

**Научная электронная библиотека ГПНТБ России:
используемые технологии и тенденции развития**

**RNPLS&T'sci-tech e-library:
Technologies applied and trends for development**

К. А. Колосов

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

Kirill Kolosov

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

В докладе приводится обзор актуальных технологий, используемых при загрузке документов в научную электронную библиотеку, обслуживании поисковых запросов, показе документов в режиме защищенного просмотра и обмене данными с внешними интеграторами. Подробно описывается технология выгрузки данных для передачи в Национальную электронную библиотеку. Рассматриваются новые подходы взаимодействия программных решений ИРБИС с системой электронного архива Xerox и с системой защищенного просмотра документов Vivaldi.

The author reviews today's technologies of loading documents into sci-tech electronic libraries, retrieval inquiries processing, document protected viewing, and data exchange with outer integrators. Data upload into the National E-library is examined in detail. New approaches towards IRBIS software solutions for interaction with Xerox e-archive and Vivaldi document control system for data protected reviewing are discussed.

Научная электронная библиотека ГПНТБ России (НЭБ ГПНТБ России), созданная с конце 2006 года, регулярно пополняется книгами, оцифрованными Центром сканирования ГПНТБ России, а также статьями из журнала «Научные и технические библиотеки», документами из сборников научных трудов, издаваемых ГПНТБ России, и материалами конференций «Крым» и «ЛИБКОВ». В процессе отбора книг, подлежащих оцифровке и размещению в НЭБ ГПНТБ России, формируются тематические коллекции, позволяющие читателям НЭБ ГПНТБ России эффективнее искать интересующие их документы в огромном массиве электронных документов.

В 2010 году в ГПНТБ России был установлен аппаратно-программный комплекс электронного архива, использующий программный продукт Docushare фирмы Xerox. В настоящее время Xerox Docushare является технологической основой процессов подготовки, размещения, хранения и показа электронных документов НЭБ ГПНТБ России. На рис. 1 показана структурная схема НЭБ ГПНТБ России, на которой представлены программные модули внутренней сети ГПНТБ России и взаимодействующие с ними интернет порталы.

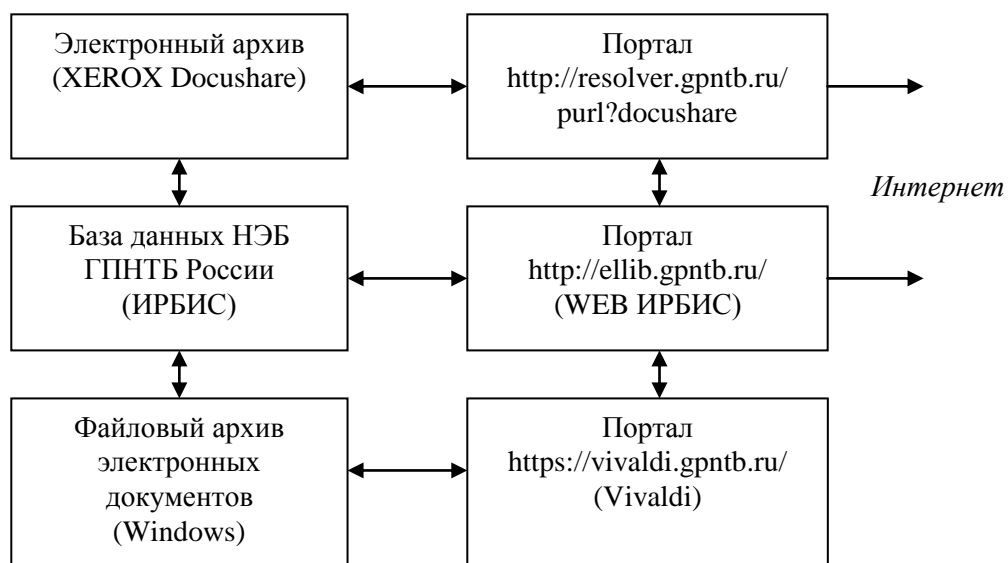


Рис. 1. Обобщенная структурная схема НЭБ ГПНТБ России

В настоящее время НЭБ ГПНТБ России содержит более 13000 документов суммарным объемом более 1 млн. 150 тыс. страниц. В электронный архив ГПНТБ России за период с 2011 по 2013 год было отсканировано и загружено 3734 изданий суммарным объемом 718100 страниц. В таблице 1 приведена статистика пополнения электронного архива за последние 3 года.

