

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего профессионального образования  
«Российский государственный гуманитарный университет»  
(РГГУ)

УДК 3:002.6  
№ госрегистрации 01201364001  
Инв. №

УТВЕРЖДАЮ  
Ректор РГГУ

\_\_\_\_\_ Е.И. Пивовар

"\_\_" \_\_\_\_\_ 2013 г.

ОТЧЕТ  
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

Исследование современных процессов документационного обеспечения управления и развития электронного документооборота.

по теме:

**Сравнительный анализ форматов файлов электронных документов постоянного (долговременного) хранения**

Начальник Управления  
по координации вузовских  
проектов и программ

\_\_\_\_\_ Л.Н. Простоволосова  
(подпись, дата)

Руководитель темы

\_\_\_\_\_ П.А. Кюнг  
(подпись, дата)

Москва  
2013

## СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель темы \_\_\_\_\_ П.А. Кюнг, канд. ист. наук  
(подпись, дата)

Исполнители темы \_\_\_\_\_ А.Д. Козлов, канд.техн.наук  
(подпись, дата)

\_\_\_\_\_ И.Г. Силина, канд.ист.наук  
(подпись, дата)

## РЕФЕРАТ

Отчет с приложениями 63 стр., 9 табл., 7 прил., источников 72.

Ключевые слова:

ФОРМАТ ФАЙЛА, ЭЛЕКТРОННЫЙ ДОКУМЕНТ, АРХИВ, ХРАНЕНИЕ ДОКУМЕНТОВ, ЭЛЕКТРОННЫЙ АРХИВ, СИСТЕМА ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА, АРХИВОХРАНИЛИЩА, МАСТЕР-КОПИЯ, НОРМАТИВНАЯ БАЗА, ТЕХНИЧЕСКИЕ СРЕДСТВА, ОБОРУДОВАНИЕ, ОЦИФРОВКА, ОЦИФРОВАНИЕ, МИГРАЦИЯ, МЕТАДАННЫЕ.

Научно-исследовательская работа проводилась в рамках государственного контракта № 016-3 от 31 мая 2013 г. «Исследование современных процессов документационного обеспечения управления и развития электронного документооборота. «Научный доклад «Сравнительный анализ форматов файлов электронных документов постоянного (долговременного) хранения» (в рамках реализации в 2012 г. мероприятий Федеральной целевой программы «Культура России (2012–2018 гг.)»).

Исследование обобщает и систематизирует данные, полученные в результате исследований по теме: Исследование современных процессов документационного обеспечения управления и развития электронного документооборота. «Научный доклад «Сравнительный анализ форматов файлов электронных документов постоянного (долговременного) хранения».

Объект исследования включает в себя форматы файлов электронных документов

Целью научно-исследовательской работы является выявление форматов файлов электронных документов наиболее пригодных для постоянного (долговременного) хранения информации в государственных и муниципальных архивах, архивах организаций с подготовкой рекомендаций.

Методы исследования: Сравнительный анализ форматов наиболее распространенных файлов электронных документов постоянного (долговременного) хранения на основе ряда характеристик, позволяющих определить пригодность рассматриваемых форматов для долговременного хранения информации, а также на основе проведенных экспериментов с преобразованием форматов.

Результаты: Определены форматы файлов, обеспечивающих постоянное (долговременное) хранение документированной информации (электронных документов) в архивах федеральных органов исполнительной власти, организаций, в государственных и муниципальных архивах. Исследование показало, что для постоянного (долговременного)

хранения информации наиболее приспособлены следующие форматы файлов электронных документов:

а) текстовые и форматы представления данных: XML (включая XSD/XSL/XHTML), PDF/A-1 (ISO 19005-1) (\*.pdf); PDF/A-2 (ISO 19005-2:2011) (\*.pdf);

б) графические: TIFF (без компрессии), JPEG2000 (lossless) (\*.jp2), PNG (\*.png), SVG (no Java script binding) (\*.svg) (векторное изображение);

в). аудио форматы: AIFF (PCM) (\*.aif, \*.aiff), WAV (PCM) (\*.wav), AES3 (LPCM) (\*.aes)

г). видео форматы: Motion JPEG 2000 (ISO/IEC 15444-4) (\*.mj2), AVI (uncompressed, motion JPEG) (\*.avi), QuickTime Movie (uncompressed, motion JPEG) (\*.mov)

д). базы и банки данных: Comma Separated Values (\*.csv); Delimited Text (\*.txt); SQL DDL

Рекомендации по внедрению: Разработка государственных требований к проектированию систем электронного документооборота органов государственной исполнительной власти, рекомендаций по работе государственных и муниципальных архивов с электронными документами при их приеме на постоянное хранение.

Область применения: Федеральное архивное агентство, государственные и муниципальные архивы, архивы организаций и службы делопроизводства федеральных органов государственной власти, органов местного самоуправления, государственных и негосударственных организаций.

## СОДЕРЖАНИЕ

ОПРЕДЕЛЕНИЯ	6
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	9
ВВЕДЕНИЕ	10
ОСНОВНАЯ ЧАСТЬ	17
ЗАКЛЮЧЕНИЕ	20
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	23
ПРИЛОЖЕНИЕ А «НАУЧНЫЙ ДОКЛАД «СРАВНИТЕЛЬНЫЙ АНАЛИЗ ФОРМАТОВ ФАЙЛОВ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ ПОСТОЯННОГО (ДОЛГОВРЕМЕННОГО) ХРАНЕНИЯ»	31
ПРИЛОЖЕНИЕ Б «СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ ФОРМАТОВ ТЕКСТОВЫХ ДАННЫХ»	50
ПРИЛОЖЕНИЕ В «СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ ФОР- МАТОВ ПРЕДСТАВЛЕНИЯ ДАННЫХ»	52
ПРИЛОЖЕНИЕ Г «СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ ГРА- ФИЧЕСКИХ ФОРМАТОВ (РАСТРОВЫЕ БЕЗ ПОТЕРЬ)»	56
ПРИЛОЖЕНИЕ Д «СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ ГРАФИЧЕСКИХ ФОРМАТОВ (ВКЛЮЧАЯ СЖАТИЕ С ПОТЕРЯМИ И АНИМАЦИОННЫЕ)»	58
ПРИЛОЖЕНИЕ Е «СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ ГРА- ФИЧЕСКИХ ФОРМАТОВ (ВЕКТОРНАЯ ГРАФИКА)»	60
ПРИЛОЖЕНИЕ Ж «ПРИМЕРНАЯ ОЦЕНКА РАСХОДОВ НА ПРОГРАММНОЕ И АППАРАТНОЕ ОБЕСПЕЧЕНИЕ»	

## ОПРЕДЕЛЕНИЯ

В настоящем отчете о НИР использованы следующие определения терминов:

Архиватор – это программа для процедуры сжатия или упаковки электронных документов. Почти каждый архиватор одновременно является программой для сжатия данных. С другой стороны, любая программа для сжатия данных может рассматриваться как архиватор. Эффективность сжатия является важнейшей характеристикой архиваторов. От нее зависит размер создаваемых архивов. Чем меньше архив, тем меньше места требуется для его хранения. Для передачи нужна меньшая пропускная способность канала передачи или затрачивается меньшее время. Преимущества архивов очевидны, если учесть, что данные уменьшаются в размере и в 2 раза, и в 5 раз.

Архивный документ – материальный носитель с зафиксированной на нем информацией, который имеет реквизиты, позволяющие его идентифицировать, и подлежит хранению в силу значимости указанного носителя и информации для граждан, общества и государства.

Единица учета электронного фонда пользования – файл или совокупность файлов, составляющих образ дела, за единицу измерения файла принимается байт (Мб, Гб, Тб).

Единица хранения электронных документов – электронное дело, содержащее электронные документы и метаданные к ним, сформированное в соответствии с номенклатурой дел.

Информационная система – информация в электронной форме и совокупность обеспечивающих ее обработку информационных технологий и технических средств.

Конвертация (электронных документов) - процесс перемещения документов с одного носителя на другой или из одного формата в другой. ГОСТ 15489–1–2007 «Управление документами. Общие требования».

Мастер-копия – «идеальная» копия, сделанная с подлинника документа, эталон. В терминах данных методических рекомендаций – первая цифровая копия (копия-оригинал, подлинник), сделанная с подлинника архивного документа и не подвергавшаяся никакой обработке, в том числе многократной перезаписи (миграции, репликации) на другой носитель информации. Мастер-копия является неприкосновенной и не предназначена для использования.

Метаданные – данные, описывающие контекст, содержание и структуру документов и этапы работы с документами на протяжении определенного времени;

Метаданные – данные, описывающие контекст, содержание, структуру документов и управление документами в течение времени. ГОСТ 15489–1–2007 «Управление документами. Общие требования».

Миграция (электронных документов) – действие по перемещению записей из одной системы в другую систему при сохранении их аутентичности, целостности, надежности и удобства в использовании;

Миграция цифровой информации – процесс периодического переноса цифровых материалов с одной аппаратно-программной конфигурации на другую, или с одного поколения компьютерных технологий на следующее (перевод в другой формат, из одной операционной системы в другую, с одного языка программирования на другой и пр.) с сохранением всех исходных функциональных характеристик, обеспечивающий долговременный доступ к цифровым объектам независимо от смены технологий. В настоящее время общепризнано, что миграция является неотъемлемой функцией цифровых архивов. Согласно ГОСТ 15489–1–2007 «Управление документами. Общие требования»: Миграция – действие по перемещению документов из одной системы в другую с сохранением аутентичности, целостности, достоверности документов и их пригодности для использования.

Обеспечение сохранности – процессы и операции по обеспечению технической и интеллектуальной аутентичности документов в течение времени. ГОСТ 15489–1–2007 «Управление документами. Общие требования».

Оцифрованный документ – документ, полученный в результате конверсии бумажных и других аналоговых документов в цифровую форму. Первичное средство хранения (первичная упаковка) – папка, конверт, коробка, футляр, кожух, предназначенные для физической защиты CD (DVD и т.п.) от механических повреждений.

Оцифровка (англ. digitization, оцифрование) – описание объекта, изображения или аудио- видеосигнала (в аналоговом виде) в виде набора дискретных цифровых замеров (выборок) этого сигнала/объекта, при помощи той или иной аппаратуры (сканеров, цифровых фотоаппаратов и т.п.), т.е. преобразование документа из традиционной, присущей ему формы, в цифровую (компьютерную) в виде электронного файла (файлов) данных, пригодных для записи на электронные носители. Результатом оцифровки документов являются – медиафайлы (графические файлы, аудиофайлы, видеофайлы и т.п.), которые могут быть воспроизведены различными цифровыми устройствами.

Репликация – метод сохранения цифровой информации, предполагающий создание одной или нескольких полных копий (клонов) цифровых материалов. Репликация требует специальных мероприятий по поддержке всех копий в актуальном состоянии.

Управление документами — совокупность действий по созданию, использованию, хранению и уничтожению документов в организациях, осуществляющих деловую деятельность планомерно и эффективно, в целях доказательства проведения деловых (управленческих) операций. ГОСТ 15489–1–2007 «Управление документами. Общие требования».

Файл (англ. file) - блок информации на запоминающем устройстве компьютера, имеющий определённое логическое представление (начиная от простой последовательности битов или байтов и заканчивая объектом сложной СУБД)

Формат - спецификация структуры данных, записанных в компьютерном файле

Цифровой документ — зафиксированная на цифровом носителе информация, которая записывается, сохраняется, передается и представляется в приемлемой для человека форме с помощью цифровых технологий и устройств.

Электронная копия документа – копия, изготовленная в цифровой форме. ГОСТ 7.48–2002. «Консервация документов. Основные термины и определения».

Электронный документ – 1) документ, в котором информация представлена в электронно-цифровой форме; 2) электронный документ - документированная информация, представленная в электронной форме, то есть в виде, пригодном для восприятия человеком с использованием электронных вычислительных машин, а также для передачи по информационно-телекоммуникационным сетям или обработки в информационных системах. Федеральный закон от 27.07.2006 N 149-ФЗ (ред. от 02.07.2013) «Об информации, информационных технологиях и о защите информации»



## ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ПО – программное обеспечение.

CRC (Cyclical Redundancy Code) - Циклический избыточный код

HTML (Hyper Text Markup Language) - Язык разметки гипертекста

JPEG (Joint Photographic Engineers Group) - Объединенная группа специалистов по компьютерной обработке фотографических изображений

OCR (Optical Character Recognition) - Оптическое распознавание символов

PDF/A-1; PDF /A-2 (Portable Document Format - Archive) - Варианты для архивного хранения формата переносимого документа

SHA-1 (Standard Hash Algorithm 1) - Стандартный алгоритм вычисления хэш-функции №1

TIFF (Tagged Image File Format) - Тегированный формат файлов изображений

XML (Extensible Markup Language) - Расширяемый язык разметки

## ВВЕДЕНИЕ

В настоящее время наблюдается непрерывный рост документирования информации в электронной форме. При этом необходимо отметить недостаточность структурированной и достоверной информации о приспособленности определенных форматов файлов, содержащих электронные документы к долговременному хранению в архивах вне электронных систем, в которых они были созданы. Данная ситуация вызывает закономерную обеспокоенность специалистов, отвечающих за хранение информации, т.к. они не в состоянии предсказать состояние массивов электронных данных, поступающих в государственные и муниципальные архивы, на долговременную перспективу. К сожалению, на данный момент в России отсутствуют научно-исследовательские разработки в этой области, доступные архивным специалистам. Поэтому сравнительная оценка форматов файлов с целью определения их пригодности для постоянного (долговременного) хранения информации является, несомненно, актуальной.

