщательная защита от размагничивания.

Использование твердотельных накопителей (SSD, флэш-карт и т.д.) также пока не надежно. Данные накопители имеют ограничение на количество циклов перезаписи (3000-10000), повышенный износ в связи с этим, высокую стоимость гигабайта информации по сравнению с жесткими дисками и оптическими дисками и невысокий объем хранения данных [10]. Проблему повышенного износа пытаются преодолеть с помощью технологии энергонезависимой памяти FRAM (Ferroelectric Random Access Memory) для которых количество циклов перезаписи оценивается величиной порядка 10^{14} [11]. Однако и эти носители не позволяют хранить большие объемы данных, зато отличаются высокой стоимостью. Время гарантированного хранения данных на SSD и FRAM оценивается в 10 лет.

Таким образом, промышленные средства хранения цифровых данных на данный момент не могут достигнуть максимального срока хранения информации, такого как на бумаге или в виде микрофильмов: до 500 лет при идеальных условиях хранения.

Кроме проблемы хранения цифровых данных на конкретном цифровом носителе, существует также проблема технологического старения. Поэтому с достаточно высокой вероятностью через 100 лет невозможно будет прочитать данные с современных цифровых носителей из-за отсутствия в будущем устройств для их чтения, даже если информация каким-то чудом на них сохранится.

Из анализа современных технологий (например, [12, 13]) складывается впечатление, что производители не очень и заинтересованы в долговременном существовании тех или иных носителей, средний срок существования технологий от момента появления до почти полного исчезновения с рынка оценивается в 10-15 лет (магнитные ленты, дискеты, CD-R, DVD-R и др.). Затем новые технологии вытесняют более старые, и производителям экономически невыгодно будет поддерживать устаревшие техно-

логии. Причем срок «оборота технологий» в информационных технологиях имеет устойчивую тенденцию к сокращению.

2. Анализ проблемы обновления аппаратно-программной среды хранения цифровых данных

Кроме собственно износа и старения физических носителей информации, существует тесно связанная с ней проблема обновления аппаратно-программной среды долговременного хранения цифровых данных [15, 16].

Ситуация здесь напоминает движение по замкнутому кругу: старые операционные системы (ОС) и другое системное программное обеспечение снимается с поддержки, новое требует большей производительности процессоров, объемов памяти и т.д., т.е. требуется обновление аппаратной среды. Старые ОС и другое программное обеспечение (ПО) не могут функционировать на новых аппаратных средствах и т.д.

Например, распространенная ОС Windows компании Microsoft придерживается следующей стратегии развития ОС: первые 5 лет осуществляется полная поддержка версии ОС (выпускаются обновления, которые могут установить любые зарегистрированные пользователи Windows). Следующие 5 лет может быть осуществлена расширенная поддержка под

конкретного заказчика. Далее ОС снимается с поддержки и не гарантируется ее работоспособность на новых аппаратных средствах.

Кратко процесс, связанный с обновлением аппаратно-программной среды можно проиллюстрировать Табл. 1. В таблице приведены проблемы, возникающие на 20 летнем сроке хранения, в то время как срок хранения, например, так называемых документов по личному составу составляет в настоящее время 50 лет.

Тем самым хранение цифровых данных необходимо обеспечить в условиях неизбежно меняющейся среды хранения. При этом сделать эту среду либо доверенной, либо использовать какие-то средства обеспечения целостности цифровых данных и их метаданных.

Как при старении носителей, так и при обновлении аппаратно-программной среды возникает необходимость корректной миграции цифровых данных. Сложность организации миграции данных в том, что высока вероятность потери данных по халатности или злему умыслу.

