

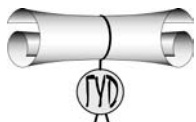
Оцифровка архивных документов в центральных архивах Москвы

*Тихонов Владимир Иванович, к.и.н., заместитель
директора Информационного центра Главного
архивного управления города Москвы*

Главное архивное управление города Москвы (Главархив Москвы) – уникальное в Российской Федерации архивное учреждение. В рамках единой организации оно совмещает функции органа исполнительной власти города Москвы в сфере архивного хранения документов с функциями их непосредственного хранения и использования. Кроме административных подразделений Главархив Москвы включает в себя семь центральных архивов Москвы всех возможных профилей¹, хранящих в составе своих фондов почти 12,5 млн архивных дел, а также функциональные центры, специалисты которых обеспечивают выполнение работ, требующих высококвалифицированных знаний и навыков – Информационный центр, Выставочный центр, Центр публикации и использования архивного фонда, Центр микрографии и реставрации документов, Центр документационного обслуживания населения и др. По сути, структура и функции Главархив Москвы сближают его со многими зарубежными национальными архивами.

Основная цель работы государственных и национальных архивов заключается в обеспечении граждан, общества и органов власти ретроспективной документированной информацией. На это направлена деятельность как самих архивов, так и органов власти, управляющих архивной сферой. Одним из эффективных способов, повышающих оперативность и качество обслуживания пользователей архивными документами, является использование в работе архивов информационно-коммуникационных технологий. Главархив Москвы, находясь на острие информационного взаимодействия с жителями Москвы, уже многие годы развивает информационные системы, способствующие повышению эффективности работы своих подразделений.

¹ Это – Центральный исторический архив Москвы (ЦИАМ), Центральный архив города Москвы (ЦАГМ), Центральный архив общественно-политической истории Москвы (ЦАОПИМ), Центральный архив научно-технической документации Москвы (ЦАНТДМ), Центральный архив электронных и аудиовизуальных документов Москвы (ЦЭ и АДМ), Центральный московский архив-музей личных собраний (ЦМАМЛС), Центральный архив документов о трудовой деятельности граждан Москвы (ЦАДТДГМ).

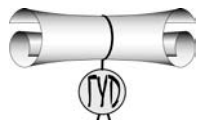


В последние десятилетия, стремительное развитие информационных технологий открыло для архивной отрасли некоторые совершенно неожиданные для нее перспективы. Сканирование и оцифровка документов оказались вдруг наиболее технологичными способами копирования и распространения архивной информации, а включение электронных образов документов в информационно-поисковые системы превратило их в наиболее продвинутые средства работы с фондом пользования архивными материалами. В последние десять лет в России не осталось, наверное, ни одного архива, который не попытался бы поставить технологии оцифровки на пользу обществу и государству.

В Главархиве Москвы первые опыты по оцифровке документов и созданию электронного фонда пользования начались в Центральном архиве аудиовизуальных документов Москвы (ЦААДМ). Это был естественный и закономерный процесс, так как оцифровка была признана практически единственной технологией, направленной на обеспечение сохранности аудиовизуальной информации и создания копий их страхового фонда. С 2004 г. в ЦААДМ проводят оцифровку фонодокументов, хранящихся на магнитной ленте. В настоящее время оцифровано уже 1004 ед. хр. из общего фонда в 4629 ед. хр. (22 %) В качестве кодека используется РСМ (Pulse Code Modulation), преобразующий звуковые сигналы в цифровой формат несжатого (т.е. без потери качества) звука. Оцифровка проводится с частотой дискретизации 44,1 КГц и разрядностью 16 бит. До 2011 г. запись фонда пользования проводилась на оптические диски CD. В настоящее время используются диски DVD, причем, каждый фонодокумент записывается на два отдельных диска.