Представленное исследование проведено в рамках изучения современных процессов документационного обеспечения управления и развития электронного документооборота. Результаты научно-исследовательской работы предназначены для определения форматов файлов, обеспечивающих постоянное (долговременное) хранение документированной информации (электронных документов) в архивах федеральных органов исполнительной власти, организаций, в государственных и муниципальных архивах.

Субъектами использования результатов научно-исследовательской работы являются Федеральное архивное агентство, государственные и муниципальные архивы, архивы организаций и службы делопроизводства федеральных органов государственной власти, органов местного самоуправления, государственных и негосударственных организаций.

Состав и объем работ в соответствии с техническим заданием:

1. Определение эффективности и целесообразности использования форматов файлов для архивного хранения текстовых, графических и других электронных документов.
2. Сравнение форматов файлов электронных документов по следующим критериям:
  - разработчик и права собственности;
  - год разработки формата файлов; год его последней модернизации; частота обновлений программного кода; наличие версий;
  - открытый/ закрытый программный код;
  - описание и сравнение математических алгоритмов сжатия;
  - предназначение формата файлов (в т.ч.: класс задач и предметная область; присутствие в пакетах прикладного программного обеспечения);

- наличие и доступность конвертеров;
- устойчивость при множественных репликациях (на основе анализа суммы программных кодов).

3. Проведение технико-экономической оценки и сравнительного анализа затрат на архивное хранение электронных документов в указанных форматах файлов.

4. Сравнительный анализ объемов информации, подлежащей архивному хранению, при условии хранения электронных архивных документов в указанных форматах файлов.

Результаты работы соответствуют и логически дополняют следующие требования, стандарты и ГОСТы:

- Распоряжения Правительства Российской Федерации от 02.10 2009 г. № 1403-р «Технические требования к организации взаимодействия системы межведомственного электронного документооборота с системами электронного документооборота федеральных органов исполнительной власти»;

- Приказа Минкомсвязи России от 02.09.2011 № 221 «Об утверждении Требований к информационным системам электронного документооборота федеральных органов исполнительной власти, учитывающих, в том числе, необходимость обработки посредством данных систем служебной информации ограниченного распространения»;

- Правил организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской Академии наук (утв. приказом Министерства культуры и массовых коммуникаций Российской Федерации от 18 января 2007 г. № 19);

- Рекомендаций по подготовке федеральными органами власти перечней документов, создание, хранение и использование которых должно осуществляться в форме электронных документов при организации внутренней деятельности (утв. приказом Росархива от 29.04.2011 № 32);

- ГОСТ Р ИСО 15489-1-2007 «СИБИД. Управление документами. Общие требования»;

- ГОСТ Р ИСО 23081-1-2008 «СИБИД. Управление документами. Процессы управления документами. Метаданные для документов»; Часть 1?

- ГОСТ Р ИСО 22310-2009 «СИБИД. Информация и документация. Руководство для разработчиков стандартов, устанавливающих требования к управлению документами»;

- ГОСТ Р 53898-2010 «Системы электронного документооборота. Взаимодействие систем управления документами. Требования к электронному сообщению»;

- ГОСТ Р 54471-2011 «Системы электронного документооборота. Управление документацией. Информация, сохраняемая в электронном виде. Рекомендации по обеспечению достоверности и надежности»;
- ГОСТ Р 54989-2012/ISO/TR 18492:2005 «Обеспечение долговременной сохранности электронных документов» (вводится в действие с 01.05.2013);
- МГС ГОСТ 7.-32-2001 «СИБИД. Отчет о научно-исследовательской работе. Структура и правила оформления».

Анализ нормативных документов, указанных в техзадании позволил сформулировать основные установленные требования к форматам электронных документов долговременного хранения:

1. Форматы файлов электронных документов являются частью системы классификации документной информации в архивах. В данном аспекте, форматы выступают в качестве признака классификации, определяющего условия воспроизведения и считывания информации пользователем в долговременной перспективе<sup>1</sup>.
2. При оценке технического состояния электронных документов, а также их страховых копий устанавливается возможность считывания информации с материального носителя, соответствие технических характеристик этих документов сведениям, содержащимся в сопровождающих актах технического состояния. Исходя из данного пункта, формат файла, определенный для постоянного (долговременного) хранения должен соответствовать техническим возможностям органов государственного власти, государственных и муниципальных архивов по воспроизведению информации в долговременной перспективе<sup>2</sup>.
3. Проверку технических параметров электронных документов проводят с использованием специальной аппаратуры и программного обеспечения, предназначенных для работы с соответствующего вида архивными документами. Данный пункт предполагает, что программное обеспечение для воспроизведения файлов электронных документов должно быть доступно организациям- хранителям документов, что предполагает недопустимость использования в качестве утвержденных форматов файлом, форматы, образуемые дорогостоящим, малораспространенным или неподдерживаемым программным обеспечением. Данный вывод подтверждается положением, изложенным в п 4.4.3. нижеуказанных Правил «Электронные документы принимаются на хранение в сопровождении программных

---

<sup>1</sup> П. 2.10.Правила организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской академии наук, утвержденных Приказом Минкультуры РФ от 18.01.2007 N 19(Зарегистрировано в Минюсте РФ 06.03.2007 N 9059)

<sup>2</sup> П. 2.11.6.1. Правила организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской академии наук, утвержденных Приказом Минкультуры РФ от 18.01.2007 N 19(Зарегистрировано в Минюсте РФ 06.03.2007 N 9059)

средств, позволяющих их воспроизвести, и с необходимым комплектом сопроводительной документации»<sup>3</sup>.

4. СЭД ФОИВ должна обеспечивать отображение следующих форматов файлов: pdf, rtf, doc, tiff. СЭД ФОИВ допускается обеспечение отображения иных форматов файлов, используемых в деятельности ФОИВ<sup>4</sup>. Данный пункт задает основные ориентиры исследования форматов файлов электронных документов, образующихся в ФОИВ и передаваемых на хранение в архивы различного уровня.

Согласно ГОСТ Р ИСО 15489-1-2007. Национальный стандарт Российской Федерации. Система стандартов по информации, библиотечному и издательскому делу. Управление документами. Общие требования(утв. Приказом Ростехрегулирования от 12.03.2007 N 28-ст) устанавливает следующие требования к документам, предполагающие определенные запросы к форматам файлов электронных документов:

#### 7.2.5. Пригодность для использования

7. Пригодным для использования является документ, который можно локализовать, найти, воспроизвести и интерпретировать.

#### 8.3.5. Конвертирование и миграция

8. Системы управления документами следует проектировать так, чтобы документы оставались аутентичными, надежными и пригодными для использования независимо от любых изменений в системе управления документами (в том числе при конвертировании форматов, миграции между техническим оборудованием, операционными системами или специальными программными приложениями) в течение всего периода их хранения.

#### 9.6. Хранение и обращение с документами

9. Организации должны разрабатывать и применять инструкции по конвертированию и миграции документов из одной системы управления документами в другую.»

Электронные системы управления документами должны быть спроектированы так, чтобы документы оставались доступными, аутентичными, достоверными и пригодными для использования независимо от любых изменений в системе на протяжении всего периода их хранения и использования. Изменения в электронной системе управления документами могут включать миграцию, конвертирование, воспроизведение в других форматах или любые другие будущие способы воспроизведения документов.

---

<sup>3</sup> 2.11.6.2. Правила организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской академии наук, утвержденных Приказом Минкультуры РФ от 18.01.2007 N 19(Зарегистрировано в Минюсте РФ 06.03.2007 N 9059)

<sup>4</sup> Требования к информационным системам электронного документооборота федеральных органов исполнительной власти, учитывающих в том числе необходимость обработки посредством данных систем служебной информации ограниченного распространения, утвержденные Приказом Минкомсвязи РФ от 02.09.2011 N 221

Электронное сообщение должно состоять из XML-документа установленной в настоящем стандарте структуры и с заданным составом элементов и их атрибутов и, возможно, набора дополнительных файлов, являющихся неотъемлемой частью сообщения.

В Рекомендациях по подготовке федеральными органами исполнительной власти перечней документов, создание, хранение и использование которых должно осуществляться в форме электронных документов при организации внутренней деятельности, утвержденных Приказом Росархива от 29.04.2011 N 32, приведен рекомендуемый Перечень видов управленческих документов, создание, хранение, использование которых в федеральных органах исполнительной власти осуществляется в форме электронных документов при организации внутренней деятельности. Данный перечень предполагает присутствие в системах федеральных органов исполнительной власти следующих типов форматов документов: pdf, doc, rtf, txt, xml.

В свою очередь «Система стандартов по информации, библиотечному и издательскому делу. Процессы управления документами. Метаданные для документов. Часть 1. Принципы. ГОСТ Р ИСО 23081-1-2008» (утв. Приказом Ростехрегулирования от 13.11.2008 N 310-ст) устанавливает основы для создания, управления и использования метаданных об управлении документами, а также объясняет принципы, лежащие в основе управления ими.

Структура описания документа состоит из описания:

- а) его физической или технической структуры;
- б) его логической структуры, то есть отношений между элементами данных, составляющих документ.

Все метаданные о документе формируют новый документ, а именно документ метаданных, которым также нужно управлять. Документ метаданных, равно как и сами метаданные, необходимо сохранять до тех пор, пока существует оригинал документа.

В соответствии с ГОСТ Р 53898-2010 «Системы электронного документооборота. Взаимодействие систем управления документами. Требования к электронному сообщению» устанавливается формат, состав и содержание электронного сообщения, обеспечивающего информационное взаимодействие систем управления документами. Согласно стандарту электронное сообщение «должно состоять из XML-документа установленной в настоящем стандарте структуры и с заданным составом элементов и их атрибутов и, возможно, набора дополнительных файлов, являющихся неотъемлемой частью сообщения. Как электронное сообщение в целом, так и составные его части, могут быть независимо подписаны электронными цифровыми подписями и другими аналогами собственноручной подписи и (или) зашифрованы».

В свою очередь ГОСТ Р 54471-2011 «Системы электронного документооборота. Управление документацией. Информация, сохраняемая в электронном виде. Рекомендации по обеспечению достоверности и надежности» описывает порядок внедрения и эксплуатации систем управления информацией и документами, которые могут рассматриваться как надежно, заслуживающим доверия образом, хранящие электронную информацию. При рассмотрении требований к файловым форматам и сжатию данных стандарт следует опираться на пп. 3.2.4, 3.2.2, 5.5.2, 5.10, 6.10 и 7.2.3. Политика должна содержать сведения о допустимых файловых форматах, которые могут быть использованы для каждого типа информации (см. 3.2.2). Благодаря реализации на практике политики использования утвержденных файловых форматов и (если таковые применяются) технологий сжатия данных могут быть успешно выполнены необходимая миграция данных или альтернативные процедуры, обеспечивающие долговременное использование хранимой информации (см. п.3.2.4), а положения, регламентирующие хранения версий информации и документов, должны быть включены в политику. При этом обеспечение долговременной сохранности информации напрямую зависит от задокументированных процедур обеспечения долговременной сохранности информации. Стандарт требует, чтобы такие процедуры учитывали требуемые сроки хранения и ожидаемый срок службы систем хранения данных. Если срок хранения превышает ожидаемый срок существования систем хранения данных, то необходимо задокументировать планы миграции в новые системы (см. п. 5.10) Дополнительную информацию об этом также содержит ISO/TR 18492.

Надежным методом решения потенциальной проблемы миграции является обеспечение хранения электронных файлов в стандартном для отрасли формате либо для поддержания программ просмотра (viewers) для каждого формата, используемого для хранения информации (см. п. 6.10). В случае же конверсии информации из одного файлового формата в другой сведения о конверсии должны быть сохранены в составе контрольной информации (см. п.7.2.3.). Например, электронный документ, созданный текстовым процессором, может быть преобразован в графический формат без изменения текста документа. С одной стороны, это не слишком сильно отличается от копирования файла, однако если форматирование имеет значение для информационного контента, то существует вероятность того, что информационный контент преобразованного файла можно будет считать изменившимся.

Таким образом, проведенный анализ нормативных актов выявил общие требования, устанавливаемые к файлам электронных документов, предназначенных для долговременного хранения:

1. Устойчивость к многократным миграциям.

2. Устойчивость к многократным репликациям
3. Доступность программного обеспечения.
4. Существующие нормативные документы предусматривают конкретно следующие форматы файлов: pdf, tiff, doc, rtf, xml. Нормативные документы предусматривают также возможность присутствия в электронных системах органов государственной власти и иных форматов файлов электронных документов.

В процессе исследования была разработана методология оценки форматов файлов электронных документов на основе критериев, изложенных в техзадании, а также определенных в ходе исследования в качестве принципиальных. Это позволило достигнуть верифицируемых результатов, выраженных в конкретных предложениях по использованию форматов файлов для постоянного (долговременного) хранения электронных документов.

В соответствии с Техническим заданием, основная часть исследования представлена в виде научного доклада, прилагаемого к настоящему отчету (Приложение А), содержащего текстовую часть, таблицы сравнений форматов файлов электронных документов в Приложениях Б, В, Г, Д, Е, В, Ж. В Отчете НИР представлены: методология, основные этапы исследования, результаты исследования с рекомендациями.