Также проблема миграции касается не только переноса самих цифровых данных, но и их метаданных [17]. Если метаданные: индексы, метаданные, классификаторы, рубрикаторы, связи с другими документами и др. не могут быть корректно перенесены, то, по сути, миграция цифровых данных выльется в повторное создание ЭА в новой аппаратно-программной среде с построением заново всех метаданных,

Табл. 1. Проблемы обновления аппаратно-программной среды хранения цифровых данных

Срок хранения	Проблемы программной среды	Проблемы аппаратной среды
3 года	1) истечение сроков сертификатов электронной подписи (ЭП)	
5 лет	1) предельный срок действия сертификатов ЭП по 63-ФЗ [14] 2) необходимы существенные доработки ПО ЭА	1) выход из строя носителей информации (CD, DVD, HDD, флэш-накопителей и др.)
10 лет	1) снимается с поддержки ОС 2) устаревают программы просмотра файлов различных форматов 3) прекращается поддержка закрытых форматов данных	1) предельный срок службы SSD и FRAM 2) проблемы с модернизацией техники
20 лет	1) значительные ограничения при использовании устаревшей ОС и устаревших программ просмотра 2) ОС можно использовать только в виртуальной среде 3) возможны проблемы даже с открытыми форматами данных 4) возможны изменения стандартов криптозащиты, как следствие невозможность проверки ЭП	1) предельный срок службы аппаратных средств хранения

что является задачей почти не выполнимой. Некорректная миграция может привести к потере семантики цифровых данных.

3. Алгоритм инвентаризации цифровых носителей

Как видно из обзора проблем, все имеющиеся на данный момент типы носителей информации недостаточно надежны для хранения данных десятилетиями, а тем более столетиями. Более того, из-за процесса технологического старения через несколько десятилетий не останутся устройств, обеспечивающих чтение актуальных на данный момент носителей информации.

Можно утверждать, что решение проблемы «старения» лежит, во-первых, в избыточности хранения информации, во-вторых, в регулярной проверке и переносе информации на новые носители данных.

Избыточность хранения данных должна быть обеспечена как хранением данных ЭА непосредственно в БД на жестком диске, так и хранением на внешних носителях копий данных ЭА. В качестве такой копии может выступать как резервная копия БД, так и копии данных, вытесненных на внешние носители.

Во всех случаях для БД ЭА необходимо организовать регулярное резервное копирование БД на внешние носители.

В качестве копии данных могут выступать как внешние носители с резервной копией БД, так и совокупность внешних носителей с резервной копией БД и внешних носителей с данными. При этом должно создаваться не менее двух копий данных ЭА, причем хранить их следует в разных помещениях, а в идеальном случае в разных зданиях, удаленных друг от друга. В случае использования в качестве резервной копии компакт-дисков рекомендуется создавать не менее трех копий данных.

Дополнительно может быть реализовано устойчивое к внешним воздействиям решение (зеркало, или, для особо ценных документов, – резервный центр обработки данных (ЦОД)), т.е. хранение точной копии (копий) документов. Это означает, что необходимо реализовать децентрализованное хранение копий данных с разными мандатами доступа для оперативного и административного персоналов ЭА.

В качестве решения описанных в данной статье проблем, авторами исследования предложен алгоритм инвентаризации цифровых носителей данных.

Формальная постановка задачи выглядит следующим образом.

Дано:

- 1) Множество цифровых данных $DD = \{DD_i\}$
- 2) Множество носителей информации $DDisk = \{DDisk_j\}$, хранящих DD
- 3) Исходный уровень сохранности (характеризует исправность носителя, его наличие, наличие гарантированной работоспособности) для $DDisk - SV_0$.

Найти:

- 1) Алгоритм, реализующий сохранность DD при изменении сохранности SV_0 цифровых носителей $DDisk$.

Решение:

Регулярная проверка и перенос (в терминах ГОСТ Р ИСО 15489-1-2007 [18] – конвертация или миграция) цифровых данных DD на новые носители должны обеспечить защиту от отказов и физической деградации цифровых носителей информации $DDisk$. Назовем такую процедуру инвентаризацией носителей.

Данная процедура должна включать проверку целостности данных DD на носителе, оценку оставшегося времени хранения данных на носителе и, при необходимости, перенос данных DD на новый носитель.

В случае выявления нарушения целостности данных на носителе в ходе проверки новая копия данных создается из других копий данной информации. Периоды проверки носителей данных выбираются, исходя из типа носителей информации, но в любом случае период хранения данных на неизменяемом носителе не должен превышать трех лет, т.е. раз в три года каждый носитель информации должен быть проверен и при необходимости заменен.