Таблица 1

| Год | Число отсканированных изданий из фондов ГПНТБ России | Суммарное число страниц |
|------|--|-------------------------|
| 2011 | 679 | 118040 |
| 2012 | 987 | 200005 |
| 2013 | 2068 | 400055 |

Размещение электронного документа в электронном архиве ГПНТБ России включает следующие технологические этапы:

- регистрация издания в системе;
- сканирование издания;
- распознавание текстов;
- атрибутирование (аннотирование) с использованием заимствования библиографического описания из каталога ИРБИС;
- размещение в электронном хранилище;
- запись ссылки (URL) размещенного издания в каталог ИРБИС.

На рисунке 2 показана последовательность технологических этапов подготовки и размещения документа в электронном архиве ГПНТБ России.

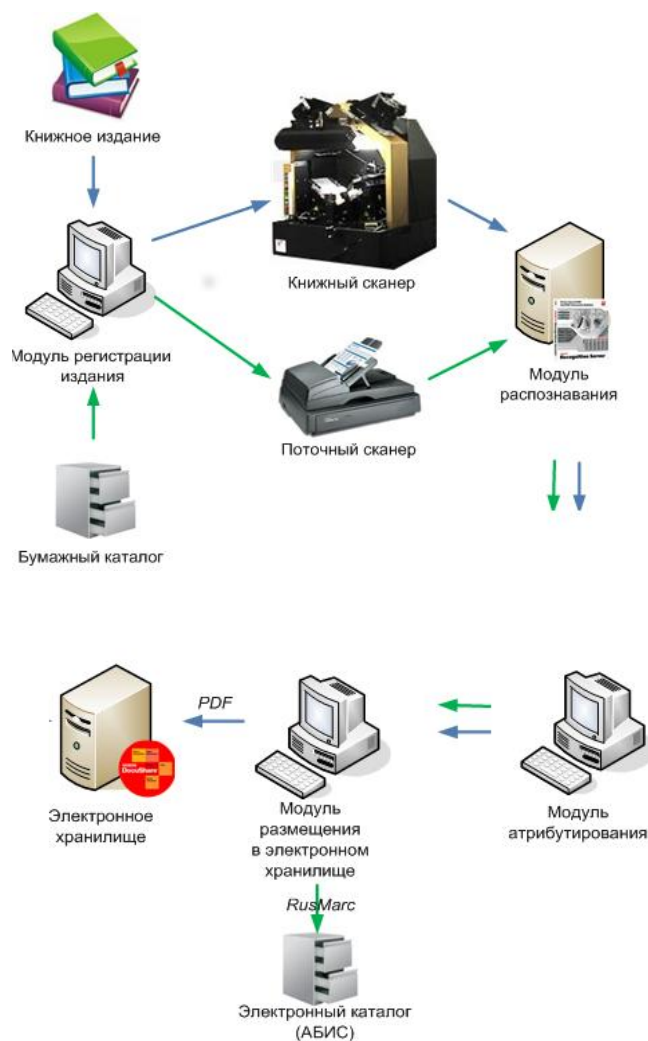


Рис 2. Технологические этапы подготовки и размещения документа в электронном архиве ГПНТБ России

Назначение основных модулей электронного архива показано в таблице 2.

Таблица 2

| № | Технологический модуль | Назначение модуля | Автоматизируемый процесс |
|---|--|--|---|
| 1 | Модуль регистрации | Модуль предназначен для регистрации задания на размещение ресурса в модуле электронного хранилища, а также последующей работы с данным заданием до передачи задания на размещение ресурса в модуль атрибутирования и импорта | <ul style="list-style-type: none"> • Сканирование ресурса; • Распознавание и верификация ресурса |
| 2 | Модуль атрибутирования и импорта | Модуль предназначен для импорта атрибутов библиографического описания ресурса и размещения ресурса в модуле электронного хранилища | <ul style="list-style-type: none"> • размещение распознанного и верифицированного ресурса в модуле электронного хранения |
| 3 | Модуль постоянного хранения оригиналов | Модуль предназначен для постоянного хранения отсканированных файлов TIFF ресурса | <ul style="list-style-type: none"> • постоянное хранение отсканированного ресурса |