## ОСНОВНАЯ ЧАСТЬ

Выбор средств кодирования для хранения информации зависит, в первую очередь, от характеристик технических средств хранения, а также от вида информации, потребностей доступа пользователей к информации и к соответствующим программным средствам.

Основной массив хранимой информации представляет собой текстовые данные различного объёма. По наименованиям единиц хранения подобные источники нарративной информации составляют более 90%. В настоящее время к такой информации, помимо текстов на различных языках человеческого общения, относятся и тексты компьютерных программ как на языках программирования, так и непосредственно в двоичных кодировках. Тексты, записанные в алфавитных системах, зачастую включают значительные фрагменты графической информации (иллюстрации). Особняком стоят тексты, записанные в неалфавитной (например, иероглифической) форме, которые по свойствам хранения ближе к чисто графической информации. Текстовая информация может храниться в оригинальном виде, а также в сжатой форме при помощи различного вида архиваторных программных средств.

Исходя из целей исследования и технического задания, а также результатов анализа нормативной базы, для исследования были отобраны следующие форматы файлов:

1. Текстовые и представления данных
2. Табличные
3. Графические
4. Аудио и Видео

Сравнение проведено по следующим параметрам:

1. Расширение
2. Стандартизация
3. Дата создания
4. Дата последних изменений
5. Открытость формата
6. Назначение формата
7. Наличие бесплатного программного обеспечения
8. Самодостаточность документа (сам документ содержит все необходимое для его корректного отображения)
9. Возможность полнотекстового поиска

10. Наличие «опасных» возможностей (встроенного кода, шифрования, средств управления правами доступа и т.д.
11. Сжатие графических объектов с потерями
12. Возможность конвертации из других форматов

Параметры были определены в соответствии с техническим заданием, нормативной базой и принципиальными параметрами электронных систем, определенными рабочей группой.

Были проведены эксперименты с наиболее распространенными и приемлемыми для постоянного (долговременного) хранения форматами. Для них были применены методики устойчивости к многократной архивации, через проверку путем многократного архивирования и разархивирования выбранных файлов наиболее распространенными программами архивации.

Для сравнения эффективности хранения документов в форматах PDF и DjVu проводился эксперимент по преобразованию исходного текста в формате MS WORD независимо в форматы PDF и DjVu. Далее оценивались коэффициенты сжатия объема текстовых файлов после преобразования в графические форматы; они сводились в два отдельных числовых массива. В третий массив сводилось отношение коэффициентов сжатия объема текстовых файлов при использовании двух графических форматов. По статистическим данным трёх таблиц рассчитывались несмещённые оценки выборочных средних значений по формулам:

$$A_{\text{ср}} = \left( \sum_{i=1}^n a_i \right) / n ,$$

где  $a_i$  – экспериментальное значение параметра из таблицы,

$A_{\text{ср}}$  – полученное среднее значение,

$n$  – размер таблицы.

Качество просмотра определялось путем оценке воспроизведения документов на устройствах визуализации с различными разрешениями и работающих на различных программных платформах.



## ЗАКЛЮЧЕНИЕ

Достигнутые результаты НИР, выполненной по теме «Исследование современных процессов документационного обеспечения управления и развития электронного документооборота. «Научный доклад «Сравнительный анализ форматов файлов электронных документов постоянного (долговременного) хранения» основаны на обобщении значительного объема современных данных по цифровым форматам, нормативным документам, стандартам и гостам, а также опытных исследований ряда форматов для оценки важных показателей по сжатию, кодированию, миграции. Исследование соответствует современным требованиям отечественных нормативных, методических документов, а также достигнутым результатам в зарубежной практике.

Исходя из анализа разнородной цифровой информации можно рекомендовать к использованию следующие основные рекомендованные форматы с точки зрения их надежности и долговечности:

1. Текстовые форматы файлов и форматы представления данных обладающие высокой степенью надежности и обеспечивающие долгосрочность хранения информации:

- XML (includes XSD/XSL/ XHTML)
- PDF/A-1 (ISO 19005-1) (\*.pdf)
- PDF/A-2 (ISO 19005-2:2011) (\*.pdf)

Исходя из проведенных сравнений между форматами PDF и его производным PDF/A, можно утверждать, что первый больше пригоден для оперативного обмена и краткосрочного хранения электронных документов, в свою очередь как PDF/A (с учетом потенциально большего размера единичного документа), имея статус международного стандарта, гарантирует, что даже через продолжительное время, вне зависимости от окружения и операционной системы, любой пользователь сможет открыть документ в данном формате, располагая ПО-просмотрщиком. В целом это соответствует концепции архива электронных документов.

В некоторыми условиями и, в крайнем случае, могут быть предложены в категории текстовых форматов файлов:

- Plain text (ISO 8859-1encoding)
- PDF (\*.pdf) (embedded fonts)
- Rich Text Format 1.x (\*.rtf)
- HTML (include a DOCTYPE declaration)
- SGML (\*.sgml)

- Open Office (\*.sxw/\*.odt)
- OOXML (ISO/IEC DIS 29500) (\*.docx)
- EPUB (unencrypted) (\*.epub)

2. Графические форматы рекомендованы и отвечают требованиям долговременного архивного хранения:

2.1. Растровые изображения:

- TIFF (без компрессии/uncompressed)
- JPEG2000 (lossless) (\*.jp2)
- PNG (\*.png)

2.2. Векторные изображения:

- SVG (no Java script binding) (\*.svg)

2.3. Виртуальные форматы данных:

- X3D (\*.x3d)

3. Рекомендованы форматы сохранения аудио и видео информации:

3.1. Аудио форматы:

- AIFF (PCM) (\*.aif, \*.aiff)
- WAV (PCM) (\*.wav)
- AES3 (LPCM) (\*.aes)

3.2. Видео форматы:

- Motion JPEG 2000 (ISO/IEC 15444-4) (\*.mj2)
- AVI (uncompressed, motion JPEG) (\*.avi)
- QuickTime Movie (uncompressed, motion JPEG) (\*.mov)

4. Базы и банки данных рекомендованы к сохранению в форматах:

- Comma Separated Values (\*.csv)
- Delimited Text (\*.txt)
- SQL DDL

В редких случаях могут быть использованы форматы:

- DBF (\*.dbf)
- OpenOffice (\*.sxc/\*.ods)
- OOXML (ISO/IEC 29500)(\*.xlsx)

Стоит отметить, что форматы файлов по презентациям не имеют формата с достаточно высокой надежностью обеспечения долговременного хранения, однако можно рекомендовать с определенными оговорками:

- OpenOffice (\*.sxi/\*.odp)
- OOXML (ISO/IEC DIS 29500) (\*.pptx)

Форматы других презентационных файлов, в т.ч. PowerPoint (\*.ppt), не отвечают требованиям долговременного хранения.

Можно рекомендовать следующие технические параметры и параметры контроля к форматам файлов при их передачи на архивное хранения предусматривают:

1. Отсутствие файлов с вирусами (особенно это относится к форматам файлов: DOC, XLS, MDB, PDF, PPT, ZIP, EXE).

2. Передаваемые на архивное хранение архивные файлы не могут быть полностью или частично: зашифрованы, защищены паролем, сжаты собственным алгоритмом сжатия, кодированы.

3. Архивирование файлов рекомендуется в незашифрованном виде ZIP (\*.zip) – наиболее соответствующий формат для передачи данных.

оценку технико-экономической эффективности внедрения

Технические требования к серверному оборудованию, системам хранения электронного контента определяются, исходя из реальных объемов имеющегося цифрового контента, перспектив его наращивания и нужд в осуществлении миграций, распечатки, объемов обработки электронных документов. Наиболее критичные требования к объёму серверной памяти предъявляются в случае архивного хранения видеoinформации с высоким качеством изображения. Аналогично, и передача хранимых видеофайлов из архива пользователю требует максимальной пропускной способности линий связи. По сравнению с видеoinформацией аналогичными требованиями архивов текстов и статических изображений можно пренебречь. Предлагаемые форматы создаются преимущественно на открытом программном обеспечении и позволяют минимизировать статьи расходов на приобретение программного обеспечения.

Стратегия выбора надежных форматов файлов также осуществляется исходя из возможности использовать форматы независимые от платформы, независимых поставщиков, неспециализированные, открытые и хорошо стабильно поддерживаемые форматы.

оценка научно-технического уровня выполненной НИР в сравнении с лучшими достижениями в данной области

Доклад в полной мере раскрывает структуру и сущность цифровых форматов, эффективность их использования в обеспечении долговременной сохранности электронных документов в архиве, чем имеющиеся аналитические данные зарубежных коллег. Например, в марте 2012 г. Университет Флориды (США) подготовил для Цифрового архива Флориды результаты своего исследования только в виде списка ранжирования цифровых форматов без детального описания нюансов.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

### ФЕДЕРАЛЬНЫЕ ЗАКОНЫ И ПОДЗАКОННЫЕ АКТЫ

1. Федеральный закон от 22.10.2004 № 125-ФЗ «Об архивном деле в Российской Федерации» (ред. Федеральных законов от 04.12.2006 № 202-ФЗ, от 01.12.2007 № 318-ФЗ, от 13.05.2008 № 68-ФЗ, от 08.05.2010 № 83-ФЗ, от 27.07.2010 № 227-ФЗ).
2. Федеральный закон от 14.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации», (ред. Федеральных законов от 27.07.2010 № 227-ФЗ, от 06.04.2011 № 65-ФЗ, от 21.07.2011 № 252-ФЗ, от 28.07.2012 № 139-ФЗ).
3. Федеральный закон от 09.02.2009 № 8-ФЗ «Об обеспечении доступа к информации о деятельности государственных органов и органов местного самоуправления». (ред. Федерального закона от 11.07.2011 № 200-ФЗ).
4. Федеральный закон от 27.07.2010 № 210-ФЗ «Об организации предоставления государственных и муниципальных услуг». (ред. Федеральных законов от 06.04.2011 № 65-ФЗ, от 01.07.2011 № 169-ФЗ, от 11.07.2011 № 200-ФЗ, от 18.07.2011 № 239-ФЗ, от 03.12.2011 № 383-ФЗ, с изм., внесенными Федеральным законом от 27.06.2011 № 162-ФЗ).
5. Федеральный закон от 30.03.2011 № 63-ФЗ «Об электронной подписи» (в ред. Федеральных законов от 01.07.2011 № 169-ФЗ, от 10.07.2012 № 108-ФЗ).
6. Соглашение от 9.12. 2010. «О единых принципах регулирования в сфере охраны и защиты прав интеллектуальной собственности».  
Постановление Правительства Российской Федерации от 15 июня 2009 г. № 477 «Об утверждении Правил делопроизводства в федеральных органах исполнительной власти» (Собрание законодательства Российской Федерации, 2009, № 25, ст. 3060; 2011, № 37, ст. 5263);
7. Постановление Правительства Российской Федерации от 22 сентября 2009 г. № 754 «Об утверждении Положения о системе межведомственного электронного документооборота» (Собрание законодательства Российской Федерации, 2009, № 39, ст. 4614; 2012, № 38, ст. 5102).

### НОРМАТИВНЫЕ И МЕТОДИЧЕСКИЕ ДОКУМЕНТЫ

1. Правила организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской ака-

демии наук. Утв. приказом Министерства культуры и массовых коммуникаций Российской Федерации от 18.01.2007 № 19 (Бюллетень нормативных актов федеральных органов исполнительной власти, 2007, № 20; «Российская газета», 20.05.2009, № 89).

2. Приказ Министерства Культуры и массовых коммуникаций Российской Федерации от 10 сентября 2007 г. № 1273 «Об утверждении форм учетных и иных документов по организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской Академии наук».

