



ПЕРЕДМОВА

Випуск дайджесту присвячено стану та проблемам електронних інформаційних технологій.

У публікації «Экономика обеспечения долговременного доступа к электронной информации: введение в проблему» розповідається про економічний аспект забезпечення доступу до електронної інформації.

У публікації «Сохранение электронной информации: зачем нам нужен зарубежный опыт» розповідається про міжнародний досвід досліджень, як неоцінне джерело збереження цифрової інформації.

У публікації «Анализ предпочтительности цветовой модели в информационно–измерительном комплексе» розповідається про вплив кольорів на зображення.

У публікації «Сохранение электронного контента в России и за рубежом» розповідається про конференцію.

У публікації «США: опубликован отчет о преимуществах и рисках формата PDF/A-3» розповідається про специфікацію файлового формату PDF/A-3.

У публікації «Швеция: Национальные Архивы учат специалистов использовать формат PDF/A» розповідається про курс навчання формату PDF/A.

У публікації «Штат Виктория, Австралия: Пересмотр стандарта VERS» розповідається про перегляд стандарту «Управління електронними документами».

У публікації «Датская стратегия обеспечения долговременной сохранности электронных документов» розповідається про стратегію Датських Національних Архівів щодо архівації електронних документів.

У публікації «Проект E-Ark получил 6 миллионов фунтов на спасение электронных данных» розповідається про виділення коштів на проект довготривалого збереження архівних документів і знань Європи.

У публікації «Отзыв об отчете РГГУ «Сравнительный анализ форматов файлов электронных документов постоянного (долговременного) хранения»» наведено коментар до аналізу форматів файлів електронних документів довготривалого зберігання.

У публікації «Национальные Архивы США опубликовали пересмотренное руководство по архивным форматам» розповідається про переглянуте керівництво щодо передачі електронних документів.

У публікації «США: Подана заявка на патент на метод архивации электронного контента, подписанного электронными цифровыми подписями» розповідається про намагання запатентувати рішення що знаходяться в загальному спілкуванні.

У публікації «На сайте Британского института стандартов начато публичное обсуждение основного терминологического стандарта ИСО в области информации и документации» розповідається про викладений на сайті проект нового стандарту ISO 5127.

У додатку наведено «Аналіз сучасного стану систем перевірки україномовних текстів».



ЭКОНОМИКА ОБЕСПЕЧЕНИЯ ДОЛГОВРЕМЕННОГО ДОСТУПА К ЭЛЕКТРОННОЙ ИНФОРМАЦИИ: ВВЕДЕНИЕ В ПРОБЛЕМУ

Автор: Мурована Т. А., Москва, ответственный секретарь Российского комитета программы ЮНЕСКО «Информация для всех»

Электронная информация сегодня превратилась в фундаментальный и ценный ресурс. Ее сохранение – задача крайне актуальная, поскольку информация является основой экономики, науки, исследовательской и образовательной деятельности, государственной политики, индустрии творчества, а также сектора, связанного с сохранением культурного наследия.

Сохранение электронной информации привлекает к себе внимание как техническая задача и, в меньшей степени, как политическая задача; тем не менее до сих пор ее почти не рассматривали как задачу экономическую.

Московская декларация о сохранении электронной информации, итоговый документ Международной конференции «Сохранение электронной информации в информационном обществе: проблемы и перспективы» (Москва, 3–5 октября 2011 г.), говорит о том, что в сфере экономики приоритетными являются следующие направления деятельности:

- разработка базовых принципов организации финансирования деятельности по долгосрочному сохранению электронной информации силами традиционных институтов памяти;
- включение оценки усилий, подходов и решений с точки зрения их экономичности в качестве обязательного компонента стратегии и деятельности в сфере сохранения электронной информации; повышение уровня информированности в этой области.

Однако для размышлений об экономическом аспекте сохранения необходимо изменить привычный для институтов памяти подход. Если специалисты таких институтов говорят о том, что без сохранения информации не будет доступа к ней, то экономисты предупреждают, что без выраженного спроса на доступ к сохраненным информационным ресурсам не будет и предложения услуг по обеспечению их сохранности.

Наиболее подробной и всесторонней работой в сфере экономики обеспечения долговременного доступа к электронной информации можно считать итоговый отчет американской Рабочей группы по устойчивому обеспечению долговременной сохранности и доступа к электронной информации (*Blue Ribbon Task Force on Sustainable Digital Preservation and Access*) «Устойчивая экономика для цифровой планеты. Обеспечение долговременного доступа к электронной информации» [http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf]. В подготовке отчета принимали участие

Национальный научный фонд США, Фонд Эндрю Меллона, Библиотека Конгресса, Объединенный комитет по информационным системам Великобритании, Национальное управление архивов и документации, Совет по библиотечным и информационным ресурсам.

Согласно этому документу, сохранение электронной информации с экономической точки зрения – это деятельность, для которой принимаются четкие решения относительно текущего распределения ресурсов, имея в виду достижение долгосрочных целей.

Рабочая группа сосредоточила свое внимание на тех материалах, которые имеют долговременное общественное значение:

- *научный дискурс*: опубликованные документы научного характера;
- *исследовательские данные*: первичные источники исследований, а также результаты этих исследований первого порядка;
- *культурный контент в коммерческой собственности*: цифровой контент, имеющий культурное значение и находящийся под защитой авторских прав;
- *коллективно создаваемый веб-контент*: электронный контент, создаваемый в интерактивном режиме, результат сотрудничества и участия потребителей.

Все электронные ресурсы как экономические продукты обладают четырьмя обязательными характеристиками:

1. Спрос на обеспечение сохранности электронной информации есть *спрос производный* (расходы на сохранение составляют часть стоимости цифрового продукта, поэтому решение о сохранении того или иного документа будет приниматься на основе выявленной ценности электронных материалов в долговременной перспективе);
2. Электронные материалы являются *обесцениваемыми ресурсами длительного пользования* (их ценность меняется с течением времени в связи с физическим износом носителя, порчей файла, устареванием форматов, отсутствием воспроизводящего оборудования);
3. Электронные ресурсы являются *неконкурентными в потреблении* и создают возможность незаконного бесплатного пользования ими (использование одними потребителями не уменьшает возможности использования другими, т. к. после того, как одна из сторон сохранила ресурсы, они практически всегда предназначены для всех. В этом случае мотивы одной-единственной стороны к несению всех затрат ослабевают, поскольку все остальные стороны могут пользоваться результатом, не приложив к его созданию никаких усилий);
4. Процесс обеспечения сохранности электронной информации имеет *длительную протяженность во времени и зависит от предшествующих решений* (любое принятое сегодня решение формирует будущие обстоятельства и определяет диапазон будущих возможностей выбора).

Экономический анализ области обеспечения сохранности этих материалов позволяет выявить структурные особенности, влияющие на стратегии обеспечения сохранности в целом:

- длинный горизонт инвестирования;

- смешанный состав заинтересованных сторон;
- разбалансированность и недостаточность стимулов;
- отсутствие ясности относительно роли и сферы ответственности заинтересованных сторон.

В результате анализа таких особенностей были предложены три императива устойчивого сохранения цифровой информации:

1. Формулировка убедительного ценностного предложения

Обосновывая необходимость сохранения информации, стоит в первую очередь говорить о ее использовании. Стороны, заинтересованные в электронной информации, как правило, принадлежат различным сообществам. Интересы будущих пользователей мало учитываются при отборе материалов, предназначенных для хранения. Доверенные общественные институты – библиотеки, архивы, музеи, профессиональные организации и проч. – могут сыграть важную роль полномочных представителей будущих поколений своих пользователей.

2. Четкое определение стимулов обеспечения сохранности информации на благо общества

Отсутствие очевидных мотивов помешает своевременности действий по обеспечению сохранности информации. Разбалансированность стимулов, побуждающих участников процесса к деятельности, наблюдается между теми сообществами, которые получают от него выгоду (следовательно, имеют стимул к сохранению), и теми, кто должен это сохранение обеспечивать (кто владеет ресурсом или контролирует его), но не имеет стимула к этому. Для смягчения этой ситуации можно использовать политические механизмы:

- финансовые стимулы и другие методы поощрения частных владельцев, сохраняющих электронную информацию на благо общества;
- в необходимых случаях предоставление полномочий на сохранение;
- пересмотр положений закона об авторских правах, что позволило бы уполномоченным организациям, представляющим долговременные интересы общества, сохранять электронные материалы, находящиеся в частной собственности.

Препятствия на пути к эффективному децентрализованному управлению процессом могут быть устранены за счет использования создателями информации неисключительных лицензий, дающим право третьей стороне на сохранение информации.

3. Определение роли и обязанностей заинтересованных сторон на протяжении всего жизненного цикла электронного документа

Самые сильные стимулы будут неэффективными без четко сформулированного соглашения о роли и обязанностях всех действующих сторон – тех, кто создает информацию, кто ею владеет, тех, кто ее сохраняет, и тех, кто делает ее доступной для использования. Каждая организация, создающая и использующая информацию, должна разработать и внедрить политику и алгоритм обеспечения сохранности информации, в том числе:

- отбор материалов, имеющих долгосрочную ценность;

- подготовку данных к архивированию;
- должна иметь протоколы, по которым можно будет осуществлять плавную и безопасную передачу цифровых ресурсов за пределы организации и обмен ею между институтами.

Все заинтересованные стороны вынуждены принимать ряд серьезных решений относительно управления информационными ресурсами, которые они создают, которым придают добавленную стоимость, которые используют и распространяют для дальнейшего использования.

Каждый сектор должен оценить долгосрочную ценность имеющихся у него цифровых ресурсов и определить соотношение между преимуществами и затратами на сохранение этих данных, передачу их другим сторонам, которые будут распоряжаться ими в долгосрочной перспективе или уничтожат данные ресурсы.

Эти три элемента являются ключевыми при разработке стратегий, которые должны оказаться достаточно гибкими, чтобы их можно было бы адаптировать к изменениям в использовании, технологиях, заинтересованных сообществах, которые будут происходить на протяжении жизненного цикла электронной информации. Каждый из элементов заставляет тех, кто принимает решения, устранять неопределенность, оставлять возможность выбора будущим участникам процесса, преодолевать препятствия на пути к устойчивости.

Из этих императивов следуют пять условий, соблюдение которых делает обеспечение сохранности устойчивым:

- признание преимуществ, которые обеспечиваются сохранением информации, теми, кто принимает решения;
- определение процесса отбора электронных документов, обладающих долгосрочной ценностью;
- разработка системы стимулов для тех, кто принимает решения о сохранении информации в общественных интересах;
- соответствующие организация и управление деятельностью по обеспечению сохранности;
- разработка механизмов текущего эффективного распределения ресурсов деятельности по обеспечению сохранности электронной информации.

В отчете предлагается программа действий, предусматривающая самые разнообразные меры и сферы их применения. Она закладывает основу устойчивого обеспечения сохранности, поскольку создает необходимые условия для этого. В краткосрочной перспективе приоритетными названы следующие направления деятельности:

Организационные действия:

- развитие государственно-частных партнерств;
- обеспечение организаций подготовленным персоналом – от информационных специалистов до юристов и представителей бизнеса;
- создание и обеспечение с течением времени устойчивого функционирования цепочек управления между организациями;
- достижение эффектов масштаба и диверсификации;

— решение проблемы невозмещаемого доступа.

