ПЕРЕДМОВА

Випуск дайджесту присвячено досвіду архівних установ світу щодо довгострокового зберігання електронних інформаційних ресурсів.

У публікації «Современные технологии создания страховых фондов документации» розглянуто перспективні технології створення страхових фондів документації в сучасних умовах. Визначено основні технології запису і відтворення бінарних даних з використанням мікрофільму та штрихового кодування. Наведено варіанти оцінки ефективності гібридних технологій мікрофільмування.

У публікації «Электронные архивы: возможные решения проблем долгосрочного хранения данных» систематизовані проблеми, що виникають при довгостроковому зберіганні електронних документів, запропоновані можливі варіанти їх рішення, випробувані авторами при створенні ряду систем електронних архівів.

У публікації «США: Где ученым следует держать свои данные?» розповідається, що федеральні органи починають випускати офіційні політики які включають рекомендації щодо зберігання, у разі невиконання вимог нових політик можна позбутися можливості отримання додаткових грошей за грантом.

У публікації «Особенности оцифровки документов в современных архивах» розповідається про те, що в даний час все більше уваги приділяється питанням збереження культурних цінностей. У цьому зв'язку затверджуються державні програми, в рамках яких, за задумом їх творців, застосування нових технологій, інноваційних підходів, а також світового досвіду дозволить забезпечити збереження культурної спадщини, історично значимих документів.

СОВРЕМЕННЫЕ ТЕХНОЛОГИИ СОЗДАНИЯ СТРАХОВЫХ ФОНДОВ ДОКУМЕНТАЦИИ

Источник: http://www.tusur.ru/filearchive/reports-magazine/2014-32-2/48.pdf
Авторы: Н. Е. Проскуряков, С. Ю. Борзенкова, Е. Е. Евсеев, О. В. Чечуга

Рассмотрены перспективные технологии создания страховых фондов документации в современных условиях. Определены основные технологии записи и воспроизведения бинарных данных с использованием микрофильма и штрихового кодирования. Приведены варианты оценки эффективности гибридных технологий микрофильмирования.

Ключевые слова: исходный электронный документ, микрофильм, методы записи и воспроизведения бинарных данных, сканирование, декодирование, штрих-код.

Задача создания страховых фондов документации.

В настоящее время в России стремительными темпами растут объемы сканирования и оцифровки бумажной документации предприятий и организаций, библиотечных и архивных фондов. Утверждены различные государственные документы, концепции и программы, нацеленные на увеличение электронного документооборота. Практически во всех федеральных органах исполнительной власти и органах исполнительной власти субъектов Российской Федерации завершается переход на использование в своей деятельности электронных документов.

Вместе с тем часть документов, относящихся к особо ценным и особо важным, требует обеспечения их длительного и надежного хранения. Эта задача в настоящее время решается системой Единого российского страхового фонда документации (далее – ЕРСФД) с применением традиционного для этих целей носителя информации – микроформы. Ежегодно увеличивающийся объем электронных документов уже сегодня ставит перед системой ЕРСФД решение вопросов по разработке принципов и методов долговременного сохранения электронных массивов информации.

Положение о ЕРСФД, утвержденное Постановлением Правительства Российской Федерации от 26.12.1995 г. №1253-68, допускает фиксацию массивов конструкторской, технологической, проектной, нормативной, научной, историко-культурной и другой документации, относящейся к ЕРСФД, не только на микроформах, но и на других компактных носителях информации.

Как показали информационные исследования, проводимые регулярно на протяжении последних лет ФГУП «НИИ Репрографии» (г. Тула), в настоящее время для долгосрочного сохранения различных видов информации в ведущих зарубежных странах применяется два основных подхода – микрофильмирование и оцифровка [1].

Между сторонниками и противниками этих направлений ведутся горячие научные споры. Особую актуальность приобретает вопрос долгосрочного сохранения электронной информации.

Информационное страхование бумажных документов с помощью классических технологий оптического микрофильмирования, несмотря на некоторый спад объемов, по прежнему продолжает осуществляться практически во всех странах. Но объективное возрастание в жизни общества роли электронного документооборота и стремительное нарастание объема документов, создаваемых, обрабатываемых и хранимых в электронной форме, диктуют необходимость развития новых подходов и технологических решений, таких как гибридные электронно-микрографические технологии.

Внедрение данных технологий в практику создания долговременно хранимых страховых информационных ресурсов происходит практически повсеместно. Преимущества электронного документооборота хорошо известны – это высокая оперативность поиска и доступа к документам, экономия времени и расходных материалов, возможность обмена документами по различным электронным каналам связи, снижение бюрократической волокиты и т.д.

Однако повсеместное внедрение электронного документооборота влечет за собой ряд серьезных проблем, важнейшей из которых является проблема долгосрочной сохранности электронных документов в целях их информационного страхования и архивирования. Без решения этого вопроса невозможно гарантировать сохранение и доступность для потомков цифрового интеллектуального, научного и культурного наследия цивилизации.

Возможности долгосрочного хранения электронных документов ограничены частой сменой поколений цифровых носителей и поддерживающих их аппаратно-программных платформ, которые склонны к быстрому устареванию и исчезновению. В поисках выхода из сложившейся ситуации мировым научным сообществом предлагаются различные варианты обеспечения длительности существования электронных документов в цифровой среде.

Самыми распространенными решениями являются миграция документов в новые программные среды и форматы, периодическая многократная перезапись на новые носители, а также эмуляция, т.е. имитация старой программной оболочки на новых операционных системах и оборудовании.

Однако оба данных подхода (миграция и эмуляция) принципиально не выходят за рамки цифровой среды, которая по самой своей природе достаточно динамична, изменчива и нестабильна. Для обеспечения постоянной миграции и эмуляции требуются большие финансовые, организационные и трудовые ресурсы. Кроме этого, проведенные эксперименты показали, что указанные процессы не обеспечивают защиты

информации от потерь при частой перезаписи и переформатировании, т.е. не дают гарантии того, что она сохранится в неизменном оригинальном виде.

Поэтому в настоящее время ученые и специалисты обращаются к исследованию и разработке других, более надежных и экономичных стратегий архивирования важнейшей электронной информации с использованием таких технологий долговременного хранения, которые не требуют постоянного обновления и поддержки. И здесь на помощь человечеству снова приходит микрофильм, проверенный и испытанный аналоговый носитель, обладающий огромным потенциалом.

Разработка гибридных способов сохранения информации. В международном стандарте по микрографии долгосрочное сохранение цифровой информации определяется в широком смысле как «действия, необходимые для поддержания доступа к цифровым данным после отказа носителя или смены технологии». По сути, управление хранением цифровых данных состоит в управлении рисками утраты цифровой информации со временем.

Цель управления хранением – обеспечить долговечность цифровой информации в приемлемой форме и гарантировать ее целостность. Для достижения этой цели лучше всего подходит архивный микрофильм как технологически независимый носитель, обеспечивающий гарантированное хранение информации сроком до 500 лет, а также ее неизменность и устойчивость за счет минимального вмешательства в процесс хранения.

Но как совместить аналоговый носитель – микрофильм, и цифровое содержание электронных документов? Для этого в микрографии необходимо осуществить интеграцию цифровых и аналоговых технологий. Принципиальная возможность такой интеграции появилась в начале 70-х годов прошлого века с изобретением COM-систем – устройств, позволяющих экспонировать электронную текстовую и графическую цифровую информацию из компьютера на микроформы. Сейчас на современном мировом рынке насчитывается около 20 моделей COM-систем ведущих мировых производителей.

Эти системы различаются по принципу записи, типам микроформ, с которыми работают, форматам принимаемых исходных файлов и другим техническим характеристикам, однако все они способны записывать цифровую информацию из компьютера на пленочные носители. Последним достижением в производстве COM-систем стала разработка лазерной цветной системы, способной качественно и с высокой скоростью вести запись цифровой информации на цветной микрофильм.

При этом продолжают совершенствоваться и существующие, хорошо зарекомендовавшие себя на рынке COM-системы. Так, фирмой Microbox была представлена новая версия изделия Polycom, способная работать с электронными образами документов до формата A0 включительно и в связи с этим являющаяся наиболее пригодным аппаратом для создания СФД для различных отраслей промышленности.

COM-системы вместе со сканерами микрофильмов по праву можно назвать ключевым звеном современных электронно-микрографических технологий, своего рода мостом между цифровым и аналоговым мирами. Несколько лет назад несовершенства и недостатки отдельных моделей, а также общая увлеченность стремительным развитием технологий оцифровки дали повод некоторым ученым считать, что микрофильм как носитель безнадежно устаревает, а COM-системы необходимы только для локального применения при сохранении специфических видов электронных документов.

Однако неудачи различных стратегий долгосрочного цифрового сохранения заставили исследователей пересмотреть свои взгляды и снова обратиться к традиционному микрофильму, теперь уже как к носителю для сохранения цифровой информации, долгосрочный и стабильный потенциал которого может быть усилен возможностями современных COM-систем.

COM-устройства коренным образом изменили способ создания архивных микрофильмов. Вместо использования для создания изображения оптической съемки эта технология считывает бинарные данные оцифрованного изображения и записывает положение каждого пикселя на пленку с помощью лазера (напрямую) или подобных устройств. Вариантом этой технологии являются записывающие устройства, способные переносить на микрофильм изображение с монитора – это стало возможным благодаря разработкам новых графических карт и специальных мониторов с очень высоким разрешением экрана. Современные COM-устройства могут принимать большую часть распространенных электронных текстовых и графических форматов, а новые аппараты позволяют улучшить качество вывода при работе с самыми различными оригинальными вводимыми изображениями.

Важная роль COM-систем в современном сохранении цифровых материалов подтверждается официальным принятием и введением в действие в 2009 г. международного стандарта ISO 11506 «Архивирование электронных данных. Компьютерный вывод на микрофильм (COM) и запись на оптический диск (COLD)». Данный стандарт впервые в мировой практике нормативно закрепляет стратегию долгосрочного архивного сохранения цифровой информации с помощью компьютерной записи на микрофильм для долгосрочного сохранения и на лазерный оптический диск для оперативного использования. Данный стандарт приобретен нашим институтом, переведен на русский язык и используется в работе.

В настоящее время в мире реализуется множество проектов сохранения цифровой информации с использованием COM-систем. Известно, что данные устройства широко применяются в библиотеке Конгресса США, различных отраслях Германии, Японии, Швеции, Франции, Великобритании и множестве других инновационных проектах по долгосрочному сохранению цифровой информации в ведущих странах мира.

Что касается России, то, по приблизительным подсчетам, в настоящее время в нашей стране находится в эксплуатации около 50 COM-систем

различных типов и производителей. Основными потребителями этих устройств являются организации и учреждения, участвующие в создании и наполнении единого российского страхового фонда документации, а также другие организации, осознающие важность долгосрочного страхового сохранения своих информационных активов.

Российский рынок такого рода оборудования представляется достаточно развитым. На нем представлены практически все основные мировые производители COM-оборудования, включая «большую тройку» ведущих немецких компаний – SMA, Zeutschel и Microbox.

Российская наука не стоит в стороне от указанных проблем. Так, в нашей стране именно ФГУП «НИИ Репрографии» на протяжении последних лет в интересах национальной безопасности государства теоретически обосновывает, нормативно и методически закрепляет, а также внедряет современные гибридные электронно-микрографические технологии создания, сохранения и использования ЕРСФД, которые позволяют интегрировать традиционные (микрографические) и современные (электронные) способы создания страховых фондов документации различного назначения.