3. Требования к информационным системам электронного документооборота федеральных органов исполнительной власти, учитывающих, в том числе необходимость обработки посредством данных систем служебной информации ограниченного распространения (утверждены приказом Минкомсвязи России от 02.09.2011 № 221) («Российская газета», 21.11.2011 № 261)

4. Рекомендации по подготовке федеральными органами исполнительной власти перечней документов, создание, хранение и использование которых должно осуществляться в форме электронных документов при организации внутренней деятельности (утверждены приказом Росархива от 29.04.2011 № 32);

## ГОСТЫ И МЕЖДУНАРОДНЫЕ СТАНДАРТЫ

1. ISO 12234-2:2001. Электронная обработка изображений фотоснимков. Сменное запоминающее устройство. Часть 2. Формат данных изображения TIF/EP
2. ISO 12651:1999. Электронная обработка изображений. Словарь.
3. ISO 12652—1 (Part 1—2) Electronic imaging.
4. ISO 12653—1:2000. Электронная обработка изображений. Контрольное задание для черно-белого сканирования офисных документов. Часть 1. Характеристики (Тест-объект для сканирования черно-белых офисных документов — Часть 1: Характеристики).
5. ISO 12653—2:2000. Электронная обработка изображений. Контрольное задание для черно-белого сканирования офисных документов. Часть 2. Метод применения (Электронное формирование изображения — Тест-объект для сканирования черно-белых офисных документов. Методы применения).
6. ISO 14721:2003 — Reference Model for an Open Archival Information System (OAIS).
7. ISO 19005-1:2005 Стандарт PDF/A-1
8. ISO 19005-2:2011 Стандарт PDF/A-2



9. ISO 29861. Прикладные системы управления документами. Контроль качества сканирования цветных учрежденческих документов.
10. ISO 37.080: Document imaging applications.
11. ISO 5127—2002. Информация и документация. Словарь.
12. ISO 9001—2000. Системы менеджмента качества. Требования.
13. ISO/IEC 10995:2011. Информационные технологии. Цифровые запоминающие среды для обмена и хранения информации. Метод испытания для оценки срока хранения в архиве оптических носителей.
14. ISO/IEC 14473:1999. Информационные технологии. Офисное оборудование. Минимальная информация, подлежащая регламентации для устройств сканирования изображений.
15. ISO/IEC 29341—9—13:2008. Информационные технологии. Архитектура устройств UPnP. Часть 9—13. Протокол управления устройствами обработки изображений. Сервис сканирования.
16. ISO/TR 12037:1998. Электронная обработка изображений. Рекомендации для удаления информации, записанной на оптических носителях одноразовой записи.
17. ISO/TR 12654:1997. Электронная обработка изображений. Рекомендации для управления системами электронной записи для записи документов, которые могут понадобиться в качестве свидетельских показаний на оптических дисках WORM.
18. ГОСТ 15971—90. Системы обработки информации. Термины и определения.
19. ГОСТ 6.10.4—84. Унифицированные системы документации. Придание юридической силы документам на машинном носителе и машинограмме, создаваемым средствами вычислительной техники. Основные положения.
20. ГОСТ Р 33.1.02—2008. Единый российский страховой фонд документации. Страховые копии кинодокументов и фотодокументов. Общие технические условия.
21. ГОСТ Р 33.3.02—2008 Единый российский страховой фонд документации. Страховые копии документации, являющейся национальным научным, культурным и историческим наследием. Общие требования к условиям хранения.
22. ГОСТ Р 33.505—2003 Единый российский страховой фонд документации. Порядок создания страхового фонда документации, являющейся национальным, научным, культурным и историческим наследием, М., 2003.
23. ГОСТ Р 50922—2006. Защита информации. Основные термины и определения.
24. ГОСТ Р 53898-2010 «Системы электронного документооборота. Взаимодействие систем управления документами. Требования к электронному сообщению».

25. ГОСТ Р 54471-2011 «Системы электронного документооборота. Управление документацией. Информация, сохраняемая в электронном виде. Рекомендации по обеспечению достоверности и надежности».
26. ГОСТ Р 54989-2012/ISO/TR 18492:2005 «Обеспечение долговременной сохранности электронных документов» (вводится в действие с 01.05.2013).
27. ГОСТ Р 6.30—2003. Унифицированные системы документации. Унифицированная система организационно-распорядительной документации. Требования к оформлению документов.
28. ГОСТ Р ИСО 15489-1-2007 «СИБИД. Управление документами. Общие требования».
29. ГОСТ Р ИСО 22310-2009 «СИБИД. Информация и документация. Руководство для разработчиков стандартов, устанавливающих требования к управлению документами».
30. ГОСТ Р ИСО 23081-1-2008 «СИБИД. Управление документами. Процессы управления документами. Метаданные для документов».
31. МГС ГОСТ 7.-32-2001 «СИБИД. Отчет о научно-исследовательской работе. Структура и правила оформления».
32. РД 50—34.698—90. Методические указания. Информационная технология. Комплекс стандартов и руководящих документов на автоматизированные системы. Автоматизированные системы. Требования к содержанию документов.
33. РД 50—680—88. Методические указания. Автоматизированные системы. Основные положения.
34. РД 50—682—89. Методические указания. Информационная технология. Комплекс стандартов и руководящих документов на автоматизированные системы. Общие положения.

#### МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ

1. Рекомендации о порядке учета, оперативного хранения и отбора на постоянное хранение электронных документов. М., ВНИИДАД, 2005.
2. Методические рекомендации по организации хранения, комплектования, учета и использования электронных документов в государственных архивах. М., ВНИИДАД, 2007.
3. Методические рекомендации по работе ЦАДЭНМ с документами на электронных носителях. М., 2006.
4. Инструкция по созданию и организации страхового фонда и фонда пользования копий особо ценных документов. М., РГАДА, 1995.

5. Методические рекомендации по организации хранения, комплектования, учета и использования электронных документов в государственных архивах. М., ВНИИДАД, 2007.
6. Рекомендации по созданию оцифрованных копий фонда пользования фото и фонодокументов. М., РГАНТД, 2008.
7. Методические рекомендации по признанию результатов интеллектуальной деятельности единой технологией. (Утверждены 1.04.2012. Министерство образования и науки Российской Федерации).  
Рекомендации по обеспечению сохранности информации, записанной на оптических дисках (Тестирование выборочного массива документов федеральных архивов). М., РГАНТД, Федеральное Архивное агентство, Авторы, 2011.
8. Рекомендации по обеспечению сохранности информации, записанной на оптических дисках (тестирование выборочного массива документов федеральных архивов). М., РГАНТД, 2011.
9. Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом / Ю.Ю. Юмашева. – М.: ВНИИДАД, 2012. – 125 с.
10. Технический отчет ISO/TR 13028–2010 «Информация и документация – Руководство по организации оцифровки документов» (Information and documentation – Implementation guidelines for digitization of records).
11. Технический отчет ISO/TR 15801:2009 «Управление документацией - Информация, сохраняемая в электронном виде – Рекомендации по обеспечению достоверности и надежности» (Document management – Information stored electronically – Recommendations for trustworthiness and reliability).
12. Оцифрование архивных документов. М., ВНИИДАД, 1998.
13. Отчет ВНИИДАД по теме «Изучение опыта оцифровки (сканирования) документов Архивного фонда в архивных учреждениях Российской Федерации», 2011.
14. Рекомендованные цифровые форматы, предназначенных для долгосрочного хранения цифровой информации» Март 2012 г. («Recommended Data Formats for Preservation Purposes...» [Электронный ресурс.] / State University Library Services/ Florida Virtual Campus. Режим доступа к ресурсу: URL: <http://fclaweb.fcla.edu/uploads/recFormats.pdf/> (дата обращения: 22.10.2013).).
15. Les archives électroniques. Manuel pratique. 2002 (Франция. «Электронные архивные документы. Практическое руководство» Дирекция архивов Франции, Париж, 2002)

16. Guidelines for the Management, Appraisal and Preservation of Electronic Records. Vol. II, Procedures, Ch. 5 Preserving Electronic Records, 1999 (Великобритания. «Руководство по управлению, экспертизе ценности и сохранению электронных документов», том II, Процедуры, глава 5 «Обеспечение сохранности электронных документов»)

## ИССЛЕДОВАНИЯ

1. Войниканис Е.А., Калятин В.О. База данных как объект правового регулирования: Учебное пособие для вузов. М., «Статут», 2011.
2. Залаев г.З. Анализ и классификация электронных документов//Вестник архивиста, 1999, № 2–3(50–51).
3. Корчемкина О.А. Понятие и признаки базы данных как объекта права. // Российский юридический журнал. 2012, № 1.
4. Ларин М.В. Актуальные проблемы обмена документированной информацией в государственном управлении // Документация в информационном обществе: унификация и стандартизация межведомственного и корпоративного документооборота. Доклады и сообщения IX Международной научно-практической конференции, 5-6 декабря 2002 г. М., 2003.
5. Ларин М.В. Обеспечение сохранности электронных документов [Электронный ресурс.] / 1с. Консалтинг. Режим доступа к ресурсу: URL: <http://consulting.1c.ru/journal-article.jsp?id=332>. (дата обращения: 1.06.2013).
6. Михайлов О. А. Электронные документы и архивы: проблемы приёма, обеспечения сохранности и использования. МАКС Пресс, 2002.
7. Сохранение электронной информации в информационном обществе. Сборник материалов Международной конференции (Москва, 3–5 октября 2011 г.) [Электронный ресурс.] / Российский комитет программы ЮНЕСКО «Информация для всех». Режим доступа к ресурсу: [http://www.ifapcom.ru/files/News/Images/2012/dig\\_pres\\_is\\_rus.pdf](http://www.ifapcom.ru/files/News/Images/2012/dig_pres_is_rus.pdf). (Дата обращения: 10.09.2013).
8. Тихонова И.Г., Ганичева С.А., Добрусина С.А., Великанова Т.Д. Особенности хранения оптических компакт-дисков в условиях архивов и библиотек. М., 2003.
9. Юмашева Ю.Ю. Создание электронных копий архивных документов: проблемы и возможные решения. Нормативно-технические и технологические аспекты создания электронных копий документов. Материалы Проблемного семинара и круглого стола в рамках 18-ой Международной Конференции «Крым 2011». «Библиотеки и информационные ре-

сурсы в современном мире науки, культуры, образования и бизнеса». 4–12 июня 2011, г. Судак, Автономная республика Крым. М., ЗАО «ДиМи-Центр», 2011.

10. Helen Heslop, Simon Davis and Andrew Wilson, «An Approach to the Preservation of Digital Records», National Archives of Australia, December 2002. Режим доступа к ресурсу: <http://old.hki.uni-koeln.de/teach/ss07/DL/material/Australia.pdf> (Дата обращения: 07.11.2013).

11. Migration: Context and Current Status Digital Preservation Testbed White Paper (December 2001). Голландия: «Тестовый стенд по обеспечению сохранности цифровой информации»

## ПРИЛОЖЕНИЕ А

### НАУЧНЫЙ ДОКЛАД

#### «СРАВНИТЕЛЬНЫЙ АНАЛИЗ ФОРМАТОВ ФАЙЛОВ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ ПОСТОЯННОГО (ДОЛГОВРЕМЕННОГО) ХРАНЕНИЯ»

Выбор средств кодирования для хранения информации зависит, в первую очередь, от характеристик технических средств хранения, а также от вида информации, потребностей доступа пользователей к информации и к соответствующим программным средствам.

Основной массив хранимой информации представляет собой текстовые данные различного объёма. По наименованиям единиц хранения подобные источники нарративной информации составляют более 90%. В настоящее время к такой информации, помимо текстов на различных языках человеческого общения, относятся и тексты компьютерных программ как на языках программирования, так и непосредственно в двоичных кодировках. Тексты, записанные в алфавитных системах, зачастую включают значительные фрагменты графической информации (иллюстрации). Особняком стоят тексты, записанные в неалфавитной (например, иероглифической) форме, которые по свойствам хранения ближе к чисто графической информации.

Текстовая информация может храниться в оригинальном виде, а также в сжатой форме при помощи различного вида архиваторных программных средств. При этом, что касается «чистого» текста, при архивировании и разархивировании потери информации исключены. Эффективность архивирования текстовых файлов приведена в табл. 1

Таблица 1 Сопоставление степеней сжатия форматов текстовых данных

Кодировка исходного текстового файла	MS WORD	RTF	TXT
Средняя степень сжатия объёма исходного файла в результате архивирования, раз	5 - 6	3	2

Графическая информация с высоким разрешением, особенно полихромные и полутоновые изображения, в свою очередь, архивному сжатию практически не поддаётся. Для её хранения, как правило, применяются растровые форматы (наиболее популярны из них JPEG, TIFF, BMP и т.д.). Эти форматы обеспечивают хранение реалистичных сцен с большой глубиной цвета.

Зарекомендовал себя в отечественной и зарубежной практике хранения электронных документов графической информации формат TIFF (англ. Tagged Image File Format) — формат хранения растровых графических изображений с большой глубиной цвета. Спецификация: [Электронный ресурс.] / Adobe Partners. Режим доступа к ресурсу: URL: <http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf/> (дата обращения: 01.11.2013). Формат \*.tiff является на сегодняшний день оптимальным с точки зрения качества сохранения визуальной (растровой) информации – он широко используется в полиграфии высокого качества. Электронный документ в данном формате с разрешением не ниже 300 dpi и последующим сохранением (без компрессии) в формате \*.tiff – лучший вариант создания электронного изображения, предназначенного как для длительного хранения, так и для дальнейшего использования, включая варианты преобразований файла<sup>5</sup>.

Многостраничный \*.tiff (мульти-Tiff) – версия известного формата TIFF для хранения и представления многостраничных файлов. Создается специальным программным обеспечением - Multi-Page TIFF Editor (<http://www.yukudr.com/multipageimage/index.ru.php>), многочисленными бесплатными программами, входит к комплект форматов многих графических редакторов. Однако для открытия \*.tiff-файлов необходимо использовать программы для просмотра многостраничных документов (например, Evince (<http://projects.gnome.org/evince/>), или функцию листания в стандартной программе просмотра изображений Windows, т.к. программы для просмотра изображений покажут только первую страницу многостраничного TIFFa<sup>6</sup>.

В зависимости от условий целесообразно создавать файл с максимальным разрешением в 300 dpi в цветном режиме в формате \*.jpg. Электронные копии в формате \*.jpg создаются в широком круге графических редакторов. JPEG (англ. Joint Photographic Experts Group, по названию организации-разработчика - URL: <http://www.jpeg.org/>, URL: <http://www.martinreddy.net/gfx/2d/JPEG.txt>) — один из популярных графических форматов, применяемый для хранения фотоизображений и подобных им изображений. Несмотря на свое широкое распространение, данный формат имеет очевидные недостатки, которые можно игнорировать при представлении информации, к примеру, в Интернет, но которые весьма существенны при формировании фонда пользования (особенно для требовательных исследователей). Например, сохраненное в формате \*.jpg при максимально возможном качестве (в Adobe PhotoShop – 10) и без компрессии изображение в случае восстано-

---

<sup>5</sup> «Рекомендованные цифровые форматы, предназначенные для долгосрочного хранения цифровой информации» («Recommended Data Formats for Preservation Purposes...» [Электронный ресурс.] / State University Library Services/ Florida Virtual Campus. Режим доступа к ресурсу: URL: <http://fclaweb.fcla.edu/uploads/recFormats.pdf/> (дата обращения: 22.10.2013).