Процесс переноса информации должен предусматривать возможность слияния данных с разных носителей, данное условие появляется из-за постоянного увеличения объемов всех видов носителей данных. Поэтому носители следующего поколения, как более емкие могут вмещать информацию с нескольких «старых» носителей.

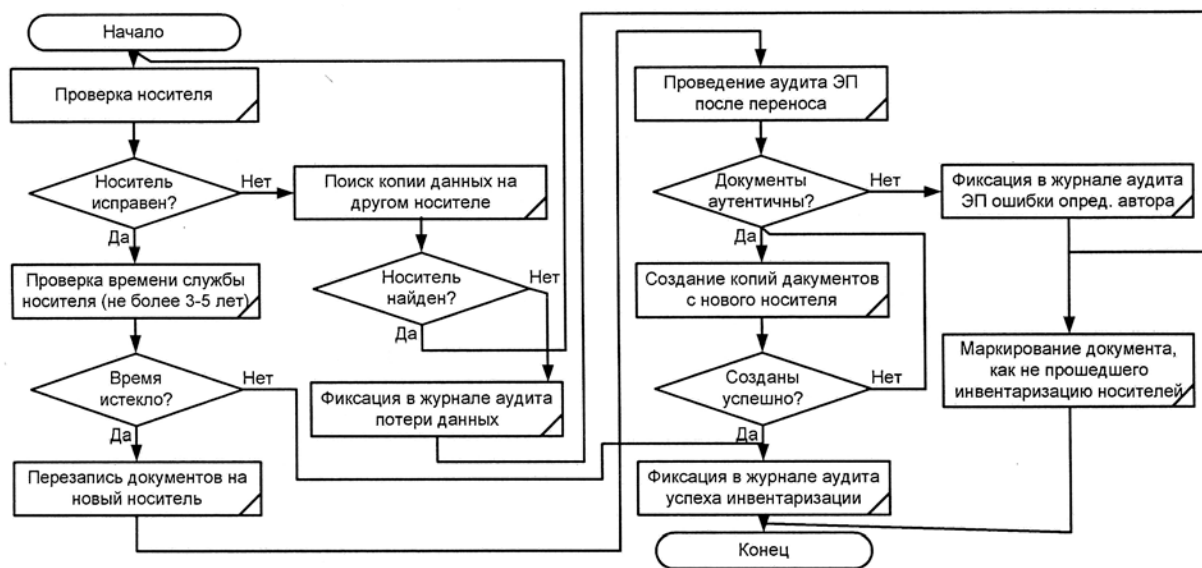


Рис. 1. Алгоритм инвентаризации цифровых носителей

Предложенный алгоритм в общем виде представлен на Рис. 1. Только в случае его реализации в рамках конкретной системы долговременного хранения цифровых данных можно будет говорить о сохранности данных *DD* при «старении» носителей *DDisk*.

Вся информация о ходе и результате инвентаризации должна отображаться в системном журнале инвентаризации носителей ЭА. Также информация по проверке аутентичности до и после переноса данных должна быть отражена в системном журнале инвентаризации ЭП.

Проблема обновления аппаратно-программной среды хранения цифровых данных, так же как и проблема «старения» носителей приводит к необходимости организации периодической миграции данных.

Можно утверждать, что миграция данных должна быть неотъемлемой частью технологии создания любой информационной системы, предназначенной для долговременного хранения цифровых данных.

Для того, чтобы правильно организовать процесс миграции данных, необходимо ответить на следующий вопрос. Что должно подвергаться миграции: только ли сами документы из БД ЭА или же еще связанные с ними метаданные, классификаторы, индексы, журналы аудита и др.?

Ответом на вопрос «что переносить» должна быть продуманная математическая модель цифровых данных при долговременном хранении. В модели должны быть прописаны все составные части данных и их назначение.

Например, классификаторы и индексы являются неотъемлемой частью цифровых данных, поскольку определяют контекст их использования: предметную область, структуру, логику хранения и классификации, связи с другими данными и т.д. По мнению авторов, потеря этих данных при миграции может оказаться критичной, документ будет вырван из контекста использования, и понять его принадлежность какой-либо тематике будет проблематично.