Данные гибридные технологии позволяют долгосрочно сохранять на микрофильме определенные виды цифровой информации, в частности текстовую, фотографическую и чертежно-графическую документацию, созданную как путем оцифровки бумажных оригиналов, так и непосредственно в ЭВМ. Исследования, проводимые в данной области, опираются на твердую государственную поддержку, высокую научную квалификацию сотрудников НИИ Репрографии, передовой зарубежный опыт и парк современного электронно-микрографического оборудования (COM-системы, сканеры микроформ), позволяющего проводить различные эксперименты, отрабатывать технологические схемы и моделировать цепочки взаимодействия новых устройств в условиях функционирования системы СФД. При этом сотрудниками НИИ Репрографии осуществляется регулярный мониторинг зарубежной информации по проблеме исследований, осуществляется ее сбор, накопление и анализ.

Благодаря COM-системам открываются новые возможности в области долгосрочного сохранения цифровой информации. Современные инновации в сфере COM-систем существенно расширяют сферу их применения.

Так, по результатам последних зарубежных исследований теоретически обоснован и экспериментально подтвержден новый подход к сохранению цифровой информации на микрофильмах. Идея такого подхода заключается в следующем.

Любой цифровой документ состоит из набора двоичных данных – битовой информации. Эта битовая информация может быть закодирована в виде двухмерного штрих-кода, состоящего из информационных точек, а далее представлена в виде двухмерного растрового изображения.

Изображение при помощи COM-системы сохраняется на микрофильме. При необходимости восстановления информации штрих-кодовые данные считываются с микрофильма сканирующим устройством, а затем декодируются, в результате чего происходит восстановление оригинального электронного документа.

Значение этой технологии заключается в том, что впервые появилась теоретически обоснованная и технологически реализуемая возможность долгосрочно сохранять на микрофильме любую цифровую информацию и документацию.

При этом тип электронного документа не имеет значения, так как все цифровые файлы состоят из набора двоичных данных и соответственно могут быть представлены в виде двухмерных графических штрих-кодов.

Помимо уже осуществляемого сохранения цифровой цветной и черно-белой чертежно-графической, текстовой и фотографической документации, применение данного метода открывает казавшиеся ранее невозможными перспективы сохранения на микрофильмах цифровой аудиовизуальной документации, программных продуктов, трехмерной документации CAD-приложений и др., т.е. любого типа цифровых данных.

Сейчас предлагаются различные варианты этого подхода, такие как гибридное хранение, т.е. совместная запись на микрофильм как самого оригинала изображения документа, так и его цифрового штрих-кода, использование цветного микрофильма, что позволит повысить объем записываемых кодированных данных благодаря использованию трехцветных слоев и т.д. Однако принципиальная схема технологии остается такой, как на рис. 1.

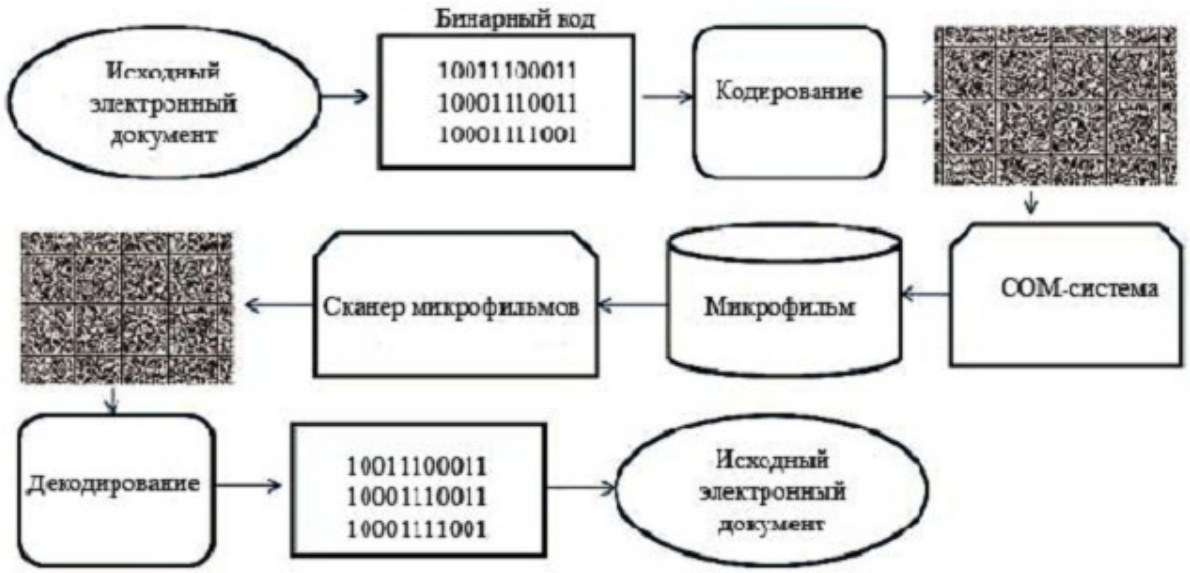

Рис. 1 – Схема сохранения бинарной информации на микрофильме

Исходный цифровой документ любого типа с помощью программных алгоритмов представляется в виде двухмерного штрих-кодового растрового

изображения, которое может восприниматься COM-системой. Затем данное изображение экспонируется COM-системой на микрофильм, который направляется на хранение. Далее с использованием сканера микрофильмов микрофильм сканируется, отсканированное штрих-кодовое изображение декодируется и происходит восстановление оригинального электронного документа (файла).

Необходимо заметить, что алгоритм кодирования / декодирования снабжен механизмом коррекции ошибок Рида–Соломона, аналогичным тому, который используется при записи / считывании оптических дисков, что повышает надежность считывания и декодирования штрих-кодовой информации.

Предлагается использовать для этих целей следующий вариант такого подхода. Хранение должно осуществляться гибридным способом, т.е. на микрофильм записываются как само аналоговое изображение, так и его цифровой код. По своей природе микрофильм позволяет считывать информацию и человеку, и машине, поэтому он может использоваться как гибридный носитель, сочетая аналоговую и цифровую информацию.

В качестве конкретного носителя предлагается цветной микрофильм производства Ilfochrome Micrographic [2]. Для хранения данных на цветной пленке есть свои основания, главное из которых заключается в том, что при хранении можно использовать все три цветовых слоя, благодаря чему увеличится объем сохраняемых данных. Двухмерный штрих-код, в который преобразовываются оригинальные документы, – это растровое изображение, в котором каждая растровая точка преставляет собой состояние. Одна растровая точка служит бинарным описанием состояния (максимальная или минимальная оптическая плотность) или описанием состояния более высокого порядка (несколько уровней плотности).

По данным экспериментальных исследований, в которых для записи цветного микрофильма использовалась цветная лазерная COM-система нового поколения Archive Laser Recorder, была достигнута достаточно высокая плотность записи информации. Так, при размере точек 15 мкм на шестисотметровом рулоне цветной пленки 35 мм можно сохранить 22 гигабайта данных. При размере точки 12 мкм – 38 гигабайт. При 9 мкм – примерно 70 гигабайт на одном рулоне.

Кажется, что такой объем не составляет конкуренции таким носителям, как, например, жесткий диск. Но не стоит забывать, что при хранении цифровой информации вместимость не всегда является определяющим фактором, особенно по сравнению с долговечностью и стабильностью.

Однако такая технология является достаточно затратной, так как для записи требуются цветная пленка, цветной лазер (цветные COM-устройства) и химико-фотографическая обработка цветной пленки, что достаточно дорого. Сканирующее оборудование, необходимое для считывания цветной

пленки, также является более сложным и дорогим, чем аналогичное оборудование для черно-белых материалов.

Соответственно если цвет решающего значения не имеет, рациональнее использовать черно-белый микрофильм. Тогда в качестве носителя используется обычный черно-белый микрофильм, а исходные электронные документы (их бинарные данные) кодируются с помощью двухмерного черно-белого графического штрих-кода. Затем эти данные трансформируются в изображение и сохраняются (экспонируются) на микропленку (рис. 2). При воспроизведении бинарных данных микрофильм сканируется, а изображение декодируется с помощью расшифровки отсканированного штрих-кода. В результате снова получается поток бинарных данных, из которых восстанавливается исходный электронный документ.

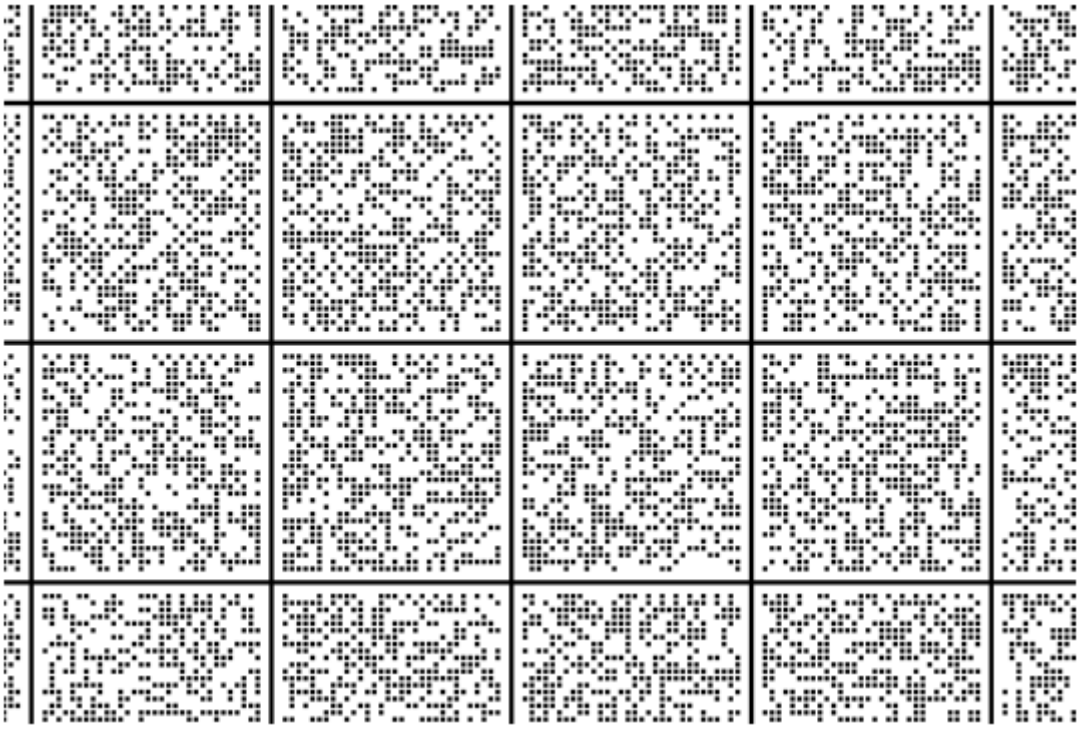

Рис. 2 – Увеличенный фрагмент черно-белого штрих-кодового изображения

Черно-белые штрих-коды позволяют добиться относительно высокой плотности записи информации. Выяснилось, что на одном 16-мм микрофильме длиной 30,5 м в штрих-кодах можно сохранить 7200 изображений формата А4 или 45,32 мегабайт информации (на 35-мм микрофильме соответственно в 2 раза больше). В данном случае стоимость хранения 1 мегабайта составит $0,28, что в долгосрочной перспективе хранения представляется наиболее оптимальной по сравнению с другими системами, особенно в сравнении ее со стоимостью миграции каждые 5–7 лет, необходимой для других форматов, и стоимостью их технической поддержки.

Так, например, хранение на современных жестких дисках обходится 0,1…0,3 доллара за 1 гигабайт, но эти технологии требуют значительных

затрат в процессе хранения, так как большое количество жестких дисков должно постоянно функционировать, чтобы поддерживать систему в рабочем состоянии. Это требует значительных затрат на электроэнергию, инфраструктуру и техобслуживание на протяжении относительно короткого срока службы.