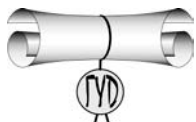
Планомерную оцифровку фотодокументов (негативов, позитивов и слайдов) в ЦААДМ начали с 2005 г. Сканирование производится с разрешением 2000 точек на дюйм в страховом формате TIFF, после чего создается дополнительная копия в пользовательском формате JPEG. К концу 2012 г. оцифровано более 11 тыс. фотодокументов, что составляет 4 % всего фонда архива. Файлы хранятся в трех экземплярах: на двух отдельных оптических дисках (TIFF-файлы на DVD, JPEG-файлы на CD), а также на файл-сервере, интегрированном с автоматизированной информационно-поисковой системой (АИПС) архива. Таким образом, пользователь системы может не только искать фотодокументы по атрибутам описания, но и просматривать их электронные образы.

Но наиболее продвинутые технологии оцифровки и электронного фонда пользования нашли применение в так называемом Видеоархивном комплексе



Главархива Москвы, который был создан в соответствии с постановлением Правительства Москвы от 18.01.2005 № 27-ПП и распоряжением Правительства Москвы от 18.07.2006 № 1381-РП. В 2007–2008 гг. велась подготовка и реализация проекта, опытная эксплуатация оборудования. С января 2009 г. Видеоархивный комплекс был введен в промышленную эксплуатацию, которую осуществляют специалисты отдела автоматизированных технологий по работе с аудиовизуальными документами ЦАЭ и АДМ². Без преувеличения можно сказать, что этот комплекс является одним из самых больших и самых сложных хранилищ цифровой аудиовизуальной информации в архивных учреждениях не только России, но и Европы. Он оснащен мощными серверами, дисковыми накопителями и ленточными библиотеками производства IBM, которые вмещают 1927 картриджей с лентами LTO-3 совокупной емкостью в 750 терабайт. Это позволит пользователям получать оперативный доступ к более чем 10 тыс. часов видеодокументов. Катастрофоустойчивость комплекса обеспечивается размещением однотипного оборудования на разных территориях Главархива Москвы, которые объединены выделенным каналом связи в 100 Мбит/с. Средства четырех видеомонтажных станций способны конвертировать, импортировать и экспортировать видеодокументы в самые разнообразные цифровые и аналоговые форматы. Но основными форматами хранения в ленточной библиотеке являются MPEG-2 (с битрейтом до 50 Мбит/с), DV (до 50,3 Мбит/с) и Uncompressed (до 175,9 Мбит/с). Помимо размещения в ленточной библиотеке, видеодокументы, после соответствующего преобразования, записывают на внешние носители (DVcam и DVD), которые передаются в традиционное архивохранилище и рассматриваются в качестве страхового фонда.

Таким образом, оцифровка видеодокументов ЦАЭ и АДМ преследует несколько целей. Во-первых, это – создание страховых копий в цифровых форматах, так как состояние видеопленок Betacam и VHS уже давно вызывает большие опасения. Во-вторых, создание автоматизированного фонда пользования. АИПС Видеоархивного комплекса позволяет вести расширенный поиск видеодокументов по любым атрибутам, но главное, с ее помощью на любое рабочее место в Главархиве Москвы можно «вызвать» и просмотреть цифровые копии видеодокументов в формате MPEG-2 низкого разрешения (1 Мбит/с). В-третьих, оборудование комплекса и хранящиеся в ленточной библиотеке полные цифровые копии видеодокументов предназначены для создания их копий по