<sup>6</sup> Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом / Ю.Ю. Юмашева. – М.: ВНИИДАД, 2012. С. 66.

ления не будет в точности совпадать с исходным изображением, формируемым при сканировании<sup>7</sup>.

В то же время имеется огромный массив графических документов, сформированных из элементарных объектов (т.н. «графических примитивов»). К таким примитивам относятся отрезки прямых и ломаные линии, многоугольники, окружности и эллипсы, а также, в некоторых случаях, кривые Безье и сплайны. При помощи графических примитивов эффективно реализуются документы типа чертежей, структурных схем, графического представления алгоритмов и т.п., для которых характерны резкие переходы яркости между фигурой и фоном (различным образом расположенные линии на однородном фоне). Для таких документов наиболее эффективны векторные форматы представления, к числу которых относятся SWG, WMF, CGM и им подобные. Однако векторные форматы, как правило, искажают цветовую схему сложного изображения.

Множество хранимых аудио документов характерно одновременно относительно большими как объёмом каждого документа, так и их номенклатурой. Такие документы, закодированные в форматах, обеспечивающих высокое качество сохранения и воспроизведения информации (например, MP3, WMF), почти не поддаются архивированию; в среднем их сжатие не превышает 5%. Объём файла в формате WMF длительностью звучания 60 минут составляет около 12-14 Мбайт, что эквивалентно документу MS WORD размером около 2500 страниц (150 печатных листов).

Видеоинформация обладает свойствами, аналогичными аудиоинформации, но значительно более выраженными. Выбор различных видов кодирования видеоинформации решающим образом сказывается на качестве изображения, что существенно влияет на объём хранимого объекта. В табл. 2 дано сравнение размеров сходных по длительности видеоресурсов в зависимости от качества хранения изображения.

Таблица 2. Сопоставление объемов форматов видео данных

Название кинофильма	Длительность воспроизведения, час	Кодировка				
		BDRip 1080p	DVD9, BDRip 720p	DVD5	DVDRip	HDRip
		Размер файла, Гбайт				
Сибирский цирюльник	3		7.8	4.3	2.04	

<sup>7</sup> Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом / Ю.Ю. Юмашева. – М.: ВНИИДАД, 2012. С. 66-67.



Утомлённые солнцем 2: Цитадель	2.5	20.4	7.6		1.46	
Гражданин Кейн (Citizen Kane)	2	27		4.3		0.745

Из данных табл.2 следует, что размер видеоресурса исключительно сильно зависит от качества хранения изображения. Практическая независимость размера от длительности воспроизведения может объясняться привязкой размера к стандартной ёмкости традиционных носителей этой информации (CD, DVD-дисков и т.д.).

Итак, даже в минимальном варианте для хранения видеодокумента требуется объём памяти, значительно превышающий размер текстовых документов. Например, суммарный объём текстов известной электронной библиотеки Максима Мошкова на 2005 год составлял всего 1.8 Гбайт. Возможности архивации видеодокументов при этом фактически отсутствуют – степень их сжатия без потери качества при восстановлении не превышает полупроцента. Следовательно, с технической точки зрения выбор программных средств для кодирования аудио- и видеоинформации определяется только требуемым качеством воспроизведения изображения и доступным объёмом архивного хранилища.

Достаточно часто встречается потребность в сохранении комплексных документов, в структуру которых входят фрагменты различной природы. Наиболее актуальны в этой связи текстовые документы с графическими вставками (фотоизображениями, диаграммами, чертежами и т.д.). Достаточно просто эта проблема решается с использованием современных текстовых форматов типа MS WORD, RTF. Документы этих форматов позволяют включать графические изображения в растровых кодировках. Разумеется, при этом хранимый размер документа увеличивается, но одновременно возможно «обтекание» изображения текстом. В случае большого количества иллюстраций в тексте оказывается более выгодным хранить документ в виде единого изображения – в этом смысле графические форматы хранения являются более универсальными. В то же время чисто графические форматы имеют существенный недостаток с точки зрения работы с текстами – функции редактирования текста, текстового поиска, объединения текстовых фрагментов при оперировании изображениями текстов исключительно затруднены. Фактически задача сводится к распознаванию графических образов; надёжные алгоритмы такого распознавания требуют для своей реализации высокопроизводительного оборудования, которое в обозримом будущем ещё не получит широкого распространения.

С другой стороны, ряд приложений при работе с документами требует анализа не только содержания текста, но и формы его представления и особенностей самого документа. К таким особенностям можно отнести шрифт (фонт) печатного документа, почерк рукописи, степень сохранности и индивидуальные дефекты документа (исправления, грязь, потёртости, царапины, выцветание и т.п.). В этом случае без графического кодирования обойтись практически невозможно.

Возможным компромиссом между этими противоречивыми требованиями можно признать межплатформенные комплексные форматы, такие как PDF, DjVu, CGM. Последний из них, правда, имеет ограниченное распространение, ибо наиболее приспособлен к работе с технологическими машиностроительными чертежами.

Вполне конкурентными могут служить форматы PDF и DjVu. Для сравнения эффективности хранения документов в этих форматах проводился эксперимент по преобразованию исходного текста в формате MS WORD независимо в форматы PDF и DjVu. Далее оценивались коэффициенты сжатия объёма текстовых файлов после преобразования в графические форматы; они сводились в два отдельных числовых массива. В третий массив сводилось отношение коэффициентов сжатия объёма текстовых файлов при использовании двух графических форматов. По статистическим данным трёх таблиц рассчитывались несмещённые оценки выборочных средних значений по формулам:

$$A_{\text{ср}} = \left( \sum_{i=1}^n a_i \right) / n ,$$

где  $a_i$  – экспериментальное значение параметра из таблицы,

$A_{\text{ср}}$  – полученное среднее значение,

$n$  – размер таблицы.

Статистически устойчивого результата удалось добиться уже при значении величины  $n$  (количество оцененных файлов), равном 20.

Документ в формате PDF может включать текстовую часть, векторную и растровую графику, элементы мультимедиа, формы и т.п. Его достоинствами являются лёгкий поиск текстового фрагмента и возможность копирования текста и изображений в редакторы типа MS WORD. В документ может быть вставлена электронная цифровая подпись. Документ в формате MS WORD легко преобразуется в формат PDF (например, бесплатно распространяемым приложением doPDF\_7.1.344). Обработка данных произведённых экспериментов показала, что текстовый документ при таком преобразовании в изображение

уменьшается в объёме до 7 раз, причём возможно использование текстов примерно на 20 языках.

Документ в формате DjVu чаще всего создаётся сканированием исходного документа и обладает свойствами, аналогичными формату PDF. Его особенность – наличие отдельного текстового слоя, позволяющего более быстро производить поиск информации в текстовой части документа. Экспериментальное сравнение одинаковых документов в этих двух форматах выявило, что документ в формате DjVu в среднем имеет объём в 2.4 раза меньше, нежели документ в формате PDF. В исключительных случаях большого количества разнообразных иллюстраций и плохой сохранности оригинального документа отклонение может достигать 16 раз. Следовательно, при ограниченном объёме информационного хранилища использование формата DjVu выглядит предпочтительнее. В случае выбора в качестве формата хранения данных формата PDF отсканированный документ из формата DjVu может быть преобразован (конвертирован) в формат PDF любым из свободно распространяемых программных средств такой конвертации. В настоящей работе в качестве инструментального средства была использована программа DjVuLibre+DjView-3.5.25.3+4.9 . Следует отметить, что сам процесс конвертации легко автоматизируется и практически не требует ручного труда сотрудника архива.

Вместе с тем следует учитывать и возможности просмотра информации при помощи различных технических средств. В случае использования стационарных или переносимых компьютеров с операционными системами семейств Windows и UNIX проблем с просмотром документов в упомянутых кодировках в настоящее время не возникает. Особым случаем является применение всё шире распространяющихся планшетных портативных устройств. Более дорогие из них, работающие на платформах iOS фирмы Apple или Android фирмы Google, обеспечивают возможность чтения документов в обеих конкурирующих кодировках. Однако следует учесть и популярность более дешёвых планшетных устройств фирмы Samsung, стандартные программные средства которых позволяют работать с информацией (файлами) в кодировке PDF, в то время как файлы с данными типа DjVu даже не распознаются. Кроме того, специальные программные средства для чтения на этих планшетах в бесплатном доступе встречаются достаточно редко (по данным на сентябрь 2013 года ссылки в Internet есть, но скачивание невозможно – программные средства в бесплатный доступ, по-видимому, ещё не выложены). Более того, именно на планшетах фирмы Samsung установленное дополнительное программное обеспечение работает неустойчиво. По этой причине для таких планшетных устройств в настоящее время имеет смысл рекомендовать использовать файлы в кодировке PDF, несмотря на их увеличенный более чем вдвое объём.

Исходя из имеющихся данных PDF («portable document format») на сегодняшний день сохраняет в достаточной мере универсальность на протяжении почти 20 лет. Изначально данный формат, созданный по инициативе компании Adobe, предназначался для электронного представления печатных материалов. Именно в этом аспекте задача формата - хранить информацию о цветовой схеме и расположении элементов, обеспечить идентичный вывод на разных мониторах и принтерах.

Безусловно, возможности формата PDF значительно расширились в последнее время:

- стало возможным вставлять в текст документа гиперссылки
- можно шифровать документ с паролем и тем самым защищать его от модификации
- появилась возможность сочетать изображение с текстовым слоем. Это дает возможность, просматривать представление документа и копировать текст из документа (версия PDF 1.4)

- возможность вставлять метаданные в виде пар «ключ-значение», стоит отметить, что каждая из которых может быть связана с какой-то частью документа (например, отдельным изображением) или с документом в целом (версия PDF 1.4)

Новые возможности формата, с точки зрения архивного хранения, крайне важны, более того, они поддерживаются форматом по сей день.

Отдельное значение, безусловно, имеет PDF/A — это подмножество формата PDF, содержащее весьма ограниченный набор возможностей представления данных. Стоит отметить, что данный формат является стандартом ISO и предназначен для долгосрочного хранения электронных документов. В целях обеспечения длительного срока хранения было осуществлено внедрение в содержимое электронного документа в формате PDF/A всей информации, необходимой для его отображения. К примеру, такой информацией, в частности, являются шрифты (как те, которые использованы в документе, так и включенные в него). В силу этой особенности тесты показали, что, размер: документа в формате PDF/A не меньше, а порой больше по размеру, чем аналогичный по содержанию документ в формате PDF.

Другое важное отличие PDF/A от PDF – запрет на использование скриптов. В случае использования обычного PDF, в документ может быть встроен вредоносный код. Многие программы для работы с PDF позволяют отключать выполнение скриптов в документах, но если настройка программного обеспечения пользователя выполнена неправильно, открытие документа может привести к нежелательным последствиям. С этой точки зрения PDF/A выглядит предпочтительней, но есть одна проблема: оба формата имеют

одинаковое расширение «.pdf», и под видом надежного PDF/A файла может скрываться зараженный документ. Поэтому настоятельно рекомендуется корректно настраивать программное обеспечение и отключать выполнение встроенных скриптов.

Есть и другие отличия. Так, документы PDF/A не поддерживают встроенного шифрования, а также вставку аудио и видео материалов. Эти особенности не столь существенны, и на практическое применение формата почти не влияют.

Несколько слов о самом PDF/A. На данный момент имеется три версии формата. Первая версия – PDF/A-1 (ISO 19005-1, 2005 год) имеет две модификации. Одна из них базовая (PDF/A-1b), вторая (PDF/A-1a) содержит дополнительные требования, направленные на упрощение поиска по документу (необходимость вставки иерархии, описания тэгов и языковых особенностей). На практике предпочтительней PDF/A-1b, так как затраты на создание документа соответствующего всем требованиям PDF/A-1a не всегда оправданы.

Вторая версия PDF/A-2 (ISO 19005-2, 2011 год) имеет ряд незначительных изменений (добавлена поддержка JPEG 2000, OpenType шрифтов и прозрачных слоев). Также существует несколько подвидов (PDF/A-2a, PDF/A-2b, PDF/A-2u), которые имеют незначительные различия. На данный момент формат PDF/A-2 распространен слабо.

Наконец, PDF/A-3 (ISO 19005-3, 2012 год) добавил возможность встраивать в документы файлы произвольных форматов. Как и PDF/A-2 почти не применяется.

В рамках исследования нашло свое подтверждение то, что документ, хранимый в формате PDF/A, в силу полного отсутствия связи с такими изменчивыми вещами как гиперссылки и мультимедийный контент, можно открыть в любой операционной системе через какое угодно длительное время с помощью приложения, поддерживающего соответствующий формат. Особое значение также имеет и тот факт, что PDF/A обеспечен статусом международного стандарта, а его поддержка со стороны разработчиков в долгосрочной перспективе оправдана. Использование целесообразно по сравнению с другими доступными форматами хранения, которые могут измениться в любой момент времени. Ситуация с отказом от поддержки документов Word старого образца до версии 2003 в Google Drive дало старт полному вытеснению (умерщвлению) формата Word 97–2003. Более того, компания Microsoft сама не оказывает последовательную поддержку и преемственность форматов в рамках ПО Word версий 97-2003 с более поздними версиями. Открытие в последних версиях Word текстовых документов ранних версий до 2003 г. включительно в 40% приводит к существенным изменениям представления текста: изменяется форматирование, происходит хаотичное «съедание» пробелов и т.п.