К тому же в отличие от других носителей, таких как жесткие диски, флеш-карты, CD- или DVD-диски, технологии считывания микрофильма очень просты и универсальны. Тогда как для воспроизведения данных с популярных электронных носителей необходимы специализированные интерфейсы и сложные технологии (оптические диски с лазерной технологией, высокоточное расположение считывающих устройств для магнитных носителей, контролирующие программы и оборудование и т.д.), для считывания данных с микрофильма необходимы только простые оптические устройства. Это выгодно отличает данный носитель от IT-систем. Если найти в будущем устаревший привод для DVD или лент, или USB-порт, совместимый с новыми компьютерными системами, будет очень сложно, то для микрофильма будет достаточно любого современного оптического устройства для формирования изображения – это может быть сканер, камера или другой аппарат.

Заключение. На основе проведенных исследований можно сделать следующие выводы:

1) В свете последних достижений науки технологический потенциал микрофильма и COM-систем в деле долгосрочного сохранения цифровой информации представляется очень существенным. Разумеется, что новые технологии требуют совершенствования, исследований и экспериментов по подбору параметров записи, отработке режимов, синхронизации оборудования, оптимизации настроек элементов системы, технико-экономические расчетов и т.д. Однако первые шаги уже сделаны, и дальнейшие исследования возможности применения данного перспективного метода обязательно будут продолжены как за рубежом, так и в нашей стране.

2) Основными моментами, определяющими направления развития работ по информационному страхованию различных видов информации за рубежом, являются следующие:

– Рост тенденции архивирования цифровой информации на микрофильме.

– Снижение доли классического оптического микрофильмирования.

– Развитие технологий цветного микрофильмирования.

– Совершенствование возможностей и улучшение технических характеристик современного микрографического оборудования, такого как COM-системы и сканеры микроформ.

3) Сегодня специалисты ведущих стран мира опять обратились к апробированной технологии обработки и сохранения информации – микрографии; правда, это теперь существенно усовершенствованная и обогащенная новыми возможностями технология.

Литература:

1. Мировой опыт создания и хранения информационных ресурсов в современных условиях / А.К. Талалаев, Е.Е. Евсеев, П.Е. Завалишин, Н.Е. Проскуряков // Изв. Тул. гос. ун-та. Технические науки. – 2013. – № 3. – С. 408–421.

2. Ilfochrome Micrographic Film [Электронный ресурс]. – Режим доступа: http://www.yumpu.com/et/document/view/549624/ilfochrome-micrographic-film, свободный (дата обращения: 07.04.2013).

ЭЛЕКТРОННЫЕ АРХИВЫ: ВОЗМОЖНЫЕ РЕШЕНИЯ ПРОБЛЕМ ДОЛГОСРОЧНОГО ХРАНЕНИЯ ДАННЫХ

Источник: http://www.isa.ru/proceedings/images/documents/2013-63-4/t-4-13_39-49.pdf

Авторы: Г. П. Акимова, М. А. Пашкин, Е. В. Пашкина, А. В. Соловьев

Аннотация. В работе систематизированы проблемы, возникающие при долгосрочном хранении электронных документов, предложены возможные варианты их решения, опробованные авторами при создании ряда систем электронных архивов. Статья является продолжением серии публикаций, посвященных проблемам создания и внедрения электронных архивов.

Введение

Как было показано ранее [1], сравнительно недавно начавшийся бум внедрения систем электронного документооборота (СЭД) в организациях не затрагивает процесса передачи завершенных документов в полноценный делопроизводственный архив. Предположительное отставание внедрения электронных архивов от оперативных информационных систем на 3–5 лет вполне объяснимо, поскольку указанный срок – это среднее время хранения документов в «оперативных» архивах или в БД СЭД до их массовой передачи в вышестоящие архивы. Кроме того, достаточно редко можно услышать об электронных архивных системах, которые позволяют длительно хранить электронные документы (сроком не менее 5 лет).

Многие организации не желают решать задачу оцифровки большого количества документов (см. [1]) или не видят перспективы результатов такой работы, даже если и используют СЭД, и продолжают работать с архивом бумажных документов. В такой ситуации можно рекомендовать использовать ЭА РК документов, как учетную систему для ускорения и упрощения задачи поиска документов, если в архивной карточке указаны

топологические данные о месте хранения документа. Решается и обратная задача – по бумажному оригиналу необходимо «поднять» всю историю работы с документом (например, выдача документа, связи данного документа с другими и т. д.). В такой ситуации электронный архив представляет собой лишь реквизитную БД, не содержащую текстовую часть документа. Такой ЭА не решает задач долговременного хранения документов, а перекладывает данную задачу на архив бумажных документов, но при этом обеспечивает эффективный способ поиска и навигации по архиву документов. Возможны варианты использования ЭА как дополнения к существующему бумажному архиву, при этом можно сохранять также и оцифрованные документы.

В предлагаемой статье рассмотрены основные способы хранения электронных документов, используемые в настоящее время, при этом особое внимание уделено документам с длительным сроком хранения. Выделены проблемы и возможные пути их решения, которые могут помочь разработчикам электронных архивов.

1. Основные понятия и определения

Электронный архив (ЭА) – структурированное хранилище неизменяемых электронных оригиналов документов (электронных изображений бумажных документов), созданное на основе законов и правил ведения архивов на конкретной территории (в конкретной стране).

Длительное хранение – хранение электронных документов не менее 5 лет. Определение не претендует на «абсолютность», так как в конкретных архивах эти сроки могут меняться. За основу срока (5 лет) взят максимальный срок хранения документов в оперативных архивах СЭД. Электронные документы могут храниться в течение десятилетий или даже столетий или «бессрочно» в зависимости от их важности.

Долговременная сохранность – «период времени, в течение которого электронные документы поддерживаются в качестве доступного и аутентичного свидетельства (доказательства)» [2].

Аутентичный электронный документ – «электронный документ, точность, надежность и целостность которого сохраняются с течением времени» [2].

Электронная подпись – «информация в электронной форме, которая присоединена к другой информации в электронной форме (подписываемой информации) или иным образом связана с такой информацией и которая используется для определения лица, подписывающего информацию» [3].

До 2012 г. вместо ЭП использовался термин ЭЦП (электронно-цифровая подпись), определявшаяся как реквизит электронного документа, предназначенный для определения лица, подписавшего документ. В контексте данной статьи ЭП является частью хранящегося в ЭА электронного документа.

Квалифицированная электронная подпись – электронная подпись, которая соответствует следующим признакам [3]:

«1) получена в результате криптографического преобразования информации с использованием ключа электронной подписи;

2) позволяет определить лицо, подписавшее электронный документ;

3) позволяет обнаружить факт внесения изменений в электронный документ после момента его подписания;

4) создается с использованием средств электронной подписи;

5) ключ проверки электронной подписи указан в квалифицированном сертификате;

6) для создания и проверки электронной подписи используются средства электронной подписи, получившие подтверждение соответствия требованиям, установленным в соответствии с» [3].

Удостоверяющий центр – «юридическое лицо или индивидуальный предприниматель, осуществляющие функции по созданию и выдаче сертификатов ключей проверки электронных подписей, а также иные функции, предусмотренные» [3].

2. Постановка задачи долговременного хранения электронного документа и проблемы ее реализации

В простейшей постановке задача формулируется следующим образом. Требуется обеспечить длительное хранение электронных документов (см. определение в [1]) в программно-аппаратной среде, причем в течение всего срока хранения должна обеспечиваться аутентичность документа. При этом предполагается, что аутентичность документа на момент передачи его в архив подтверждена, документы не искажены, сохранность документов полная. Нет ограничений на форматы данных передаваемых в ЭА документов. ЭА сертифицирован для работы со средствами ЭП.

В рамках данной статьи предполагаем, что при длительном сроке хранения истекают сроки действия сертификатов ЭП, заканчивается оперативное хранение документов в архивах подразделений, завершается поддержка версий некоторых операционных систем (ОС), например Windows, и прикладного программного обеспечения.

На первый взгляд постановка достаточно простая, однако на практике при реализации может возникнуть множество проблем, связанных с технической сложностью реализации требований именно долговременного хранения. Выделим основные проблемы, которые всегда возникают при решении поставленной задачи:

1) аутентичность документа в течение всего срока хранения;

2) «старение» носителей информации;

3) перемещение данных и сохранность метаданных;

4) интерпретируемость и отображение электронных документов.

Указанные выше проблемы, конечно, хорошо известны и неоднократно обсуждаемы в среде разработчиков ЭА (см. например, ГОСТ Р 54989–2012

[2], являющийся переводом ISO TR 18492:2005, а также иные «переведенные» ГОСТ [4, 5] и системы требований [13]). Однако, в перечисленных документах описание проблем носит скорее рекомендательный характер и формулируется как «разработчики ЭА должны продумать» те или иные вопросы. Конкретных рекомендаций не приводится из-за отсутствия более развитых и всеобщих (мировых) стандартов на правила хранения электронных документов, правила аутентификации, отсутствия стандартов на форматы хранения электронных архивных документов и средств интерпретации документов.

2.1. Сохранность аутентичности документа

Срок хранения в архиве зависит от вида документа, некоторые особо ценные документы должны храниться бессрочно (фактически столетиями), другие – десятилетиями, как документы по личному составу, или годами. Это накладывает определенные требования на технику, используемую в архивах длительного хранения, и требование сохранения аутентичности документа в течение всего срока хранения.

Для гарантии неизменности документа должны применяться как организационные меры, так и программные средства. К организационным мерам обычно относят защиту документов от несанкционированного доступа, например, непосредственно к хранилищам данных, носителям информации, коммуникационному оборудованию и др.

К программным средствам относят разграничение прав доступа на электронные документы и обеспечение контроля целостности, который реализуется с помощью хранения хеш-кодов электронных документов или использования ЭП. В последнем случае ЭП может автоматически устанавливаться программными средствами ЭА при вводе электронного документа в БД ЭА. Возможен вариант, когда в ЭА поступает электронный документ, подписанный ЭП. В этом случае перед ЭА стоит задача проверки корректности ЭП при вводе и, в дальнейшем, обеспечения неизменности документа в процессе хранения.

Задача оперативного хранения документов (не более 5 лет) решается достаточно просто, поскольку не приходится решать сопутствующие проблемы: большинство сертификатов ключей подписи не успевают исчерпать срок своего действия, мала вероятность, что изменения коснутся удостоверяющих центров (УЦ), например, прекращение деятельности УЦ.

При организации хранения документов свыше 5 лет разработчики и пользователи ЭА гарантированно столкнутся с проблемой просроченных сертификатов ключей ЭП, в том числе и корневых, т. е. сертификатов удостоверяющего центра, а в этом случае ЭП будет считаться недействительной, и при проверке будут зафиксированы ошибки, связанные с истечением срока действия сертификатов.

Еще одной проблемой, возникающей при использовании низкоразрядных ключей (до 256 бит) ЭП и накоплении огромных массивов электронных документов, является возможность получения так называемых

коллизий первого и второго рода соответственно: подделка документов в ЭА для соответствия ЭП и появление в БД ЭА разных документов с одинаковой ЭП. Несмотря на то, что эта проблема пока маловероятна (стойкость ЭП с 256-битными ключами до 10^{30} операций), в будущем, если документы хранятся десятилетиями и массив таких документов огромен – проблема может проявиться очень остро. К тому же в связи с бурным развитием техники и технологий подделка ЭП на низкоразрядных ключах через несколько лет не составит большого труда.

И, наконец, законодательство, в частности [6], допускает наличие у одного лица (организации) нескольких ключей (сертификатов) ЭП. Также прямо не запрещено использование одного ключа (сертификатом) ЭП несколькими лицами. А это, в свою очередь, может создавать путаницу при идентификации лица, подписавшего электронный документ.