| № | Технологический модуль | Назначение модуля | Автоматизируемый процесс |
|---|------------------------------------|---|---|
| 4 | Модуль электронного хранилища | Модуль предназначен для хранения размещенного ресурса, включающего в себя файл PDF и атрибуты размещаемого ресурса | <ul style="list-style-type: none"> • работа с размещенным ресурсом |
| 5 | Модуль распознавания и верификации | Модуль предназначен для распознавания файлов TIFF и сборки PDF, а также для последующей верификации файла PDF | <ul style="list-style-type: none"> • распознавание и верификация ресурса |
| 6 | Модуль защищенного просмотра | Модуль предназначен для обеспечения защищенного просмотра содержимого файла PDF, а также полнотекстового поиска и навигации по нему | <ul style="list-style-type: none"> • работа с размещенным ресурсом |

Модуль атрибутирования и импорта работает совместно с программным модулем АБИС адаптера, разработанным в ГПНТБ России. Использование АБИС адаптером протокола Z39.50 позволяет применять его для поиска и импорта библиографических описаний как при взаимодействии с ИРБИС, так и при работе с другими системами автоматизации библиотек. Важным технологическим этапом является добавление обратной ссылки (URL) размещенного в хранилище электронного документа к исходной библиографической записи в электронном каталоге или создание новой библиографической записи, содержащей URL электронного документа, в технологической базе данных. Для электронного архива могут использоваться и тот и другой вариант, в зависимости от последующей технологии показа документов пользователям. Поскольку в ГПНТБ России помимо решения Xerox Docushare используется серверная программа защищенного показа документов Vivaldi, создание копии дополненной библиографической записи в отдельной технологической базе является предпочтительней.

Документы, загружаемые в электронный архив ГПНТБ России, становятся доступными для просмотра пользователям сразу же после загрузки (за исключением скрытых коллекций документов). Пользователи могут осуществлять навигацию по коллекциям ресурсов или воспользоваться поисковыми возможностями электронного архива (<http://resolver.gpntb.ru/purl?docushare>). Одновременно с загрузкой документов в электронный архив пополняется технологическая база данных ИРБИС, содержащая дополненные библиографические описания. Эти библиографические записи используются на следующем технологическом этапе, включающем пополнение электронного каталога НЭБ ГПНТБ России и выгрузку массивов документов для их последующей передачи в Национальную электронную библиотеку (НЭБ).

Электронный каталог НЭБ ГПНТБ России ведется на основе системы автоматизации библиотек ИРБИС и является основным инструментом поиска и представления библиографических записей на портале <http://ellib.gpntb.ru>. В процессе добавления новых библиографических записей к массиву базы данных НЭБ ГПНТБ России в автоматизированном режиме производится их модификация с целью добавления гиперссылки для показа документа с использованием системы защищенного просмотра документов Vivaldi. Пользователи НЭБ ГПНТБ России, просматривая результаты поиска, могут выбирать, какой вариант просмотра документа им удобнее: Docushare на основе flash-технологии или Vivaldi с использованием специализированного бесплатного клиента.

Для обеспечения взаимодействия системы Vivaldi с ИРБИС в ГПНТБ России были разработаны следующие дополнительные программные решения:

- модуль поиска документа по гиперссылке (URL) и показа его библиографического описания;
- модуль для авторизации пользователя Vivaldi с использованием проверки сведений из базы данных читателей ИРБИС и технологии OpenID.

Оба перечисленных выше программных решения основаны на использовании сервера Z64, работающего в режиме WEB-сервиса по протоколу SRU/SRW, и использовании специальных форматных файлов ИРБИС.

С 2008 года ГПНТБ России является участником проекта Национальной электронной библиотеки (НЭБ) (<http://www.rusneb.ru>). Для этого проекта ежегодно требуется выгрузка электронных документов и их библиографических описаний, представленных в специальном формате хранения. Согласно требованиям НЭБ, действующих до настоящего времени, для каждого документа, передаваемого в НЭБ, требуется создание отдельной дисковой поддиректории, в которую выгружается распознанный документ в формате PDF и весь набор его исходных страниц в формате TIFF. Так же в поддиректорию записывается библиографическое описание документа в формате MARC21. В ГПНТБ России процедура выгрузки документов из хранилища электронного архива для передачи в НЭБ осуществляется одновременно с пополнением базы данных НЭБ ГПНТБ России в системе ИРБИС. Для этих целей был разработан специальный пакет программ, в задачи которого, в частности, входит определение местоположения TIFF файлов в хранилище электронного архива Xerox DocuShare. На рисунке 3 представлен алгоритм формирования массива документов и библиографических записей для передачи в НЭБ.