Отдельно стоит вопрос о полнотекстовых индексах. Конечно, не хочется терять такую ценную информацию, однако большинство СУБД не позволяет распорядиться полнотекстовыми индексами самостоятельно, а перестройка индекса для огромного массива данных после миграции может оказаться дорогостоящей по времени процедурой. Несовместимым может оказаться и формат индексов при переносе в другую среду хранения.

При решении данной проблемы рекомендуется либо переносить полнотекстовые индексы вместе с цифровыми данными, либо включить процедуру перестройки индексов в процесс миграции. Во втором случае переход на новую

аппаратно-программную среду долговременного хранения должен быть осуществлен с поддержкой в эксплуатации после миграции данных. Это, в свою очередь, означает допущение существования двух различных сред (систем) хранения при условии полной синхронизации цифровых данных между ними.

Процедуру миграции можно будет производить реже, если использовать преимущества виртуализации. ОС, запущенная на виртуальном компьютере, будет функционировать, даже тогда, когда она не может быть установлена на более современный компьютер.

Однако и у этого решения есть свои проблемы. Так, например, рано или поздно встанет вопрос о поддержке данной старой ОС со стороны производителя. К тому же в настоящий момент существуют ограничения на использование некоторых ОС в виртуальных средах. Например, использование IBM i (старое название OS/400) возможно только в виртуальных средах на платформе Power, на платформе Intel данная ОС работать не будет даже в виртуальной среде.

4. Возможное практическое применение математических моделей

Предложенный в работе алгоритм был доведен до программной реализации. При создании большой территориально-распределенной информационной системы – Электронный архив персонифицированного учета Пенсионного фонда Российской Федерации (ЭАПУ ПФ РФ) – алгоритм был разработан программный модуль инвентаризации различных типов носителей, в том числе съемных. Созданная информационная система функционирует с 2004 года в 80 регионах Российской Федерации. В настоящий момент обеспечивает хранение более 50% документов персонифицированного учета ПФ РФ.

Результаты применения алгоритма в рамках крупной информационной системы помогли решить одну из важных задач обеспечения долговременной сохранности цифровых данных – обеспечение физической сохранности данных. За время эксплуатации электронного архива было проведено 2 миграции без потерь данных.

Тем самым, цифровые данные были полностью сохранены.

Алгоритм также был опробован в рамках создания ряда электронных архивов документов и хранилищ цифровых данных.

В дальнейшем планируется существенно доработать реализацию данного алгоритма для повышения их производительности. Планируется также включить данный алгоритм в разрабатываемую комплексную технологию организации долговременного хранения цифровых данных для информационных систем предприятия.

Заключение

В настоящее время в мире создан ряд нормативных документов, в которых регламентируются основные аспекты долговременной сохранности цифровых данных (например, [19, 20]). Однако содержащиеся в них положения содержат постановки проблемы, но не оговаривают вопросы разработки тех или иных алгоритмов для их решения. Такой, в частности, является проблема изменения аппаратно-программной среды хранения цифровых данных, а также технологической старение носителей информации.

Для решения поставленных в данном исследовании проблем надежного долговременного хранения цифровых данных на существующих носителях предложен алгоритм инвентаризация цифровых носителей данных.

Предложенный алгоритм был реализован в виде программных модулей и прошел проверку практикой при разработке ряда информационных систем электронных архивов, в которых предполагалось долговременное хранение цифровых данных.

В дальнейших исследованиях планируется существенно повысить производительность программные реализации предложенного алгоритма, усовершенствовать сам алгоритм с точки зрения простоты реализации.

В перспективе планируется доработать созданную авторами технологию организации долговременного хранения цифровых данных, включив в нее данный алгоритм. Планируется использовать разрабатываемую технологию для информационных систем масштаба предприятия (ERP – Enterprise Resource Planning).

