

## **Электронный архив документов ВАК: результаты и проблемы четырехлетней работы**

### **VAK documentary digital archive: Results and issues of quadrennial activities**

*А. И. Бродовский, Е. М. Зайцева, Ю. И. Заславский*

*Государственная публичная научно-техническая библиотека России,  
Москва, Россия*

*Alexander Brodovsky, Ekaterina Zaytseva and Yury Zaslavsky*

*Russian National Public Library for Science and Technology,  
Moscow, Russia*

Описываются работы по разработке и ведению электронного архива документов ВАК. Приводится характеристика информационной системы архивных документов ВАК, технологии оцифровки архивных материалов. Представлены результаты работ по проекту за 2016–2019 гг. Выявлены проблемы, связанные с созданием электронного архива.

**Ключевые слова:** электронный архив документов ВАК, систематизация документов, верификация данных, оцифровка, поиск, хранение.

VAK documentary digital archive design and development activities are described. VAK archival information system and digitization technology are characterized. The project results for 2016–2019 are reported. The issues of digital archive building are highlighted.

**Keywords:** VAK documentary digital archive, document systematization, data verification, digitization, search, storage.

В рамках выполнения государственных заданий 2016-2019 гг. ГПНТБ России реализует проект по созданию электронного архива документов ВАК. Целью проекта является обеспечение функционирования и развития государственной системы научной аттестации: создание фонда электронных архивных документов ВАК, обеспечение поиска и доступа к электронным архивным документам работников Минобрнауки России и экспертных советов ВАК, повышение эффективности и качества работы Департамента аттестации научных и научно-педагогических работников Минобрнауки России и экспертной работы за счет использования в работе данных оцифрованного архива документов.

Основываясь на имеющемся опыте создания электронных архивов [1, 2, 3, 4, 5], на сложившейся нормативной и методической базе в этой области [6, 7, 8, 9, 10, 11], на практическом опыте оцифровки и создания имидж-каталогов в ГПНТБ России [12], разработчики создали информационную систему архивных документов ВАК с учетом особенностей архивных документов ВАК и требований Минобрнауки России и на протяжении четырех лет реализуют наполнение архива электронными документами. Данный электронный архив призван выполнять следующие функции: хранение электронных документов, эффективный поиск информации, оперативный доступ к документам, управление документами (печать, выгрузка).

На первом этапе проекта были выполнены следующие работы, связанные с анализом архива, формированием технических и технологических требований, разработкой технологий и системы архива:

- анализ и систематизация архива документов;
- разработка порядка верификации документов;
- определение состава аппаратного комплекса;
- разработка требований и технологии оцифровки архивных материалов и комплекса проверок качества оцифровки;
- разработка требований к поисковому интерфейсу и функциональным возможностям электронного архива;

– разработка информационной системы архивных документов ВАК на основе системы ИРБИС64;

– разработка технологии передачи и хранения файлов информационной системы архивных документов ВАК.

В последующие годы помимо модернизации программного и информационного обеспечения электронного архива проводились в основном следующие работы по обеспечению формирования электронного архива:

- дополнительная систематизация архивных материалов;
- верификация архивных документов;
- оцифровка архивных материалов;
- программно-технологическая обработка оцифрованных материалов;
- регулярная загрузка данных в информационную систему архивных документов ВАК;
- обеспечение хранения оцифрованных документов и базы данных;
- вспомогательные работы: расшивка томов архивных материалов и переплетные работы.

#### 1. Систематизация архивных материалов.

В результате анализа архива и определения критериев отбора документов в качестве базовых материалов для первичного наполнения информационной системы архивных документов ВАК в 2016 г. были выбраны наиболее востребованные материалы – приказы Минобрнауки России и решения Президиума ВАК. В 2017 г. после проведения дополнительной систематизации архивных материалов спектр охвата архивных документов был расширен. Были определены типы основных, значимых и контрольных типов архивных документов ВАК за 2015-2017 гг., которые обрабатывались в полном объеме в 2017-2019 гг.

#### 2. Верификация архивных документов.

Верификация документов, проводящаяся постоянно, включает проверки на дублетность, лакуны, правильность порядка расположения, последовательную простановку дат и номеров документов; комплекты документов проверяются на полноту подборки; удаляются лишние страницы (копии, черновые страницы); в отдельных случаях проводится перекомпоновка материалов по согласованию с Департаментом аттестации научных и научно-педагогических работников Минобрнауки России.