Технические действия:

- создание инфраструктуры, обеспечивающей возможность управления информацией во всех сферах;
- снижение общих затрат на обеспечение сохранности;
- определение оптимального уровня технического сопровождения, необходимого для реализации стратегии опционов для всех видов цифровых материалов.

Действия в сфере государственной политики:

- внесение изменений в законодательство об авторском праве с учетом потребностей обеспечения сохранности электронной информации;
- создание финансовых стимулов и стандартов передачи информации для частных владельцев, сохраняющих информацию для общественного блага;
- спонсорская поддержка государственно-частных партнерств;
- разрешение правовых вопросов, связанных с интернет-материалами;
- придание полномочий ведущим профильным организациям по защите цифровых материалов неизвестной принадлежности во избежание их невозможной утраты.

Действия в сфере образования и общественной жизни:

- развитие образования и обучения специалистов в сфере сохранения электронной информации – специальная подготовка в сфере того или иного контента;
- передовые технологии сопровождения информации, базовые знания в соответствующей научной дисциплине, технические и инженерные дисциплины, математика;
- информирование общества о необходимости своевременного обеспечения сохранности материалов.

В заключение важно отметить, что решения о сохранении материалов всегда принимаются в условиях неопределенности: в течение их жизненного цикла будут меняться технологии, политическая обстановка, инвестиционные приоритеты, общественное мнение. Тем не менее можно принять такую тактику, которая поможет принять или прояснить неопределенность, распределить ресурсы между заинтересованными сторонами и которая, кроме того, оставит открытые возможности для тех, кто будет принимать решения в будущем.

Стратегия устойчивого сохранения позволит превратить неопределенность времени и ресурсов в возможности и маневренность, способность адаптироваться к изменению приоритетов, перенаправлять ресурсы туда, где в них нуждаются больше всего.

Хотелось бы привлечь внимание к одному выводу, заключающемуся в том, что устойчивая экономика сохранения электронной информации – это не постоянный поиск финансовых средств. Электронными ресурсами необходимо активно управлять на протяжении длительного периода времени, используя наиболее эффективные методы организации и управления данными в течение всего их жизненного цикла, охватывающего создание,

описание, поддержку, безопасное хранение, использование и повторное использование. Некоторые электронные материалы требуют достаточно интенсивных мер по сохранению, которые лишь и способны обеспечить возможность их использования. Другим документам такие меры не требуются. Однако в любом случае доступ к информации завтра зависит от мер по обеспечению сохранности сегодня.

Обязательства, принятые сегодня, не могут стать обязательствами на все времена. Однако уже сегодня необходимо предпринять действия, обеспечивающие возможность выбора в будущем. Для каждого из секторов, в которых создается электронная информация, будь то государственный или частный, научный или культурный сектор, необходимо смоделировать и опробовать функциональную стратегию обеспечения сохранности информации. Все заинтересованные сообщества должны принять на себя ведущую роль в развитии общей инфраструктуры сохранения электронных материалов, которая сохранит свою устойчивость до будущих поколений. Эта задача, как и другие стоящие перед обществом задачи, требует сбалансированного подхода, учитывающего потребности и устремления как сегодняшнего, так и завтрашнего дня.



СОХРАНЕНИЕ ЭЛЕКТРОННОЙ ИНФОРМАЦИИ: ЗАЧЕМ НАМ НУЖЕН ЗАРУБЕЖНЫЙ ОПЫТ

Автор: Кисловская Г. А., Москва, Российская государственная детская библиотека

Многочисленные доклады, посвященные росту количества отечественных электронных ресурсов, генерируемых и библиотеками, и СМИ, и архивами, и музеями, не проливают, к сожалению, свет на вопрос об их долговременной сохранности а следовательно, и долговременного доступа к цифровым объектам разного происхождения.

Судьба контента, рожденного в цифровом виде, равно как и судьба цифровых объектов, появившихся в результате конверсии традиционных носителей информации, одинаково не ясна.

Хартия цифрового наследия, принятая ЮНЕСКО в 2003 г., предостерегает нас от невнимательного отношения к цифровому контенту. Занимаясь вопросами обеспечения сохранности вот уже около 20 лет, я отмечаю, что в профессиональной библиотечной среде наметилась довольно опасная тенденция декларации рисков, с которыми связано бытование цифровых объектов, и полное отсутствие каких-либо действий, которые бы эти риски снизили или устранили. Существует до сих пор и другая крайность: возводить оцифровку в ранг еще одной методики сохранности

наряду с консервацией, микрофильмированием и другими методиками, которые составляют суть Национальной программы сохранения библиотечных фондов Российской Федерации.

Заблуждение это легко можно развеять, если вспомнить о том, что один раз в два года (в лучшем случае) меняется программно-аппаратная среда, в которой создается, хранится и предоставляется в доступ цифровой объект. На эту тему есть много исследований, аргументировано обосновывающих риски, с которыми сопряжены жизненные циклы цифровых документов.

Я бы не хотела повторять в своем выступлении прописные истины. Но вот на тему, связанную с формированием Национальной программы обеспечения сохранности цифрового наследия, мне бы хотелось высказать свое мнение.

В рамках Национальной программы сохранения библиотечных фондов есть несколько различных компонентов, но до сих пор нет подпрограммы, которая бы решала вопросы обеспечения сохранности цифрового наследия. В свое время международный опыт, в первую очередь, распространяемый под эгидой ИФЛА, много дал для структурирования нашей Национальной программы. И сегодня международный опыт – неоценимый источник в формулировании основных постулатов цифровой сохранности на национальном и институциональном уровнях, которые, конечно, связаны между собой. Русскоязычный центр ИФЛА, созданный в РГБ, уделяет много времени переводу на русский язык и распространению документов, посвященных этой злободневной теме.

Мне бы хотелось привлечь ваше внимание к одному исследованию, результаты которого, как мне кажется, могли бы лечь в основу планирования нового сегмента Национальной программы.

Речь пойдет об исследовании, проведенном Британской библиотекой и Университетским колледжем в Лондоне под эгидой JISC (Joint Information Scientific Committee). Исследование называется LIFE. Его выводы были обнародованы в 2006 г. [Полную информацию о проектах серии LIFE 2005 – 2013 гг. на английском языке см. на <http://www.life.ac.uk/>. Прим. ред.] В ходе исследования были тщательно обсчитаны расходы, связанные с комплектованием, проверкой цифровых документов, созданием метаданных и т. д. Исследование интересно не только своими финансовыми расчетами, но и другими результатами, которые могли бы быть заимствованы нашими специалистами. Так, например, поскольку у нас нет закона об обязательном сборе и хранении сетевых электронных ресурсов, который в определенной степени мог бы влиять на их сохранность и доступность, можно было бы воспользоваться опытом Британской библиотеки в том, какие документы она исключала из схемы добровольного депонирования электронных документов. Понятно, что мы и сами можем определить, что включать, а что не включать в этот список, но изобретать велосипед в эпоху глобализации неэффективно и нерационально. Британская библиотека рекомендует не принимать на долговременное хранение электронные материалы, которые рекламируют конкретные коммерческие продукты, электронные открытки,

электронные игры, онлайн-документы, генерируемые за пределами страны, и др.

Общеизвестно, как важно создавать метаданные при оцифровке. Так вот, исследование показало, что 50 % всех затрат, связанных с депонированием и поддержанием жизненного цикла электронного документа, связано с созданием метаданных (технических, описательных, административных, сохраненных).

В исследовании сравниваются характеристики стоимости хранения в примитивном файловом хранилище и в DOM (Digital Object Management) системе или репозитории, совместимом с Open Archive Information System (OAIS). Так, цена хранения 1 Гб в течение первого года в обычном файловом хранении – 80,85 фунтов стерлингов, а в DOM – 27,44. Соответственно, в течение 5 лет цена снижается до 16,17 в простом файловом хранении и до 5,48 фунтов в DOM.

Исследование дает детальнейший стоимостный анализ того, сколько стоит хранение, инсталляция, поддержка софта, причем в привязке к совершенно определенному типу файлов. Параллельно хотелось бы дать еще одну ссылку на данные, полученные Digital Preservation Coalition, относительно хранения цифровых архивов большого объема (www.dpconline.org/graphics/reports/index.html#gescale), в соответствии с которыми будет наблюдаться падение цены хранения на жестких дисках примерно на 30 – 40 % при удвоении количества информации каждые 12 месяцев.

Все, кто занимался проектированием, причем не обязательно относящимся к сохранности, знают, что один из самых трудоемких процессов – это создание бюджета проекта. Вот почему цифры, которые приводятся в разных зарубежных источниках, так полезны.

При изучении такого этапа жизненного цикла электронного объекта, который называется «сохранность», исследование LIFE отметило, что при хранении файлов в репозиториях, совместимых с OAIS, не имеет значение операционная система, а потому меньше риск потери информации. Исследователи показали, что из 231773 файлов в 22 форматах не был потерян ни один. Тем не менее, некоторые форматы явно стареют, все меньше поддерживаются на программном уровне.

Наглядным примером тому – файлы в формате GIF, который был наиболее используемым графическим форматом в 2006 г. К моменту завершения первого этапа проекта LIFE этот формат, созданный в 1989 г., стал использоваться гораздо реже, поскольку файлы создавались на основе закрытых кодов, а ему на смену пришел новый – PNG (Portable Network Graphic).

Наиболее уязвимыми с точки зрения сохранности были названы файлы в форматах GIF, BMP, PCX, и хотя в коллекции депонированных электронных документов Британской библиотеки таких файлов всего 1,16 %, им предстояло пройти процедуру миграции в новые форматы. 75 % файлов (2669) были сохранены в форматах TXT и HTML. Ни один из них не был под

угрозой потери. И, тем не менее, нельзя было сказать, что все с этими файлами было в порядке. Так, сохранению файлов в формате HTML угрожали сложности, сопряженные со средой бытования этих файлов и их отношением с другими программами и языками. Поэтому проводились тесты по конвертированию в более предпочтительный формат – PNG, и исследовалась цена вопроса. В качестве средства миграции файлов был выбран PHOTOSHOP 6/0 и рассчитана цена миграции: 1 файл – 0,22 фунта; цена тестирования – 0,18. Итого 0,40 фунтов.

Одним из позитивных выводов исследования являлось то, что опасность потери цифрового контента несколько преувеличена. Но цифровые объекты немые, а потому за ними следует следить внимательно, причем отслеживать нужно весь жизненный цикл цифрового объекта.

Еще одним практическим выводом исследования был вывод о том, что использование старой кодировки HTML крайне ненадежно при работе с вебом. Поэтому при решении вопросов сохранности следует переводить формат HTML в XML, и это гарантирует доступ к файлам с помощью различных браузеров.

Однако ручное конвертирование 70000 HTML-файлов, хранящихся в Британской библиотеке, очень дорогостоящее дело. Поэтому исследование рекомендовало сделать процедуру конвертирования автоматизированной на этапе «ingest», т.е. на этапе загрузки файла на сервер библиотеки. Таким образом, наши коллеги предлагают нормализацию файлов в качестве первой ступени обеспечения сохранности.