2.2. «Старение» носителей информации

Помимо возможности подмены или потери документов в процессе переноса с носителя на носитель из-за умысла или халатности персонала, существует проблема выхода из строя самих носителей информации (дисков, лент, оптических носителей и др.). Ни один производитель подобной техники не гарантирует сохранность ее в течение десятилетий (тем более столетий), а, следовательно, встает проблема своевременной диагностики носителей электронных документов и своевременной перезаписи документов на другие носители.

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

Даже для специально предназначенных для ЭА накопителей на базе технологии UDO (Ultra Density Optical, разработка компании Plasmon) на основе ультраплотной записи [7] не подтверждена возможность их работы в течение многих десятилетий. Накопители UDO служат в ЭА медицинских изображений для хранения медицинских документов, например, медицинских карт пациентов, причем гарантированный срок хранения не превышает 5 лет. UDO представляет собой картридж 5.25 с оптическим диском внутри. Объем диска на данный момент составляет от 60 Гб до 120 Гб. Для записи может использоваться как красный лазер (650 нм), так и сине-фиолетовый (405 нм), причем во втором случае максимальный объем диска может достигать 500 Гб. Оптический диск не подвержен размагничиванию, как магнитные носители.

Магнитные ленты являются крайне неустойчивыми к внешним воздействиям, поскольку требуется их перемотка 1 раз в полгода и тщательная защита от размагничивания.

Использование твердотельных накопителей (SSD — Solid state disk, флэш-карт и т. д.) также пока ненадежно. Данные накопители имеют ограничение на количество циклов перезаписи (3000–10000), повышенный износ в связи с этим, высокую стоимость гигабайта информации по сравнению с жесткими дисками и оптическими дисками и невысокий объем хранения данных [8]. Проблему повышенного износа пытаются преодолеть с помощью технологии энергонезависимой памяти FRAM[1] (количество циклов перезаписи до 10^{14}) [9]. Однако, и эти носители не позволяют хранить большие объемы данных, зато отличаются высокой стоимостью. Время гарантированного хранения данных на SSD и FRAM оценивается в 10 лет.

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

Кроме того, при стремительном развитии вычислительной техники имеет место технологическое старение, поэтому с достаточно высокой вероятностью через 100 лет невозможно будет прочитать данные с современных магнитных и оптических носителей из-за отсутствия в будущем устройств их чтения, даже если информация каким-то чудом на них сохранится.

2.3. Перемещение данных и сохранность метаданных

Поскольку выше была затронута тема надежности хранения документов на внешних носителях, то с необходимостью возникает задача переноса архивных данных на новые носители информации, а, значит, встает вопрос об отсутствии потерь данных при проведении данной операции.

При этом проблема касается не только переноса самих документов, но и метаданных (см. [10]) документов, индексов (в том числе и полнотекстовых). Если сопутствующие данные (индексы, метаданные, классификаторы, рубрикаторы, связи с другими документами и др.) не могут быть корректно перенесены, то, по сути, перемещение данных (миграция) выльется в повторное создание ЭА в новой операционной среде (на новой платформе) с построением заново метаданных, индексов и т. д. К тому же, если документ является частью единицы классификации (дела, пачки) или связан с другими документами, данные связи также должны быть восстановлены, иначе целостность хранения может быть поставлена под сомнение.

[1] Ferroelectric Random Access Memory – сегнетоэлектрическая память с произвольным доступом.

2.4. Интерпретируемость и отображение данных

При долговременном хранении электронных документов возникает проблема интерпретируемости и отображения данных в новых информационных условиях, т. е. наличие возможности раскодировать хранимый формат электронного документа через десятилетия и показать документ в том или ином виде, например, отобразить на экране, распечатать и т. д.

Отсутствие стратегии в данном вопросе и превращение ЭА в склад разноформатных документов может спустя десятилетия привести к тому, что часть информации невозможно будет раскодировать изза отсутствия (устаревания) средств интерпретации хранимых форматов данных, а также из-за утери описания хранимых форматов, в случае использования закрытых форматов представления электронных документов. Некоторым решением проблемы может быть создание конвертеров, преобразующих старые форматы в новые, но здесь следует иметь в виду, что чем позже будет поставлена задача конвертации данных, тем менее реально будет ее решение.

2.5. Прочие проблемы хранения документов в ЭА

Список рассмотренных выше проблем, естественно, не является окончательным. В частности, он не содержит таких важных задач, как скорость работы с архивом и информационная безопасность. Про подходы к решению этих задач написано достаточно много, в том числе и методики создания моделей угроз, нарушителей и т. д. В контексте этого исследования мы не будем касаться данных вопросов.

Не менее важной задачей является хранение и обработка больших объемов данных (как реквизитов, так и документов, включая задачу первоначального наполнения ЭА и потокового ввода документов). К обзору этого вопроса мы планируем вернуться в следующих статьях, посвященных электронному архиву.

Не рассматриваются в рамках данного исследования всевозможные юридические «тонкости», связанные с подписанием документов ЭП юридическими и физическими лицами, возможности непризнания электронного документа, не представляющего юридической силы. Для простоты примем допущение, что все электронные документы, заверенные ЭП и поступающие в ЭА, прошли проверку юридической значимости в момент их заверения ЭП.

Также не рассматриваем в данной статье такой важный аспект, как обеспечение катастрофоустойчивости решения ЭА. Этот вопросом мы планируем осветить в следующих статьях, посвященных созданию электронных архивов документов.

3. Предлагаемые решения проблем долговременного хранения электронных документов

В данной главе авторы делятся опытом, полученным при создании электронных архивов длительного хранения документов.

3.1. Сохранение аутентичности документа

На настоящий момент основным решением проблемы сохранения аутентичности документа является использование ЭП. Однако сертификаты и открытые ключи ЭП обладают ограниченным сроком действия, поэтому спустя год или 5 лет при обращении к документу с просроченной ЭП можно получить сообщение о некорректности ЭП, что поставит под сомнение подлинность документа. ЭП удобно использовать в системах электронного документооборота, поскольку сроки работы с документом малы, однако в системах, обеспечивающих длительное хранение, гарантированно возникнут проблемы просроченных сертификатов и ключей подписи.

При решении таких задач рекомендуется использовать для длительного хранения только усиленную квалифицированную ЭП, заверенную квалифицированным сертификатом (см. [3]), т. е. ЭП должна содержать подтвержденный штамп времени. При этом цепочка сертификатов ключей в идеале должна обязательно содержаться внутри ЭП или передаваться в ЭА вместе с ЭП. Только в этом случае есть гарантия, что спустя десятилетия подлинность документа можно будет подтвердить, если за это время, конечно, не изменятся стандарты, и будут существовать средства проверки данной ЭП. При этом нужно учесть, что при проверке ЭП может потребоваться список отзыва сертификатов (СОС), актуальный на момент проставления подписи.

В качестве ключевой меры обеспечения аутентичности хранимых документов в ЭА авторами предлагается использовать архивную ЭП, которая автоматически вычисляется для всех электронных документов, помещаемых в ЭА. Процесс простановки такой ЭП должен быть возложен на операторов ввода, каждый из которых должен подписывать документ своей ЭП.

В организациях, работающих с ЭП, принято за правило периодически проводить смену ключей. Это означает, что все электронные документы, находящиеся в ЭА, следует переподписывать новым ключом ЭП (по сути новой ЭП), при этом старая ЭП сохраняется. Надо понимать, что такая схема не исключает подмены документов административным персоналом, эксплуатирующим ЭА, но гарантирует невозможность проведения данной операции операторами ввода. Кроме того, данная процедура не утверждена законодательно. Однако, переподписывание может быть включено в регламентные действия ЭА, например, стать частью процедуры инвентаризации архивных фондов. В этом случае необходимо тщательно защищать закрытые ключи электронной подписи, прописать подробно процедуру инвентаризации. Авторы считают, что процедура переподписывания документа электронной подписью оператора при вводе в архив должна быть закреплена законодательно и явиться основой для создания ЭА длительного хранения. Назовем данную процедуру инвентаризацией ЭП. В процессе инвентаризации ЭП подтверждается корректность ЭП документа, и он заверяется дополнительной ЭП (например, с ключом более высокой разрядности) в подтверждение факта

инвентаризации. Новая ЭП, как более криптостойкая, исключит (или, по крайней мере, существенно снизит) риск появления в будущем документов-подделок, заверенных старыми «правильными» ЭП в БД ЭА.

Процедуру инвентаризации ЭП можно запускать в автоматическом режиме от имени оператора, работающего с архивом (особенно при огромных объемах данных), предоставляя операторам в проблемных случаях (например, нечитаемость данных, ошибка проверки ЭП и др.) принимать решения по возникшей проблеме.

Мощность компьютеров постоянно увеличивается, поэтому средства взлома ЭП, использующие полный перебор, со временем могут преодолевать все большую разрядность ключа подписи. Так, на сегодняшний момент безопасными считаются ЭП с 512-битным ключом и выше, однако в 2009 г. была взломана ЭП с 768-битным ключом, но пока это возможно только за продолжительное время с использованием практически неограниченных компьютерных мощностей. Для некритичных данных можно использовать ЭП с 256-битными ключами (стойкость до 10^{30} операций).

Поэтому теоретически со временем возможна подделка документов в ЭА (коллизия первого рода), когда подбирается документ для ЭП, тем самым нарушается принцип неизменности документа в архиве. Только совместные организационные (включая политики безопасности ЭА), технические и программные способы позволят снизить вероятность взлома.

С накоплением документов при использовании низкоразрядных ключей (до 256 бит) возможна коллизия второго рода: наличие разных документов с одинаковой ЭП, что маловероятно, но теоретически возможно. Поэтому при проектировании ЭА нужно учитывать предполагаемый размер БД и возможный ее рост, чтобы предоставить адекватные средства защиты информации.

Следует обратить внимание на еще один аспект, возникающий при подтверждении аутентичности заверенных ЭП электронных документов, – сложность взаимодействия ЭА с удостоверяющим центром. Особенно часто с ним сталкиваются, когда в ЭА хранятся электронные документы, подписанные ЭП, которые выданы разными УЦ, в том числе в различных регионах РФ. В таком случае возникают ситуации, когда ЭА не может проверить ЭП поступившего документа, кроме того нет никаких гарантий хранения сертификатов ЭА самими УЦ. На данный момент решения указанной проблемы нет. В качестве одного из промежуточных решений авторы статьи предлагают непосредственно в ЭА организовать хранение всех сертификатов, списков отзыва сертификатов (СОС) и много другой дополнительной информации, на основании которой может быть проведено расследование и установлена подлинность документа. Тем самым функции УЦ переносятся в ЭА, особенно в отсутствии единой сети УЦ страны, и усложняют его.

3.2. «Старение» носителей информации

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

Поэтому решение проблемы лежит, во-первых, в избыточности хранения информации, во-вторых, в регулярной проверке и переносе информации на новые носители данных. Избыточность хранения данных должна быть обеспечена как хранением данных ЭА непосредственно в БД на жестком диске, так и хранением на внешних носителях копий данных ЭА. В качестве такой копии может выступать как резервная копия БД, так и копии данных, вытесненных на внешние носители. Надо отметить, что хранение данных в ЭА может быть устроено следующими способами: индексы и данные находятся в БД; индексы находятся в БД, данные на внешних носителях; индексы находятся в БД, часть данных находится в БД, часть вытеснена на внешние носители. Во всех случаях для БД ЭА необходимо организовать регулярное резервное копирование БД на внешние носители. В качестве копии данных могут выступать как внешние носители с резервной копией БД, так и совокупность внешних носителей с резервной копией БД и внешних носителей с данными. При этом должно создаваться не менее двух копий данных ЭА, причем хранить их следует в разных помещениях, а в идеальном случае в разных зданиях, удаленных друг от друга. В случае использования в качестве резервной копии компакт-дисков рекомендуется создавать не менее трех копий данных. Дополнительно может быть реализовано катастрофоустойчивое решение (зеркало, или, для особо ценных документов, – резервный центр обработки данных (ЦОД)), т. е. хранение точной копии (копий) документов. Это означает, что необходимо реализовать децентрализованное хранение копий данных с разными мандатами доступа для оперативного и административного персоналов ЭА.