#### 3. Оцифровка архивных материалов.

В ходе выполнения проекта была разработана технология оцифровки архивных документов. Предварительно были сформулированы требования, предъявляемые к файлам электронного архива, определены формат и технологические параметры файлов, содержащих графические образы. Технологические параметры файлов определялись с учетом общепринятых стандартов и требований сканирования, методических рекомендаций [10, 11], читабельности получаемых изображений для пользователя. В качестве основных были установлены следующие параметры, обеспечивающие высокое качество изображения при создании цифровых копий текстовых аналогов документов, используемых в системах полнотекстового поиска [11]: формат файлов – JPEG (рекомендуемый для пользовательских копий); оптическое разрешение – 300 dpi (рекомендуемое и наиболее часто используемое для оцифровки документов формата А4); глубина цвета – 8 бит (оттенки серого), рекомендуемая для страниц, содержащих текст. Были также определены требования к именам файлов: уникальность, структурированность, включение элементов метаданных.

Основные технологические этапы оцифровки определяются следующим образом:

##### 1) Первичный осмотр и структурирование исходных материалов.

Перед сканированием проводится осмотр и экспертиза архивных материалов с оценкой качества их оформления. Выявляются документы, имеющие различного рода дефекты и искажения и требующие применения не поточного, а отдельного постраничного сканирования.

##### 2) Сканирование документов.

В процессе работы используются сканеры: высокопроизводительный поточный сканер «Kodak», обеспечивающий качественную и оперативную оцифровку документов, и планетарные сканеры «Zeutschel» и «ЭЛАР ПланСкан А2», применяемые для бережного постраничного сканирования литературы. Для автоматизации процессов сканирования и обработки цифровых копий применяется программное обеспечение ЭЛАР СканИмидж, Zeutschel OmniScan, TWAIN ISIS; для

обработки и корректуры изображений также используется ABBYY FineReader и программа XnView, имеющая функцию пакетного переименования файлов.

3) Обработка и проверка полученных образов.

Осуществляется полистная проверка оцифрованных документов, корректировка и чистка файлов с устранением имеющихся дефектов, отклонений, затемнений и т.д.

4) Структурирование оцифрованного массива.

В процессе работы проводится формирование папок с именами, включающими тип содержащихся в папке документов, даты их создания и нумерацию документов.

5) Выходной контроль качества массивов графических образов.

На выходе проводится общий контроль качества результатов оцифровки, включающий комплекс проверок различного вида.

4. Разработка информационной системы архивных документов ВАК.

Основой программного обеспечения информационной системы архивных документов ВАК является САБ ИРБИС64. При создании системы использовались следующие модули и технологии САБ ИРБИС64: сервер баз данных ИРБИС64; технология создания базы данных имидж-каталога; модуль ИРБИС-Навигатор.

Базой данных системы является имидж-каталог, который представляет собой полнотекстовую базу данных, созданную на основе распознанных скан-образов текстовых документов. Для системы документом является одна страница архива ВАК. Процесс формирования базы данных системы представляет собой полностью пакетную обработку. Для этого используются специальные средства САБ ИРБИС64, включающие функцию автоматического распознавания текста.

Пользователю для работы с системой предлагается ИРБИС-Навигатор. Данный модуль представляет собой клиентское приложение, предназначенное для выполнения произвольных операций с базами данных САБ ИРБИС64 на основе интерфейсов, программируемых с помощью форматов САБ ИРБИС64. Для ИРБИС-Навигатора был разработан набор форматов, реализующих весь функционал системы, адресованный пользователю. Интерфейс предлагает для поиска две основные возможности: 1) полнотекстовый поиск страниц архива по словам; 2) поиск по папкам с возможностью последовательного просмотра архивных страниц в определенной папке. Возможен также выбор типа документа и задание года составления документа, отраженного в названии папки документов.

5. Программно-технологическая обработка оцифрованных материалов.

Программно-технологическая обработка отсканированных документов включает следующие основные этапы: 1) "скрытое" распознавание текста, включающее: разбивку текста на предложения; извлечение слов; удаление стоп-слов (неинформативных слов); 2) формирование базы данных информационной системы архивных документов ВАК, включающее: формирование записей с координатами слов; формирование словаря.