Много полезной информации можно найти в этом исследовании и в отношении веб-архивирования, к которому в России еще толком никто не приступал. Если в переводах других материалов ИФЛА, сделанных Русскоязычным центром этой Федерации, дается информация о критериях отбора сайтов для веб-архивирования и других организационных принципах этой важной работы, уже ведущейся во Франции, Швеции, Дании, то исследование LIFE делает акцент на стоимостных характеристиках этого непростого процесса, начиная с выявления и отбора сайтов и заканчивая предоставлением заархивированных сайтов в доступ, что дает возможность использовать эту методику и нам при обсчете нашей будущей политики и практики.

Программа PANDAS, разработанная Национальной библиотекой Австралии, использовалась наряду с программой HT Track для управления проектом веб-архивирования. На момент написания отчета о результатах исследования Британская библиотека архивировала ежегодно до 1000 сайтов.

Очень интересна мысль о том, кто должен отбирать сайты. Отбором занимаются subject specialists, т.е. комплектаторы, а точнее библиографы-предметники. Исследование дает описание команды, работающей над отбором. В нее обязательно входит человек, занимающийся прояснением вопросов авторского права на сайт.

Замечательной особенностью этого исследования является то, что все расчеты даны и в стоимостных и во временных категориях, что важно для

использования в российских реалиях. К примеру, при обсчете нашей будущей программы сохранности цифрового контента полезны такие данные: 5 % всех расходов при веб-архивировании на серверах сторонней компании выплачивается за предоставление доступа к контенту, а 95 % – за собственно хранение данных. Каждый архив сайта состоит из множества цифровых объектов в разных форматах, и стоимость обеспечения сохранности и доступности такого сайта очень сложно рассчитать. Тем не менее, британцы подсчитали, что стоимость этой работы в течение одного года за один сайт составляет 707 фунтов, а за 20 лет – 13 731.

Подводя итог, хочу еще раз напомнить об отсутствии в России национальной стратегии в вопросах сохранения цифрового наследия и об огромном опыте, накопленном нашими зарубежными коллегами, который доступен через Русскоязычный центр ИФЛА, работающий в интересах всего сообщества учреждений культуры России.



АНАЛИЗ ПРЕДПОЧТИТЕЛЬНОСТИ ЦВЕТОВОЙ МОДЕЛИ В ИНФОРМАЦИОННО–ИЗМЕРИТЕЛЬНОМ КОМПЛЕКСЕ

Автор: Р. С. Череватый, канд. техн. наук, доц., dla-obmenf@rambler.ru
(Россия, Тула, ТулГУ)

Рассмотрено влияние нарушения синхронизации каналов в разных цветовых пространствах на цветопередачу и цветовой контраст изображения по отношению к исходному изображению при восстановлении полноцветного изображения из многокадровой модели совмещением цветовых плоскостей.

Документы, представляющие некоторый носитель с нанесенной на них информацией, имеют важное значение для всех структур общества, начиная от межгосударственных организаций и кончая мелкими коммерческими фирмами, семьями и отдельными гражданами. Частичная или полная утрата документированной информации, ее несанкционированное изменение приводят к невосполнимым потерям как финансового, так и юридического характера. Одним из носителей, обеспечивающих длительное хранение документированной информации, является микрофильм, представляющий собой фото пленку с изображением документа.

Микрофильм как носитель информации обладает следующими достоинствами:

- кодирование документа сводится к уменьшению геометрических размеров изображения;

- для хранения микрофильма не требуется значительных производственных площадей;
- климатические условия при хранении микрофильмов могут быть менее жесткими, чем при хранении бумаги и электронных носителей;
- для нанесения информации на микрофильм не требуется сложного оборудования;
- геометрические размеры и оптические свойства микрофильма жестко регламентированы, что делает возможным применение его в качестве носителя в автоматизированных информационно-поисковых системах;
- в процессе хранения информации в нее не могут быть внесены несанкционированные изменения;
- при использовании микрофильма не требуется параллельного хранения специализированных технических средств доступа и соответствующего программного обеспечения.

Традиционные технологии микрофильмирования ориентированы на хранение текстовой и чертёжно-графической документации. Изображения подобных документов являются черно-белыми и хорошо переносятся на высококонтрастные светочувствительные материалы, применяемые в микрофильмировании, поскольку передачи промежуточных градаций серого тона здесь обычно не требуется. Вместе с этим, к настоящему времени в архивах накоплен значительный по объёму фонд документов, в которых часть информации передается с кодированием цветом (цветной документацией) и которые требуют долговременного хранения с сохранением именно цветового кодирования (топографические карты, схемы и планы с нанесенной цветовой разметкой, документация, содержащая цветные иллюстрации, графики, диаграммы и т.п.).

В то же время технологический процесс подготовки микрофильма к гарантированному длительному хранению ориентирован на черно-белые оригиналы, жестко регламентирован и слабо поддается модификациям.

Поэтому в настоящее время существует проблема, заключающаяся в отсутствии промышленной технологии микрофильмирования цветной документации, близкой к стандартному технологическому процессу микрофильмирования черно-белых документов и обеспечивающей сопоставимые с черно-белыми микрофильмами сроки хранения.

Таким образом, наличие потребности в создании страхового фонда цветной документации и отсутствие простых и надежных информационно-измерительных систем, реализующих технологический процесс микрофильмирования цветных документов, делают задачу проведения исследований по разработке методов их проектирования весьма актуальной.

Существующие технологии аналогового цветного микрофильмирования с использованием цветных светочувствительных материалов проигрывают традиционным чёрно-белым по целому ряду показателей, и, прежде всего, по стоимости используемых материалов и процесса химико-фотографической обработки в целом, а также по длительности гарантированного хранения информации.

Вариант аналогового микрофильмирования цветных оригиналов с использованием чёрно-белых светочувствительных материалов и оптическим цветоделением отличается сложностью подбора параметров экспонирования и химико-фотографической обработки, обеспечивающих полутоновое черно-белое изображение, а следовательно, корректную цветопередачу оттенков.

В качестве альтернативного подхода к решению указанной задачи используются гибридные технологии аналого-цифрового микрофильмирования, в которых изображение документа с помощью измерительно-информационного комплекса предварительно преобразуется в цифровую форму, обрабатывается средствами цифровой вычислительной техники, и лишь затем выводится на микрофильм с использованием СОМ-системы.

Если на этапе существования изображения документа в форме цифровой модели информацию о цвете удастся закодировать яркостным сигналом, то в качестве долговременного носителя изображения такого документа можно будет использовать традиционные чёрно-белые светочувствительные материалы. Это приводит к необходимости разработки математического и алгоритмического обеспечения, реализующего адекватные преобразования моделей изображений при их выводе на микрофильм и считывании с него.

Таким образом, использование чёрно-белых светочувствительных материалов в задачах микрофильмирования документации с цветовым кодированием информации требует решения научно-технических задач:

- цветоделения исходного изображения документа;
- растеризации полутоновых изображений цветовых плоскостей для вывода на микрофильм;
- фильтрации изображений отдельных цветовых плоскостей;
- совмещения цветовых плоскостей в единое цветное изображение;
- цветовой коррекции восстановленного изображения.

На практике количество цветовых составляющих ограничивают обычно тремя, выбирая в качестве первичных цветов красный, зелёный и синий (цветовая модель RGB).

Модуль формирования служебной информации может вносить в изображение дополнительную информацию, упрощающую процесс последующего восстановления цветного изображения с цветоделённой копии (метки-реперы для совмещения цветовых плоскостей, тип плоскости и т.п.).

Однако даже применение меток для совмещения цветовых плоскостей при восстановлении цветного изображения из многокадровой модели цветоделённого изображения в цветовой модели RGB не позволяет точно совместить цветовые плоскости в единое цветное восстановленное изображение с правильной цветопередачей по отношению к исходному изображению.

Это связано с тем, что в информационно-измерительном комплексе гибридного микрофильмирования пространственный сигнал, содержащийся изначально на цветном документе в виде изображения, последовательно

проходит ряд преобразований, связанных со сменой материального носителя сигнала. Каждое из преобразований обусловлено типами респондентов, участвующих в обмене информацией и может рассматриваться как своеобразный канал передачи данных.

Естественно, что каждый этап преобразований связан с потерями данных, которые в комплексе приводят к следующим явлениям:

- изменению общего контраста изображения;
- изменению цветового контраста;
- искажению цвета;
- появлению ореола на границах темного и светлого участков.

Комплекс микрофильмирования цветной документации должен включать в себя модули, приведенные на рис. 1. Сканер документов предназначен для формирования цифровой модели изображения документа оригинала, представляющей собой числовую матрицу, элементы которой характеризуют яркость и цветовой тон соответствующих точек изображения. Фактически такой сканер выполняет также первичное цветоделение оптического изображения, формируя несколько монохромных цветовых плоскостей, яркости точек которых соответствуют интенсивностям цветовых составляющих отдельных пикселей.

Выбор цветового пространства является неоднозначной процедурой, и с точки зрения дальнейших преобразований информации система RGB не обязательно оказывается наилучшей [1, 2, 3].

Например, при кодировании цвета в виде триады «цветовой тон-насыщенность-яркость» (система HSV), яркостная составляющая, представляющая собой полноценную монохромную версию изображения, может использоваться отдельно от остальных в тех случаях, когда наличие цвета не критично (быстрый поиск документа, предварительный просмотр и т.п.).



Рис. 1. Процесс формирования микрофильма

При использовании системы RGB для решения аналогичных задач пришлось бы считывать все три цветовые составляющие. С другой стороны, деление на большее количество цветовых плоскостей может повысить результирующее качество восстановленного цветного изображения по сравнению с традиционной системой RGB.

При цветоделении изображения, содержащего большой объем контуров и текста, а также выводе цветоделенных каналов на черно-белую пленку и последующем обратном процессе, деление на 4 канала (СМΥК), а не на 3 (RGB) дает в итоге более высокое качество передачи контуров и цвета изображения.

При воспроизведении изображения тремя цветными каналами текст и контурные элементы генерируются информацией из трех же каналов. Если хотя бы один из них будет смещен или искажен по всему полю изображения или даже в определенном его фрагменте, потери в зоне основного информационного содержания будут более значительны, чем при использовании 4 каналов, так как четвертый канал (К) выполняет ключевую роль в построении изображения. В результате после конвертации в СМΥК даже один-единственный канал черного (К) будет содержать основную информацию изображения (символы, подписи, контуры, абрисы объектов).

Канал Black в большинстве случаев содержит большее количество информации, чем остальные 3 канала (исключение - живописные изображения), в противовес модели RGB, где все каналы несут примерно равный объем информации, позволяя сохранять контуры и текст в лучшем качестве.

Относительно цветопередачи цветоделенного изображения даже потеря 1/3 информации (в одном из каналов RGB) более критична, чем потеря 1/4 информации (СМΥК), а искажение ахроматической компоненты в СМΥК будет меньшим, чем при цветоделении в RGB, притом искажение цвета в RGB больше, чем в СМΥК. Однако при микрофильмировании цветных оригиналов с использованием чёрно-белых светочувствительных материалов и цветоделением изображения в СМΥК возникает проблема, заключающаяся в необходимости записи четырех монохромных составляющих цветоделенного изображения (цветовое пространство СМΥК) вместо трех (RGB), что нерационально, так как увеличивается площадь записи одного изображения на микрофильм и увеличивается трудоемкость при сборке цветоделенного изображения.

