

М1:АРХИВ

ОБЩЕЕ ОПИСАНИЕ

Содержание

1. НАЗНАЧЕНИЕ СИСТЕМЫ.....	3
2. ОПИСАНИЕ СИСТЕМЫ	4
2.1. Принципы работы системы электронного архива М1:Архив.....	4
2.1.1. Описание принципов работы системы электронного архива М1:Архив.....	4
2.2. Варианты импорта данных.....	5
2.2.1. Выгрузка данных из системы-источника базовым коннектором	5
2.2.2. Интеграция с системой-источником данных напрямую.....	5
2.2.3. Принцип работы программного интеграционного комплекса АВВУУ+М1:Архив...	7
2.3. Принципы организации архивных данных.....	7
2.4. Возможности пакетной загрузки документов с обеспечением контроля целостности ..	8
2.5. Управление пользователями	8
2.6. Разграничение прав доступа.....	8
2.7. Настройка поисковых форм	9
2.8. Управление политиками хранения	9
2.8.1. Политики хранения.....	10
2.8.2. Политики удаления.....	10
2.8.3. Типы хранения	11
2.9. Основные сценарии работы пользователя с электронным архивом.....	11
2.9.1. Ручной поиск по каталогам.....	11
2.9.1. Поиск информации через поисковые формы	12
2.10. Изменение атрибутов архивного документа	12
2.11. Создание архивного документа	13
ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ	14

1. НАЗНАЧЕНИЕ СИСТЕМЫ

Система «М1:Архив» (далее Система, Система электронного архива) предназначена для управления большими объемами информации предприятий и организаций. В основе управления – надежное хранение и эффективное использование информации. Система позволяет наладить ввод документов, их индексирование, хранение, поиск и редактирование, обеспечить их защиту от несанкционированного доступа, изменения или уничтожения.

М1:Архив используется:

для создания высоконадежной архивной системы;

для организации поиска данных разных типов (текстовые документы, изображения, чертежи, медицинская информация и пр.)

Система позволяет архивировать любые типы данных, как структурированные, так и не структурированные, поэтому она может использоваться для архивирования:

- Данных систем электронного документооборота;
- Данных систем технического (инженерного) архива;
- Данных электронных архивов клиентских дел;
- Данных, полученных в результате ретроспективного сканирования;
- Данных о транзакционных операций;
- Баз данных.

2. ОПИСАНИЕ СИСТЕМЫ

M1: Архив - это масштабируемое решение для архивации данных предприятия, электронный архив. Система обеспечивает сохраняет, поддерживает и контролирует постоянный доступ к ценным информационным ресурсам. Она также дает независимость от приложений, предоставляя один унифицированный архив для всех данных приложений, размещенных в архиве на базе Системы.

M1:Архив решает две группы задач архивирования данных предприятия

1. вывод из эксплуатации устаревших информационных систем (ИС) с целью сокращения расходов на их обслуживание.
2. архивирование данных из действующих систем для уменьшить нагрузку на них

M1:Архив позволит хранить данные на более дешевых ресурсах при этом обеспечивая необходимый уровень доступности данных для пользователей.

Данные, выбираемые для архивирования из действующих систем, обрабатываются на основании правил, регламентов и бизнес-процессах организации. Например, еженедельно банк может архивировать все банковские выписки старше года.

2.1. Принципы работы системы электронного архива M1:Архив

2.1.1. Описание принципов работы системы электронного архива M1:Архив

Система M1:Архив основана и работает на общем открытом стандарте OAIS, а также поддерживает пакет международных стандартов CMIS, поэтому она полностью совместима с большинством существующих бизнес-приложений.

В соответствии с OAIS, данные передаются в M1:Архив в виде информационных пакетов (SIP). Данные пакеты содержат бизнес-объекты, которым требуется обеспечить долговременное хранение с обеспечением функций неизменности, управления политиками хранения и безопасности доступа.