Литература

1. Universal Electronic Records Management (ERM) Requirements. U.S. National Archives and Records Administration [Электронный ресурс], 2017, URL: <https://www.archives.gov/records-mgmt/policy/universalmrequirements> (дата обращения: 01.05.2020).
2. National Archives Announces a New Model for the Preservation and Accessibility of Presidential Records. U.S. National Archives and Records Administration [Электронный ресурс] – 2017 – URL: <https://www.archives.gov/press/press-releases/2017/nr17-54> (дата обращения: 01.05.2020).
3. Суровцева Н.Г. Хранение электронных документов: зарубежный опыт // Вестник культуры и искусства. – 2017. – №4(52). – С.17.-23.
4. Даниленко А.Ю., Пашкина Е.В., Пашкин М.А., Соловьев А.В. Применение технологии блокчейн в информационных системах. Часть 2. Подтверждение авторства и обеспечение целостности // Журнал «Системы высокой доступности» №1, т.14, М.: Радиотехника, 2018, С.9-12.
5. Melanie Swan. Blockchain: Blueprint for a New Economy. — O'Reilly Media, Inc., 2015. — 152 p. — ISBN 978-1-4919-2047-3.
6. Solovyev A.V. Long-term storage technology of digital documents // Lecture Notes in Electrical Engineering, vol.641, pp.901-911, 2020. ISSN 1876-1100. DOI:10.1007/978-3-030-39225-3_97.
7. 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.
8. ПРОГРАММА «Цифровая экономика Российской Федерации». УТВЕРЖДЕНА распоряжением Правительства Российской Федерации от 28 июля 2017 г. № 1632-р. М.:2017 – 88 с.
9. Архивные оптические накопители Plasmon UDO2 G-серии. ЭЛАР [Электронный ресурс] – URL: http://www.plasmon.ru/udo2_g.pdf (дата обращения: 02.05.2020).
10. Орлов, С. Наступление SSD [Электронный ресурс] // Журнал сетевых решений/LAN - 2010 - №11 – URL: <http://www.osp.ru/lan/2010/11/13005552/> (дата обращения: 02.05.2020).
11. Rzehak V. Особенности применения FRAM микроконтроллеров Texas Instruments [Электронный ресурс] // Журнал РАДИОЛОЦМАН – 2012 – URL: <http://www.rlocman.ru/review/article.html?di=113273> (дата обращения: 03.05.2020).
12. Корепанов, И. Как сохранить архив на десятилетия? // «Журнал сетевых решений/LAN» - 2008 - №03 [Электронный ресурс] – URL: <http://www.osp.ru/lan/2008/03/4899898/> (дата обращения: 03.05.2020).
13. Тихонов, В. Архивное хранение электронных документов: проблемы и решения [Электронный ресурс] // Журнал «Делопроизводство и документооборот на предприятии» - 2006 – URL: <http://www.delo-press.ru/articles.php?n=5150> (дата обращения: 03.05.2020).
14. Федеральный закон Российской Федерации от 6 апреля 2011 г. N 63-ФЗ «Об электронной подписи».
15. Афанасьева, Л.П. Автоматизированные архивные технологии // Федеральное агентство по образованию. Государственное образовательное учреждение высшего профессионального образования Российский Государственный Гуманитарный университет – М.: 2005. – С. 114.
16. Miller J. NARA to suspend development of ERA starting in 2012 [Electronic resource] – URL: <http://www.federalnewsradio.com/?sid=2204570&nid=35> (дата обращения: 04.05.2020).
17. ГОСТ Р ИСО 23081-1-2008 Процессы управления документами. Метаданные для документов.
18. ГОСТ Р ИСО 15489-1-2007 Система стандартов по информации, библиотечному и издательскому делу. Управление документами. Общие требования.
19. ГОСТ Р 54989-2012 /ISO TR 18492:2005 Обеспечение долговременной сохранности электронных документов (вступил в силу с 01.05.2013).
20. Preservation of Evidence of Cryptographically Signed Documents // BSI Technical Guideline TR-03125 – Version 1.2 – Federal Office for Information Security – 2015 – 183 p.

Соловьев Александр Владимирович. Главный научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), г. Москва, пр-т 60-летия Октября, 9. Доктор технических наук. Количество печатных трудов: 105. Область научных интересов: системный анализ, системы управления базами данных, теория надежности, математическое моделирование, долговременное хранение электронных документов. E-mail: soloviev@isa.ru

Даниленко Андрей Юрьевич. Ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), г. Москва, пр-т 60-летия Октября, 9. Кандидат физико-математических наук. Количество печатных трудов: 40. Область научных интересов: системное программирование, системный анализ, информационные технологии, электронный документооборот, информационная безопасность, защита данных. E-mail: danilenko@isa.ru