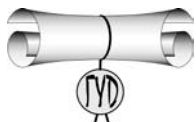


запросам пользователей без дополнительного использования оригиналов, что должно благотворно сказаться на обеспечении их сохранности. При этом, по желанию пользователя возможно проведение монтажа видеороликов из фрагментов разных документов, а также реставрация изображения и звука. Наконец, в Видеоархивный комплекс загружают новые поступления видеодокументов в архив, причем эти поступления – уже изначально в цифровых форматах. Так, что ленточные библиотеки используются, в том числе, как временные хранилища для документов, до проведения более тщательной экспертизы их ценности и последующего отбора на постоянное хранение. Остается добавить, что в настоящее время оцифровано и загружено в Видеоархивный комплекс около 180 часов архивных видеодокументов (12,5 % стоящих на учете) и несколько сотен часов новых поступлений. На повестку дня выдвигаются задачи по приобретению и внедрению в ЦАЭ и АДМ кодеков с дополнительными форматами хранения видеодокументов MPEG-4 и MXF, а также технологий оцифровки кинодокументов и фонодокументов на виниловой основе.

Не менее насущной для Главархива Москвы является задача оцифровки и создания электронного фонда пользования архивными документами на бумажной основе, которая в настоящее время решается путем привлечения сторонних организаций. В 2008 г. в соответствии с распоряжением Правительства Москвы от 11.04.2008 № 753-РП, в органах исполнительной власти Москвы предстояло реализовывать проект по созданию электронных архивов, ядром которого являлось сканирование архивных документов. Управлением информатизации города Москвы был объявлен соответствующий конкурс, победителем которого стало ЗАО «ПроСофт-М». В августе 2008 г. наступила очередь Главархива Москвы предоставить исполнителю по государственному контракту архивные документы для перевода их в оцифрованную форму.

Известно, что оцифровка архивных материалов – процесс весьма трудоемкий, а потому дорогостоящий. С самого начала проекта было ясно, что перевести в цифровой вид удастся лишь несколько тысяч архивных дел. Поэтому перед Главархивом Москвы встали две взаимосвязанные задачи: определить состав документов, подлежащих оцифровке, и способ использования подготавливаемого информационного ресурса. При решении первой задачи выбор был остановлен на фондах райисполкомов Москвы (ЦАГМ), документы которых за 1950–1980-е годы имеют важное социальное значение и по

² В феврале 2008 г. ЦААДМ и Центральный архив документов на электронных носителях Москвы (ЦАДЭНМ)

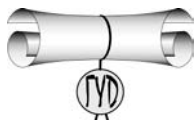


которым ежегодно исполняется почти треть из 60 тыс. социально-правовых запросов, поступающих в службу «одного окна» Главархива Москвы. Решение второй задачи виделось во включении электронных копий документов в подсистему «Научно-справочный аппарата» Интегрированной автоматизированной информационной системы (ИАИС) Главархива Москвы. Это означало не только модернизацию системы, но и проведение индексации файлов с оцифрованными образами. В частности, в систему должны были поступать сведения обо всех персоналиях и адресах зданий, которые встречались в текстах решений райисполкомов. Забегая вперед, стоит отметить, что такой подход увеличил трудозатраты при создании информационного ресурса, однако, его эффективность сразу же сказалась на существенном повышении оперативности исполнения запросов и на улучшении ситуации с обеспечением сохранности используемых архивных документов. В настоящее время приказом Главархива Москвы, при возможности выводить копии документов из ИАИС, запрещено выдавать дела из архивохранилищ.

Оцифровка документов ЦАГМ проводилась на территории архива, но на оборудовании и специалистами ЗАО «ПроСофт-М». Для этого в здании архива были оборудованы 10 рабочих мест с планетарными сканерами (9 сканеров для оцифровки документов форматом до А2 и один сканер – для формата А1). Кроме того, были организованы еще 6 рабочих мест для разбрюшеровки и последующей сбрюшеровки сканированных архивных дел. Тестовое сканирование архивных листов на сканерах с автоматической протяжкой листов показало, что данная технология негативно влияет на физическую сохранность архивных документов. В связи с этим от использования поточных сканеров отказались с самого начала.