В отличие от цветовых пространств RGB или СМΥК, которые являются набором аппаратных данных для воспроизведения цвета на бумаге или на экране монитора (цвет может зависеть от множества различных факторов), Lab однозначно определяет цвет. Lab нашел широкое применение в программном обеспечении для обработки изображений в качестве промежуточного цветового пространства, через которое происходит конвертирование данных между другими цветовыми пространствами. Благодаря характеру определения цвета в Lab возможно отдельно воздействовать на яркость, контраст изображения и на его цвет. Это

позволяет ускорить обработку изображений, Lab предоставляет возможность избирательного воздействия на отдельные цвета в изображении, усиления цветового контраста, также очень важными являются возможности, которые Lab предоставляет для борьбы с шумом на изображении, кроме того, пространство Lab более удобно для цветокоррекции, чем похожее на него пространство HSB [4, 5].

Каналы в цветовом пространстве Lab не нуждаются в точном совмещении [5], что имеет важное значение в связи с тем, что при гибридном микрофильмировании данные о цвете, содержащиеся изначально в документе, последовательно проходят ряд преобразований, связанных со сменой материального носителя сигнала, приводящих к несовпадению каналов и, следовательно, к неправильной цветопередаче. В RGB это невозможно, так как все три канала несут данные о цвете и образуют единую группу. На рис. 2 представлено оригинальное изображение, подвергавшееся рассинхронизации каналов в RGB- и Lab-версии пространства.



Рис. 2. Оригинальное изображение

На рис. 3 и 4 представлено оригинальное изображение с принудительно нарушенной синхронизацией каналов в RGB- и Lab-версии соответственно, три канала смещены относительно друг друга на 10 пикселей (красный канал в RGB-версии файла сдвинут вниз и вправо, а синий – влево; в Lab -версии файла канал a сдвинут вниз и вправо, а канал b – влево).



Рис. 3. Изображение с нарушенной синхронизацией каналов RGB



Рис. 4. Изображение с нарушенной синхронизацией каналов в Lab - версии

В результате несовмещения трех каналов, каждый из которых участвует в формировании деталей, изображение на рис. 3 стало нерезким, произошло искажение цвета, образовался ореол на границах темного и светлого участков, появились какие-то другие формы и контуры. На рис. 4

пострадал красноватый цвет, поскольку нарушилось совмещение каналов А и В, но в остальном картинка выглядит вполне приемлемо.

Таким образом, нарушение синхронизации каналов в LAB-версии не требует такого точного совмещения каналов при восстановлении полноцветного изображения, как в других цветовых пространствах, а применение цветовой модели LAB более предпочтительно при восстановлении цветного изображения из многокадровой модели цветоделённого изображения совмещением цветowych плоскостей и позволяет получить цветное восстановленное изображение с более правильной цветопередачей по отношению к исходному изображению при прочих равных условиях на каждом этапе преобразований, связанных с потерями данных, приводящих к изменению общего и цветового контраста изображения, к искажению цвета и появлению ореола на границах темного и светлого участков, а также позволяет минимизировать затраты и ускорить работу.

Список литературы

1. Иванов Р.Н. Репрография. М.: Экономика, 1986. 335 с.
2. Кузьмин И.В., Кедрус В.А. Основы теории информации и кодирования. Киев: Вища школа, 1986. 360 с.
3. Мамиконов А.Г., Кульба В.В. Синтез оптимальных систем обработки данных. М.: Наука, 1986. 280 с.
4. Дэн Маргулис Photoshop для профессионалов. Классическое руководство по цветокоррекции. М.: Интерсофтмарк, 2003. 464 с.
5. Дэн Маргулис Photoshop LAB Color. Загадка каньона и другие приключения в самом мощном цветовом пространстве М.: Интелбук, 2006. 450 с.



СОХРАНЕНИЕ ЭЛЕКТРОННОГО КОНТЕНТА В РОССИИ И ЗА РУБЕЖОМ

Источник: Сборник материалов Всероссийской конференции, Москва, 2013

Предисловие

В Москве в Доме русского зарубежья имени Александра Солженицына состоялась Всероссийская научно-практическая конференция «Сохранение электронного контента в России и за рубежом», по материалам которой выпущен сборник в котором представлены доклады и статьи участников. Организаторами конференции выступили Министерство культуры Российской Федерации, Федеральное агентство по печати и массовым

коммуникациям, Российский комитет Программы «Информация для всех» и Межрегиональный центр библиотечного сотрудничества.

Эта первая конференция всероссийского масштаба на данную тему носила междисциплинарный и межотраслевой характер. Она объединила около 150 участников – руководителей и представителей библиотек, архивов, музеев, НИИ, вузов, издательств, печатных и электронных СМИ, а также федеральных и региональных органов управления в сфере культуры, науки, образования, информационных технологий и связи, печати и массовых коммуникаций.

Конференция была направлена на повышение уровня информированности власти, специалистов и широкой общественности о вызовах, связанных с вопросами долгосрочного сохранения электронной информации, и различных аспектах этой проблемы, а также на создание импульса к формированию философии, методологии, стратегии, политики и деятельности на общероссийском и институциональном уровнях в этой сфере.

Свои приветствия в адрес организаторов и участников конференции направили Министр культуры РФ А. А. Авдеев и Ответственный секретарь Комиссии РФ по делам ЮНЕСКО Г. Э. Орджоникидзе.

Конференцию открыл заместитель Руководителя Федерального агентства по печати и массовым коммуникациям В. В. Григорьев, который подчеркнул, что данное мероприятие – это «первая в истории России всероссийская научно-практическая конференция, полностью посвящённая рассмотрению различных аспектов проблемы сохранения электронной информации, проблемы, которая сегодня волнует весь мир. Данная проблема имеет множество аспектов – философский, социокультурный, политический, управленческий, правовой, этический, экономический, психологический, методологический, методический, организационный, технологический, кадровый, однако многими эта проблема ошибочно воспринимается как сугубо техническая, технологическая, компьютерная».

На открытии конференции выступили заместитель председателя Комитета по науке, образованию, культуре и информационной политике Совета Федерации Л. Б. Нарусова, заместитель директора департамента образования, науки и информационных технологий – начальник отдела библиотек и архивов Министерства культуры Российской Федерации Т. Л. Манилова, директор Дома русского зарубежья имени Александра Солженицына В. А. Москвин. Они подчеркивали актуальность для России темы конференции, необходимость формирования стратегии, политики и законодательства в этой сфере, тщательной проработки вопроса о выявлении и распределении ролей, функций, полномочий и ответственности традиционных и новых институтов и структур, которые должны быть вовлечены в деятельность по долгосрочному сохранению электронной информации.

Проблема долгосрочного сохранения электронной информации рассматривалась на конференции с различных ракурсов и точек зрения:

международный и отечественный, федеральный и региональный опыт; профессиональная специфика различных институтов памяти и других структур, вовлеченных и/или заинтересованных в сохранении электронной информации; экономика долгосрочного сохранения, его социокультурные и правовые аспекты.

Обсуждения, состоявшиеся на конференции, ясно показали, что далеко не все российские традиционные хранители аналоговой информации – библиотеки, архивы и музеи – осознают угрозы, риски, возможности и перспективы, связанные с вопросами долгосрочного сохранения электронной информации. На политическом уровне нет должного понимания, что для сохранения электронной информации необходимо ускоренно создавать новую инфраструктуру, скорее всего – на основе инфраструктуры сохранения информации на традиционных носителях, модернизируя и укрепляя ее.

Выступления участников в ходе конференции свидетельствовали о необходимости дальнейшего всестороннего изучения проблемы сохранения электронной информации, результаты которого должны быть представлены на национальном и международном уровнях, в политическом и профессиональном контекстах.

Конференция выявила не только проблемы, но и новые профессиональные подходы к сохранению стремительно растущих объемов электронной информации, когда увеличивается многообразие форматов представления информации, многообразие типов и видов цифровых объектов, а срок их жизни из-за постоянной смены информационных технологий уменьшается.

Конференция приняла итоговый документ «Сохранение электронной информации в России: призыв к действию», в котором изложены видение обсуждаемой проблемы, предложения и обращения к основным заинтересованным сторонам.

Структурировать проблемы и выявить различные аспекты темы долгосрочного сохранения электронной информации, понять существующее в России положение дел в данной сфере и взглянуть на проблемы через призму ситуации в других странах позволило проведение в нашей стране Международной конференции «Сохранение электронной информации в информационном обществе: проблемы и перспективы», которая была организована в Москве в рамках председательства России в Межправительственной программе ЮНЕСКО «Информация для всех».

Итоговый документ этой конференции – Московская декларация о сохранении электронной информации¹ – на данный момент является одним из основополагающих международных документов в этой сфере. В нём впервые на международном уровне была ёмко структурирована проблемная ситуация и сформулированы приоритетные направления действий, которые необходимо предпринять в таких сферах, как политика, образование и просвещение, научные исследования, экономика, ИКТ-индустрия, международное/межведомственное сотрудничество и координация.

Международная конференция показала, что в большинстве стран мира функции по сохранению электронной информации еще не распределены и не закреплены за соответствующими институтами и структурами. Нет выстроенной политики в данной сфере, и, соответственно, нет необходимой нормативной основы. В то же время и в развитых странах, и в развивающихся идет поиск ответов на вопросы о том, что именно и в каких целях необходимо сохранять, кто, где, когда, на какие средства и в какой мере обязан это делать.

Зарубежный опыт представлен в сборнике материалов международной конференции, получившем название «Сохранение электронной информации в информационном обществе» и изданном на русском и английском языках². Знания, которые мы получили после проведения этой конференции, позволили нам понять ситуацию в мире: полноту, структуру, содержание проблем, основные болевые точки, направления деятельности, по которым есть прорывы и достижения. Конкретизировались и прояснились роль и функции библиотек, архивов и музеев в этой сфере, а также позиция и направления деятельности международных организаций – ЮНЕСКО, Совета Европы, ИФЛА.

Всё это стало основой для того, чтобы в ходе Всероссийской конференции определить полезные для России наработки и опыт специалистов и институтов из других стран, а также выявить специфические российские проблемы в сфере долгосрочного сохранения электронной информации и сконцентрировать на них внимание.

Надеюсь, что представленные в данном сборнике идеи, позиции, суждения и опыт будут способствовать привлечению более пристального внимания органов власти к вопросам сохранения электронной информации, развитию понимания значимости и содержания и более глубокому изучению данной проблемы со стороны институтов памяти, университетов, НИИ и других заинтересованных сторон.

Председатель Межправительственного совета и Российского комитета Программы ЮНЕСКО «Информация для всех», президент Межрегионального центра библиотечного сотрудничества

Е. И. Кузьмин

1 См. <http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/ifap/Moscow%20Declaration.pdf> (на англ. яз.) и http://www.ifapcom.ru/files/Documents/Moscow_Declaration_on_Digital_Information_Preservation_rus.pdf (на рус. яз.)

