

## Процесс оцифровки документов для создания электронного архива

Статья посвящена процессу оцифровки – наиболее удобному механизму сохранения и использования информации в современных условиях. Автор рассматривает основные моменты процесса оцифровки архивных материалов, включая анализ возможных способов перевода документов в электронный вид, категории документов для оцифровки, форматы хранения электронных документов и исследование современных стандартов оцифровки. Также в статье анализируются необходимые условия для создания виртуального архива исторической персоналии или события в современном контексте. Особое внимание автор уделяет особенностям организации электронного ресурса, исследует существующие современные проекты по данной тематике и анализирует возможные трудности, с которыми может столкнуться исследователь при создании электронного архива.

*Ключевые слова:* оцифровка, архивный документ, электронный фонд пользования, виртуальный архив.

Реалии современной жизни все чаще требуют от нас активного использования новых информационных технологий. Информационная эпоха вносит свои коррективы также и в мир диалога архивных источников и исследователя. Современные цифровые технологии позволяют осуществлять самые глобальные идеи по созданию альтернативного поля хранения – хранения документов в виртуальной плоскости.

Оцифровка архивных документов на сегодняшний день является наиболее удобным инструментом для сохранения и использования информации. Она дает возможность для безопасного и более экономичного хранения, удобного поиска и использования

информации, а также для оперативного доступа к материалам архива. Процесс оцифровки документов основывается на идее формирования единого электронного фонда пользования (далее – ЭФП). ЭФП представляет собой совокупность электронных копий архивных документов, которые записаны на цифровые носители и предназначены для использования вместо подлинников документов<sup>1</sup>. Важным моментом при создании ЭФП является необходимость регламентации процесса его создания.

По мнению заместителя руководителя Федерального архивного агентства О.В. Наумова, основные цели оцифровки документов заключаются в расширении и упрощении доступа к документам архивного фонда, обеспечении сохранности подлинников путем постепенного вывода их из оборота использования, предоставлении доступа к электронному фонду, а также в ускорении процесса предоставления государственных услуг<sup>2</sup>.

Рассмотрим основные моменты процесса оцифровки. Прежде всего, необходимо учитывать, что архивный фонд Российской Федерации хранит обширный документальный массив. Таким образом, прежде чем начать сканирование самих документов, целесообразно провести оцифровку научно-справочного аппарата архива, создать электронные описи для поиска данных. Первоначальная оцифровка описей архива дает возможность пользователю не выходя из дома познакомиться с перечнем хранящихся документов и заказать интересующие документы по Интернету. При этом необходимо также учитывать факт научно-технической редакции заголовков дел – только если опись усовершенствована, можно начинать оцифровку фонда.

Категории документов для оцифровки включают в себя не только документы, для которых существует угроза утраты информации, но также наиболее ценные, уникальные материалы и востребованные документы. Критерии документов об уникальности и ценности документов несколько размыты, а востребованность не является постоянной величиной, поэтому при выборе материалов оцифровки часто возникает вопрос: какие документы необходимо переводить в цифровой вид в первую очередь? Возможно, добиться более результативной выборки будет проще, если в процессе отбора примут участие не только члены экспертной комиссии архива, но также и привлеченные профессионалы – историки, политологи, социологи, общественные деятели и др. Процесс создания такой рабочей группы требует четкой регламентации. Следует отметить, что, несмотря на сформулированные критерии отбора<sup>3</sup>, они не являются обязательными и региональные архивы

вправе сами выбирать те документы, которые нужно оцифровать в первую очередь.

Выбор способа оцифровки важен, так как носитель и формат представления материала также несет полезную для работы исследователя информацию, поэтому очень важно передать ее в наиболее точной, сопоставимой с подлинником форме.

В непосредственной оцифровке документов важно соблюдать все меры безопасности при работе с подлинником, особенное внимание необходимо уделить при сканировании книг и древних актов (свет, принтер, поведение специалиста при оцифровке должны соответствовать принятым нормам). В данном этапе ключевую роль играют финансовые возможности проекта по оцифровке – выбор оборудования для сканирования определяет степень безопасности этого процесса для подлинника. Выбор более экономичного аппаратного устройства неизбежно влечет за собой риск нанести непоправимый вред документам, а также сокращает шансы на создание сопоставимой, полноцветно заменяющей подлинник копии<sup>4</sup>.