Регулярная проверка и перенос информации на новые носители должны обеспечить защиту от отказов и физической деградации цифровых носителей информации. Назовем такую процедуру инвентаризацией носителей. Данная операция должна включать проверку целостности данных на носителе, оценку оставшегося времени хранения данных на носителе и, при необходимости, перенос данных на новый носитель с уничтожением старого. В случае выявления нарушения целостности данных на носителе в ходе проверки новая копия данных создается из других копий данной информации. Периоды проверки носителей данных выбираются, исходя из типа носителей информации, но в любом случае период хранения данных на неизменяемом носителе не должен превышать трех лет, т. е. раз в три года каждый носитель информации должен быть проверен и при необходимости заменен. Процесс переноса информации должен предусматривать возможность слияния данных с разных носителей, данное условие

появляется из-за постоянного увеличения объемов всех видов носителей данных.

Только в этом случае можно будет говорить о сохранности данных в ЭА.

3.3. Перемещение данных и сохранность метаданных

Миграция данных должна быть неотъемлемой частью методологии создания ЭА долговременного хранения. Другой вопрос, что должно подвергаться миграции: только ли сами документы из БД ЭА или же еще связанные с ними метаданные, классификаторы, индексы и др.

Как было показано ранее ([1]), классификаторы и индексы являются неотъемлемой частью документа, поскольку определяют контекст его использования: предметную область, структуры организаций, логику хранения и классификации, связи с другими документами и т. д. По мнению авторов статьи, потеря этих данных при миграции может оказаться критичной, документ будет вырван из контекста использования, и понять его принадлежность какой либо тематике будет проблематично.

Поэтому решение по миграции данных должно включать не только миграцию самих электронных документов, но и метаданных документа, расширив описание формата долгосрочного хранения (см. п. 3.4) набором тегов, которые нужны для хранения метаданных (например, расширенное дублинское ядро[2] [12]) документа. Практическая реализация данного положения при разработке авторами статьи электронных архивов подтверждают его правильность.

Отдельно стоит вопрос о полнотекстовых индексах документа. Конечно, не хочется терять такую ценную информацию, однако большинство СУБД не позволяет распорядиться полнотекстовыми индексами самостоятельно, а перестройка индекса для огромного массива данных после миграции может оказаться дорогостоящей по времени процедурой. Несовместимым может оказаться и формат индексов при переносе в другую среду хранения. При решении данной проблемы рекомендуется либо переносить полнотекстовые индексы вместе с документами, либо включить процедуру перестройки индексов в процесс миграции. Во втором случае переход на новую инфраструктуру ЭА должен быть осуществлен с задержкой в эксплуатации после миграции данных, т. е. для организации постепенного ввода в строй новой версии ЭА, что, в свою очередь, означает допущение существования ЭА в двух различных средах хранения при условии полной синхронизации данных между версиями БД ЭА.

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

[2]Стандарт метаданных. Создан см. Dublin Core Metadata Initiative (http://dublincore.org/). В России с 01.07.2011 действует ГОСТ [12].

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

3.4. Интерпретируемость и отображение электронных документов

В информационном мире существует множество различных форматов электронных документов, но со временем многие из них перестают поддерживаться, а тем самым с течением времени трудно будет найти программное обеспечение, способное проинтерпретировать документ, сохраненный десятки лет назад в некотором формате.

Согласно [2]: «Стратегия долговременной сохранности должна обеспечить, чтобы электронные документы в будущем оставались читаемыми. Для достижения этой цели составляющий электронные документы поток битов должен быть доступен на той компьютерной системе или устройстве:

– на которой(ом) он первоначально был создан;

– на которой(ом) он в настоящее время хранится;

– которая(ое) в настоящее время используется для доступа к нему;

– которая(ое) будет использоваться для хранения электронной информации в будущем».

Первые два варианта относятся скорее к сохранности собственно носителей (см. п. 3.2). Рассмотрим теперь проблему читаемости собственно данных, расположенных на читаемом носителе (не столь важно новом или старом).

Для решения такой проблемы должен быть подобран формат хранения архивных документов, отвечающий требованиям: простой, открытый и документированный, которые в свою очередь снизили бы вероятность «не интерпретируемости» документов, сохраненных в ЭА в данном формате, в будущем.

Федеральное агентство по техническому регулированию и метрологии РФ утвердило ГОСТ Р ИСО/МЭК 26300–2010 (перевод принятого в 2006 году международного стандарта ISO/IEC 26300:2006), в котором в качестве стандартного формата для офисных приложений определен Open Document Format (ODF).

Существует поддержка ODF, начиная с MS Office 2007. Google Docs и IBM Lotus Symphony так же поддерживают ODF.

В первой версии спецификации MoReq [13] редакции 2001 г. существовал раздел, посвященный долгосрочному хранению электронных документов, в котором одним из ключевых было требование предпочтительного использования открытых, документированных форматов в противовес к проприетарным (коммерческим).

В настоящее время при использовании обычных «текстовых» форматов офисных приложений выделяют группу рисков, которые связаны с используемыми форматами файлов:

«Во-первых, это проблема скрытой информации. Потенциально любой офисный документ может содержать в себе данные о предыдущих правках, комментарии, невидимый текст, сведения о компании и авторе. Все это для окончательной редакции является лишним и не должно попадать в электронный архив.

Во-вторых, автор может использовать в документе поля, значения которых могут изменяться, что приводит к искажению всего документа. Простейший пример – поле с текущей датой. Представьте, мы распечатываем документ из архива, а он датирован сегодняшним числом. Также не следует забывать и о макросах, которые могут изменить документ.

В-третьих, документ может содержать гиперссылки на веб-страницы или на другие связанные объекты (рисунки, схемы, другие документы). Иногда это действительно необходимо для удобства пользования этим документом и для его понимания. Но при помещении такого документа в архив с этим надо что-то делать – сохранять вместе с документом копии веб-страниц, например» [14].

Для решения указанной проблемы авторы предлагают в качестве формата архивного документа использовать открытые документированные форматы XML, ODF, PDF/A, в один из них конвертировать принимаемые в архив файлы, сохраняя оригиналы файлов как приложения (в случае их заверения ЭП – сохраняя вместе с ЭП). Однако, для более строго решения необходимо законодательно утвердить правила приема документов в ЭА и их переформатирование при сдаче на длительное хранение. Тогда в процессе приема в ЭА необходимо будет перезаверить ЭП весь набор полученных файлов документов, сохраняя оригиналы документов в исходном формате и их оригинальные ЭП. Соответствующая процедура также должна быть разработана и утверждена.

Отдельно стоит вопрос работы с видео, аудио документами, презентациями, анимационными файлами, программным кодом (скрипты), исполняемыми файлами и их компонентами. Для видео и аудио документов также необходим перевод их в наиболее простые и открытые форматы (возможно, что таким стандартом станет WebM [16]), сохраняя оригиналы сдаваемых документов. Впрочем, данная тема требует отдельного исследования, авторы статьи не располагают в данном вопросе достаточным опытом.

Помимо преобразования электронных документов в форматы хранения документов, потребуется предусмотреть процедуру инвентаризации данных, в процессе выполнения которой «устаревшие» форматы электронных документов ЭА должны быть преобразованы в современные для процедуры инвентаризации данных форматы электронных документов. Например, если в ЭА хранятся видеоданные в формате программы, которая больше не

развивается, то следует провести преобразование таких видеоданных в формат программы, которая будет развиваться и поддерживаться, или преобразовать в стандарт, который поддерживается многими производителями программного обеспечения.

Отметим, что при копировании данных ЭА на внешние носители информации должны сохраняться как структура описания данных, так и описание формата хранения данных. Для хранения метаданных рекомендуется использование XML-формата.

3.5. Синхронизация электронного и бумажного архивов

На данный момент задача является значимой ввиду наличия огромных бумажных архивов. Одним из вариантов решения задачи синхронизации является использование штрих-кодов (двумерные штрих-коды способны хранить до 4 КБ информации), которым надпечатывается бумажный документ при регистрации (оцифровке) в архиве. Для особо ценных документов, особенно в случае их возможной порчи при надпечатке, может быть надпечатана штрих-кодом бумажная архивная карточка документа.

При этом перед началом создания ЭА должна быть продумана топология (например, комната-стеллаж-полка) хранилища бумажных документов, информация о которой хранится в электронном архиве как реквизит документа (вплоть до координат GPS).

Без решения данной проблемы и при наличии одновременно бумажного и электронного архивов (реквизитная БД), использование архива по назначению будет затруднительно и связано с временными издержками на поиск бумажных оригиналов и, наоборот, при отслеживании связей, окружения, классификации бумажного оригинала по ЭА (обратная задача).

4. Модель документа в ЭА долгосрочного хранения

Рассмотрев проблемы долгосрочного хранения документов и возможные пути их решения, можно скорректировать модель документа в ЭА, представленную в [1].

Графически модель документа в электронном архиве можно представить в виде графа (дерева), состоящего из взаимосвязанных семантических блоков B_i. Блоки в свою очередь представляют собой подграфы (поддеревья), также состоящие из семантических блоков следующего уровня: в любом документе всегда можно выделить заголовок, подзаголовки, повторяющиеся части, агрегаты (массивы, структуры данных), атомарные данные (листы дерева). Между документами могут существовать различные отношения (связи) [15], т. е. лес документов может быть связан в единый граф. При этом в вершинах деревьев можно указывать неявные связи с другими документами.