6. Загрузка данных в информационную систему архивных документов ВАК.

Регулярно осуществляется загрузка массивов оцифрованных документов и данных по ним в информационную систему архивных документов ВАК в Минобрнауки России. Актуализация базы данных происходит по мере оцифровки и технологической обработки отсканированных документов.

7. Обеспечение хранения оцифрованных материалов.

В целях поддержки длительной сохранности оцифрованных документов используется технология серверного хранения данных с применением специализированного хранилища для резервного копирования. Сервер "Архив ВАК" был создан как виртуальная рабочая машина в составе кластера серверов ГПНТБ России (операционная система Windows Server 2012 R2 Standard). Благодаря используемой технологии виртуализации обеспечивается повышенная отказоустойчивая работа сервера. Для обеспечения информационной безопасности доступ к данным сервера имеет ограниченное количество должностных лиц. С целью организации сохранности данных ежедневно выполняется резервное копирование на специализированное хранилище. Согласно плану резервного копирования ежемесячные архивные копии хранятся в течение года.

Поддерживается следующий порядок размещения, передачи и хранения файлов информационной системы архивных документов ВАК: сканирование порции документов архива и сохранение

файлов в рабочей папке на персональном компьютере; проверка, корректировка, очистка и структурирование файлов в рабочей папке; копирование готовых файлов в папку результатов на сервере; проверка файлов в папке результатов; передача папки результатов на программно-технологическую обработку в информационной системе, а также на хранение на сервере и создание резервной копии; актуализация базы данных архивных документов в информационной системе с обеспечением хранения на сервере и создания резервной копии; передача базы данных и отсканированных архивных документов в Минобрнауки России.

#### 8. Вспомогательные работы.

Вспомогательные работы по проекту включают расшивку томов архивных материалов и переплетные работы.

В ходе выполнения проекта получены следующие основные результаты:

- 1) Создана информационная система архивных документов ВАК.
- 2) Разработана технология оцифровки архивных документов.
- 3) Осуществлен анализ архива и систематизация архивных материалов.
- 4) В 2018 г. проведена верификация базовых архивных документов (приказов Минобрнауки России и решений Президиума ВАК) за 2000-2017 гг. (654 тома). В 2019 г. осуществлена верификация широкого спектра архивных документов за 2015-2017 гг. (507 томов).
- 5) В ходе реализации проекта в 2016-2018 гг. оцифровано и обработано 1215 томов архивных документов ВАК за 2000-2018 гг. (около 240 тыс. страниц).

Практически по каждому виду работ исполнители проекта столкнулись с определенными трудностями. Основная проблема – необходимость доработки системы электронного архива на 2-3 года реализации проекта. Эта проблема связана с тем, что основной акцент работ изначально был сделан на оцифровку данных. Уже на первом году реализации проекта пришлось приступить непосредственно к формированию электронного архива, так как на этот год была запланирована оцифровка 450 томов архивных материалов, а на предпроектное обследование и разработку электронного архива было отведено всего несколько месяцев. Это привело к необходимости дополнительной систематизации архивных материалов и модернизации электронного архива – поискового функционала, набора метаданных и структуры базы данных.

Вторая проблема связана с качеством архивных материалов, имеющих дефекты, которые затрудняют сканирование и распознавание текста (наличие скопированного текста, оформленного как оригинальный документ; присутствие штампов, подписей и рукописных комментариев, накладываемых на текст). Указанные недостатки привели к образованию массива документов, которые проходили постраничное сканирование и чистку и подвергались индексированию вручную.

Общей проблемой для электронного архива документов ВАК и других электронных архивов, содержащих юридически значимые документы, является необходимость ведения бумажного архива параллельно с электронным. При использовании электронного архива налицо выигрыш в скорости поиска и доступа к необходимым данным, но трудозатраты, связанные с ведением архивов двух видов, при этом увеличиваются. Вопрос легализации архивов, в которых электронные документы хранятся и используются в качестве подлинников, требует решения на государственном уровне. Кроме того, следует указать на недостаточную разработанность правовых, нормативных и методических требований к электронным архивам. Сдвиг в этой области должен, безусловно, произойти в связи с принятием нескольких документов на государственном и ведомственном уровне [13, 14]. Требуется решения целый ряд вопросов: разработка стандартов на создание электронных копий архивных документов, установление обязательного набора метаданных для электронных документов, определение порядка учета документов при их наличии в бумажном и электронном виде, определение условий хранения в электронном виде документов ДСП и т. д. Следует также отметить, что существенную проблему для электронных архивов могут создать мероприятия по импортозамещению программного обеспечения. Очевидно, будет необходима разработка единого отечественного программного обеспечения для сети электронных архивов. Несмотря на имеющиеся проблемы, электронные архивы в последние годы активно внедряются в практику работы государственных органов, организаций, предприятий. Но этот процесс требует финансовой и нормативно-правовой поддержки.

