10. Регламент документооборота

10. Регламент документооборота

Назначение

Раздел описывает порядок работы с корпоративным электронным архивом ООО «АЙТИ ГРУПП»: где лежат оригиналы, как документ попадает в архив, кто отвечает и как проверяется доступ. Сами конфиденциальные файлы в BookStack **не хранятся** — здесь только реестр и ссылки.

Роли

Роль	Кто	Права
Владелец	Айнур Карпиев	полный доступ ко всем категориям
Архивариус	Hermes (агент)	загрузка, разметка метаданных, реестр, отчёты
Сотрудник	по запросу	чтение категорий 08, 09 и выданных документов

Категории архива

Код	Категория	Тег Paperless	Конфиденциальность по умолчанию
01	Учредительные и регистрационные	`01. Учредительные и регистрационные`	Конфиденциальный
02	Бухгалтерия и налоги	`02. Бухгалтерия и налоги`	Конфиденциальный
03	Банки и финансы	`03. Банки и финансы`	Строго конфиденциальный
04	Договоры и клиенты	`04. Договоры и клиенты`	Внутренний
05	Судебные дела и задолженности	`05. Судебные дела и задолженности`	Конфиденциальный
06	Кадровые документы	`06. Кадровые документы`	Строго конфиденциальный
07	Лицензии и сертификаты	`07. Лицензии и сертификаты`	Внутренний
08	Корпоративные шаблоны	`08. Корпоративные шаблоны`	Внутренний
09	Отчёты и аудиты	`09. Отчёты и аудиты`	Внутренний

Обязательные метаданные (настраиваемые поля Paperless)

Организация · Категория · Номер документа · Дата документа · Контрагент · Срок действия · Статус · Уровень конфиденциальности · ID клиента или сделки Twenty CRM. Тип документа задаётся штатным полем Paperless «Тип документа».

Порядок приёма документа

1. Документ поступает из разрешённого источника: ручная загрузка утилитой, webhook-automation n8n, письмо в почтовом правиле Paperless.
2. Проверка принадлежности компании, типа файла (только PDF) и размера (до 100 МБ).
3. Вычисляется SHA-256; при совпадении checksum документ не создаётся повторно (Paperless дополнительно отклоняет дубликаты на уровне consumer).
4. Загружается в Paperless-ngx (`https://pdf.karpiev.ru`), оригинал сохраняется без изменений и не удаляется.
5. Присваиваются тег категории, тег `ITG Corporate Archive`, тип документа и настраиваемые поля.
6. Выполняется OCR (русский + английский) и индексация полнотекстового поиска.

7. Запись попадает в реестр BookStack (этот раздел) — синхронизация по тегу-маркеру.
8. Результат пишется в журнал импорта.

Инструменты (рабочее состояние на 2026-11-05)

Инструмент	Где	Назначение
`archive_upload.py`	`/opt/data/scripts/archive_upload.py`	защищённая загрузка PDF с категорией и метаданными, SHA-256, дедупликация, журнал
`archive_expiry_reminders.py`	`/opt/data/scripts/`	напоминания о сроках: 30 / 7 / 0 дней (cron Hermes `ed3d2262aaa3`, ежедневно 04:00 UTC = 09:00 Екб, доставка в Telegram «Инфраструктура»)
`sync_registry.py`	`/opt/data/itg-archive/`	синхронизация этого реестра из Paperless по тегу-маркеру
n8n workflow	`ITG — Corporate Document Intake v1.0` (id `cj5yyskp60Fp7lq2`)	webhook-приём: проверка, дедупликация, загрузка, ожидание OCR, классификация, метаданные, уведомление
Журнал импорта	`/opt/data/itg-archive/import_log.jsonl`	SHA-256, статус, ID документа

Пример защищённой загрузки

```
python3 /opt/data/scripts/archive_upload.py \
  --file /путь/документ.pdf --category "03. Банки и финансы" \
  --title "Выписка СберБизнес 10.2026" --date 2026-10-31 \
  --type "Банковская выписка" --secret 4 --expiry 2027-10-31
```

Токен Paperless — файл `/opt/data/.secrets/paperless_token` (права 600), в реестр и логи не попадает.

n8n workflow: приём по webhook

- Адрес: `https://n8n.karpiev.ru/webhook/itg-doc-intake` (POST, JSON).