Акимова Галина Павловна. Ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), г. Москва, пр-т 60-летия Октября, 9. Кандидат технических наук. Количество печатных трудов: 65. Область научных интересов: системное программирование, системный анализ, информационные технологии, влияние человеческого фактора, информационно-аналитические системы, электронный документооборот, электронный архив. E-mail: akimova@isa.ru

Пашкин Матвей Александрович. Научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), г. Москва, пр-т 60-летия Октября, 9. Количество печатных трудов: 20. Область научных интересов: системное программирование, информационные технологии, информационно-аналитические системы, электронный архив. E-mail: pashkin@isa.ru

Подрабинович Андрей Александрович. Ведущий программист Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), г. Москва, пр-т 60-летия Октября, 9. Количество печатных трудов: 15. Область научных интересов: системное программирование, проектирование и создание методов и программных средств управления электронными документами, защита информации. E-mail: podrabinovich@isa.ru

Туманова Ирина Владимировна. Ведущий программист Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), г. Москва, пр-т 60-летия Октября, 9. Количество печатных трудов: 10. Область научных интересов: системное программирование, информационные технологии, электронный документооборот, электронный архив. E-mail: tumanova-irin@mail.ru

Long-term digital media inventory algorithm

A. V. Solovyev, A. Yu. Danilenko, G. P. Akimova, M. A. Pashkin,
A. A. Podrabinovich, I. V. Tumanova

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

Abstract. The article describes the problems of reliability and technological aging of digital storage media, as well as problems of changing in the hardware and software environment for storing digital data, that occur over a long keeping period (decades). A brief overview of various types of digital storage media with an assessment of the terms of reliable guaranteed data keeping. The problem of the forced change in the hardware and software environment for storing digital data is described that occurs due to the limited life of operating systems and software of storage systems, as well as due to the limited life of hardware of information systems. To solve the problems presented, an algorithm for the inventory of digital media in the organization of long-term keeping is proposed. The main aspects and rules for the application of the presented algorithm are agreed. Aspects of the practical application of the developed algorithm are described. It is concluded that the proposed algorithm, tested by practice, allows us to solve the problems posed in the study. In conclusion, the prospects for the development and use of the proposed algorithm are considered.

Keywords: long-term keeping, reliability, metadata, inventory algorithm, digitalization, data migration, technological aging.

DOI 10.14357/20718632200304

References

1. 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/universalmrequirements> (last access: 01.05.2020).
2. 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> (last access: 01.05.2020).
3. Suvorovtseva N.G. Hraneniye elektronnykh dokumentov: zarubezhnyy opyt [Storage of electronic documents: foreign experience] // Vestnik kultury i iskusstva [Bulletin of culture and art]. – 2017. – №4(52). – P.17.-23.
4. Danilenko A.Yu., Pashkina E.V., Pashkin M.A., Solovyev A.V. Primeneniye tehnologii blockchain v informacionnykh sistemah. Chast 2. Podtverzhdeniye avtorstva i obespecheniye celostnosti [The use of blockchain technology in information systems. Part 2. Confirmation of authorship and integrity.] // Systemy visokoy dostupnosti [High Availability Systems]. 2018. T. 14. № 1. P. 9–11.
5. Melanie Swan. Blockchain: Blueprint for a New Economy. — O'Reilly Media, Inc., 2015. — 152 p. — ISBN 978-1-4919-2047-3.