Система M1:Архив импортирует данные в виде SIP-пакетов. Преобразование данных в SIP-пакеты происходит до того, как Система получит данные. Для преобразования данных используются инструменты извлечения, преобразования, загрузки (ETL) или коннекторы, которые доступны для ABBYY FineReader Server, ABBYY FlexiCapture, OT Documentum, SAP.

2.2. Варианты импорта данных

Предусматривается 2 варианта загрузки данных в архив:

Выгрузка базовым коннектором из системы источника SIP пакетов данных (XML с метаданными + вложения) и загрузка в «M1:Архив».

Интеграция напрямую с системой с использованием специальных коннекторов для работы с ABBYY FineReader Server, ABBYY FlexiCapture, OT Documentum, SAP.

2.2.1. Выгрузка данных из системы-источника базовым коннектором

Если система-источник данных не имеет возможности выгрузить архивируемые данные в формате SIP архивов, то используется ETL средство загрузки данных.

2.2.2. Интеграция с системой-источником данных напрямую

Наполнение выделенных под существующие в организации информационные системы (ИС) ресурсов происходит очень быстро и не всегда эти ресурсы можно безболезненно расширить. Решением в этом случае может являться перенос части данных в архивные системы. M1:Архив позволяет автоматизировать процесс переноса данных из самых разных ИС в унифицированное хранилище, сохраняя при этом целостность данных и обеспечивая в дальнейшем возможность их использования.

Для решения этой задачи могут использоваться следующие программно-информационные комплексы

M1:Архив + ABBYY;

M1:Архив + OT Documentum.

Варианты применения программно-информационного комплекса M1:Архив + ABBYY:

- 1) На практике достаточно часты случаи, когда в системах хранятся нераспознанные сканы документов. Если такая необходимость появилась, то можно выполнить извлечение данных в процессе активного архивирования и уложить их в структуры вместе с другими уже существующими данными. При этом нет необходимости менять существующие ИС (расширять атрибутивный состав). Для этого можно использовать программно-интеграционный комплекс M1:Архив + ABBYY.

- 2) Программно-интеграционный комплекс М1:Архив + АBBYУ может также помочь, когда в организации встает вопрос вывода из эксплуатации устаревших систем с сохранением данных. Не всегда данные можно перевести в существующие ИС без потерь и доработок ИС. Если часть информации, переносимой из устаревших ИС, представлена на нераспознанных сканах, интегрированное решение М1:Архив + АBBYУ может решить задачу извлечения и сохранения необходимых атрибутов.
- 3) Программно-интеграционный комплекс М1:Архив + АBBYУ может быть также использован при ретроспективном сканировании (переводе текстовых данных из графического вида в цифровой машиночитаемый вид путем сканирования и автоматического распознавания, либо вручную). Программно-интеграционный комплекс М1:Архив + АBBYУ может быть использован в рамках процесса ретроспективного сканирования, например, для оцифровки архивных фондов. При оцифровке одного типа изданий (например, одного типа периодического издания) может быть настроен не только процесс распознавания текста, но и процесс извлечения выходных данных печатного издания, что сможет сэкономить большие объемы ручного труда по атрибутированию каждой единицы хранения.
- 4) Программно-интеграционный комплекс М1:Архив + АBBYУ может решить задачу конверсии старой технической документации, выполненной в бумажном виде, в электронный вид. Решение обеспечит как распознавание текста, так и извлечение выходных данных документов, а также классификацию и хранение документации в структуре в соответствии с требованиями, предъявляемыми к архивному хранению технической документации.

Варианты применения программно-информационного комплекса М1:Архив + OT DOCUMENTUM :

- 1) Программно-интеграционный комплекс М1:Архив + OT DOCUMENTUM может использоваться для архивации данных из устаревших приложений. Это позволит избавиться от необходимости поддержки устаревших бизнес-приложений для обеспечения возможности доступа к хранящимся в них данным. Таким образом, может быть достигнута унификация хранения неизменяемой информации с возможностью оперативного доступа к ней со стороны бизнес-пользователей и контролирующих органов.