### Список источников

1. **Рындин А. А.** Архив без пыльных полок или способы организации архива документов предприятия // ЕСМ-Journal: журнал о системах электронного документооборота (СЭД). – Режим доступа: <https://esm-journal.ru/card.aspx?ContentID=1912029> (дата обращения: 28.06.2019).
2. **Юмашева Ю. Ю.** Архивы электронных документов: проблемы и возможные решения // Власть. – 2015. – № 3. – С 61-66.
3. **Залаев Г. З., Каленов Н. Е., Цветкова В. А.** Оцифровка документов в научных архивах и библиотеках: вопросы и ответы // НТИ. Сер.1. – 2016. – № 2. – С. 14-21.
4. **Тихонов В. И.** Архивное хранение электронных документов: проблемы и решения // Делопроизводство и документооборот на предприятии, февраль 2006. – Режим доступа: <http://www.delopress/articles.php?n=5150> (дата обращения: 28.06.2019).
5. **Евстигнеева Г. А.** Качество оцифровки – проблемы и решения // Современная библиотека. – 2012. – № 5. – С. 58-61.
6. **Ларин М. В.** Электронные документы: нормативно-методическое обеспечение // Вестник архивиста. – 2008. – № 4. – С. 106-124.
7. **ГОСТ Р ИСО 30300-2015.** Система стандартов по информации, библиотечному и издательскому делу. Информация и документация. Системы управления документами. Основные положения и словарь. – Введ. 2016-07-01. – М.: Стандартинформ, 2016. – 14 с.
8. **ГОСТ Р ИСО 15489-1-2007.** Система стандартов по информации, библиотечному и издательскому делу. Управление документами. Общие требования. – Введ. 2007-07-01. – М.: Стандартинформ, 2007. – 20 с.
9. **Правила** организации хранения, комплектования, учета и использования документов Архивного фонда Российской Федерации и других архивных документов в органах государственной власти, органах местного самоуправления и организациях (утверждены приказом Министерства культуры РФ от 31 марта 2015 г. № 526) // Электронный фонд правовой и нормативно-технической документации. – Режим доступа: <http://docs.cntd.ru/document/420266293> (дата обращения: 28.06.2019).
10. **Методические** рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом / Ю. Ю. Юмашева. – М.: ВНИИДАД, 2012. – 125 с.
11. **Методика** контроля качества сканирования бумажных документов: методическое пособие и техническое руководство / С. М. Тимиргалиев, Н. И. Черновалова, О. В. Баркова, Е. В. Ларкин, В. В. Котов, С. Н. Клещарь, Ю. И. Заславский. – М.: ДиМи-Центр, 2012. – 53 с.
12. **Бродовский А. И.** Создание имидж-каталогов в ГПНТБ России средствами ИРБИС // Научные и технические библиотеки : ежемес. науч.-практ. журн. для специалистов библиотеч.-информ. и смеж. отраслей. – 2009. – № 8. – С. 5-16.
13. **План мероприятий по направлению «Нормативное регулирование»** программы «Цифровая экономика Российской Федерации» (утвержден Правительственной комиссией по использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности (протокол от 18 декабря 2017 года № 2)) – Режим доступа: <http://static.government.ru/media/files/P7L0vHUjwVJPIncHrMZQqEEeVqXAcwXR.pdf> (дата обращения: 03.11.2019).
14. **Типовые функциональные требования к системам электронного документооборота и системам хранения электронных документов в архивах государственных органов** (Проект для утверждения приказом Федерального архивного агентства) – Режим доступа: <https://regulation.gov.ru/projects#npa=96479> (дата обращения: 03.11.2019).