2 Электронная версия сборника опубликована на сайте Российского комитета Программы ЮНЕСКО «Информация для всех» по адресу: http://www.ifapcom.ru/files/News/Images/2012/dig_pres_is_rus.pdf



США: ОПУБЛИКОВАН ОТЧЕТ О

ПРЕИМУЩЕСТВАХ И РИСКАХ ФОРМАТА PDF/A-3

Автор: Батч Лазорчак (Butch Lazorchak) электронный архивист Библиотеки Конгресса США

Нашему сообществу специалистов, занимающихся курированием электронных материалов, повезло в том, что стоящие перед нами проблемы, как правило, не угрожают чьей-либо жизни. Тем не менее, когда мы начинаем о чём-то спорить, эмоциональная дискуссия и страстная защита своих позиций представителями всех сторон гарантированы.

Так случилось и с публикацией в октябре 2012 года спецификаций файлового формата PDF/A-3. Об этом мы сразу сообщили на нашем блоге, и тема тут же стала «горячей».

Семейство международных стандартов PDF/A специфицирует основанные на формате Portable Document Format (PDF) файловые форматы, включающие такой механизм представления электронных документов, который обеспечивает сохранение во времени их статического визуального представления независимо от инструментов и систем, используемых для создания, хранения и отображения файлов. «Статическое визуальное представление» в конечном счёте означает, что файлы в формате, соответствующем спецификациям PDF/A, являются самодостаточными и не используют внешних ссылок или данные в ином, отличающемся от PDF формате.

Первая редакция спецификаций PDF/A (PDF/A-1) была опубликована в сентябре 2005 года и с тех пор регулярно обновлялась. Версия спецификаций PDF/A-3 была воспринята сообществом с определённым беспокойством, поскольку она добавляет к своим предшественникам всего одну, но очень существенную функциональную возможность. Спецификация PDF/A-2 допускает вложение других файлов, при условии, что вложенные файлы соответствуют спецификациям PDF/A. Редакция PDF/A-3 разрешила вложение файлов любого формата.

Первичный документ, содержащийся в файле формата PDF/A-3, по-прежнему будет устойчив к рискам, связанным с сохранением в течение очень длительного времени. В то же время спецификации PDF/A-3 не требуют рассматривать вложенные файлы как архивный контент, создавая тем самым ряд возможных технических и политических проблем для учреждений, обеспечивающих длительную сохранность электронных материалов.

Рабочая группа по стандартам и практике (Standards and Practices Working Group) «Национального альянса попечения об электронных материалах» (National Digital Stewardship Alliance, NDSA), безусловно, видит эти проблемы, и она сочла, что сообществу будет полезен документ, в

котором рассмотрены особенности формата PDF/A-3 и проанализировано его возможное влияние на практику работы архивных учреждений.

В результате 20 февраля 2014 года «Национальный альянс попечения об электронных материалах» опубликовал отчет объемом 25 страниц «Преимущества и риски использования файлового формата PDF/A-3 архивными учреждениями» (The Benefits and Risks of the PDF/A-3 File Format for Archival Institutions, http://www.digitalpreservation.gov/ndsaworking_groups/documents/NDSA_PDF_A3_report_final022014.pdf).

В отчете дана взвешенная оценка затрат и преимуществ в случае широкомасштабного использования формата PDF/A-3, особенно если в архивные учреждения начнет поступать контент в этом формате. В документе содержится базовая информация о процессе разработки спецификаций, выделены конкретные ситуации, в которых возможно использование данного формата, и предложен ряд положений, которые архивные учреждения могли бы включить в свои политики.

Например, авторы отчета полагают, что допустимость наличия вложенных файлов в документах формата PDF/A для учреждений, занимающихся сохранением культурно-исторического наследия, будет опираться на очень детальные правила взаимодействия фондообразователей и архивных учреждений, устанавливающие допустимые форматы вложенных объектов, а также порядок действий, обеспечивающий полное понимание архивным учреждением взаимосвязи основного PDF-документа и вложенных файлов.

Кроме того, в отчете отмечается, что сложность формата PDF и существенные различия между приложениями, используемыми для создания и отображения PDF-файлов, предполагают, что формат PDF/A-3 может быть применим в рамках контролируемых потоков рабочих процессов (workflow), однако может оказаться неподходящим в качестве универсального контейнерного формата.

Безусловно, введение столь спорной новой функциональной возможности в последнюю редакцию спецификаций семейства PDF/A должно подтолкнуть сообщество сохраняющих культурно-историческое наследие учреждений к тому, чтобы играть более стратегическую, более активную и заметную роль в процессе разработки стандартов, и, в частности, спецификаций PDF/A.

Данный отчет является самым новым в серии публикаций и мероприятий NDSA, способствующих пониманию ряда проблем, связанных с обеспечением сохранности электронных материалов и их курирования. Мы будем рады Вашим замечаниям по отчету о формате PDF/A-3, равно как и предложениям относительно вопросов, по которым NDSA может помочь сообществу специалистов, занимающихся курированием электронных материалов.



ШВЕЦИЯ: НАЦИОНАЛЬНЫЕ АРХИВЫ УЧАТ СПЕЦИАЛИСТОВ ИСПОЛЬЗОВАТЬ ФОРМАТ PDF/A

Источник: сайт Национальных Архивов Швеции
<http://riksarkivet.se/sok-arrangemang?item=106469>

Автор: Наташа Храмцовская

В России в последнее время стало модно, не вникая глубоко в суть дела, вписывать в разного рода регламенты и рекомендации по межведомственному электронному взаимодействию и архивному хранению положения об использовании стандарта PDF/A. Наши шведские коллеги подходят к этому вопросу серьезнее, что иллюстрирует приведенное ниже объявление, появившееся на сайте Национальных Архивов страны 20 февраля 2014 года. В нём говорится следующее:

«Национальные Архивы (Riksarkivet) организуют курс обучения по формату PDF/A продолжительностью 3,5 часа для тех, кто работает в области управления информацией и архивного дела. Этот курс посвящен проблемам и возможностям данного формата с точки зрения обеспечения долговременной сохранности электронных документов.

PDF/A является одним из форматов, рекомендуемых для целей долговременного хранения разработанными Национальными Архивами правилами PA-FS 2009:2. Данный формат стандартизирован Международной организацией по стандартизации (ИСО) и рекомендуется в качестве формата для долговременного хранения даже физическим лицам. Хотя PDF/A имеет долгую историю, есть и немало неопределенностей, связанных с самим форматом и с его применением. Поскольку PDF/A получает все большее распространение на национальном и на международном уровне, от специалистов требуется большее его понимание и компетентность при его использовании.

Целью данного курса – дать слушателям представление о PDF/A как об одном из инструментов сохранения электронных объектов; объяснить, что он из себя представляет и как используется на практике; разницу между PDF/A-1, PDF/A-2 и PDF/A-3, что такое уровни соответствия " a ", " b " и " b ", и когда, как и зачем использовать эти версии формата. Курс также познакомит слушателей с проблемами, которые существуют у формата PDF/A.

Целевая аудитория: Этот курс адресован как тем, кто на практике работает с электронными информационными ресурсами, так и архивистам, руководителям, менеджерам и ИТ-разработчикам, а также тем, кто принимает решения, по вопросам сохранения информации, управления архивами и ИТ-системами (например, о закупке и разработке).»

Такого рода курс был бы полезен и многим нашим специалистам. В него также стоило бы, с моей точки зрения, включить рассказ о программных средствах, умеющих создавать файлы в формате PDF/A и/или контролировать корректность формата.



ШТАТ ВИКТОРИЯ, АВСТРАЛИЯ: ПЕРЕСМОТР СТАНДАРТА VERS

Источник: Сайт Управления государственных документов штата Виктория (Австралия) <http://prov.vic.gov.au/government/recordkeeping-standards-project/vers-standard-renewal>

Автор: Наташа Храмцовская

11 февраля 2014 года Управление государственных документов австралийского штата Виктория (Public Record Office Victoria, PROV) сообщило на своём сайте о том, что запланировано проведение пересмотра фундаментального для «Стратегии штата Виктории по управлению электронными документами» (Victorian Electronic Records Strategy, VERS, <http://prov.vic.gov.au/government/vers>) стандарта «Управление электронными документами» (PROS 99/007 Management of Electronic Records, <http://prov.vic.gov.au/government/vers/implementing-vers/standard-2>), известного как «стандарт VERS».

Для справки: VERS - это инициатива Управления государственных документов австралийского штата Виктория (Public Record Office of Victoria). Стандарт VERS состоит из 5 различных спецификаций. В основе его лежит стандарт формата, определяющий технические компоненты, необходимые для создания документов в виде VEO (VERS encapsulated objects - инкапсулированные VERS-объекты) средствами текущей программной системы для управления документами. Документы преобразуются в формат, который можно читать, экспортировать и импортировать, и который устойчив по отношению к изменениям в программных системах.

VERS первоначально был опубликован в 1999 году (таким образом, это были ранние, новаторские спецификации). Стандарт был пересмотрен в 2003 году, и была выпущена 2-я версия, которая остается действующей.

Стандарт VERS имеет программу соответствия, и возможна сертификация по любой из 5 спецификаций. Соответствующие VERS программные продукты перечислены на сайте VERS. (из статьи Барбары Рид, см. <http://rusrim.blogspot.ru/2010/08/3.html>).

На сайте опубликовано письмо старшего менеджера по вопросам стандартов и политики Управления Андрю Во (Andrew Waugh), в котором том пишет, обращаясь ко всем заинтересованным лицам:

Хочу пригласить Вас принять участие в проекте обновления стандарта VERS.

В течение 2013-2014 годов Управление государственных документов штата (PROV) осуществляет проект по пересмотру стандарта VERS (PROS 99/007 «Управление электронными документами»). Задача проекта - сохраняя достигнутый уровень обеспечения сохранности электронных материалов, улучшить следующие характеристики стандарта:

- Упростить процесс создания VERS-объектов (и, соответственно, сделать его дешевле).

Мой комментарий: речь идёт о преобразовании электронных документов в контейнерный формат для архивного хранения. Коллеги из Минкомсвязи и ВНИИДАД, это Вам на заметку!

- Сделать стандарт намного короче и проще для понимания;
- Сделать стандарт более подходящим для долговременного хранения своей информации органами исполнительной власти штата.

Для нас очень важны Ваши идеи относительно того, как можно улучшить стандарт. Мы были бы признательны, если бы Вы посмотрели прилагаемый документ («Предложения по обновлению стандарта VERS для публичного обсуждения», <http://prov.vic.gov.au/wp-content/uploads/2014/02/VERStdRevisionStrawmanPublicComments.doc>) и прислали свои замечания и предложения.

К содержащимся в этом документе предложениям следует относиться как к «стартовой точке» обсуждения, а не как к готовым рекомендациям. Хотя документ содержит ряд конкретных вопросов, касающихся этих предложений, мы приветствуем любые Ваши замечания и предложения. Вы, в частности, можете:

- Предложить другие способы, с помощью которых мы могли бы достичь наших целей;
- Высказать своё мнение о том, будут ли сформулированные в документе предложения способствовать достижению наших целей, особенно если вы считаете, что их будет проще реализовать на практике, чем ныне действующий стандарт;
- Предложить свои улучшения этих предложений.

Имейте, пожалуйста, в виду, что реальные изменения, которые появятся в новой редакции стандарта, будут опираться на полученные отзывы и дальнейшее изучение вопросов. Они, скорее всего, будут значительно отличаться от тех, что приведены в предложениях.