Существуют и определенные риски с форматом PDF/A. Так, применительно к формату PDF/A следует отметить, что целостность и неизменность неподписанного документа в данном формате не может быть гарантирована и более того не заявляется как особенность формата. Иными словами, несмотря на то, что данный формат позиционируется как обеспечивающий долгосрочное хранение, изменение содержимого документа остается возможным и что странно не является отклонением от нормы, если оно не зашифровано. Существует также следующий нюанс: для каждого конкретного документа, формат которого заявлен как PDF/A, невозможно заведомо утверждать, что это действительно так. Таким образом, необходима верификация на соответствие требованиям формата для каждого конкретного документа, и если на этапе размещения в архиве или после очередного изменения она не будет проведена, то задача по принятию электронного документа на долгосрочное хранение потенциально провалена.

Особое значение приобретает использование PDF и его репликаций в отношении образов документов в виде растрового изображения. Предполагается, что текста поверх него нет, то есть в документе хранится исключительно отсканированный растр — изображение, текст на котором непонятен компьютеру, а понятен только человеку. В исключительных случаях поверх растрового изображения может быть расположен текстовый слой, частично или целиком наполненный либо вручную человеком, либо с помощью системы распознавания текста (в зависимости от формата графического файла). Можно предположить, что документ содержит метаданные, так или иначе связанные с видом документа и его содержимым (например, если это договор, метаданные могут содержать информацию о поставщике, дате выставления, сумме, сроках и т.д.). Очевидно, что графический образ документа должен храниться не меньше бумажного оригинала, а зачастую значительно дольше, поскольку значимость и важность его в контексте организации-владельца всегда держится выше нулевой отметки. Кроме того, образ бумажного документа изменение не подразумевается.

На основании перечисленных особенностей, которыми обладает образ документа, можно смело заявлять, что одинаково пригодны оба формата, поскольку для каждого конкретного документа не будут использованы возможности формата такие, которые не поддерживаются стандартом PDF/A. Последний допускает и метаданные, и текстовый слой поверх изображения, и даже подписание документа с целью защитить его от модификации. А необходимостью внедрения шрифтов в каждый документ можно пренебречь: поскольку речь идёт об образах, и текста нет, соответственно, нет и шрифтов, и добавочный вес в этом отношении будет мал.

Аналогично, с каждым годом увеличивается количество документов, включённых в разнообразные базы данных – база данных также является документом со структурой, состоящей из полей разной информационной природы. Выбор кодировки для хранения базы зависит в основном от её структуры.

Если база представляет собой один файл, то её имеет смысл хранить в кодировке, соответствующей этому типу файла – как правило, в виде двоичных данных. При хранении базы, записи которой явно состоят из полей разной структуры, наиболее часто употребимы окажутся кодировки RTF и PDF – в случае, если в базе нет аудио- и видеофрагментов (почти невозможно представить приложения для использования подобной базы). В случае реляционной базы данных (самый частый в настоящее время вариант) запись в базе как единое целое, как правило, не хранится, а формируется их полей различных таблиц по запросу. При этом вполне возможно хранение информации отдельных таблиц в наиболее адекватных случаю ранее описанных кодировках. Проблема заключается лишь в «сборке» готовых информационных фрагментов с выдачей их пользователю в требуемом порядке.

Отдельная задача встаёт в связи с частотой пользовательского обращения к информационному источнику. Для систем архивного хранения информации характерным является режим именно долговременного хранения данных в наиболее удобной для информационного поиска системе классификации метаописаний единиц хранения. Информационная коллекция пополняется, однако обращение к этому массиву по «читательскому» запросу происходят достаточно редко. Для таких архивных хранилищ основная проблема – угроза переполнения принципиально ограниченного объёма хранилища. Следовательно, хранение информации целесообразно осуществлять в максимально архивированном, «сжатом» виде. Пример подобного архива – архив Департамента Герольдии Правительствующего Сената Российской Империи (Санкт-Петербург) или Российский Государственный Военно-исторический архив (Москва). В настоящее время собрания этих архивов практически не пополняются, а запросы к отдельным архивным делам отсутствуют десятилетиями. Аналогичная ситуация складывается с Центральным архивом Министерства обороны России, хотя его собрание продолжает регулярно пополняться.

В случае «оперативного», частого пользования данными из информационного источника объём хранимой информации для пользователя является менее критическим, и её следует хранить не в сжатом виде, а в кодировке, наиболее адекватной как природе самой информации, так и предметным запросам пользователя. Это позволяет существенно сократить как время доступа к информации, так и время её неформального анализа и обработки человеком-заказчиком данных из информационной системы.

Современные алгоритмы сжатия (архивации) данных пригодны для работы практически со всеми используемыми видами информации. При выполнении условия минимальной потери информации в процессе архивации и последующего восстановления данных при разархивации степень сжатия данных слабо зависит от непосредственно структурно-математических особенностей алгоритмов сжатия и, как следствие, самих программ-архиваторов. Необходимое требование к современным алгоритмам и программам архивации – точное восстановление оригинала документа из его архивного образа. Используемые в настоящее время популярные архиваторные средства (WinRAR, ARJ и т.д.) этому требованию удовлетворяют. Получение многочисленных копий документа с одного оригинала, очевидно, даёт тождественные результаты, и потому архивация в этом смысле устойчива. При множественной репликации архивированного документа (получении очередной копии исходного документа восстановлением из архивного оригинала) для текстового источника документы восстанавливаются точно, т.е. погрешности не накапливаются. Для аудио- и видеодокументов погрешность возникает только при первой итерации архивирования-разархивирования, однако дальнейшего накопления ошибок не происходит. Первичное же искажение информации, оцениваемое по размеру оригинального и восстановленного документа (в байтах) отсутствует. Формальная оценка искажений восстановленного изображения затруднена, однако неформально может быть признано весьма незначительно. В случае совместного хранения информационных структур различной природы (например, в базах данных), архивация оказывается вполне доступной в силу упомянутой относительной универсальности архивирующих программ. Отдельному исследованию и решению подлежит только проблема комплексирования разархивированной информации, т.е. восстановления существовавших оригинально связей между разнородными информационными структурами, особенно в случаях хранения в архивированном виде реляционных баз данных.

С точки зрения математической теории сложности алгоритмов, задачу хранения информации следует оценивать как по временной, так и по объёмной сложности. Исследование объёмной сложности алгоритмов сжатия, т. е. принадлежности зависимости требуемого объёма запоминающих устройств (как в процессе архивации, так и при хранении сжатых данных) от размера задачи (изменения размера архивируемых данных) определённому классу функций, показывает, что алгоритмы архивации имеют линейную сложность. Это приводит к выводу, что в перспективе не следует ожидать больших проблем, связанных во «взрывным» характером требуемых характеристик к объёмам технических устройств архивного хранения электронной информации.



Вместе с тем, временная сложность соответствующих алгоритмов имеет как линейную, так и нелинейную составляющие. Для одного документа временная сложность алгоритмов архивирования и разархивирования линейна, т. е. время архивирования документа большого объема асимптотически пропорционально этому объему. В то же время при работе с документом, имеющим многочисленные информационно-ссылочные связи с другими документами архива, количество таких связей растёт квадратично, что существенно замедляет процессы обработки таких структур при росте количества архивных документов. Непосредственно изменением алгоритмов сжатия эту проблему решить с ростом объёмов и структурированности архивов невозможно. Потенциальным путём преодоления этой проблемы может явиться попытка введения иерархической структуры архивного хранения информации с разбиением коллекции данных на группы. Связь документа с другим документом в этом случае можно обрабатывать последовательным доступом к группе целевого документа, а уже затем – к подгруппе документов и т.д. вплоть до самого документа-цели. Это является аналогией используемых структур библиотечной классификации документов (УДК и др.). Если подобные методы ввести в алгоритмы архивирования связей документов, по существу возрастет не только скорость поиска документа в архиве, но и время, требуемое на процесс архивирования-разархивирования систем документов сложной структуры.

Сравнительный анализ объемов информации, подлежащей архивному хранению, при условии хранения электронных архивных документов в указанных форматах файлов проведен через миграцию из одного формата в другой.

За основу опытных исследований сравнительного анализа объемов информации изначально был взят файл, созданный в наиболее доступной версии текстового процессора Microsoft Word 97-2003 с текстом и таблицами. Инструментарием для миграции информации из одного формата в другой стало программное обеспечение - Word 2007.

Основные характеристики файла в формате \*.doc:

Версия: Microsoft Word 97-2003

Объем: 53 КБ

Страниц: 3

Слов 647

Знаков без пробелов: 4654

Знаков с пробелами: 5203

Абзацев: 104

Строк: 242

Во время миграции из формата \*.doc (53 КБ) в другие форматы объем файла только в 1 случае из 5 в значительной степени был уменьшен (в формате \*.odt файл занимает в 3,3 раза меньше дискового пространства). В остальных форматах файл в среднем в 3 раза увеличивается (Таблица 3). Наименьшее увеличение мы видим при миграции в формат \*.xml (в 2,2 раза), а наибольшее в формат \*.pdf (PDF/A) (в 3,7 раз).

Таблица 3. Сравнительный анализ объемов информации при миграции файлов из текстового формата \*.doc Microsoft Word 97-2003, 2007

Формат файла	Объем после миграции	Изменение от исходного файла
*.doc Microsoft Word 97-2003, 2007	<u>53 КБ</u>	Исходный файл
*.odt	16 КБ	в 3,3 раза уменьшился
*.xml	115 КБ	в 2,2 раз увеличился
*.rtf	144 КБ	в 2,7 раза увеличился
*.pdf	168 КБ	в 3,2 раза увеличился
*.pdf (PDF/A)	195 КБ	в 3,7 раза увеличился

Проведенный опыт миграции файла из формата \*.odt показал, что размерность файла в данном случае значительно увеличивается. Объемов файлов увеличился в среднем в 9,6 раз (наименьший показатель увеличения показала миграция в формат \*.doc (3,4 раза), а наибольший в формат \*.pdf (PDF/A) (в 18,4 раза) (Таблица 4). Уменьшения размерности файлов в данном примере мы не наблюдаем.

Таблица 4. Сравнительный анализ объемов информации при миграции файлов из текстового формата \*.odt

Формат файла	Объем при миграции	Изменение от исходного файла
*.odt	<u>16 КБ</u>	Исходный файл
*.doc (*.doctx) Microsoft Word 97-2003, 2007	54 КБ	в 3,4 раза увеличился
*.xml	134 КБ	в 8,4 раза увеличился
*.rtf	123 КБ	в 7,7 раз увеличился
*.pdf	160 КБ	в 10 раз увеличился
*.pdf (PDF/A)	294КБ	в 18,4 раза увеличился

Объемы информации при сохранении информации \*.xml формата в другие показал значительные изменения в сторону уменьшения объемов в форматах \*.odt (в 8,2 раза) и

\*.doc (2,2 раза) (Таблица 5). Увеличение же произошло в среднем в 1,3 раза. Наибольшее увеличение, как ожидаемо, произошло при миграции в формат \*.pdf (PDF/A).

Таблица 5. Сравнительный анализ объемов информации при миграции файлов из текстового формата \*.xml

Формат файла	Объем при миграции	Изменение размера от исходного файла
*.xml	<u>115 КБ</u>	Исходный файл
*.odt	14 КБ	в 8,2 раза уменьшился
*.doc (*.doctx) Microsoft Word 97-2003, 2007	53 КБ	в 2,2 раза уменьшился
*.rtf	119 КБ	в 1,04 раза увеличился
*.pdf	161 КБ	в 1,4 раза увеличился
*.pdf (PDF/A)	297КБ	в 2,6 раз увеличился

Очевидно, что весьма стабильно при миграции в выбранном программном обеспечении ведет себя формат \*.doc и \*.xml (Таблица 6). При этом показательна достаточная устойчивость показателей изменения объемов информации при миграциях из одного формата в другой.

Таблица 6. Матрица изменений объемов информации при миграции (во сколько раз)

Форматы файлов исходные/Форматы файлов миграции	*.doc Microsoft Word 97-2003	*.odt	*.xml	*.rtf	*.pdf	*.pdf (PDF/A)
*.doc Microsoft Word 97-2003	Исходный	< 3,3	>2,2	>2,7	>3,2	>3,7
*.odt	>3,4	Исходный	>8,4	>7,7	>10	>18,4
*.xml	< 2,2	< 8,2	Исходный	>1,04	>1,4	>2,6

Тенденции к сохранению объемов информации при миграциях отмечаются в большинстве случаев за исключением миграций из форматов \*.odt и \*.xml в \*.pdf (PDF/A) (Таблица 6).