Сканирование осуществлялось с разрешением в 300 точек на дюйм, с сохранением результирующих файлов в форме TIFF. Выборочный контроль качества оцифровки проводили как работники ЗАО «ПроСофт-М», и архивисты. При хорошем качестве цифруемого материала сканирование проводилось в «бинарном режиме». Однако в половине случаев пожелтевшая бумага документов и нечеткость машинописи потребовали проводить сканирование в «грациях серого», что значительно увеличивало объем файлов с оцифрованными образами документов. Так что, в зависимости от исходных документов (формат которых мог достигать размера А0 и больше), объем файла мог достигать 50-60 Мбайт.

были слиты в единый Центральный архив электронных и аудиовизуальных документов Москвы (ЦАЭ и АДМ).

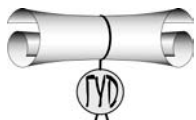


По условиям государственного контракта, оцифрованный материал исполнитель перевозил на свою территорию, где силами собственных сотрудников осуществлялась его индексация. Описание оцифрованных образов проводилось на уровне каждого документа (распоряжения райисполкома) и по согласованным с Главархивом Москвы правилам, основанным на «Методических рекомендациях по описанию и классификации документной информации для создания АИПС "Организационно-распорядительные документы Московского городского совета и Исполнительного совета народных депутатов за 1931–1991 гг.»» (Мосгорархив, 2003).

Собственно оцифровка архивных документов ЦАГМ заняла около четырех месяцев. А вот этап индексации оказался наиболее продолжительным и растянулся почти на полгода. Результаты многомесячной работы передавались Главархиву Москвы постепенно, по мере готовности. Данный информационный ресурс представлял собой электронные копии документов, записанные на диски DVD в двух экземплярах. Каждая партия переданных DVD-дисков сопровождалась CD-дисками, содержащими базу данных с индексами и описанием архивных документов. Передача осуществлялась по описям дисков с электронными копиями архивных документов, которые включали следующие статьи описания:

- номер диска по описи;
- идентификация, шифр диска (включавший номер описи, номер диска по описи и обозначение его экземпляжности – с рабочими или резервными копиями архивных документов);
- номера архивного фонда, описи, архивных дел, копии документов которых записаны на диск);
- количество оцифрованных дел, копии документов которых записаны на диск;
- количество электронных копий архивных документов, записанных на диск;
- количество компьютерных папок, составляющих электронный массив;
- количество компьютерных файлов, составляющих электронный массив;
- общий объем файлов в мегабайтах и байтах
- дата записи электронного массива на диск.

В итоговую запись к описи вносилось количество: дисков, электронных копий архивных документов, оцифрованных дел, компьютерных папок и файлов.

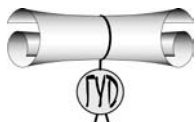


После приема носителей с оцифрованными образами документов наступал следующий этап работы по подготовке информационного ресурса к использованию. В Информационном центре Главархива Москвы с помощью специализированного программного проводилась проверка технического состояния всех поступивших дисков. Практика показала, что на проверку одного DVD-диска с записью нескольких сотен графических файлов необходимо 30-40 минут. В результате многомесячной работы было выявлено около 2% дисков, оказавшихся в неудовлетворительном состоянии, которые исполнителю пришлось заменить. При этом нередко отмечалась выбраковка до десятка дисков подряд (принадлежащих к одной серийной партии), что подтверждает одно из базовых правил подготовки электронных документов к долговременному хранению – запись рабочих и резервных экземпляров документов на электронные носители разных фирм-производителей.

Завершающим этапом формирования архивного информационного ресурса был импорт электронных копий документов и их метаданных (индексной информации) в ИАИС Главархива Москвы. На этой стадии стандартными средствами СУБД выявлялись и исправлялись ошибки в индексации и в привязке оцифрованных образов к архивному шифру документов. Окончательное мнение о полноте и качестве электронной картотеки по документами райисполкомов Москвы должны были высказать ее конечные пользователи – работники ЦАГМ и управлений Главархива Москвы, которые приступили к ее эксплуатации и, в целом, остались довольны ее поисковыми возможностями.