6. Solovyev A.V. Long-term storage technology of digital documents // Lecture Notes in Electrical Engineering, vol.641, pp.901-911, 2020. ISSN 1876-1100. DOI:10.1007/978-3-030-39225-3_97.
7. Bakanova, N.B., Solovyev, A.V. Problemy dolgovremennoy sokhrannosti bol'shikh dannykh [Problems of long-term keeping of big data] // Informatsionnyye tekhnologii i vychislitel'nyye sistemy [Information Technology and Computing Systems], №2, 2019. pp. 44-53. DOI 10.14357/2071863219020.
8. 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.
9. Arhivniye opticheskiye nakopiteli Plasmon UDO2 G-serii. ELAR [Archive optical drives Plasmon UDO2 G-series. ELAR] – Access mode: http://www.plasmon.ru/udo2_g.pdf (last access: 02.05.2020).
10. Orlov, S. Nastupleniye SSD [Offensive of SSD] // «Jurnal setevih reshenii/LAN» [Network decision journal], 2010, №11. Access mode: <http://www.osp.ru/lan/2010/11/13005552/> (last access: 02.05.2020).
11. Rzhak, V. Osobennosti primeneiya FRAM mikrokontrolerov Texas Instruments [Texas Instruments FRAM microcontroller application features] // RADIOLOZMAN [RADIOPILOT], 2012. Access mode: <http://www.rlocman.ru/review/article.html?di=113273> (last access: 03.05.2020).
12. Korepanov, I. Kak sohranit' arhiv na desyatletiya? [How to keep archive for decades?] // «Jurnal setevih reshenii/LAN» [Network decision journal], 2008, №03. Access mode: <http://www.osp.ru/lan/2008/03/4899898/> (last access: 03.05.2020).
13. Tikhonov, V. Arhivnoye hranenie elektonnih dokumentov: problemi i resheniya [Archival storage of electronic documents: problems and solutions] // Deloproizvodstvo i dokumentooborot na predpriyatii [Paperwork and document management at the enterprise], 2006. Access mode: <http://www.delo-press.ru/articles.php?n=5150> (last access: 03.05.2020).
14. Federal law of the Russian Federation 6 april 2011. №63-FZ «On the electronic signature».
15. Afanasyeva L.P. Avtomatizirovannye arhivniye tehnologiiu [Automated Archive Technologies] // Federal'noye agentstvo po obrazovaniyu. Gosudarstvennoye Obrazovatelnoye 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.
16. Miller J. NARA to suspend development of ERA starting in 2012 [Electronic resource] – Access mode: <http://www.federalnewsradio.com/?sid=2204570&nid=35> (last access: 04.05.2020).
17. GOST R ISO 23081-1-2008 Processy upravleniya dokumentami. Metadanniye dlya dokumentov [Document management processes. Metadata for documents].
18. GOST R ISO 15489-1-2007 Sistema standartov po informacii, bibliotechnomu i izdatel'skomu delu. Upravleniye dokumentami. Obshie trebovaniya [System of standards for information, library and publishing. Document management. General requirements].
19. GOST R 54989-2012 /ISO TR 18492:2005 Ensuring long-term preservation of electronic documents (entered into force with 01.05.2013).
20. Preservation of Evidence of Cryptographically Signed Documents // BSI Technical Guideline TR-03125 – Version 1.2 – Federal Office for Information Security – 2015 – 183 p.

Solovyev A. V. Chief Researcher, Department 94 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

Danilenko A. Yu. Leading Researcher, Department 94 FRC CSC RAS. Moscow, prosp. 60-let Oktyabrya, 9. Candidate of Physical and Mathematical Sciences. Number of publications: 40. Area of scientific interests: system programming, system analysis, information technology, electronic document management, information security, data protection. E-mail: danilenko@isa.ru

Akimova G. P. Leading Researcher, Department 94 FRC CSC RAS. Moscow, prosp. 60-let Oktyabrya, 9. Candidate of Technical Sciences. Number of publications: 65. Area of scientific interests: system programming, system analysis, information technologies, the influence of the human factor, information and analytical systems, electronic document management, electronic archive. E-mail: akimova@isa.ru

Pashkin M. A. Researcher, Department 94 FRC CSC RAS. Moscow, prosp. 60-let Oktyabrya, 9. Number of publications: 20. Area of scientific interests: system programming, information technologies, information and analytical systems, electronic archive. E-mail: pashkin@isa.ru

Podrabinovich A. A. lead programmer, Department 94 FRC CSC RAS. Moscow, prosp. 60-let Oktyabrya, 9. Number of publications: 15. Area of scientific interests: system programming, design and creation of methods and software for managing electronic documents, information security. E-mail: podrabinovich@isa.ru

Tumanova I. V. lead programmer, Department 94 FRC CSC RAS. Moscow, prosp. 60-let Oktyabrya, 9. Number of publications: 10. Area of scientific interests: system programming, information technology, electronic document management, electronic archive. E-mail: tumanova-irin@mail.ru