При длительном хранении документа кроме классификаторов и индексов [1], являющихся неотъемлемой частью электронного документа и проходящих вместе с ним возможные миграции данных, документ дополняется содержимым документа, преобразованным в один из форматов

долгосрочного хранения (открытых, документированных форматов) XML, ODF, PDF/A. Поэтому модель документа в ЭА преобразуется в следующую (оператор «+» в данной записи в отличие от [1] заменен на операцию «объединения», так как речь идет не о фактическом сложении, а об объединении множеств различных данных):

$$DAr = \cup_{(i=1,N)}(B_i) = ArCard \cup OdfD \cup OrD \cup FTIdx \cup CLIdx,$$

где:

ArCard — архивная карточка документа (состоит из набора реквизитов, которые могут задаваться древовидной схемой) – изменяемая часть электронного документа, может меняться форма карточки, а также состав ее реквизитов. Однако изменение значений реквизитов, по крайней мере тех, которые получены из оригинала документа, запрещено, либо выполняется только уполномоченными лицами. Оперативно могут изменяться только значения реквизитов, определяющих нумерацию в данном конкретном архиве, топологию (размещение физического оригинала), служебную информацию: шифры, аннотация и т. д.;

OdfD преобразованное к формату долгосрочного хранения содержимое оригиналов документов – неизменяемая часть электронного документа, создается при приеме документов в ЭА, OdfD заверяется ЭП (в общем случае несколькими) при приеме в ЭА;

$$OdfD = OdfDoc \cup (\cup_{i=1,N1} OdfPic_i) \cup (\cup_{j=1,N2} Sign_j),$$

где:

OdfDoc – собственно преобразованное к формату долгосрочного хранения содержимое сдаваемых документов,

OdfPic – набор (1 – N1) графической информации (растровые и векторные изображения, элементы презентаций и др.), подлежащей преобразованию из сдаваемых документов в графические форматы долгосрочного хранения (TIFF, JPEG, PDF/A), при этом OdfDoc содержит ссылки на графические материалы,

Sign – набор ЭП (1–N2), заверяющих преобразованный документ (содержит в себе сертификаты подписавших, цепочку сертификатов, сертификаты удостоверяющих центров (УЦ));

OrD – оригиналы документов (электронные оригиналы документов или оцифрованные изображения оригинальных бумажных документов, которые далее также будем обозначать как оригиналы) – неизменяемая часть

электронного документа (может включать ЭП, проставленные, например, в системе электронного документооборота – см. [1]);

FTIdx – полнотекстовый индекс, полученный на основе индексирования реквизитов и текстов документа – изменяемая часть электронного документа (строится на основе полнотекстового анализа оригиналов документов), представляет собой набор всех слов оригиналов документов, приведенных к единственному числу, именительному падежу (для существительных), неопределенной форме (глаголов) и т. д. Является необязательной частью документа, ссылки на элементы FTIdx содержаться в OdfDoc;

CLIdx — вектор связок между электронным документом и классификаторами

$<CLIdx_1, \ldots, CLIdx_k, \ldots, CLIdx_K>$ (k = 1, K) – изменяемая часть электронного документа, так как набор связок может изменяться или дополняться. Является необязательной частью документа, ссылки на элементы CLIdx могут содержаться в OdfDoc. В простейшем случае представляет собой набор позиций классификаторов, с которыми связан архивный документ. В случае долговременного хранения данная часть документа является информацией о классифицировании и среде хранения (окружении) документа. О классификаторах и их видах было рассказано в статье [1].

Заключение

Данная статья явилась попыткой систематизировать знания и опыт, полученные при разработке архивных систем, в частности электронных архивов долговременного хранения документов для Пенсионного фонда РФ (в эксплуатации с 2001 г., сроки хранения документов до 100 лет в зависимости от типов документов), АКБ «Газпромбанк» (в эксплуатации с 1997 года, сроки хранения – десятки лет), коммерческих и государственных предприятий. Авторы систематизировали проблемы, возникающие при долговременном хранении электронных документов, с которыми неизбежно столкнется разработчик подобных систем.

Тема является весьма актуальной, поскольку электронные документы (по факту в ПФ РФ, ФНС и др.) начинают активно замещать документы бумажные, а значит, при длительных сроках хранения должна быть обеспечена их сохранность. Общая тенденция развития говорит о том, что в ближайшее время вытеснение бумажных документов станет массовым явлением, и подходы к их хранению должны быть выработаны уже сейчас.

В работе выделены проблемы, которые возникают при решении задачи долгосрочного хранения документов, приведены варианты их решения, доказана необходимость предлагаемых решений для обеспечения сохранности документов длительного хранения. Предлагаемые решения по хранению электронных документов предполагают избыточность хранения данных: хранение нескольких копий, оригиналов и переформатированных документов, наличие процедур инвентаризации носителей, ЭП и данных.

Авторы постарались показать, что современный ЭА — это не нечто неизменное, статичное, а постоянно изменяющаяся во времени структура, работая с которой необходимо регулярно обновлять носители информации, следить за действующими форматами данных, и, возможно, проводить обновление программной части ЭА.

Литература

1. Акимова Г. П., Пашкин М. А., Пашкина Е. В., Соловьев А. В. Архивные хранилища и электронные архивы документов, основные постулаты и проблемы разработки / Труды Института системного анализа РАН (ИСА РАН). Т.62. Вып.4. М.: Красанд/URSS, 2012, С. 3–13.

2. ГОСТ Р 54989–2012 /ISO TR 18492:2005 Обеспечение долговременной сохранности электронных документов.

3. Федеральный закон Российской Федерации от 6 апреля 2011 г. № 63-ФЗ «Об электронной подписи».

4. ГОСТ Р 54471–2011/ISO/TR 15801:2009 Системы электронного документооборота. Управление документацией. Информация, сохраняемая в электронном виде. Рекомендации по обеспечению достоверности и надежности.

5. ГОСТ Р ИСО 15489–1–2007 Система стандартов по информации, библиотечному и издательскому делу. Управление документами.

6. 1-ФЗ «Об электронной цифровой подписи» от 10 января 2002 г.

7. Оптические накопители Plasmon G-серии. Электронная публикация. [http://www.plasmon.ru/g-seria.shtm].

8. Наступление SSD // Журнал сетевых решений / LAN. № 11. 2010. Электронная публикация: [http://www.osp.ru/lan/2010/11/13005552/].

9. Volker Rzehak. Особенности применения FRAM микроконтроллеров Texas Instruments // Журнал РАДИОЛОЦМАН. Апрель. 2012. Электронная публикация. [http://www.rlocman.ru/review/article.html?di=113273].

10. ГОСТ Р ИСО 23081–1–2008. Процессы управления документами. Метаданные для документов.

11. ГОСТ Р 34.10–2001. Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной цифровой подписи.

12. ГОСТ Р 7.0.10–2010 (ИСО 15836:2003) «НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ. Система стандартов по информации, библиотечному и издательскому делу. НАБОР ЭЛЕМЕНТОВ МЕТАДАННЫХ ”ДУБЛИНСКОЕ ЯДРО“».

13. Типовые требования к автоматизированным системам электронного документооборота. Спецификация MoReq. Версия 5.2.1. Март. 2001. Электронная публикация [http://www.cornwell.co.uk/moreq.html].

14. Макаров С. Хранение е-документов: как угнаться за ИТ? Электронная публикация. [http://www.cnews.ru/reviews/index.shtml?2011/02/08/426535].

15. Белова А. Н., Соловьев А. В. Построение баз данных взаимосвязанных документов / Труды Института системного анализа РАН (ИСА РАН). Т. 62. Вып. 3. М.: Красанд/URSS, 2012, С. 25–30.
16. Ходаковский К. Google представила новый открытый видеостандарт. Электронная публикация [http://www.3dnews.ru/news/Googlepredstavlyaet-noviy-otkritiy-videostandart/].
17. Обзор 10 облачных хранилищ данных. Электронная публикация [http://topobzor.com/obzor-10-oblachnyx-xranilishh-dannyx/.html].
18. Резервное копирование в «Облачное хранилище». Электронная публикация [http://habrahabr.ru/company/selectel/blog/168249/].
19. Шамшина П. Ю., Шамшина Т. А. Риски информационной безопасности и аппаратно-программные средства защиты для облачных хранилищ данных. Рижский институт транспорта и связи. Латвия. Электронная публикация [http://mosi.ru/ru/conf/news/riski-informacionnoy-bezopasnosti-i-apparatno-programmnogo-sredstva-zashchity-dlya].

США: ГДЕ УЧЕНЫМ СЛЕДУЕТ ДЕРЖАТЬ СВОИ ДАННЫЕ?

Источник: Сайт издания «Хроники высшего образования» http://chronicle.com/article/Where-Should-You-Keep-Your/231065/

Федеральные органы, занимающиеся выдачей грантов, начали, наконец, выпускать официальные политики по вопросам хранения информации.

Выдающие гранты федеральные органы уже давали понять, что подаваемые запросы на гранты должны включать планы по предоставлению другим ученым доступа к научно-исследовательским данным. До сих пор этого не было, однако, ясность с тем, как и где исследователи должны хранить свои данные, в составе которых могут быть, например, и конфиденциальная персональная медицинская информация, и колоссальные массивы полученных со спутников изображений, с учетом возможности совместного использования должна быть определена.

В прошлом исследователи должны были сами заботиться о себе. Хотя ключевые органы финансирования научной деятельности уже ряд лет выставляли требования о возможности совместного использования данных, университеты не спешат помогать ведущим исследователям в выполнении этих обязательств. В результате многие держат данные на личных жестких дисках и самостоятельно отвечают на запросы.

Хорошей новостью является то, что федеральные органы начинают выпускать официальные политики – включающие рекомендации по

хранению. Плохая новость заключается в том, что в случае невыполнения требований новых политик можно лишиться возможности получения дополнительных денег по гранту.

Исследователи обычно не слишком охотно делились данными, опасаясь злоупотреблений либо боясь утратить собственные конкурентные преимущества в научной работе. В 2003 году, стремясь максимизировать использование собранных на государственные деньги данных, Национальные институты здравоохранения (National Institutes of Health, NIH) потребовали, чтобы каждое крупное предложение о финансировании включало в себя план по обмену данными.

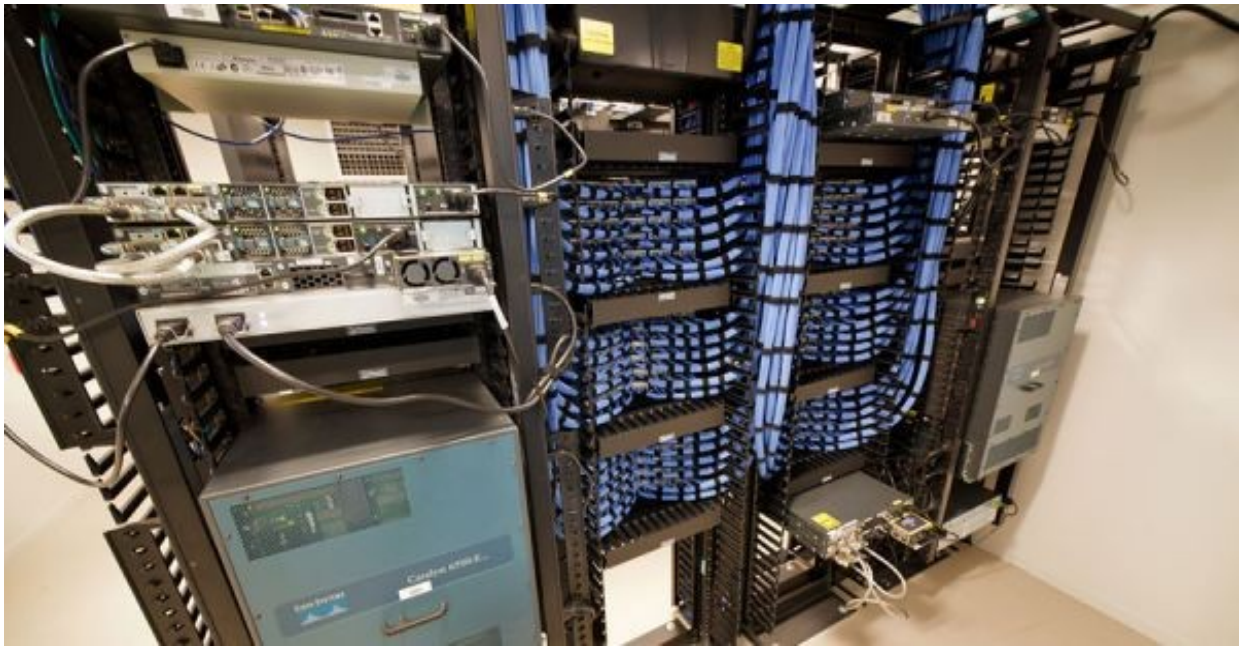

Фото: Roy Niswanger

Их примеру в 2011 году последовал Национальный научный фонд (National Science Foundation, NSF), сделав планы управления данными обязательной частью всех заявок. Большой толчок был дан в 2013 году, когда Управление Белого дома по вопросам науки и технической политики (Office of Science and Technology Policy) распорядилось, чтобы все федеральные органы исполнительной власти, тратящие на научные исследования более 100 миллионов долларов в год, разработали планы обеспечения публичной доступности полученных данных (см. https://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research).