- 2) Программно-интеграционный комплекс М1:Архив + ОТ DOCUMENTUM может использоваться для архивации определенных типов документов в процессе активного архивирования и расположения их в архиве в определенной структуре в зависимости от пожеланий Заказчика. В одних случаях есть потребность переноса документов в той же структуре, в которой они хранятся в исходной СЭД, например, по номенклатуре дел. В других случаях необходимо располагать заархивированные документы согласно реальной структуре хранения бумажных документов в архивном фонде. Система Архив:М1 позволяет это сделать.

2.2.3. Принцип работы программного интеграционного комплекса АBBYY+М1:Архив

В базовом варианте система «Планета Архив» получает через модуль загрузки информацию, которая извлекается в результате обработки документа:

При интеграции с АBBYY FineReader Server – исходное изображение (pdf, jpeg, tiff и т.д.), текстовый файл с результатами полнотекстового распознавания (doc, txt, odf и т.д.) и служебную информацию о документе.

При интеграции с АBBYY FlexiCapture - исходное изображение (pdf, jpeg, tiff и т.д.), перечень извлеченных по шаблону данных (в виде xml-файла) и служебную информацию о документе.

Для каждого обработанного документа в системе создается карточка с метаданными (их набор зависит от типа документа) и набор вложений, в которых есть исходное изображение документа. В зависимости от типа документа, в карточке могут быть закладки с перечнем специфичной информации, получаемой при обработке документа с помощью АBBYY FlexiCapture (например, перечень товаров в товарной накладной).

2.3. Принципы организации архивных данных

Все метаданные документов и системных объектов системы хранятся в базе данных PostgreSQL, при этом основные (системные) атрибуты хранятся в реляционной структуре, а прикладные атрибуты хранятся в специальной колонке JSONB представляющей из себя индексированный объект JSON. Это позволяет гибко управлять схемой данных и хранить реквизиты документов, которые даже не определены в типе документа, но при необходимости определять типизацию атрибутов и выполнять автоматическую валидацию при загрузке данных. Контент документов хранится отдельно и для

типа может быть настроено хранение контента на файловой системе или в объектном хранилище, поддерживающем протокол S3.

2.4. Возможности пакетной загрузки документов с обеспечением контроля целостности

В системе предусмотрены автоматические фоновые процессы, которые в соответствии с методологией OAIС обеспечивают загрузку входящих пакетов с возможностью промежуточного преобразования и фильтрации данных. Логи загрузки данных в систему сохраняются после каждого продуктивного запуска процедуры в виде специальных системных документов, по ним можно четко определить, на каких именно записях были сгенерированы ошибки и предупреждения, и сколько всего документов было загружено в систему. Кроме того, есть специализированный механизм сверки импорта, входными данными для него является файл CSV с записями, выгруженными из системы-источника. Система сверяет данные в файле CSV с данными, присутствующими в архиве, и выдает заключение по целостности и полноте загруженных в систему пакетов данных. Также проверке может подлежать контент документов, при этом сверяется чек-сумма, вычисленная по определенному алгоритму.

2.5. Управление пользователями

Система позволяет хранить пользователей как в СУБД, так и настраивать интеграцию с сервером каталогов LDAP.

Для аутентификации по-умолчанию создаваемое приложение настроено на каталог пользователей в базе данных PostgreSQL. Также есть возможность выполнить настройку на один или несколько каталогов LDAP, например, Active Directory.

2.6. Разграничение прав доступа

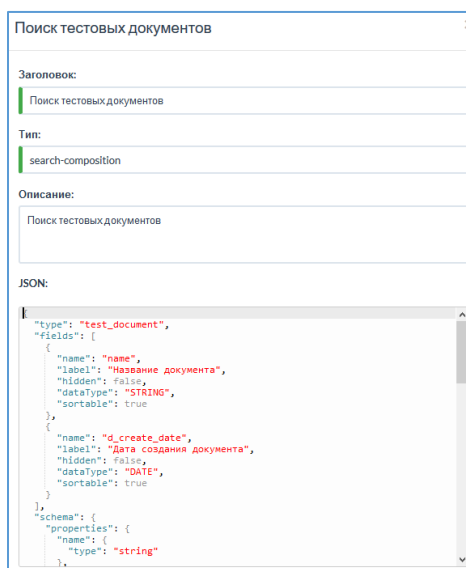
Авторизация на выполнение специфических операций (например, администрирование и конфигурирование системы) происходит на прикладном уровне в соответствии с ролями, на которые назначен пользователь.