Таблица 6. Матрица фактических объемов информации при миграции (в КБ)

Форматы файлов ис-	*.doc	*.odt	*.xml	*.rtf	*.pdf	*.pdf
--------------------	-------	-------	-------	-------	-------	-------

ходные/Форматы файлов миграции	Microsoft Word 97-2003					(PDF/A)
*.doc Microsoft Word 97-2003	Исходный 53	16	115	144	168	195
*.odt	54	Исходный 16	134	123	160	294
*.xml	53	14	Исходный 115	119	161	297

Использование свободного программного обеспечения OpenOffice для создания и миграции документов имеет как положительные так и отрицательные стороны. Безусловно, при создании из файла формата \*.odt - \*.xml, \*.pdf (PDF/A), даже \*.doc Microsoft Word 97-2003 размеры файлов в большей части свое значительно меньше (Таблица 7). Существенно «легче» стал файл в формате \*.pdf (PDF/A).

Таблица 7. Матрица фактических объемов информации при миграции формата \*.odt с помощью различного программного обеспечения (в КБ)

Файл в формате *.ODT (16 КБ) /Форматы файлов миграции	*.doc Microsoft Word 97-2003	*.xml	*.rtf	*.pdf	*.pdf (PDF/A)
*.odt (OpenOffice 4)	36	71	81	170	176
*.odt (Microsoft Word 2007)	54	134	123	160	294

При более детальном рассмотрении мы видим, что \*.pdf (PDF/A) созданный Microsoft Word 2007 имеет версию: как PDF/A-1A так и PDF/A-1B. С помощью же OpenOffice 4 (Writer) создается файл в формате: PDF/A-1A. PDF/A-1B ставит целью обеспечение надёжного воспроизведения внешнего вида документа. PDF/A-1A включает все требования стандарта PDF/A-1B и дополнительно требует, чтобы была включена структура документа, ставя при этом целью обеспечение возможности поиска и переназначения (преобразования) содержимого документа, что очень актуально с точки зрения современных требований к архивированию электронной документации.

Кроме того, необходимым условием в OpenOffice 4 (Writer) для создания файла в XML формате является наличие установленного окружения Java (JRE). Программное обеспечение распространяется свободно, но это потребует дополнительных технических

настроек системы. Родной же формат XML OpenOffice 4 - \*.stw еще не получил значимой поддержки и описания. К примеру, в Microsoft Word 2007 данный формат файла открывается как пустой документ – информация не отображается.

Расходы на хранение информации напрямую зависят от необходимого программного аппаратного обеспечения, необходимого для воспроизведения конкретных типов файлов. Примерные расчеты представлены в Приложении «Ж». Определение стоимости хранения информации будут определяться по следующим параметрам: формат файла - стоимость программного обеспечения : объем информации – стоимость носителей информации. В связи с постоянным удешевлением носителей информации повышению их емкости, основными факторами, определяющими выбор форматов файлов для постоянного (долговременного) хранения являются доступность программного обеспечения и считываемость формата файла в долговременное перспективе.

Достиженные результаты основаны на обобщении значительного объема современных данных по цифровым форматам, нормативным документам, стандартам и ГОСТам, а также опытных исследований ряда форматов для оценки важных показателей по сжатию, кодированию, миграции. Исследование соответствует современным требованиям отечественных нормативных, методических документов, а также достигнутым результатам в зарубежной практике.

Исходя из анализа разнородной цифровой информации можно рекомендовать к использованию следующие основные рекомендованные форматы с точки зрения их надежности и долговечности:

1. Текстовые форматы файлов и форматы представления данных обладающие высокой степенью надежности и обеспечивающие долгосрочность хранения информации (Приложение Б; Приложение В):

- XML (includes XSD/XSL/ XHTML)
- PDF/A-1 (ISO 19005-1) (\*.pdf)
- PDF/A-2 (ISO 19005-2:2011) (\*.pdf)

Исходя из проведенных сравнений между форматами PDF и его производным PDF/A, можно утверждать, что первый больше пригоден для оперативного обмена и краткосрочного хранения электронных документов, в свою очередь как PDF/A (с учетом потенциально большего размера единичного документа), имея статус международного стандарта, гарантирует, что даже через продолжительное время, вне зависимости от окружения и операционной системы, любой пользователь сможет открыть документ в данном

формате, располагая ПО-просмотрщиком. В целом это соответствует концепции архива электронных документов.

В некоторыми условиями и, в крайнем случае, могут быть предложены в категории текстовых форматов файлов (Приложение Б; Приложение В):

- Plain text (ISO 8859-1encoding)
- PDF (\*.pdf) (embedded fonts)
- Rich Text Format 1.x (\*.rtf)
- HTML (include a DOCTYPE declaration)
- SGML (\*.sgml)
- Open Office (\*.sxw/\*.odt)
- OOXML (ISO/IEC DIS 29500) (\*.docx)
- EPUB (unencrypted) (\*.epub)

2. Графические форматы рекомендованы и отвечают требованиям долговременного архивного хранения:

2.1. Растровые изображения (Приложение Г, Приложение Д):

- TIFF (без компрессии/uncompressed)
- JPEG2000 (lossless) (\*.jp2)
- PNG (\*.png)

2.2. Векторные изображения (Приложение Е):

- SVG (no Java script binding) (\*.svg)

2.3. Виртуальные форматы данных:

- X3D (\*.x3d)

3. Рекомендованы форматы сохранения аудио и видео информации:

3.1. Аудио форматы:

- AIFF (PCM) (\*.aif, \*.aiff)
- WAV (PCM) (\*.wav)
- AES3 (LPCM) (\*.aes)

3.2. Видео форматы:

- Motion JPEG 2000 (ISO/IEC 15444-4) (\*.mj2)
- AVI (uncompressed, motion JPEG) (\*.avi)
- QuickTime Movie (uncompressed, motion JPEG) (\*.mov)

4. Базы и банки данных рекомендованы к сохранению в форматах:

- Comma Separated Values (\*.csv)
- Delimited Text (\*.txt)
- SQL DDL

В редких случаях могут быть использованы форматы:

- DBF (\*.dbf)
- OpenOffice (\*.sxc/\*.ods)
- OOXML (ISO/IEC 29500)(\*.xlsx)

Форматы файлов по презентациям не имеют формата с достаточно высокой надежностью обеспечения долговременного хранения, однако можно рекомендовать с определенными оговорками:

- OpenOffice (\*.sxi/\*.odp)
- OOXML (ISO/IEC DIS 29500) (\*.pptx)

Форматы других презентационных файлов, в т.ч. PowerPoint (\*.ppt), не отвечают требованиям долговременного хранения.

Можно рекомендовать следующие технические параметры и параметры контроля к форматам файлов при их передачи на архивное хранения предусматривают:

1. Отсутствие файлов с вирусами (особенно это относится к форматам файлов: DOC, XLS, MDB, PDF, PPT, ZIP, EXE).

2. Передаваемые на архивное хранение архивные файлы не могут быть полностью или частично: зашифрованы, защищены паролем, сжаты собственным алгоритмом сжатия, кодированы.

3. Архивирование файлов рекомендуется в незашифрованном виде ZIP (\*.zip) – наиболее соответствующий формат для передачи данных.

Технические требования к серверному оборудованию, системам хранения электронного контента определяются, исходя из реальных объемов имеющегося цифрового контента, перспектив его наращивания и нужд в осуществлении миграций, распечатки, объемов обработки электронных документов. Наиболее критичные требования к объёму серверной памяти предъявляются в случае архивного хранения видеоинформации с высоким качеством изображения. Аналогично, и передача хранимых видеофайлов из архива пользователю требует максимальной пропускной способности линий связи. По сравнению с видеоинформацией аналогичными требованиями архивов текстов и статических изображений можно пренебречь. Предлагаемые форматы создаются преимущественно на открытом программном обеспечении и позволяют минимизировать статьи расходов на приобретение программного обеспечения.

Стратегия выбора надежных форматов файлов также осуществляется исходя из возможности использовать форматы независимые от платформы, независимых поставщиков, неспециализированные, открытые и хорошо стабильно поддерживаемые форматы.

В исследовании в полной мере раскрывает структуру и сущность цифровых форматов, объективность их использования в обеспечении долговременной сохранности электронных документов в архиве, чем имеющиеся аналитические данные зарубежных коллег. Например, в марте 2012 г. Университет Флориды (США) подготовил для Цифрового архива Флориды результаты своего исследования только в виде списка ранжирования цифровых форматов без детального описания нюансов.



ПРИЛОЖЕНИЕ Б

СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ ФОРМАТОВ ТЕКСТОВЫХ ДАННЫХ

Название	Текстовый файл	OpenDocument Text	Microsoft Word			WordPerfect	Rich Text Format	OpenOffice.org Writer
			DOC	DOCX	DOCМ			
Расширение	<b>ТХТ</b>	<b>ODT</b>	<b>DOC</b>	<b>DOCX</b>	<b>DOCМ</b>	<b>WPD</b>	<b>rtf</b>	<b>sxw</b>
Стандартизация		ISO/IEC 26300, ГОСТ Р ИСО/МЭК 26300-2010		ЕСМА-376, ISO/IEC 29500				
Дата создания		2005	1997	2006		Нет данных	Конец 1980-х	2000-2002
Дата последних изменений		2011	2012	2012		Нет данных	2008	С 2005 заменен на OpenDocument Text
Открытость формата	Открытый	Открытый	Проприетарный (частично открыт)	Открытый (однако, поддерживает вставку объектов закрытых форматов)		Проприетарный	Проприетарный	Открытый
Назначение формата	Предназначен для хранения текстовых данных в чистом виде. Не поддерживает вставку изображений. Форматирование только специальными символами.	Открытый формат файлов документов, в основе которого лежит язык разметки XML. Формат активно развивается.	Формат документов, разработанный Microsoft.	Открытый формат файлов документов, в основе которого лежит язык разметки XML. Разработан Microsoft.	DOCX с поддержкой макросов и скриптов.	Текстовый формат редактора WordPerfect	Межплатформенный формат хранения размеченных текстовых документов.	Устаревший формат текстовых документов использовавшийся в ранних версиях пакета OpenOffice.org

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Б

<b>Наличие бесплатного программного обеспечения</b>	Да	Да	Частично (закрытые части стандарта могут поддерживаться не полностью)	Частично (несмотря на открытость формата, реализация полной поддержки в бесплатном ПО затруднена)	Частично (поддерживается неофициально и не полностью)	Да (некоторые особенности могут быть не реализованы)	Да	
<b>Самодостаточность документа (сам документ содержит все необходимое для его корректного отображения)</b>	Частично (содержит только текст)	Частично (не содержит шрифты, документ может отображаться по-разному в различных программах или ОС)	Да	Да	Нет данных	Частично (сам формат самодостаточен, но лишь несколько программ поддерживают его полностью)	Частично (не содержит шрифты, документ может отображаться по-разному в различных программах или ОС)	
<b>Возможность полнотекстового поиска</b>	Да	Да	Да	Да	Да	Да	Да	
<b>Наличие «опасных» возможностей (встроенного кода, шифрования, средств управления правами доступа и т.д.)</b>	Нет	Могут использоваться (работа макросов и скриптов, встроенных в документ, зависит от конкретного приложения)	Могут использоваться	Нет	Могут использоваться (макросы, скрипты)	Могут использоваться (макросы PerfectScript)	Частично (скрипты и макросы отсутствуют, но имеется дыра в безопасности при встраивании WMF файлов)	Могут использоваться (работа макросов и скриптов, встроенных в документ, зависит от конкретного приложения)
<b>Сжатие графических объектов с потерями</b>	Графические объекты не поддерживаются	Нет	Нет	Нет	Нет данных	Нет (WMF) / Да (в остальных случаях)	Нет	
<b>Возможность конвертации из других форматов</b>	Да (для сохранения форматов могут потребоваться специальные инструменты)	Да	Частично (в связи с закрытостью некоторых частей стандарта)	Да	Частично (скрипты и макросы практически не конвертируются)	Частично	Да	Да

ПРИЛОЖЕНИЕ В

СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ ФОРМАТОВ ПРЕДСТАВЛЕНИЯ ДАННЫХ

Название	HTML	XML	Electronic Publication)	FictionBook	DjVu	Portable Document Format	
Расширение	<b>.htm, .html</b>	<b>.xml</b>	<b>.epub</b>	<b>.fb2</b>	<b>.djvu, .djv</b>	<b>PDF</b>	<b>PDF/A</b>
Стандартизация	ISO/IEC 15445, W3C HTML 4.01					ISO 32000	ISO 19005-1:2005 ISO 19005-2:2011 ISO 19005-3:2012
Дата создания	1995	1998	2007	2004	1998	1993 (открыт в 2008)	2005
Дата последних изменений	2011	2008	2011	2008	2006	2011	2012
Открытость формата	Открытый	Открытый	Открытый	Открытый	Частично открытый (открыт формат файла и декодер)	Открытый	Открытый
Назначение формата	Основной язык разметки, применяемый для создания веб-страниц и других документов, которые могут быть открыты веб-браузером	Универсальный язык разметки предназначен для описания различных данных	Свободный открытый формат для электронных книг	Формат для электронных книг с несложной версткой	Технология сжатия изображений с потерями, разработанная специально для хранения сканированных документов. Документ фактически остается изображением, а для получения текста необходимы дополнительные действия.	Межплатформенный формат файлов документов	Стандарт для долгосрочного архивного хранения электронных документов
Наличие бесплатного программного обеспечения	Да	Да	Да	Да	Да	Да	Да