Мало кто будет спорить с идеей сделать собранные на деньги налогоплательщиков данные широко доступными для научного анализа, однако многих исследователей раздражает необходимость составления планов управления данными, представляющих собой ещё один

административный компонент в и без того забюрократизированном процессе подачи заявок на гранты.

К счастью, директива Белого дома описывает элементы, которые должны быть включены во все политики обеспечения публичного доступа, и настоятельно призывает все федеральные органы исполнительной власти сделать такие политики совместимыми друг с другом. Политики применимы к научным данным в электронном формате, но не к лабораторным образцам и иным физическим объектам. Также из области охвата политик исключаются лабораторные журналы, данные предварительных анализов и аналогичные рабочие материалы.

Федеральные органы учитывают качество плана управления данными в составе заявки в ходе её оценки, что является мощным стимулом для ученых приложить серьезные усилия к разработке такого плана. Федеральные политики позволяют исследователям выделять часть полученных по гранту средств на управление и хранение данных.

Не все политики федеральных органов точно указывают, когда полученные в рамках работ по гранту научные данные должна стать публично доступными. Некоторые политики увязывают этот момент со временем публикации итоговой научной статьи, чтобы из исследователей преждевременно не «выдаивались» их собственные данные.

Руководители программ финансирования несут ответственность за отслеживание выполнения получателями грантов их обязательств по обеспечению коллективного использования данных. Для этого используется «кнут»: не исполняющие требований исследователи могут лишиться финансирования. Ведущим исследователям настоятельно рекомендуется размещать свои данные в существующих общедоступных электронных хранилищах. NIH составили в помощь исследователям список таких хранилищ, который планируется расширить. NSF направляет исследователей в конкретные хранилища океанографических и климатических данных.

Некоторые федеральные органы разрабатывают шаблоны, помогающие заявителям написать приемлемый план управления данными. У других имеются контрольные списки элементов, которые должны быть включены в такие планы. Исследователи должны быть бдительны: несмотря на требуемую сверху схожесть, в требованиях к планам разных органов (а иногда и различных департаментов одного органа) имеются различия, и необходимо убедиться в том, что при подготовке плана используются «правильные» руководства. Например, некоторые подразделения NSF требуют предоставления доступа только к электронным данным, в то время, как отделение океанологии NSF требует предоставления доступа также к взятым пробам.

Конечно же, некоторые данные не подлежат публичному раскрытию. Обязательные требования распространяются только на несекретные исследования, и это позволяет федеральным органам учитывать вопросы конфиденциальности и неприкосновенности частной жизни (например, в

рамках биомедицинских исследований), а также имущественные интересы и права интеллектуальной собственности.

В то же время, появилась помощь и для раздраженных следователей, которые не могут найти подходящий шаблон федерального органа. Располагающийся на сайте университета Калифорнии свободно доступный ресурс «DMP Tool» (Data Management Planning Tool - «Инструмент планирования управления данными», https://dmptool.org/) позволяет в интерактивном режиме пройти этапы подготовки плана управления данными для более чем десятка выдающих гранты организаций. Для его использования нужно создать учетную запись, но инструмент можно использовать даже в том случае, если Ваш университет не является учреждением-партнером. Инструмент предлагает ввести информацию, необходимую для заполнения плана. Что он не делает, так это не дает рекомендаций о том, как выбирать формат, хранить, распространять и обеспечивать долговременную сохранность данных. В этих вопросах Вам, возможно, сможет помочь Ваша университетская библиотека.

Попробуйте начать с веб-сайта библиотеки Вашего кампуса. У многих библиотек есть специальная страница, содержащая связанные с научными исследованиями руководства и рекомендации по различным вопросам, в том числе по подготовке планов управления данными. Поищите на сайте Вашей университетской библиотеки «план управления данными» и посмотрите, что будет найдено. По моему опыту, эти страницы могут быть полезными, но их порой непросто отыскать. Если Вы не можете сами найти нужную веб-страницу, обратитесь к сотруднику справочной группы библиотеки.

Подобная страница, вероятно, будет содержать руководство «для начинающих» по планам управления данными, ссылки на требования выдающих гранты учреждений, а также ссылки на хранилища данных по различным научным дисциплинам. Там же может найтись ценная информация о той поддержке, которую Ваше учреждение предоставляет для написания плана управления данными и для форматирования, хранения, распространения и обеспечения долговременной сохранности данных.

Многие библиотеки играют ведущую роль в этих усилиях. Для неспециалистов хранение и подготовка данных могут быстро стать проблемой, и полезно иметь под рукой кого-то знающего, кто сможет провести Вас через все тонкости наименования и форматирования данных для облегчения доступа к ним других заинтересованных лиц. Федеральные органы планируют ввести стандарты для этих так называемых «метаданных».

Исследователи также могут самостоятельно изучать хранилища данных. Два места, с которых стоит начать, это справочники Open Access Directory (http://oad.simmons.edu/oadwiki/Data_repositories) и Re3data.org (http://www.re3data.org/). их перечислены сотни хранилищ данных из разных областей знаний, от истории искусства до зоологии.

Тем исследователям, у которых до сих пор вызывает отторжение дополнительная работа, связанная с обеспечением коллективного

использования данных, следует смотреть на неё как на способ повышения своей репутации и репутации своего учреждения в научном сообществе. Вы выигрываете в престиже, когда люди начинают использовать составленные Вами наборы данных. Подготовку набора данных также можно подать как элемент ожидаемой отдачи, которую нужно описать в заявке на грант. В будущих заявках Вы сможете в своей биографии упомянуть такие наборы данных как научный продукт.

ОСОБЕННОСТИ ОЦИФРОВКИ ДОКУМЕНТОВ В СОВРЕМЕННЫХ АРХИВАХ

Источник: http://www.pcweek.ru/ecm/article/detail.php?ID=154329
Автор: Ольга Звонарева

В настоящее время все больше внимания уделяется вопросам сохранности культурных ценностей. В этой связи утверждаются государственные программы, в рамках которых, по замыслу их создателей, применение новых технологий, инновационных подходов, а также мирового опыта позволит обеспечить сохранность культурного наследия, исторически значимых документов. Поэтому сегодня архивы России осуществляют перевод бумажных документов в электронный вид. О том, что послужило началом этой масштабной работы, с какими трудностями сталкиваются архивисты и как решают вопросы сохранности оцифрованных документов, заместитель руководителя Федерального архивного агентства (Росархива) ***Олег Наумов*** *рассказал корреспонденту PC Week/RE* ***Ольге Звонаревой.***

PC Week: Расскажите, пожалуйста, как начинался процесс оцифровки архивных документов, каковы особенности этой работы?

Олег Наумов: Задача массового перевода архивных документов в электронный вид была поставлена в программе “Информационное общество (2011 – 2020)”. Это совсем непростая задача. К тому же одно дело – оцифровка документов, а другое – оцифровка научно-справочного аппарата (НСА), без которого не найти нужного документа. Поэтому, когда года два назад появилась возможность нормального финансирования работ в рамках федеральной целевой программы “Культура России 2012 – 2018”, в первую очередь стали создавать в электронном виде НСА, позволяющий эффективно искать документы.

PC Week: Какое программное обеспечение используется при этом?

О. Н.: Задолго до принятия программы “Информационное общество” Росархив начал разработку общероссийского стандартизированного ПО организации учета документов – программный комплекс “Архивный фонд”. Система, являющаяся собственностью Российской Федерации, внедрена на уровне федеральных, региональных и муниципальных архивов. С ее использованием создан центральный фондовый каталог, размещенный на портале “Архивы России”. Изначально “Архивный фонд” создавался как учетный аппарат, но потом его стали использовать как поисковый. Сейчас в ряде архивов используются специализированные поисковые системы.

PC Week: Опыт какого архива был основополагающим в деле оцифровки описей?

О. Н.: Первым, кто у нас оцифровал все описи, исключительно в силу сложившихся обстоятельств, стал Российский государственный Исторический архив в Санкт-Петербурге. Когда встал вопрос о его переезде, описи были отсканированы, и их электронные образы использовались на всех этапах перемещения, т. е. отслеживалась каждая коробка, каждое дело.

Это был хороший опыт, но с точки зрения развития информационных технологий он был не очень удачным. Хотя появилась возможность ознакомиться с ними в читальном зале и в онлайн-режиме, коллекция графических образов не позволяет проводить автоматизированный поиск. Куда большие удобства предоставляют пользователям описи, переведенные в формат базы данных, с возможностью простого и расширенного поиска. Росархив начал активно финансировать эти работы. К настоящему времени сделано порядка 20 – 30% описей федеральных архивов.

PC Week: С какими трудностями сталкиваются архивы при оцифровке документов?

О. Н. Первая проблема – огромный объем. Общий объем Архивного фонда РФ составляет 494 млн. дел, из них 9% приходится на федеральные архивы.

Но самое сложное – это определить, какие именно документы следует оцифровывать. Самый простой ответ – наиболее востребованные. А как определить эту востребованность? Сегодня востребовано одно, завтра — другое. И в советское время было введено понятие особо ценных архивных фондов и документов. На них создавался страховой фонд на микропленке, а также фонд пользования. Однако сегодня эти документы практически не востребованы. То есть критерии ценности и востребованности документов советской эпохи совершенно не работают сейчас. Спрос меняется постоянно. Определить наверняка, какие документы будут наиболее востребованными, практически невозможно.

Немаловажной является и техническая сторона вопроса. Например, возник огромный спрос на составление собственных родословных. Дело хорошее и нужное. Стали активно заказывать эти дела: метрические книги, ревизские сказки. Но они абсолютно не приспособлены для такого массового

использования. Сканировать их как наиболее востребованные – нонсенс. Из книги толщиной в 80 см человеку нужна только одна страница. А чтобы удовлетворить этот спрос, нужно сканировать всю тысячу страниц. Да еще и с оборотом. Но, с другой стороны, если отсканировать эти документы, то подлинники выводятся из оборота и гарантируется их физическая сохранность. Они все уникальные. И самое главное, несчастные хранители (среди которых немало женщин) не будут таскать на руках все эти тонны бумаг.

Однако тут мы упираемся в форму предоставления информации. Ведь пользователю куда интереснее было получить не образ, а расшифровку и БД. А массив этих документов огромен. Только в РГАДА в фонде 350 “Ландратские книги и ревизские книги” – пять с лишним тысяч дел, около 3 млн. стр. Сколько людей нужно привлечь? В Перми нашли хороший выход: привлекли финансы, которые выделялись для обеспечения рабочих мест. Работа выполнялась людьми на дому. Успешно решили проблему безработицы. И в итоге у них получилось сделать то, что больше никому в России, на моей памяти, сделать не удалось.

Имеющийся опыт сканирования в федеральных архивах позволяет сделать два вывода. Во-первых, создавать электронный фонд пользования без создания НСА – занятие малопродуктивное. Получается огромная куча сканов, где невозможно найти нужный. Во-вторых, сканированию, особенно массовому, целесообразно подвергать законченные комплексы – фонды или описи. При этом дела должны сканироваться целиком, дабы избежать конфликтов в случае ошибок в нумерации. В отдельных случаях, в силу уникальности и значимости документов, вне зависимости от того, есть на оборотной стороне листа какой-либо текст или нет, нужно оцифровывать лист, включая оборот. Тогда и вопрос о том, что мы, возможно, что-то утаили, отпадает.

Есть проблема с выбором оборудования. Где-то нужен простой сканер, где-то он должен быть сложнее. К примеру, для оцифровки ландратских книг специально заказывали “глубокую колыбель”. Иначе не скопировать, потому что толщина корешка некоторых книг достигает 80 см.