При оцифровке документа делается минимум две копии подлинника – рабочая копия и мастер-копия. Оба материала должны быть замаркированы и зарегистрированы в специальном реестре. Благодаря такой системе регистрации поиск отсканированного документа будет более быстрым и комфортным для пользователя, а учетные данные позволят легко отслеживать документ в общей информационной системе архива.

Среди форматов хранения изображений наибольшее распространение у архивистов получили форматы TIFF и JPEG. Формат TIFF своей популярностью обязан возможности сохранения качества изображения благодаря алгоритмам сжатия данных без потерь. Формат JPEG обладает возможностью высокой степени сжатия, но при этом происходит потеря качества изображения. Изменения могут быть незаметны невооруженным взглядом, но сжатое изображение будет отличаться резкими контрастами или пикселями. Поэтому JPEG нельзя использовать как промежуточный формат при обработке изображений. В JPEG допустимо сохранять лишь окончательный вариант. Что касается других разновидностей электронных документов и наиболее предпочтительных форматов их хранения, то существуют текстовые документы (ms-word, txt, pdf, html, xml, rtf), видео (mov, avi, mpeg, mp4) и аудио (wav, aiff, mp3), чертежи (autocad), графики и схемы (tiff, psx), БД в виде электронных таблиц и реляционных баз данных (xls, xml, html, mdb). Для фотодокументов также приемлем формат pdf. Полученный массив данных может использоваться для передачи по цифровым каналам,

сохраняться на цифровом оптическом носителе (однократно записываемые CD-R, DVD-R, многократно записываемые CD-RW, DVD-RW, DVD-RAW диски), магнитном (жесткие диски, гибкие магнитные диски, магнитные ленты), Flash-накопителях и т. д.

Требования к качеству получаемых цифровых копий четко не сформулированы. Отсутствуют критерии оценки качества электронных копий бумажных документов: параметры изображения, тоновое воспроизведение (контрастность), яркость, шум, точность цветопередачи, резкость, разрешение, геометрические искажения и др. За рубежом уже используется ряд документов, опыт которых в области оценки качества оцифрованных документов был бы полезен для России. Национальный стандарт США ANSI/AIIM MS44 «Руководящие указания для контроля качества сканеров изображений» устанавливает и раскрывает базовую терминологию, основные параметры и критерии оценки качества сканирования, а также подходы к их практическому измерению. На его основе с учетом современных требований в 2000 г. были разработаны международные стандарты ISO для оценки качества сканирования черно-белых документов.

После оцифровки подлинник возвращается обратно на хранение в архив, а сделанные копии пополняют электронный фонд пользователя и становятся доступными для исследователей.

Поиск решения проблемы хранения и оперативного использования архивных документов начался еще в середине 1990-х гг., когда Библиотека Конгресса США приступила к тотальной оцифровке имеющегося собрания микрофильмов (как самостоятельной коллекции и как основного носителя страхового и пользовательского фондов). В процессе оцифровки американские специалисты столкнулись с необходимостью выработки единых подходов к реализации процессов перевода информации с материальных носителей в электронный вид, т. е. возникла необходимость в регламентации процессов оцифровки.

Современные стандарты оцифровки по уровню их регламентации можно разделить на 3 категории стандартов: международный, национальный и стандарт организации. Международный стандарт ISO<sup>5</sup> разрабатывается группой ученых, его использование дает технологическое, экономическое и социальное преимущества, но не является обязательным ни для одной страны – участницы организации ISO. Национальный стандарт является обязательным для использования органами государственной власти различных уровней, он разрабатывается для улучшения процедуры оцифровки с учетом особенностей законодательства, стандартов документа-

ции в определенной стране. Стандарт организации подчеркивает специфику конкретной компании в процессе перевода документа в электронный вид и его последующее хранение и использование. Наиболее известным национальным стандартом, регламентирующим процессы оцифровки, является «Стандарт оцифровки S6» (S6: Digitisation Standard)<sup>6</sup>, введенный в действие национальным комитетом по стандартизации Австралии и Новой Зеландии в 2006 г.<sup>7</sup> К сожалению, в России нет аналогичного документа, регламентирующего процессы оцифровки и создания ЭФП.