В 2011–2012 гг. работы по оцифровке архивных документов были продолжены. На этот раз, кроме райисполкомов выбор пал также на фонд Мосгорисполкома (Моссовета): на дела, содержащие постановления, распоряжения, протоколы и другие распорядительные и нормативные документы. В итоге в 2008, 2011 и 2012 гг. были оцифрованы 8126 дел по фондам шести райисполкомов (из 30), совокупным объемом более 1,6 млн страниц. То обстоятельство, что к 2011 г. в Главархиве Москвы была полностью завершена электронная картотека «Решения и распоряжения Мосгорисполкома», сократило объемы работ по индексации документов Мосгорисполкома и позволило довести этот оцифрованный массив до 8712 дел (более 2,5 млн страниц).

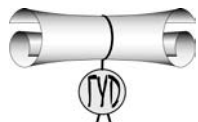
С 2011 г. изменились форматы записи оцифрованных образов и электронные носители, на которых они принимались от исполнителя (ЗАО «Электронный архив»). Теперь использовался формат JPEG и внешние жесткие диски большой емкости. В настоящее



время файлы с электронными копиями документов ЦАГМ хранятся в Информационном центре Главархива Москвы в трех экземплярах: в двух экземплярах на 2792 дисках DVD (райисполкомы) и 24 внешних жестких дисках (Мосгорисполком и райисполкомы) и один экземпляр на 238 дисках BD-R в роботизированном дисковом накопителе, интегрированном с ИАИС. Совокупный объем сохраняемых оцифрованных образов, таким образом, составляет почти 25 Тбайт данных.

Конечно, 17 тыс. оцифрованных дел – это слишком незначительный процент от всего объема Архивного фонда города Москвы. До 2016 г. планируется повести в оцифрованный вид еще более 100 тыс. дел, в том числе дела с научно-технической документацией (ЦАНТДМ). Но и этого окажется недостаточно для вывода из регулярного использования всех социально и культурно значимых архивных документов. По нашим оценкам, при сохранении существующих темпов оцифровки (читай, существующих объемов финансирования этих работ) потребуется не менее сотни лет на то, чтобы перевести в электронную среду лишь наиболее востребуемую часть фондов центральных архивов Москвы.

Впрочем, оцифровка архивных документов – это только начальный этап создания полноценных систем доступа к архивной информации. Одновременно с этим необходимо решать не менее сложную и дорогостоящую задачу сохранения оцифрованного контента. И проблема здесь не только в выборе надежных и емких систем хранения сотен терабайт данных (выбор аппаратных и программных компонентов, электронных носителей), но и в последующих, неоднократных переводах информационно-поисковых систем на новые технологические платформы. Не только практика, но методология хранения электронных (оцифрованных) документов в нашей стране находится в зачаточном состоянии. Цели и задачи электронных архивов, создаваемых в организациях и органах власти, определяются, как правило, на ближайшую перспективу, без учета требований долговременного хранения документов. Да, и эти требования никто не торопится формулировать, не говоря уже об их нормативном оформлении. Такой подход уже в недалеком будущем, через 15-20 лет, грозит практически полным исчезновением баз данных с документами, вышедшими из оперативного использования. Повторится судьба электронных информационных ресурсов 80–90-х годов прошлого века. Те же немногие ценные базы данных, в том числе с оцифрованными архивными документами, ждет дорогостоящий перевод на более современные платформы. Однако издержки на



проведение миграции данных можно было бы снизить, если уже сегодня при создании таких информационных систем ориентироваться на широко распространенное программное обеспечение и внутреннюю организацию баз данных, облегчающую экспорт документов. Таким образом, без стандартизации и унификации принципов и подходов к созданию информационных систем с ценными электронными документами, на государственном ли уровне или на уровне профессиональных ассоциаций, будет практически невозможно обеспечить сколь-нибудь длительное сохранение цифрового богатства современной России.