Авторизация на выполнение операций в контексте конкретного документа или объекта (например, редактирование, удаление документа, изменение прав доступа) происходит в соответствии с ACL документа и пересечения групп, на которые назначен пользователь, и групп, прописанных в ACL.

Специфичной является обработка прав доступа на чтение документов при выполнении поисковых запросов и выборок. Для нее используется механизм Row Level Security СУБД PostgreSQL: в специальной колонке прописывается массив пользователей и групп, имеющих доступ на чтение к документу, и в момент выполнения любого запроса к данным СУБД накладывает специальный фильтр, в котором автоматически вычисляется пересечение массива групп пользователя с массивом прав читателей документа. Таким образом на уровне СУБД ограничивается доступ всех пользователей на получение каких-либо данных по документу и существенно ускоряется выполнение запросов, так как нет необходимости выполнять дополнительную фильтрацию результатов запроса по правам. Из этой особенности вытекает необходимость хранить информацию о пользователях в базе данных (как минимум соответствие имени пользователя и групп), даже если первоисточником данных о пользователях является LDAP каталог.

2.7. Настройка поисковых форм

Основная настройка поисковой формы производится в формате JSON, текст файла настройки доступен прямо из системы.



```
{"type": "test_document",  
  "fields": [  
    {  
      "name": "name",  
      "label": "Название документа",  
      "hidden": false,  
      "dataType": "STRING",  
      "sortable": true  
    },  
    {  
      "name": "d_create_date",  
      "label": "Дата создания документа",  
      "hidden": false,  
      "dataType": "DATE",  
      "sortable": true  
    }  
  ],  
  "schema": {  
    "properties": {  
      "name": {  
        "type": "string"  
      }  
    }  
  }  
}
```

2.8. Управление политиками хранения

Если данные попадают в систему, а политики хранения не применяются, данные останутся в архиве до бесконечности. Параметры хранения определяют, когда данные в хранилище должны быть уничтожены. Управление политиками хранения поддерживает управляемый процесс утилизации для менеджера хранения.

Управление политиками хранения осуществляется при помощи настроек следующих параметров хранения:

- 1) Политик хранения
- 2) Политик удаления
- 3) Типов хранения

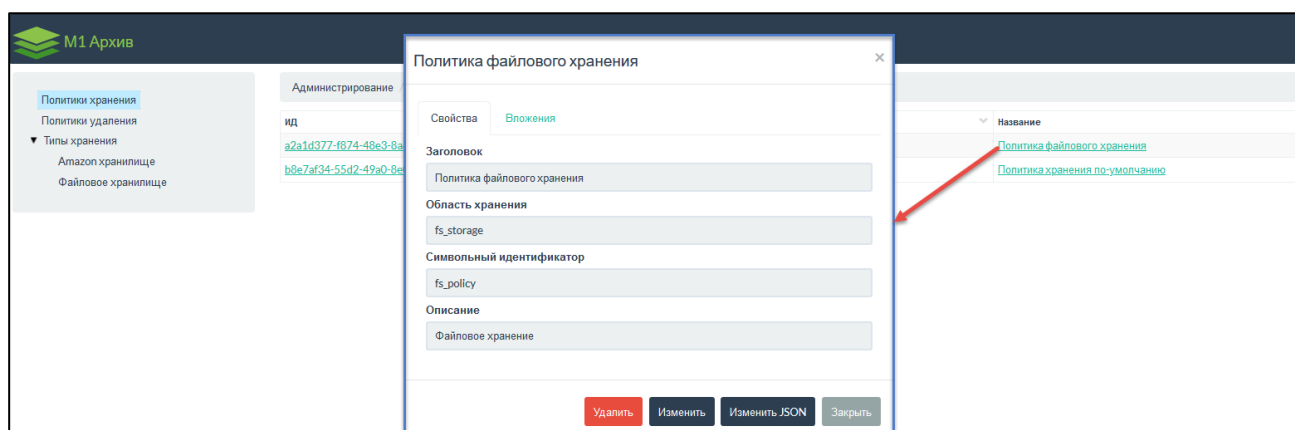
2.8.1. Политики хранения

Параметр настройки «Политики хранения» предназначен для глобального управления политиками хранения архивных файлов. По-умолчанию в системе предусмотрены 3 политики хранения:

Политика файлового хранения

Политика локального S3

Политика хранения по-умолчанию (на данный момент – Политика файлового хранения).

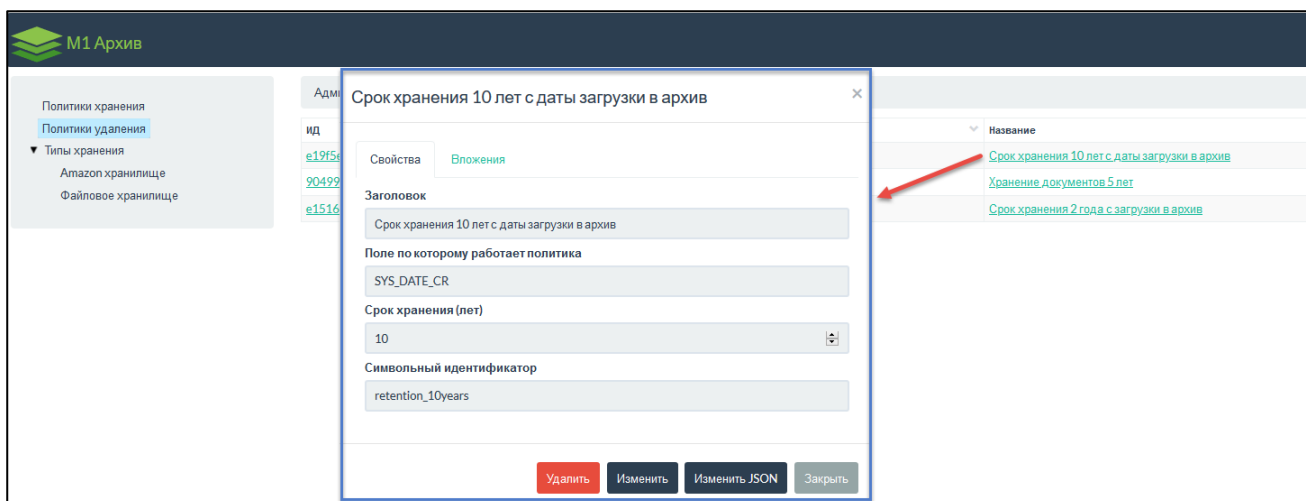


2.8.2. Политики удаления

Параметр настройки «Политики удаления» предназначены для глобального управления политиками удаления архивных файлов. На данный момент, политика удаления может работать по 2-м полям (то есть производить расчёт конечного срока):

SYS_DATE_CR – дата создания файла;

SYS_DATE_MOD – дата изменения файла.

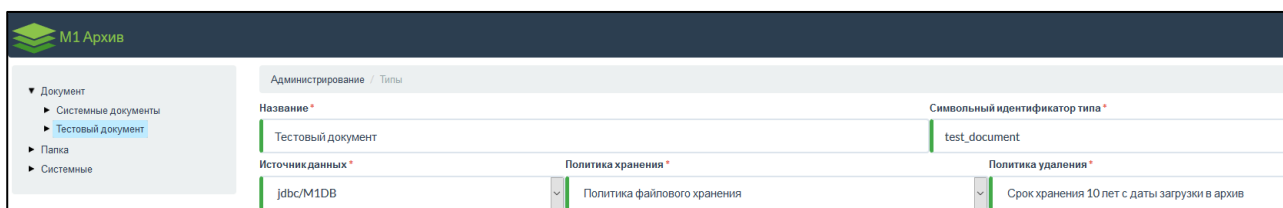


2.8.3. Типы хранения

Параметр настройки «Типы хранения» предназначен для глобального управления типами хранения. По умолчанию в системе заведено 2 типа хранения:

Файловое хранилище - определяет настройки для файлового хранилища.