В 2012 г. сотрудниками *Всероссийского научно-исследовательского института документоведения и архивного дела (ВНИИДАД)* и Федерального архивного агентства (Росархив) были разработаны «Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом». Моделью для этих рекомендаций стали «Стандарт оцифровки S6» и «Технические рекомендации» FADGI. В целях продолжения работ в данном направлении представляется целесообразным выработать Стратегию разработки и обновления отраслевых нормативно-методических актов, регламентирующих различные направления деятельности в области информатизации архивного дела, а также создать регламент и перспективный план по ее реализации. Разработчиками были выделены ключевые моменты, которые необходимо учитывать при подготовке регламентирующего документа.

1. Структура и содержание разработанных и введенных в действие документов международных и национальных систем стандартизации обусловлены спецификой правовой базы той страны, где они возникли. Поэтому их прямой перевод и использование в Российской Федерации требуют взвешенного подхода.

2. Необходимо создание не одного нормативного документа, регламентирующего вопросы оцифровки, а комплекса нормативно-правовых актов, описывающих требования ко всем этапам и аспектам этого процесса.

3. Структура и общее содержание этой нормативно-методической документации могут быть построены на зарубежных аналогах, выбранных экспертами отрасли в качестве примеров, наиболее близких и адекватных российским условиям. При этом отечественные разработки должны в полном объеме отражать специфику всех процессов и аспектов оцифровки (включая требование к оборудованию, персоналу, процедурам и менеджменту качества), присущую России<sup>8</sup>.

Оцифровка документов – необходимая мера для организации оперативного доступа к документам архива. Наличие определенно-

го массива документов позволяет создать новый, комфортный для пользователя, научно полезный информационный ресурс – виртуальный архив. Примеры таких виртуальных хранилищ можно увидеть на сайте Российского государственного архива литературы и искусства.

Виртуальный архив И.А. Бунина представляет собой ресурс с удобной классификацией представляемых документов и комфортным для пользователя интерфейсом. Документы разделены на три группы: «Рукописи», «Вырезки из газет и журналов», «Изобразительные материалы» и представляют собой сканированные копии подлинников фондов Российского государственного архива литературы и искусства и Архива русского зарубежья в г. Лидсе<sup>9</sup>. Разработчики проекта обозначают ключевые моменты по реализации подобных проектов: финансовые вопросы, проблема разобщенности архивных документов и необходимость переговоров с хранилищами разных учреждений и даже стран для сбора необходимого материала и, наконец, решение вопросов юридического характера – так как на электронную публикацию материалов также распространяется авторское право. Ресурс представляет собой положительный пример реализации проекта по созданию виртуального архива. Сайт Российского государственного архива литературы и искусства также предлагает ознакомиться с такими прекрасными примерами электронных хранилищ, как ресурс, посвященный документам Отечественной войны 1812 г. и Первой мировой войны.

Создателями электронного хранилища под названием «Воссоединенный виртуальный архив Осипа Манделъштама» стали Оксфордский университет и Манделъштамовское общество. Разработчики проекта ставили целью выявить, описать и разместить в Интернете все или максимально большое количество сохранившихся творческих и биографических материалов Осипа Манделъштама независимо от места их физического расположения<sup>10</sup>. При этом в проекте объединяются рукописи, транскрипты текстов, а также комментарии к ним.