PC Week: Какова основная цель оцифровки в настоящее время?

О. Н. Их три. Первая – расширение и облегчение доступа к документам Архивного фонда. Вторая – обеспечение сохранности подлинников путем вывода их из оборота и предоставление доступа к электронному фонду пользования. Третья – упрощение предоставления государственных услуг. Сейчас появляется много индивидуальных запросов граждан на документы по личному составу. Архивы начинают их также оцифровывать и использовать электронные копии для подготовки ответов, что значительно ускоряет работу.

PC Week: А как архивы оцифровывают документы?

О. Н. Процесс идет по-разному: где-то делают сами, используя подручные средства или приобретая разнообразную технику. Второй путь –

это привлечение сторонних организаций. Но все-таки на самый главный вопрос однозначного ответа пока нет: к чему мы стремимся, оцифровывая документы?

PC Week: И никто им не задается?

О. Н. Почему, все задаются. Но как найти единственно верный ответ? Конечно, преимущество использования оцифрованных документов, очевидно. Это и, обеспечение сохранности, и удаленный доступ, и простота предоставления информации, и простота изготовления копий – полный спектр плюсов. Другое дело, как этого достичь? Вот вы пользователь, вы зашли на сайт и определили, что нужные вам документы находятся на хранении. Допустим, в Перми. Зашли на сайт по ссылке, нашли описи дел. Интересующее дело есть. И как его получить? Хорошо, если оно уже оцифровано. А если нет? Ехать самому в архив и заказывать подлинники? Или же архив должен оцифровывать все дела? Но это невозможно. Значит, по сути, остается только один вариант. Создание в каждом архиве возможности оперативного изготовления электронных копий по требованию пользователя. Причем это не обязательно может быть собственное подразделение. Это может быть и аутсорсинг. Но такая услуга должна быть оперативно предоставлена. Но изготавливать эти копии, как показывает опыт, имеет смыл только тогда, когда уже есть электронная система, которая позволит качественно и структурировано разместить эти материалы. Вот тогда это заработает. Когда мы это сделаем, и сможем ли мы это сделать, скажу честно, не знаю.

PC Week: Наверное и до утверждения программы “Информационное общество” в 2010 г. работы по оцифровке уже проводились?

О. Н. Да. Были отдельные проекты. Например, Электронный архив Коминтерна или коллекция документов СВАГ. Как правило, они реализовывались совместно с зарубежными партнерами. Собственных средств для их осуществления не хватало. Дело шло не просто. Были и технические, и методические, и организационные, и технологические сложности. Но опыт накапливался. И сейчас он очень пригодился.

PC Week: То есть уже после утверждения программы стали определять, какими должны быть основные критерии при оцифровке документов?

О. Н. Эта программа, скажем так, позволила архивистам привлечь внимание властей к тому, что проблема есть и надо ее решать. Готовить и утверждать программы по оцифровке. И привлекать под это финансы. Но на сегодняшний день нет ни одного универсального критерия. Сложно определить единый подход для всех, потому что архивы хранят совершенно разную документацию.

PC Week: То есть каждый архив фактически определял критерии сам?

О. Н. Фактически да. В принципе, это четыре критерия, они общепринятые, но не скажу, что они самые правильные. Это востребованность документов, обеспечение сохранности, облегчение работы сотрудников архива, улучшение условий обслуживания пользователей. Плюс еще ускорение работ. Если у меня на сайте есть электронный НСА, то пользователь приходит в архив подготовленным, просмотрев описи в Интернете. Если искомые документы уже оцифрованы, он имеет возможность посмотреть их у себя на компьютере, не заходя в архив, или без задержки получить в читальном зале.

PC Week: Получается, каждый архив должен сам определить, что ему оцифровывать в первую очередь?

О. Н. Совершенно верно. Это должна быть исключительно индивидуальная программа каждого архива. Нельзя сделать единую. Это зависит от многих факторов: от финансовых возможностей, от степени развития и внедрения ИТ-технологий, от конкретного спроса. Так, у кого-то очень востребованы документы по личному составу, и необходимо удовлетворять потребностям граждан. Хотя с архивной точки зрения это документы не постоянного, а временного срока хранения – 75 лет. Но это облегчает работу и помогает людям, потому что человек ждет ответа на запрос не месяц, а получает искомую информацию через два дня. Значит, надо в это вкладывать силы и средства.

PC Week: Определяя приоритеты документов, которые подлежат оцифровке, архивы согласовывают их с Росархивом?

О. Н. Если говорить о региональных архивах, то нет. Это дело каждого архива. Они присылают нам свои программы по информатизации, по оцифровке. Мы с ними знакомимся и даем свои советы. В меру своих сил им помогаем, выделяем средства ФЦП. Другое дело – федеральные архивы. Здесь позиция Росархива имеет определяющее значение. Но и тут нет универсального критерия. Для удовлетворения все возрастающего спроса на генеалогическую информацию Росархив выделил более 40 млн. руб. на перевод в электронный вид ландратских книг, хранящихся в РГАДА. Отвечая на устойчивый интерес общества к истории нашего отечества в недавнем прошлом, в июне этого года был запущен сайт “Документы советской эпохи”, где размещены образы документов личного фонда И. В. Сталина и Политбюро ЦК. Были подготовлены интернет-проекты, посвященные 1150-летию российской государственности, 400-летию окончания Смуты, 200-летию Отечественной войны 1812 года. Естественно, там были размещены образы наиболее важных и востребованных документов.

PC Week: Значит теперь в читальном зале архива будут выдавать только электронные копии, а не подлинные документы?

О. Н. В идеале нужно, чтобы человек нашел искомый документ на сайте, заказал дело и ему выдали отсканированные документы. И подлинник

остается нетронутым. Однако возникают случаи, когда необходимо посмотреть подлинник. К примеру, какой бы хорошей ни была техника, водяных знаков она не передаст. А по ним можно установить дату документа. Но таких случаев один на тысячу.

PC Week: Получается, архив предоставил пользователю государственную услугу, но после оцифрованную копию документа необходимо сохранить?

О. Н. Конечно, в этом и заключается преимущество сканирования. Раньше были микрофильмы, ксерокс, машинопись. И архивист, сделав копии, отдавал их пользователю. То есть копии документа у него не оставалось. В случае с оцифровкой копия остается. Но для нее должно быть четко определено место хранения. И если впоследствии будет второй запрос, не будет надобности поднимать дело. Нужно накопить опыт такой работы.

PC Week: Каким образом определяются сроки оцифровки массивов документов?

О. Н. Сроки зависят от трех причин: от технических и финансовых возможностей и от состояния документов. Есть документы, которые сброшюрованы так, что даже самой современной техникой получить текст полностью не удается. То есть придется их расшивать, чего очень не любят ни документы, ни архивисты. Сканировать. Потом обратно сшивать. Это требует времени. Также сроки зависят от объема дел, от количества листов в них. Не так сложно отсканировать стопку бумаг стандартного формата, а вот карту размером 3х8 м – уже сложнее. По учетным документам это один лист. Но сканируется он только частями. А после в ПО надо это все “сшивать”, подгонять и смотреть, как этот документ будет выглядеть.

PC Week: Как идет работа по реализации 89-го пункта федеральной целевой программы “Культура России (2012—2018): формирование архивных электронных ресурсов и их предоставление в сети Интернет”?

О. Н. В рамках реализации этого пункта происходит оцифровка описей. Также осуществляется оцифровка документов, в том числе ландратских книг. Причем стараемся и регионам помогать: приблизительно пять-шесть регионов в год мы включаем в эту программу и выделяем от 10 до 15 млн. руб. Создаем интернет-выставки и т. д. Всего на реализацию работ по этой программе в год выделяется 67 млн.

PC Week: Распределение средств зависит от степени важности и значимости документов?

О. Н. Да. Архивами подаются заявки в определенной форме на участие, мы эти заявки рассматриваем. Также и фирмы, которые считают, что предлагают интересные вещи, подают заявки. Мы их рассматриваем, выбираем интересные перспективные решения, составляем общий план и осуществляем закупку на конкурсной основе. В прошлом году по этому мероприятию было 52 госконтракта. То есть каждую неделю мы заключали контракт.

PC Week: Вы затронули тему уже хранящихся в архиве документов, начиная с XI века, но ведь в архивы поступают и новые бумажные документы и это огромный массив. Как проводится работа с ними? Как решается вопрос по их оцифровке?

О. Н. Для нас важно только одно: относятся ли эти документы к составу архивного фонда или не относятся. То есть подлежат они постоянному сроку хранения или нет. А дальше с точки зрения использования абсолютно не важно, документы ли это XI или XXI века. В этом плане они равны. Другое дело, что на документы XI века у нас, естественно, есть страховой фонд и фонд пользования. И эти уникальные документы крайне редко выдаются на руки. Имеется специальный Государственный реестр уникальных документов РФ, который размещен на сайте Росархива и содержит их описания и электронные образы.

PC Week: В этой связи наверняка проще принять на хранение документы, переведенные в электронный формат? К примеру, вступили в силу изменения в законодательстве относительно кадровой документации, и архивы, наверное, могут принимать ее в электронном виде?

О. Н. В принципе да. Но особого смысла государственному архиву принимать современные документы по личному составу сейчас нет. Зачем? Рассчитывать пенсию? Есть Пенсионный фонд, у которого налажен специализированный учет граждан, начиная с 2000 г.

PC Week: То есть архивы не принимают документацию, хранящуюся 75 лет?

О. Н. Нет, она у нас хранится с давних времен. И сейчас мы активно ведем переговоры с тем же Пенсионным фондом, чтобы нам этот срок для бумажных документов сократили с 75 до 15 лет. Но Пенсионный фонд боится, что их электронная система может рухнуть.

А те документы, которые со временем будут поступать к нам в электронном виде – проблема очень большая и совершенно отдельная. Это вопросы, касающиеся электронного **документооборота**, электронных документов. Как их принимать, как хранить, это отдельная тема.

PC Week: Пока еще такие документы не поступают?

О. Н. Нет, поступают. Материалы переписи населения и сельскохозяйственной переписи ГАРФ принял в электронном виде. Другое дело, что эти переписи только в электронном виде и существуют. Но массового приема делопроизводственной документации пока нет. Тут довольно много нерешенных вопросов.

PC Week: А как быть с документами ликвидированных организаций? Их принимают на хранение?

О. Н. Принимаем. Правда, конкурсные управляющие обязаны в случае банкротства и ликвидации компании найти деньги на описание документов ликвидированных организаций, но не всегда находят.

PC Week: Под руководством Росархива находятся 15 федеральных архивов. Региональные архивы в настоящий момент выведены из-под контроля вашего ведомства?

О. Н. Да, прямого контроля над ними нет. Они подчиняются исполнительным органам субъектов федерации. Где-то это самостоятельные архивные службы. Где-то они, как и мы, входят в министерства культуры. Где-то – в министерства юстиции или непосредственно в аппарат губернатора. Но той вертикали власти, которая была в Советском Союзе и просуществовала до 2004 г., к сожалению, уже не существует. Пока мы сохраняем методическое единство. Чем можем, помогаем региональным архивам. Представители Росархива принимают участие в ежегодно проводимых в каждом федеральном округе научно-методических советах. Раз в год проводится Совет по архивному делу, где собираются архивисты со всей страны. То есть связи остаются. Мы шутим, что архивист – это такая профессия, при которой в любом городе можно оказаться без денег и вещей, прийти к коллегам архивистам, и они тебя примут, накормят и отправят домой. Есть такое братство, которое, на самом деле, не свойственно ни одной другой профессии.

PC Week: Спасибо за беседу.

ЗМІСТ