Amazon хранилище - определяет настройки для хранилища по протоколу S3.



2.9. Основные сценарии работы пользователя с электронным архивом

Поиск информации в системе возможен следующими способами:

Ручной поиск по каталогам – поиск необходимых данных вручную по директориям (папкам).

Поиск информации через поисковые формы – ввод запроса в поисковую форму и поиск по электронному архиву.

2.9.1. Ручной поиск по каталогам


Ручной поиск по каталогам предназначен для поиска информации в случае, если пользователь знает (предполагает) в какой папке (по какому пути) находится необходимая ему информация. Для поиска документа необходимо в левой части рабочего стола пользователя перейти на закладку «Папки», затем выбрать требуемый каталог. В правой части экрана откроется список документов, относящихся к выбранному каталогу.

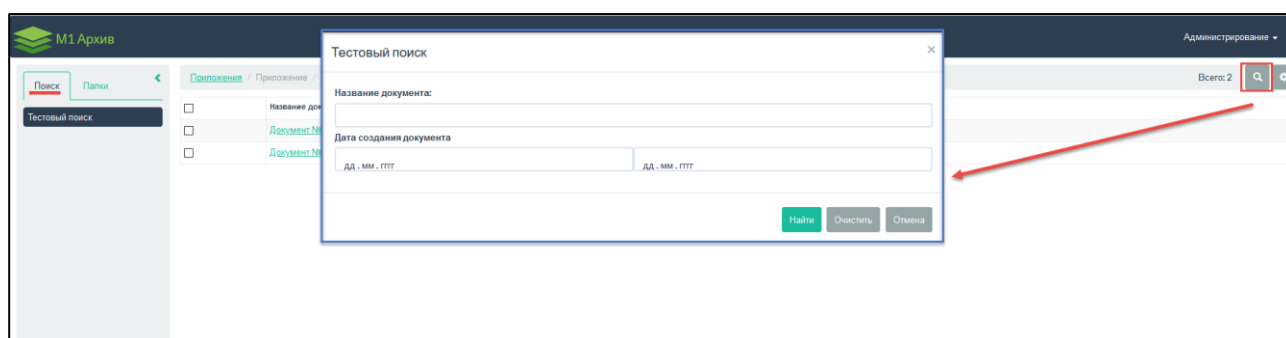
Поиск		Папки	
<ul style="list-style-type: none"> Папки <ul style="list-style-type: none"> Папка для тестовых документов 			
Приложения / Приложение / Тестовый поиск			
<input type="checkbox"/>	Название документа		Дата создания документа
<input type="checkbox"/>	Документ №2		10/03/2021
<input type="checkbox"/>	Документ №1		01/03/2021

2.9.1. Поиск информации через поисковые формы

Поиск информации через поисковые формы предназначен для поиска информации в архиве через запросы в преднастроенной поисковой форме.

Для поиска документа необходимо в левой части рабочего стола пользователя перейти на закладку «Поиск», затем выбрать требуемый вид документа, в котором тре-

буется произвести поиск, после чего нажать кнопку  для вызова поисковой формы. В поисковой форме необходимо заполнить поисковые поля.



2.10. Изменение атрибутов архивного документа

В случае, если в атрибутах документа допущена ошибка, имеется возможность внести изменения в карточку документа. Для этого необходимо открыть карточку требуемого документа и нажать на кнопку «Изменить». В открывшейся карточке необходимо внести исправления в требуемое поле, после чего сохранить изменения, нажав на кнопку «Сохранить».