Ваше мнение чрезвычайно ценно для нас. Пожалуйста, пришлите Ваш отзыв Эндрю Во, старшему менеджеру по стандартам и политике, руководителю проекта обновления стандарта VERS.

Многие органы исполнительной власти и поставщики сделали значительные инвестиции в системы, соответствующие требованиям действующего стандарта. Чтобы эти инвестиции не обесценивались, Управление будет продолжать принимать и сохранять объекты, соответствующие текущей редакции стандарта независимо от итогов процесса его обновления.

С уважением,

Эндрю Во (Andrew Waugh) старший менеджер по стандартам и политике.

Мой комментарий: На данном примере я хотела бы обратить внимание руководства Росархива на то, что попытки «сбавать» за несколько месяцев силами узкой группки специалистов готовые документы, которые должны

кардинальным образом изменить практику управления документами и архивного дела – бесперспективная затея, даже если к этой работе будут привлечены лучшие силы отрасли.

Австралийцы работают над своим стандартом более 15 лет, они уже обкатали на практике две его редакции – и, тем не менее, считают необходимым консультироваться со всеми заинтересованными сторонами и думать об экономических последствиях изменений.

Коллеги из Росархива, учитесь – и сразу закладываете в свои конкурсы проведение публичных обсуждений, пилотных проектов, а также доработку проектов нормативных документов и рекомендаций по их итогам. Так будет дороже, - но не дороже, чем выбрасывать деньги на ветер безо всякой пользы.



ДАТСКАЯ СТРАТЕГИЯ ОБЕСПЕЧЕНИЯ ДОЛГОВРЕМЕННОЙ СОХРАННОСТИ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ

Источник: сайт Национальных Архивов Дании http://www.sa.dk/media%284826,1033%29/Strategy_for_archiving_digital_records.pdf

Автор: Наташа Храмцовская

Сейчас настал в нашей стране, как мне кажется, тот момент, когда уже недостаточно ругать неудачные документы, разрабатываемые уважаемыми, но растерявшими за последние годы свой потенциал учреждениями – теперь нужны конструктивные идеи и предложения. С моей точки зрения, нет смысла продолжать попытки создания детальных инструкций и регламентов; начинать нужно с определения стратегии и высокоуровневых требований, которые во многом можно списать с хороших зарубежных документов.

По этой причине я хотела бы привлечь внимание к опубликованной в январе прошлого года «Стратегии Датских Национальных Архивов в области архивации электронных документов» (Strategy for archiving digital records at the Danish National Archives, http://www.sa.dk/media%284826,1033%29/Strategy_for_archiving_digital_records.pdf, на английском языке). Это маленький 6-страничный документ, в котором архивная служба страны фиксирует принципиальные решения, определяющие дальнейшее развитие программ архивного хранения электронных документов.

В первую очередь формулируется главная задача: обеспечить долговременную сохранность электронных документов таким образом, чтобы сохранялась их аутентичность и чтобы их можно было находить и использовать.

Первым принципиальным выбором является выбор стратегии обеспечения долговременной сохранности – миграции либо эмуляции. Национальные Архивы Дании выбрали миграцию. Следствием такого выбора является обязанность создателей документов провести их миграцию в небольшое число четко специфицированных форматов для архивного хранения, выбранных Национальными Архивами. При этом Национальные Архивы понимают связанные с миграцией риски, и считают, что миграцию в архивные форматы желательно проводить как можно раньше после создания документов.

Далее, Национальные Архивы Дании идут по пути обеспечения сохранности независимо от информационных систем. При таком подходе возможность использования и интерпретации данных не должна зависеть от непрерывной доступности той системы, в которой данные были первоначально созданы – должна быть возможность всё это делать и в других системах. При этом датские архивисты понимают, что абсолютной независимости от систем и технологий быть не может и что рано или поздно электронные документы определенного периода становятся столь зависимыми от технологий своего времени, что становится необходимой их миграция.

В отличие от традиционного архивного дела, объектом сохранения являются не первоначальные артефакты (оригиналы), а содержащаяся в них информация – однако при этом должна обеспечиваться её аутентичность.

Конкретная реализация стратегии опирается на:

- Раннее выявление и отбор систем, данные из которых будут передаваться на архивное хранение. Органы власти обязаны информировать Национальные Архивы о новых ИТ-системах. Архивисты проводя экспертизу ценности данных в этих системах, и, если данные признаются заслуживающими длительного хранения, то система должна получить одобрение Национальных Архивов. Ключевым условием получения такого одобрения является возможность экспорта данных в формате, принимаемом Национальными Архивами.

В Дании отсутствуют жёсткие нормативные требования о передаче определенных типов данных в архивы, поэтому архивисты могут проводить экспертизу гибко, принимая решения на основе реальной ценности этих данных.

Кроме того, Национальные Архивы регулярно контролируют использование государственными органами их ИТ-систем.

- Частую передачу материалов в независимом от конкретной информационной системы формате. Для Дании типична передача материалов на архивное хранение по истечении 5 лет. В отличие от бумажных документов, электронные документы передаются на архивное хранение вне зависимости от того, продолжается ли их активное использование ведомством или нет. Датские архивисты подчеркивают необходимость вовремя захватывать соответствующую

техническую документацию – пока эти знания все ещё сохраняются в государственном органе.

Таким образом, Национальные Архивы осознанно идут на то, что копии одних и тех же данных хранятся как у них (возможно, в нескольких экземплярах), так и в государственном органе.

Документы принимаются в виде структур файловой системы, а данные – в виде реляционных баз данных.

Центральные органы государственной власти обязаны сдавать свои документы в Национальные Архивы. У региональных и местных властей есть альтернатива – они могут создать собственный архив, но в любом случае при сдаче материалов должны соблюдаться установленные Национальными Архивами требования.

Физическая сохранность электронных документов обеспечивается путем сохранения в территориально-удаленных друг от друга хранилищах экземпляров документов, записанных на ряде различных носителей информации.

- Постоянное планирование действий, обеспечивающих долговременную сохранность, и периодическая миграция в новые форматы, подходящие для долговременного хранения.

В случае выбора Национальными Архивами нового формата для длительного хранения, в течение нескольких ближайших лет должна быть проведена миграция имеющихся материалов в этот формат.



ПРОЕКТ E-ARK ПОЛУЧИЛ 6 МИЛЛИОНОВ ФУНТОВ НА СПАСЕНИЕ ЭЛЕКТРОННЫХ ДАННЫХ

Источник: сайт ITPro.co.uk <http://www.itpro.co.uk/storage/21528/e-ark-gets-6-million-to-save-digital-data>

Проект E-Ark (от European Archival Records and Knowledge Preservation – «Обеспечение долговременной сохранности архивных документов и знаний Европы») по созданию электронного «ковчега» для изначально-электронных и бумажных архивных документов, получил денежную инъекцию от Евросоюза. Европейская комиссия выделила на него 6 миллионов фунтов стерлингов.

В рамках этого проекта, которым руководит группа, базирующаяся в университете английского города Портсмута, изучаются способы архивации электронных данных из различных видов систем, встречающихся в Европе. Участники проекта надеются, что их подход станет «золотым стандартом» архивации данных в рамках Европы. Команда из Портсмута заявила, что их

система обеспечит жизнеспособность во времени существующих электронных архивов, в том числе и «больших данных».

Проект объединит усилия специалистов по электронной сохранности как самого Портсмутского университета, так и дюжины крупных партнеров, в число которых входит ряд основных национальных архивов Евросоюза.

В ходе проекта E-Ark будет проанализирован имеющийся передовой опыт с целью создания пилотного сервиса архивации, обеспечивающего сохранение аутентичности документов и их пригодности к использованию. Проект охватит три основных направления деятельности архива – приём на хранение, обеспечение сохранности и создание условий для повторного использования информации.

В течение трех лет в общеевропейском масштабе будет отрабатываться стандартный процесс архивации, поддерживаемый руководствами и сводами рекомендуемой практики для данных различных видов, поступающих из различных источников, в том числе из систем управления документами и баз данных.

Д-р Джанет Делв (Dr Janet Delve) и профессор Дэвид Андерсон (David Anderson) из Портсмутского университета отметили колоссальность стоящей перед проектом задачи, которая с каждым днем становится всё сложнее.

«Масштабы данной проблемы огромны. Мы имеем дело с накопленными за многие годы почти 30 странами данными, которые хранились с использованием множества различных методов и в различных системах», - сказала д-р Делв. «С появлением электронного правительства и инициатив в области «открытых данных» архивам теперь нужно справляться с хранением огромных объемов электронных материалов. Масштабы проблемы растут из-за колоссального количества электронных данных, ежедневно производимых самыми разнообразными организациями, такими как банки, учреждения здравоохранения и национальные архивы».

Д-р Делв отметила, что цель проекта заключается в снижении риска утраты информации вследствие плохих методов ведения и архивации документов за счет разработки единого общего подхода. По её словам, «Данный подход должен быть воспроизводимым и масштабируемым, с тем, чтобы удовлетворять потребности многих видов организаций, государственных и частных, больших и малых; и он должен быть способен поддерживать сложные типы данных, такие, как веб-страницы и большие данные».

Профессор Андерсон сказал, что ключевой проблемой, которую предстоит решить, являются различия в правовых системах и в национальных традициях управления документами. По его словам, задача создания и развития инфраструктуры, которой могли бы пользоваться организации различных типов из всех стран Европы, представляет собой огромную головоломку из сотен кусочков, которые следует изучить и проанализировать. «Мы будем брать лучшие элементы от различных рассмотренных нами систем, и нашей целью является создание того, к чему,

как нам известно, вызывают крупные организации и архивисты», подчеркнул Андерсон.

Проект официально открыт 4 февраля 2014 года на церемонии, прошедшей в Высшем техническом институте (Instituto Superior Técnico) в Лиссабоне, Португалия. Помимо пяти национальных архивов, в числе организаций-участников есть четыре ведущих научно-исследовательских учреждения, три поставщика программных решений и услуг в области архивации, два государственных органа и два международных ассоциации, представляющие сообщества, которые могут получить отдачу от проекта, такие, как владельцы и поставщиков данных.

Рене Милман (Rene Millman)



ОТЗЫВ ОБ ОТЧЕТЕ РГГУ «СРАВНИТЕЛЬНЫЙ АНАЛИЗ ФОРМАТОВ ФАЙЛОВ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ ПОСТОЯННОГО (ДОЛГОВРЕМЕННОГО) ХРАНЕНИЯ»

Автор: Наташа Храмцовская

Отчет по теме: «Сравнительный анализ форматов файлов электронных документов постоянного (долговременного) хранения» в рамках НИР «Исследование современных процессов документационного обеспечения управления и развития электронного документооборота» был подготовлен Российским государственным гуманитарным университетом (РГГУ) по заказу Росархива в 2013 году.

Документ объёмом 62 странице выложен на сайте Росархива по адресу <http://archives.ru/sites/default/files/rekomendation-rggu-format-2013.pdf>.

Первое, что режет глаз в отчете, это наличие морально устаревшей терминологии бумажной эпохи, к которой авторы не рискнули подойти критически. В отчете приводятся следующие определения:

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.