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ В

<b>Самодостаточность документа (сам документ содержит все необходимое для его корректного отображения)</b>	Нет (не содержит шрифты; хранит только ссылки на изображения и другие встроенные данные)	Зависит от ПО (изначально не предназначен для хранения документов)	Да (представляет собой архив с необходимыми файлами)	Нет ( не содержит шрифты)	Да	Не содержит шрифты	Да
<b>Возможность полнотекстового поиска</b>	Да	Да	Зависит ПО	Да	Да (при наличии текстового слоя)	Да (при наличии текстового слоя)	Да (при наличии текстового слоя)
<b>Наличие «опасных» возможностей (встроенного кода, шифрования, средств управления правами доступа и т.д.</b>	Могут использоваться	Зависит от ПО	Зависит от ПО	Нет	Нет	Могут использоваться	Нет*
<b>Сжатие графических объектов с потерями</b>	Нет (формат сам по себе не предназначен для хранения изображений )	Нет (формат сам по себе не предназначен для хранения изображений )	Нет	Нет	Да (в некоторых случаях значительно ухудшается качество графических объектов)	Нет	Нет
<b>Возможность конвертации из других форматов</b>	Нет (существующие способы на практике не применимы)	Нет (существующие способы на практике не применимы)	Частично (на практике могут возникнуть трудности)	Частично (поддерживаются только простые средства форматирования)	Частично (официальный конвертер запатентован и открыт лишь частично)	Да	Да

\*Документы в формате PDF/A и PDF имеют одинаковое расширение .pdf и под видом безопасного документа может скрываться зараженный файл

ПРИЛОЖЕНИЕ Г

СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ ГРАФИЧЕСКИХ ФОРМАТОВ (РАСТРОВЫЕ БЕЗ ПОТЕРЬ)

Название / растровые без потерь	BMP	GIF	PNG	PCX	PSD	XCF
Расширение	<b>.bmp</b>	<b>.gif, .gfa, .giff</b>	<b>.png</b>	<b>.pcx</b>	<b>.psd</b>	<b>.xcf</b>
Стандартизация			ISO/IEC 15948, ETF RFC 2083			
Дата создания	Начало 1990-х	1987	1996	1985	Начало 1990-х	1996
Дата последних изменений	1995	1989	Почти не менялся	1991	2013	2013
Открытость формата	Да*	Частично**	Да	Нет данных	Нет	Да
Назначение формата	Простой формат хранения растровых изображений, разработанный компанией Microsoft.	Один из первых графических форматов, созданных для сети Интернет. Поддерживает не более 256 цветов и в настоящее время почти не применяется. Поддерживает анимацию.	Растровый формат хранения графической информации, использующий сжатие без потерь по алгоритму Deflate. Создан для замены GIF.	Устаревший формат растровой графики. Поддерживает только RGB палитру. Формат был распространен во времена DOS.	Закрытый графический формат, используемый программой Adobe Photoshop.	Открытый графический формат, созданный специально для программы GIMP. Не рекомендуется для обмена данными, так как его спецификация может меняться.
Наличие бесплатного программного обеспечения	Да	Да	Да	Да	Частично (поддерживается неофициально и не полностью)	Да
Самодостаточность документа (сам документ содержит все необходимое для его корректного отображения)	Да	Да	Да	Да	Да	Да
Возможность полнотекстового поиска	Нет (не хранит текст)	Нет (не хранит текст)	Нет (не хранит текст)	Нет (не хранит текст)	Нет	Нет

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Г

<b>Наличие «опасных» возможностей (встроенного кода, шифрования, средств управления правами доступа и т.д.)</b>	Нет	Нет	Нет	Нет	Нет	Нет
<b>Сжатие графических объектов с потерями</b>	Может использоваться	Нет	Нет	Нет	Может использоваться	Может использоваться
<b>Возможность конвертации из других форматов</b>	Да	Да	Да	Да	Да	Да

\*Официально формат не является открытым, однако он не запатентован и полностью документирован.

\*\*Официально формат GIF может свободно использоваться только для некоммерческого применения. Метод кодирования запатентован.

ПРИЛОЖЕНИЕ Д

СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ ГРАФИЧЕСКИХ ФОРМАТОВ  
(ВКЛЮЧАЯ СЖАТИЕ С ПОТЕРЯМИ И АНИМАЦИОННЫЕ)

Название / включая сжатие с потерями и анима- ционные	PGF	TIFF	JPEG XR	JPEG / JP2 / JPEG- LS	EXR	APNG
Расширение	<b>.pgf</b>	<b>.tiff, .tif</b>	<b>.jxr, .hdp, .wdp</b>	<b>.jpg, .jpeg, .jpe</b>	<b>.exr</b>	<b>.apng, .png</b>
Стандартизация			ITU-T Rec. T.832 (12/2009), ISO/IEC 29199-2:2010	ISO/IEC 10918, ITU-T T.86		
Дата создания	2000	1986	2009	1992	1999	2008
Дата последних изменений	Нет данных	2002	2010	2012	2012	
Открытость фор- мата	Да, но не является свобод- ным.	Нет	Да, но не является сво- бодным.	Да, но запатенто- ван	Да	Да
Назначение форма- та	Растровый формат хранения изображений, разработанный для замены JPEG. Формат слабо распространен.	Гибкий графически фор- мат. Поддерживает сжатие без потерь. Права на дан- ный формат принадлежат Adobe Systems.	Формат кодирования и файловый формат для фотографий, разработан- ный и запатентованный корпорацией Microsoft. Эффективнее JPEG, но слабо распространен.	Один из популяр- ных графических форматов, приме- няемый для хра- нения фотоизо- бражений. При сильном сжатии заметно искажает изображение.	Формат графиче- ских файлов для хранения изображе- ний с широким диа- пазоном яркости (HDRI). Применяет- ся в основном в ки- нопроизводстве.	Формат изображе- ний, осно- ванный на формате PNG и пре- дусматри- вающий возмож- ность хра- нения ани- мации. Слабо рас- пространен.
Наличие бесплат- ного программного обеспечения	Да (очень мало)	Да (некоторые особен- ности могут не поддержи- ваться)	Да	Да	Да	Да

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Д

Самодостаточность документа (сам документ содержит все необходимое для его корректного отображения)	Да	Да	Да	Да	Да	Да
Возможность полнотекстового поиска	Нет	Нет	Нет	Нет	Нет	Нет
Наличие «опасных» возможностей (встроенного кода, шифрования, средств управления правами доступа и т.д.)	Нет	Нет	Нет	Нет	Нет	Нет
Сжатие графических объектов с потерями	Может использоваться	Может использоваться	Может использоваться	Да	Может использоваться	Нет
Возможность конвертации из других форматов	Да	Да	Да	Да	Да	Да



ПРИЛОЖЕНИЕ Е

СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОСНОВНЫХ ГРАФИЧЕСКИХ ФОРМАТОВ  
(ВЕКТОРНАЯ ГРАФИКА)

Название / векторные	AI	CDR	EMF/PICT.	PS	XPS (это аналог PDF, как графический формат не применяется)	EPS	SWF (как графический формат не применяется)	VRML	X3D	SVG
Расширение	.ai	.cdr	.emf, .emz	.ps	.xps, .oxps	.eps, .epsf, .epsi	.swf	.wrl, .wrz	.x3d, .x3dv, .x3dz и др.	.svg, .svgz
Стандартизация					ECMA-388			ISO/IEC 14772-1:1997	ISO/IEC 19775/19776/19777	
Дата создания	Нет данных	Нет данных	1993	1982	2006	Нет данных	1996	1994	2006	2001
Дата последних изменений	Нет данных	Нет данных	Нет данных	1997	2009	Нет данных	Нет данных	1997	2013	2011
Открытость формата	Нет	Нет	Частично (некоторые особенности скрыты)	Да, но не является свободным	Да	Да, но не является свободным	Нет	Да	Да	Да
Назначение формата	Векторный формат хранения графической информации, является проприетарным. Разработан компанией Adobe Systems для представления одностраничных векторных изображений.	Файл проекта, созданный в программе CorelDRAW, который содержит изображение или растровый рисунок. Данный формат файла разработан компанией Corel для использования в собственных программных продуктах.	Универсальный формат векторных графических файлов для Windows приложений.	Язык описания страниц, в основном используемый в настольных издательских системах. Для хранения графики применяется редко.	Открытый формат фиксированной разметки на базе XML разработанный компанией Microsoft. Функционал направлен исключительно на документооборот (документ проще и легче PDF).	Подмножество PS, в котором оставлены только средства для работы с графической информацией.	Проприетарный видео формат для флеш-анимации, векторной графики, видео и аудио в сети Интернет.	Стандартизированный формат файлов для демонстрации трёхмерной интерактивной векторной графики, чаще всего используется в сети Интернет.	Стандарт ISO, предназначенный для работы с трёхмерной графикой в реальном времени. X3D является расширением VRML, включающим анимацию двуногих персонажей, NURBS, GeoVRML и др.	Язык разметки масштабируемой векторной графики. Поддерживает как неподвижную, так и анимированную интерактивную графику.

ПРОДОЛЖЕНИЕ ПРИЛОЖЕНИЯ Е

Наличие бесплатного программного обеспечения	Да (некоторые особенности могут не поддерживаться)	Да (некоторые особенности могут не поддерживаться)	Да	Да	Да	Да	Нет	Да	Да	Да
Самодостаточность документа (сам документ содержит все необходимое для его корректного отображения)	Да	Да	Да	Да	Да	Да	Да	Да	Да	Да
Возможность полнотекстового поиска	Нет	Нет	Нет	Да	Да	Нет	Нет	Нет	Нет	Нет
Наличие «опасных» возможностей (встроенного кода, шифрования, средств управления правами доступа и т.д.)	Нет	Нет	Да (в формате имеется уязвимость)	Нет	Нет	Нет	Да	Нет	Нет	Нет
Сжатие графических объектов с потерями	Нет	Может применяться	Нет	Нет	Может применяться	Нет	Может применяться	Нет	Нет	Нет
Возможность конвертации из других форматов	Да	Нет	Да	Да	Да	Да	Нет	Да	Да	Да

Примечание:

3DS (устаревший закрытый формат трехмерной графики) в таблице не проводится.

ПРИЛОЖЕНИЕ Ж  
ПРИМЕРНАЯ ОЦЕНКА РАСХОДОВ НА ПРОГРАММНОЕ И АППАРАТНОЕ ОБЕСПЕЧЕНИЕ

**Расходы на базовую конфигурацию**

Расходы на базовую конфигурацию представлены в таблице 1

Таблица 1 Расходы на базовую конфигурацию

№	Наименование ПО	К-во, шт.	Цена, руб.	Сумма, руб.
1	ОС MS Windows XP Professional	5	5138	25690
2	ОС MS Windows 2003 Server	1	26339	26339
3	Антивирус NOD32 Enterprise Edition на 10 пользователей	1	14160	14160
4	Microsoft office 2003 Basic	5	6409.52	32047.6
5	1С:Бухгалтерия 8. Комплект на 5 пользователей	1	18000	18000
6	Agnitum Outpost Firewall PRO 2008	1	699 руб.	699
Итого:				116935.6

## Расходы на адаптацию

Установкой и настройкой базового программного обеспечения и прикладного программного обеспечения, а так же настройкой сети и сетевого оборудования будет заниматься технический специалист.

В таблице 2 показаны расходы на оплату труда тех.специалиста.

Таблица 2 Расходы на адаптацию

№	Вид работы	Трудоемкость на 1 операцию, часы	К-во операций, шт.	К-во часов на данный вид работы	Цена за 1 час работы	Цена, руб.
1	Установка и настройка ОС	2	6	12	60	720
2	Установка и настройка прикладного ПО	2	6	12	60	720
3	Настройка сети в офисе	4	1	4	60	240
4	Настройка сервера и принтеров в офисе	2	1	2	60	120
Итого:						1800

## Расходы на приобретение аппаратных средств

В таблице 3 и 4 показаны затраты на приобретение аппаратных средств.

Таблица 3 Конфигурация ПК для сотрудников

№	Наименование	Кол-во шт.	Цена Руб.	Сумма Руб.
1.	Материнская плата Asus P5B-V	5	3272	16360
2	Процессор Socket – 775 Core 2 Duo E4400 2.0 GHz Box	5	4528	22640
3	Оперативная память DDRII 1024 Мб PC5300	5	804	4020
4	Жесткий диск SATA II 160 Gb Seagate 7200 8Mb	5	1782	8910
5	DVD-RW Nec	5	943	4715
6.	Монитор TFT 17” Acer AL1716As черный	5	6453	32265
7	Клавиатура Mitsumi Classic	5	393	1965
8	Мышь Logitech RX-250	5	315	1575
9	Блок бесперебойного питания PowerCom WAR-600	5	1452	7260
10	Сетевой фильтр Sven Optima Pro (8 розеток)	5	212	1060
			Итого:	100770

Таблица 4 Конфигурация сервера

№	Наименование	Кол-во шт.	Цена Руб.	Сумма Руб.
1.	Материнская плата Asus P5B-V	1	3272	3272
2	Процессор Socket – 775 Core 2 Quad Q6600 2.4 GHz	1	8599	8599
3	Оперативная память DDRII 1024 Mb PC6400 Kingston	2	1634	3268
4	Жесткий диск SATA II 250 Gb Seagate 7200 16Mb Raid Edition	4	3400	13600
5	DVD-RW Asus SATA	1	1099	1099
6	Блок бесперебойного питания PowerCom WAR-1000A	1	2478	2478
7	Сетевой фильтр Sven Optima Pro (8 розеток)	1	212	212
			Итого:	32528

Итого для приобретения аппаратных средств потребуется 133928 рублей.