Документ №1
✕

Свойства
Вложения
Версии
Комментарии

Заголовок *

Наименование *


Дата создания *

Описание

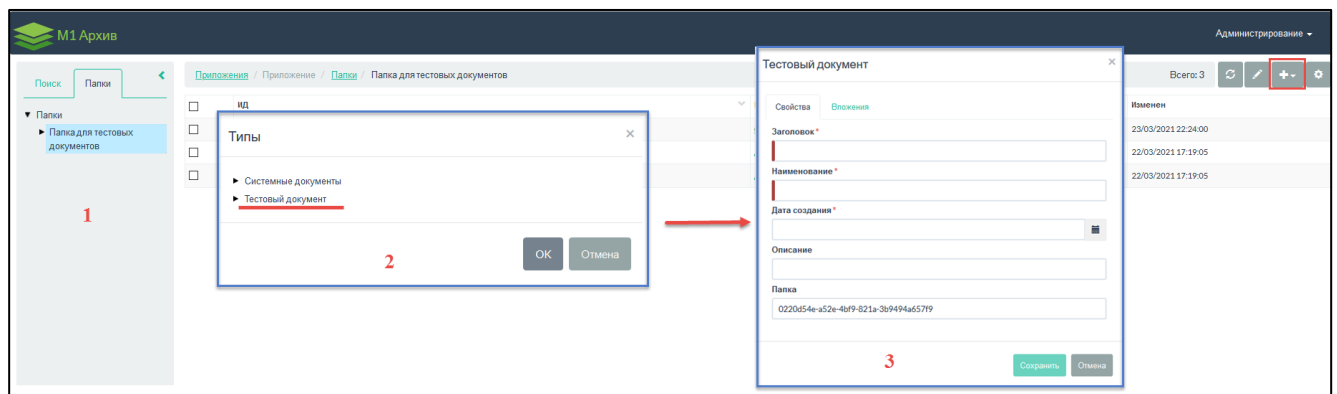
Папка

Сохранить
Отмена

2.11. Создание архивного документа

Одной из особенностей системы M1:Архив является возможность создания документа вручную в интерфейсе системы. Для создания документа необходимо перейти на закладку «Папки», выбрать папку, в которой предполагается создание документа, а затем нажать на кнопку  справа. Далее, в открывшемся меню *Типы* выбрать нужный тип и нажать кнопку «ОК». После этого, в открывшейся карточке нового документа, необходимо заполнить поля с атрибутами, прикрепить вложение и нажать кнопку «Сохранить». Документ появится в выбранной ранее папке.

Далее, в открывшемся меню *Типы* выбрать нужный тип и нажать кнопку «ОК». После этого, в открывшейся карточке нового документа, необходимо заполнить поля с атрибутами, прикрепить вложение и нажать кнопку «Сохранить». Документ появится в выбранной ранее папке.



ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

№	Термин	Определение
1.	ACL	Access Control List - список управления доступом, который определяет, кто или что может получать доступ к архивной записи/документу
2.	AIP	Archival Information Package - пакет заархивированных данных в системе М1:Архив
3.	Amazon S3	Amazon Simple Storage Service — онлайн веб-служба, предлагаемая Amazon Web Services, предоставляет возможность хранения и получения любого объема данных в любое время из любой точки сети, так называемый файловый хостинг.
4.	DIP	Dissemination Information Package - пакет распространения информации, формат, в котором данные предоставляются пользователю в ответ на его запрос
5.	ETL	Extract, Transform, Load – извлечение, преобразование, загрузка. Один из основных процессов в управлении хранилищами данных, который включает в себя: <ul style="list-style-type: none">– извлечение данных из внешних источников;– их трансформация и очистка, чтобы они соответствовали потребностям бизнес-модели;– загрузка данных в хранилище данных.
6.	OAIS	Open Archival Information System – открытая архивная информационная система
7.	SIP	Submission Information Package – пакет передаваемой информации
8.	СХД	Система хранения данных
9.	Структурированная информация	Информация, хранимая в СУБД и/или структурированная необходимым образом
10.	Неструктурированная информация	Содержимое или данные, хранимые в не структурированном виде – документы, тексты, изображения, аудио и т.д.