Реальность такова, что электронные документы далеко не все «бумагоподобны». Скажем, базу данных трудно «сформировать в дело» (да и зачем?). Более того, для электронного документооборота характерен (и уже реально применяется) учет на уровне документа, а не на уровне дела. Что

касается объёма электронного документа в байтах, то это одна из наиболее бессодержательных его характеристик – тем более, что этот объём во времени будет изменяться как вследствие миграции электронных документов в новые форматы, так и вследствие пополнения метаданных документа.

Что касается единиц хранения, то авторы забыли про обычную логику. Это должны быть именно те объекты, которые, как говорится, можно взять в руки (носители информации), а не некие виртуальные сущности!

Приведено, на мой взгляд, неудачное, устаревшее определение термина «файл»:

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).

Для сравнения процитирую определение из терминологической базы Международного совета архивов

(см. <http://www.ciscra.org/mat/termdb/term/2548>).

Компьютерный файл – набор данных, обрабатываемый компьютером как единый объект.

Авторы удивили тем, что ввели отдельные определения понятий «цифровой документ» и «электронный документ», которые в современной отраслевой литературе означают одно и то же, при этом в законодательстве Российской Федерации предпочтительно используется термин «электронный документ». С моей точки зрения, это способно лишь запутать читателя.

Постановка задачи во Введении дана достаточно хорошо, но всё же от РГГУ я ожидала большей гибкости. Например, не могу согласиться со следующим утверждением на стр.12:

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

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 файлов, форматы, образуемые дорогостоящим, малораспространенным или неподдерживаемым программным обеспечением.

Коллегам стоило бы знать, что дешёвое и распространённое программное обеспечение существует только для офисных форматов. Скажем, для таких распространённых объектов, как базы данных (ряду которых по закону уже установлен постоянный или длительный срок хранения) стандартных форматов и дешёвого массового ПО нет. Его по сути дела нет для аудио и видеофайлов, отличающихся огромным разнообразием структуры, не говоря уже о САПР-системах, геоинформационных системах и

т.п. Вместо непрактичных заявлений о недопустимости нестандартных форматов не лучше ли было сказать о предпочтительности использования определенных форматов, хорошо поддерживаемых существующим матобеспечением?

Ссылка при этом на оторванное от жизни положение 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, согласно которому «Электронные документы принимаются на хранение в сопровождении программных средств, позволяющих их воспроизвести, и с необходимым комплектом сопроводительной документации», говорит лишь о том, что кабинетным учёным иногда полезно выходить в реальный мир. В реальном мире просто так передать архиву лицензионное ПО нельзя, не нарушив при этом права правообладателя – не говоря уже о том, что архивы явно не готовы стать музеями разнокалиберного программного обеспечения и необходимой для его использования программно-аппаратной инфраструктуры. В реальном мире федеральное или региональное правительство может просто приказать архивам принимать всё, что им дают, и в таком виде, в каком дают – примеры тому уже есть.

Следующий пункт 4 откровенно удивляет. По мнению авторов. «СЭД ФОИВ должна обеспечивать отображение следующих форматов файлов: pdf, rtf, doc, tiff». На практике практически все СЭД позволяют хранить объекты любых форматов. Часть из них может просматриваться с использованием встроенной программы просмотра, а остальные – с использованием иного программного обеспечения, установленного на компьютере пользователя. Возникает вопрос о том, почему не упомянуты такие современные форматы, как ODF и OOXML (в.ч. .docx), как можно в современных условиях считать необязательной поддержку форматов электронных таблиц и презентаций?

В документе приведены многочисленные цитаты из различных стандартов и нормативных документов. В целом такой подход, наверное, оправдан, поскольку стандарты для многих практиков недоступны. В то же время авторам следовало отнестись к этим документам более критически – положение дел в области форматов меняется быстро, и большинство цитируемых ими документов либо были созданы в бумажную эпоху несведущими в электронных технологиях людьми, либо вышли в свет более 5 лет тому назад. Соответственно, они как минимум частично морально устарели.

Говоря об общих требованиях к файловым форматам (стр.15), я бы все-таки акцентировала внимание не на «устойчивости к многократным миграциям» - это дело непонятное, поскольку результат миграции зависит не только от формата, но и от используемого программного обеспечения, от квалификации проводящего миграцию специалиста, от наличия и надёжности системы контроля качества миграции. Главным является то, что

использование подходящих форматов позволяет резко сократить частоту миграций и, соответственно, снизить риски и расходы.

Предложенная в документе методика сравнения форматов вполне адекватная и учитывает практические все основные факторы (стр.17-18). При этом материал о проведенных экспериментальных исследованиях все-таки лучше было вынести в приложения. Для той аудитории, которой адресован документ, приведенные формулы не нужны, а специалисты по математической статистике их, скорее всего, раскритикуют. К тому же на данный момент в отношении большинства форматов накоплен настолько огромный международный опыт, что очень скромные по масштабу собственные эксперименты мало что к нему добавляют.

Что в сухом остатке? Здесь уже не столь хорошо. В сухом остатке получаем (см. фото ниже) рекомендации, позаимствованные – местами без перевода на русский язык! - из рекомендаций по форматам для долговременной сохранности, подготовленных университетом штата Флорида (Recommended Data Formats for Preservation Purposes in the Florida Digital Archive, University of Florida, <http://ufdc.ufl.edu/AA00017119/00011>):

В некоторых условиях, в крайнем случае, могут быть предложены в категории текстовых форматов файлов:

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

Сравните:

and submitting files to the Florida Digital Archive. It is not intended to suggest that these file formats are allowable formats for ETDs or any other official publication of any Florida university. (See the notes following this table.) Source: FCLA, Mar. 2012.

Media	High Confidence Level	Medium Confidence Level	Low Confidence Level
Text	<ul style="list-style-type: none"> - Plain text (encoding: US-ASCII, UTF-8, UTF-16 with BOM) - XML (includes XSD/XSL/XHTML, etc.; with included or accessible schema and character encoding explicitly specified) - PDF/A-1 (ISO 19005-1) (*.pdf) - PDF/A-2 (ISO 19005-2:2011) (*.pdf) 	<ul style="list-style-type: none"> - Cascading Style Sheets (*.css) - DTD (*.dtd) - Plain text (ISO 8859-1 encoding) - 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) 	<ul style="list-style-type: none"> - PDF (*.pdf) (encrypted) - Microsoft Word (*.doc) - WordPerfect (*.wpd) - DVI (*.dvi) - All other text formats not listed here
Raster	<ul style="list-style-type: none"> - TIFF (uncompressed) 	<ul style="list-style-type: none"> - BMP (*.bmp) 	<ul style="list-style-type: none"> - MrSID (*.sid)

Эх, ну кто же у нас не знает очень широко распространенного в России формата SGML? Кто из государственных органов не использует в широких масштабах форматы EPUB и JPEG2000 (но не JPEG)? Список-то у американцев хороший и обширный (хотя и не бесспорный), но он отражает их, а не нашу реальность...

Список литературы показывает, что авторы из русскоязычных публикаций знают только стандарты, нормативные документы и публикации ВНИИДАД (в которых, как они сами признали во введении, в систематическом виде ничего по данному вопросу сказано до сих пор не было). Их знакомство с зарубежными публикациями также поверхностное – названа одна единственная свежая публикация (которая, собственно, и легла в основу результатов отчета) да ещё две старенькие, к тому же не самые авторитетные работы. Не упомянуты, скажем, ни переведенные на русский язык спецификации MoReq и MoReq2, ни документы Национальных Архивов США и стран Евросоюза...

В целом, учитывая голод на методическую литературу и актуальность вопроса, я бы оценила подготовленный РГГУ документ на «четверочку». В данном случае, на мой взгляд, важнее его очевидная потенциальная полезность для разработки государственных органами своей нормативно-методической базы, чем не менее очевидные недостатки.



НАЦИОНАЛЬНЫЕ АРХИВЫ США ОПУБЛИКОВАЛИ ПЕРЕСМОТРЕННОЕ РУКОВОДСТВО ПО АРХИВНЫМ ФОРМАТАМ

Источник: блог Национальных Архивов <http://blogs.archives.gov/records-express/2014/02/05/revised-format-guidance-issued/>

Мы рады сообщить о публикации нового руководства для федеральных органов исполнительной власти о передаче на архивное хранение электронных документов постоянного срока хранения.

Бюллетень Национальных Архивов 2014-04 «Пересмотренное руководство по форматам, приемлемым для передачи на архивное хранение электронных документов постоянного хранения (Revised Format Guidance for the Transfer of Permanent Electronic Records, <http://www.archives.gov/records-mgmt/bulletins/2014/2014-04.html>) заменяет все предыдущие руководства, устанавливающие файловые форматы, которые могут использоваться при передаче электронных документов на постоянное хранение в Национальные Архивы. Данный бюллетень был подготовлен во взаимодействии с различными заинтересованными сторонами, включая Административно-бюджетное управление администрации Президента (OMB), Федеральный

совет по документам (Federal Records Council), федеральные органы исполнительной власти и общественность.

Данный бюллетень содержит высокоуровневые требования к электронным документам для федеральных органов исполнительной власти и включает приложение (см. <http://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>), в котором перечислены файловые форматы, подходящие для различных категорий электронных документов. Оба документа выложены на веб-сайте Национальных Архивов в разделе рекомендаций по передаче документов на архивное хранение, см. <http://www.archives.gov/records-mgmt/policy/transfer-guidance.html>.

В ближайшие дни мы объявим о проведении онлайн-брифингов, посвященных новому руководству.



США: ПОДАНА ЗАЯВКА НА ПАТЕНТ НА МЕТОД АРХИВАЦИИ ЭЛЕКТРОННОГО КОНТЕНТА, ПОДПИСАННОГО ЭЛЕКТРОННЫМИ ЦИФРОВЫМИ ПОДПИСЯМИ

Автор: Наташа Храмцовская

Пока основная масса учёных и практиков бесечно ведёт свои споры по поводу методов обеспечения долговременной сохранности электронного контента на конференциях, в журналах и в социальных сетях, некоторые более «подкованные» по части бизнеса американцы без лишнего шума уже пытаются патентовать такого рода решения.