- Авторизация: заголовок `X-ITG-Intake-Key`, значение — в `/opt/data/.secrets/intake_webhook_key` (600).
- Тело: `{"filename": "...pdf", "source": "webhook", "content_base64": "...", "title": "..."}`.
- Разрешённые источники: manual, webhook, telegram, crm, paperless.
- Поведение: дубликат → уведомление без создания документа; ошибка → уведомление в Telegram владельцу.
- **Статус: протестирован, по умолчанию выключен.** Включение — только по решению владельца (обработка почты и массовый приём не включаются без проверки фильтров).

Контроль сроков

Для документов с заполненным полем «Срок действия» напоминания формируются за 30 дней, за 7 дней и в день окончания. Бессрочные документы (пустое поле) в напоминания не попадают, повторная отправка в один день исключена.

Права доступа

- Архив доступен только авторизованным пользователям Paperless; анонимный доступ отключён (`pdf.karpiev.ru` отдаёт 302 на страницу входа).
- Категории 03 (банки) и 06 (кадры) — строго конфиденциальные: доступ только владельцу.
- BookStack не содержит копий файлов; ссылки ведут на Paperless и требуют входа.
- API-токены хранятся в файлах с правами 600 вне веб-корня; в реестр и журналы секреты не попадают.
- Пользователь Paperless `hermes-archive` — суперпользователь только для задач разметки; пароль не задан (вход только по API-токену).

Интеграция Twenty CRM (этап 1)

- В Paperless используется поле «ID клиента или сделки Twenty CRM» — там хранится UUID записи CRM.
- В CRM хранятся только ссылки и метаданные (`https://pdf.karpiev.ru/documents/<id>/details`), копии оригиналов не дублируются.
- Существующие бизнес-процессы CRM не изменялись: связи создаются вручную или отдельным этапом.

- API Twenty (`https://crm.itgsystem.online/rest/...`) доступен по токену из конфигурации MODX; из контейнера Hermes прямой TLS до хоста не проходит (маршрутизация RU-origin) — вызовы выполняются через VPS-хост или n8n.

Резервное копирование

- Ежедневный vzdump LXC 106 в 03:00 на отдельное физическое хранилище `hdd-seagate` (Proxmox job `paperless-106-nightly`, zstd, keep-daily=7, keep-weekly=4).
- Дополнительно: дамп БД + `/opt/paperless_data` + конфигурация перед изменениями — `/var/lib/vz/dump/paperless-prearchive-20261105T142151Z` (права 600, SHA256SUMS).
- Проверка пригодности: пробное восстановление дампа в отдельную временную БД с подсчётом документов/тегов/типов/полей, затем удаление временной БД (выполнено 2026-11-05).

Отсутствующие документы (требуется загрузка владельцем)

Категория	Документ	Статус
01	Устав, свидетельства ОГРН/ИНН, листы записи ЕГРЮЛ	нет файла на сервере
02	Налоговые декларации и бухгалтерская отчётность 2026	нет файла на сервере
03	Выписки СберБизнес	нет файла на сервере
03	Справки и выписки Т-Банка	нет файла на сервере
05	Справки о задолженности, судебные документы	нет файла на сервере
06	Кадровые документы	нет файла на сервере
07	Лицензии и сертификаты	нет файла на сервере

Загрузка выполняется утилитой `archive_upload.py` (см. выше) или через workflow n8n после его включения.

Журналы

- Импорт: `/opt/data/itg-archive/import_log.jsonl`.
- Сводный реестр: этот раздел BookStack (синхронизируется из Paperless).
- Ошибки автоматизации: уведомление в Telegram (личный чат владельца) и журнал executions n8n.

Revision #2
Created 2026-11-05 14:30:15 UTC by Айнур
Updated 2026-11-05 14:56:06 UTC by Айнур