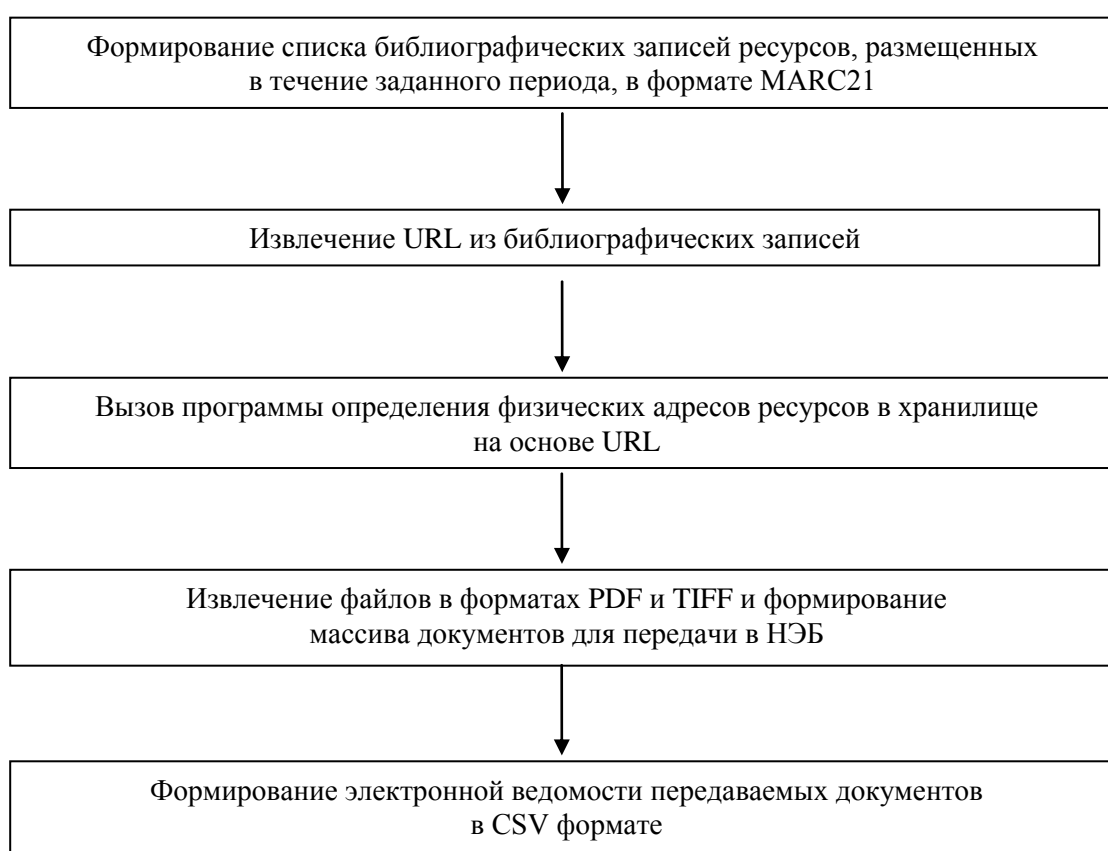


Рис. 3. Алгоритм формирования массива документов и библиографических описаний для передачи в НЭБ

В настоящее время представленный алгоритм реализован с использованием набора отдельно вызываемых программных модулей. В дальнейшем планируется произвести интеграцию модулей в единый модуль, встроенный в интерфейс одного из модулей АРМ ИРБИС, поскольку задача по передаче в НЭБ массивов данных, хранящихся в электронном архиве Xerox DocuShare, возникает для целого ряда библиотек, использующих аналогичные решения для своих электронных библиотек.

Литература:

1. Новое решение Xerox для библиотек сделает редкие издания доступными для всех. [Электронный ресурс] / Режим доступа: <http://www.xerox.ru/ru/press-centre/release/2011/05/10/428810/>
2. Ефимова О.А. Интеграционные решения Xerox для научных библиотек: полнотекстовые электронные библиотеки и технологии формирования библиотечных электронных информационных ресурсов. [Электронный ресурс] / Режим доступа: http://www.benran.ru/SEMINAR/SEM/Sb_09/doc_62.html