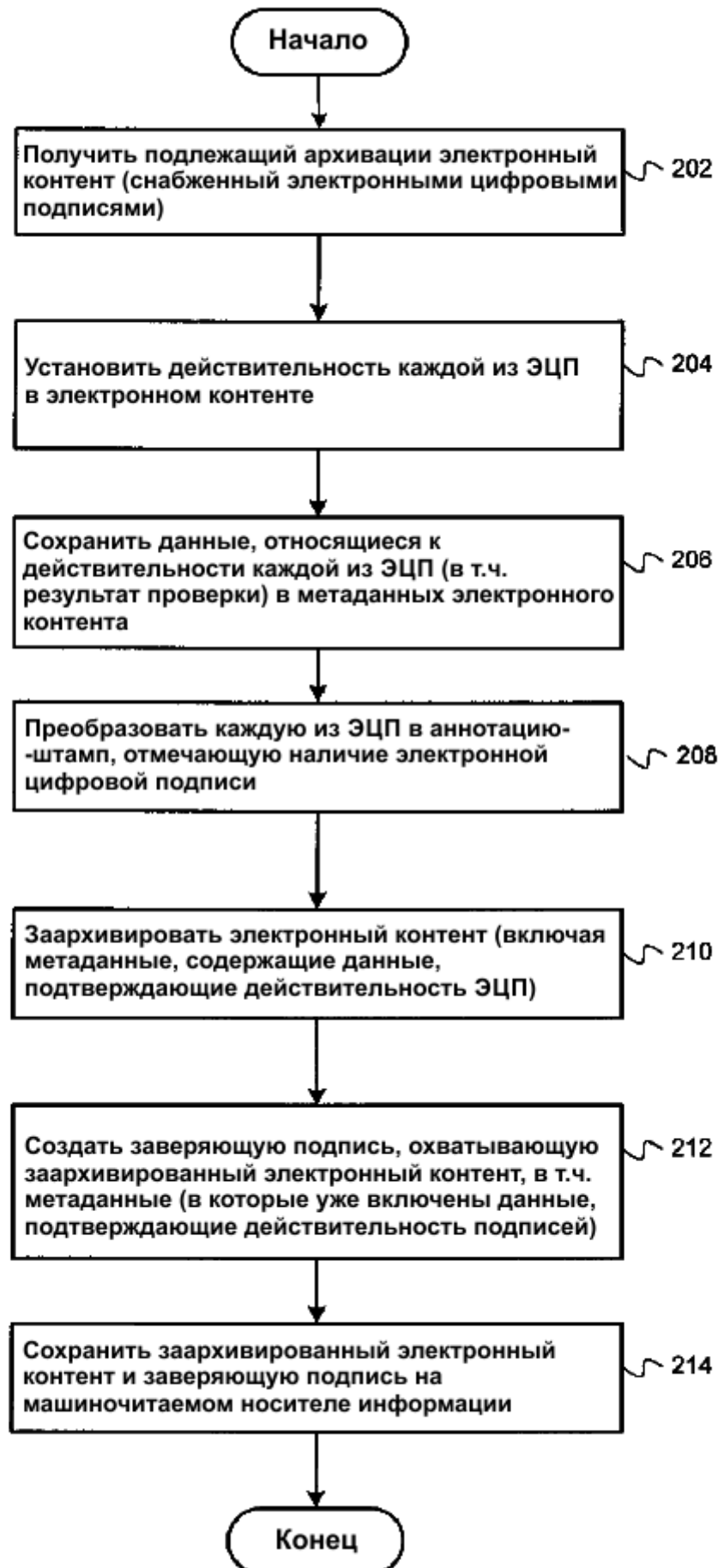
Так, 30 января 2014 года была опубликована поданная изобретателями из Калифорнии Суджатой Дасом (Sujata Das) осенью 2013 года заявка №20140032927 на получение американского патента на метод архивации электронного контента, подписанного электронными цифровыми подписями (Archiving electronic content having digital signatures, <http://www.freepatentsonline.com/20140032927.pdf>).

В приложенном описании можно найти целый ряд интересных идей по более детальной реализации предложенного алгоритма.

Конечно, неприятно, что кое-кто пытается «приватизировать» таким образом широко обсуждающиеся идеи. Наверное, и нашему уполномоченному органу государственной власти следует поактивнее поработать над вопросом длительного хранения документов, подписанных ЭЦП/УЭП, и, может быть, также запатентовать в России соответствующий алгоритм).

Идеологию предложенного метода отражает приведенная ниже блок-схема:

200





НА САЙТЕ БРИТАНСКОГО ИНСТИТУТА СТАНДАРТОВ НАЧАТО ПУБЛИЧНОЕ ОБСУЖДЕНИЕ ОСНОВНОГО ТЕРМИНОЛОГИЧЕСКОГО СТАНДАРТА ИСО В ОБЛАСТИ ИНФОРМАЦИИ И ДОКУМЕНТАЦИИ

Источники: сайт Британского института стандартов / сайт ИСО

<http://drafts.bsigroup.com/Home/Details/52442>

http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber

<https://www.iso.org/obp/ui/#iso:std:iso:5127:ed-1:v1:en>

19 февраля 2014 года на сайте Британского института стандартов был выложен для публичного обсуждения проект новой редакции стандарта ISO 5127 «Информация и документация – Словарь» (Information and documentation – Vocabulary), подготовленный рабочей группой WG4 «Терминология в области информации и документации» технического комитета ТС 46 Международной организации по стандартизации (ИСО).

Обсуждение продлится до 27 апреля 2014 года. Познакомиться с проектом можно по адресу <http://drafts.bsigroup.com/Home/Details/52442>.

Этот стандарт содержит базовые термины в области информации и документации, и их определения. Его назначение – способствовать обмену знаниями и информацией.

В настоящее время действует редакция стандарта 2001 года. Основное содержание этого документа доступно здесь:

<https://www.iso.org/obp/ui/#iso:std:iso:5127:ed-1:v1:en>



АНАЛІЗ СУЧАСНОГО СТАНУ СИСТЕМ ПЕРЕВІРКИ УКРАЇНОМОВНИХ ТЕКСТІВ

Автор: Мазничко А. Б., НДІ мікрографії

Для аналізу були обрані найбільш розповсюджені системи автоматизованої перевірки текстів українською мовою. Деякі з них функціонують у складі текстових редакторів. Для їх тестування використовувались два перевірені тести:

1) текст, що містить 34 помилки з орфографії та пунктуації;

2) текст, що містить 68 граматичних та лексичних помилок.

Результати аналізу наведені у таблиці.

Таблиця

Найменування, офіційна веб-сторінка	Ціна в грн. або різновид ліцензії ¹⁾	Кількість знайдених ²⁾ помилко	Стислий опис
1	2	3	4
Автономні програми			
LibreOffice http://ru.libreoffice.org/	LGPL, MPL	3	Редактор з можливістю перевірки орфографії.
AbiWord http://abisource.com/	GNU GPL	4	Редактор з можливістю перевірки орфографії; висока швидкість роботи.
OpenOffice.org Writer http://www.openoffice.org/ru/	LGPL	2	Редактор з можливістю перевірки орфографії.
Orfo Switcher http://warstar.ru/prog/2186-orfo-switcher-2.35-mlrus.html	Donateware	2	Редактор та окремий модуль для перевірки орфографії.
Багатофункціональна система перевірки правопису "Орфо" http://www.orfo.ru/	575 грн 200 грн – додатковий модуль для української мови. 282 грн – за 1 ліцензію	Н ³⁾	Порівняно з модулем перевірки орфографії в Microsoft Office 2010, ОРФО містить показ всіх форм будь-якого слова, доповнені та оновлені орфографічні словники, підтримка морфологічного словника користувача.
MS Office MS Word 2010 http://www7.buyoffice.microsoft.com/	1002 грн пакет "Персональний" для одного комп'ютера.	3	Перевірка української та російської орфографії та граматики, переноси. Пакет містить: Word, Excel, PowerPoint, OneNote
ПУТА 5.0 SP2 http://www.prolingoffice.com/product/9	480 грн (версія для одного користувача)	2	Перевірка орфографії в українських та російських текстах. Розставлення переносів в українських та російських словах. Словник синонімів (тезаурус) українських та російських слів. Граматичний контроль українських та російських текстів. Функціонує у складі Microsoft Office XP/2003/2007/2010
Babylon 9.0 http://russian.babylon.com/	875 грн	0	Переклад слів. Повнотекстовий переклад. Автоматичне визначення мови. Перевірка орфографії, правопису. Інтеграція в браузері та інші популярні програми. Додаткові словники (які можна скачати або створювати свої).
SPELLINK Multilingual Pro http://www.allvirtualware.com/	393 грн	Н ³⁾	UkrSpeller – модуль перевірки української орфографії. Документація англійською мовою.

Продовження таблиці			
1	2	3	4
Онлайн перевірка			
http://uareferats.com/index.php/speller	–	0	Перевірка орфографії
http://translate.meta.ua/ua/orthography/	–	4	Перевірка орфографії
http://slovari.yandex.ru/~spellchecker/	–	3	Перевірка орфографії
http://zoho.com/	–	0	Перевірка орфографії
http://www.languagetool.org/	–	0	Перевірка орфографії
<p>¹⁾ Розшифрування назв ліцензій: LGPL – безкоштовна ліцензія на програмне забезпечення, видана Фондом Вільних Програм. Вона проектувалася як компроміс між суворо копілефтною GNU GPL та дозволяючими ліцензіями. MPL – (Mozilla Public License) одна з ліцензій на вільне ПЗ. GNU GPL – (GNU General Public License, відкрита ліцензійна угода GNU) Мета GNU GPL — надання користувачеві прав копіювати, змінювати та розповсюджувати програми, і зобов'язань, згідно з якими користувачі всіх похідних програм теж отримають зазначені вище права. Donateware – модель ліцензування, при якій користувачу надається цілком працююче ПЗ з можливістю зробити добровільні внески розробнику.</p> <p>²⁾ У колонці «Кількість знайдених помилок» наведено сумарну кількість правильно знайдених помилок в обох тестах.</p> <p>³⁾ Виробник не надає безкоштовну версію для тестування.</p>			

На підставі аналізу результатів тестування можна дійти таких висновків:

1. Усі протестовані системи базуються на виявленні наявності слова у словнику (словник може коригувати користувач).
2. Протестовані системи не виявили жодної пунктуаційної або лексичної помилки.
3. Усім системам, що перевірялись, властивий порівняно однаково низький рівень якості – кількість правильно знайдених помилок не перевищувала 4. Системи також помилково позначали слова (від 5 і більше) як порушення норм правопису, хоча, насправді, слова не містять відхилень від правил.
4. Можна вважати перспективним використання систем “РУТА 5.0” та “Орфо”, як допоміжних для поверхневого пошуку орфографічних помилок, зважаючи на такі причини:
 - а) розробники мають найбільший досвід роботи з україномовними текстами;
 - б) наявність документації та технічної підтримки українською або російською мовами.
 Але і цим програмам властиві вищенаведені недоліки.

ЗМІСТ

Передмова.....	1
Экономика обеспечения долговременного доступа к электронной информации: введение в проблему.....	2
Сохранение электронной информации: зачем нам нужен зарубежный опыт.....	7
Анализ предпочтительности цветовой модели в информационно–измерительном комплексе.....	11
Сохранение электронного контента в России и за рубежом.....	18
США: опубликован отчет о преимуществах и рисках формата PDF/A-3	22
Швеция: Национальные Архивы учат специалистов использовать формат PDF/A.....	24
Штат Виктория, Австралия: Пересмотр стандарта VERS.....	25
Датская стратегия обеспечения долговременной сохранности электронных документов.....	27
Проект E-Ark получил 6 миллионов фунтов на спасение электронных данных.....	29
Отзыв об отчете РГГУ «Сравнительный анализ форматов файлов электронных документов постоянного (долговременного) хранения»..	31
Национальные Архивы США опубликовали пересмотренное руководство по архивным форматам.....	35
США: Подана заявка на патент на метод архивации электронного контента, подписанного электронными цифровыми подписями.....	36
На сайте Британского института стандартов начато публичное обсуждение основного терминологического стандарта ИСО в области информации и документации.....	38
Аналіз сучасного стану систем перевірки україномовних текстів.....	38