Поиски документов – важнейший этап при создании такого рода проектов. Этот процесс осложняется тем, что некоторых из ранее известных и введенных в научный оборот коллекций, возможно, более не существует. Попав в продажу в 1990-е гг., они меняли владельцев, даже имена которых не всегда возможно установить. В иных случаях владельцы оказываются не в состоянии обнаружить те или иные автографы или документы. Это вполне объясняется тем, что комплектование архивно-рукописных фондов документами

деятелей русского зарубежья осуществлялось в основном путем дарения и не носило систематического характера<sup>11</sup>. Разработчики отмечают, что то же самое случается и с государственными архивами. Так, в Национальном архиве Франции на протяжении ряда лет не могли найти матрикуляционные документы О.Э. Мандельштама (они были заново обнаружены в апреле 2008 г.)<sup>12</sup>.

Важно отметить, что виртуальный архив О.Э. Мандельштама не только являет собой удачный пример реализации подобного проекта, но также, благодаря подробному описанию процесса создания такого архива на сайте, представляет своеобразное учебное пособие для продолжателей и исследователей процесса оцифровки.

Анализ реализованных российских проектов по созданию виртуальных хранилищ исторических материалов подчеркивает проблему разобщенности документов и сложность их поиска. Таким образом, при подготовке проекта важно акцентировать внимание на возможном сотрудничестве с зарубежными архивами.

Таким образом, возрастает роль совместных работ, совместных проектов по созданию единого виртуального поля хранения материалов одной тематики.

При оцифровке необходимо уделять внимание процессу выборки документов для перевода их в цифровой вид, а также подбору качественного сканирующего оборудования для обеспечения безопасного для подлинников процесса оцифровки. Однако на данный момент одним из самых серьезных вопросов в этой тематике остается необходимость регламентации процесса оцифровки (включая подробное описание процесса выборки материалов и закрепление терминологической базы). Без соответствующего, утвержденного законом стандарта процесс создания виртуального архива остается трудоемким и недостижимым для большинства исследователей.

#### Примечания

<sup>1</sup> Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом. [Электронный ресурс] URL: [http://archives.ru/documents/rekomend\\_el-copy-archival-documents/razdel-2.shtml](http://archives.ru/documents/rekomend_el-copy-archival-documents/razdel-2.shtml) (дата обращения: 13.05.2016).

<sup>2</sup> Особенности оцифровки документов в современных архивах. [Электронный ресурс] URL: <https://www.pcweek.ru/ecm/article/detail.php?ID=154329> (дата обращения: 13.05.2016).

<sup>3</sup> Методические рекомендации по электронному копированию...

- <sup>4</sup> Юмашева Ю.Ю. Архивы и «цифровая гонка вооружений» // Историческая информатика. 2013. № 3. С. 93.
- <sup>5</sup> ИСО – Международная организация по стандартизации. Разработчик и издатель международных стандартов. [Электронный ресурс] URL: <http://www.iso.org/iso/ru/> (дата обращения: 13.05.2016).
- <sup>6</sup> Отчет о научно-исследовательской работе по теме 2.2.4 «Разработка проекта отраслевого стандарта создания электронных копий архивных документов», Плана научно-исследовательской и опытно-конструкторской работы, выполняемой на основе государственного задания Федерального архивного агентства на 2014 г. № 89 от 26.12.2013 (первый этап) «Исследование и анализ зарубежной нормативно-методической документации, регулирующей вопросы оцифровки архивных документов» / Ю.Ю. Юмашева. М.: ВНИИДАД, 2012. С. 84–163.
- <sup>7</sup> Там же. С. 20.
- <sup>8</sup> Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом / Ю.Ю. Юмашева. М.: ВНИИДАД, 2012.
- <sup>9</sup> Объединенный электронный архив Ивана Бунина. [Электронный ресурс] URL: <http://www.bunin-rgali.ru/> (дата обращения: 13.05.2016).
- <sup>10</sup> Воссоединенный виртуальный архив Осипа Манделъштама. [Электронный ресурс] URL: <http://mandelstam-world.info/intro.php> (дата обращения: 13.05.2016).
- <sup>11</sup> Попов А.В. Русское зарубежье и архивы: Документы российской эмиграции в архивах Москвы: проблемы выявления, комплектования, описания и использования (Материалы к истории русской политической эмиграции. Вып. 4). М.: РГГУ, 1998. С. 150–151.
- <sup>12</sup> Воссоединенный виртуальный архив Осипа Манделъштама.