

archive.alai.no — Paperless-ngx Setup & Operations

archive.alai.no — Paperless-ngx Setup & Operations

URL: <https://archive.alai.no> **Backend:** Paperless-ngx (image `ghcr.io/paperless-ngx/paperless-ngx:latest`) **Host:** Azure VM `4.223.110.181` (alai-admin) **Container:** `alai-paperless-1` (with redis, gotenberg, tika sidecars) **MC reference:** [#9546](#), [#9982](#) (DR backup TODO)

“ Document management system za sve ALAI-srodne legalne, ugovorne, partnerske, istraživačke i finansijske dokumente. OCR, full-text search, taxonomy.

Access requirements

CF stack (oba sloja) traže `92.221.168.61/32` (ALAI LAN egress) u bypass listama. Vidi [CF IP Access Rules — ALAI LAN Bypass](#).

Iz Mac Studio sa aktivnim VPN-om: bind interface `192.168.68.65` (Deco LAN) zaobilazi VPN routing:

```
curl --interface 192.168.68.65 https://archive.alai.no/...
```

Mac Air i ostali bez VPN-a: direktno radi.

API authentication

Paperless koristi DRF Token auth.

Token za admin user (root@localhost) sačuvan lokalno na Mac Studio:

```
~/config/alai/paperless-token.env (mode 600)
```

```
PAPERLESS_TOKEN=c9ec30192db3c95802349335edea4bca864a937a
PAPERLESS_BASE=https://archive.alai.no
PAPERLESS_BIND_INTERFACE=192.168.68.65
```

Svi API zahtjevi:

```
Authorization: Token c9ec30192db3c95802349335edea4bca864a937a
```

Regenerate token (ako compromised — Django shell preko docker exec):

```
ssh -b 192.168.68.65 -i ~/.ssh/azure_alai alai-admin@4.223.110.181 \
'docker exec alai-paperless-1 python manage.py shell -c "
from rest_framework.authtoken.models import Token
from django.contrib.auth import get_user_model
u = get_user_model().objects.get(username="admin")
Token.objects.filter(user=u).delete()
print(Token.objects.create(user=u).key)
"'
```

Schema (taxonomy)

Setup-ovan 2026-04-28 preko `/tmp/paperless-setup.sh`. ID-evi mogu varirati po instanci — koristi `name__iexact` za lookup.

Document Types (14 base, currently 25 active)

Contract, LOI, NDA, Registration, Insurance Policy, Research Paper, Invoice, Receipt, Email Archive, Identity Document, Tax Document, Financial Statement, Meeting Notes, Pitch Deck — plus historical types from prior usage. Numbers grow naturally; verify current via API.

Tags (23 base, currently 39 active, color-coded)

Cross-cutting (cilj): `legal`, `research`, `kuran-19`, `partnership`, `regulator`, `contract`, `nda`, `loi`, `invoice`, `registration`, `urgent`, `signed`, `pending-signature`

Company tags: ALAI, Drop, Bilko, Tok, Lobby, LumisCare, Plock, ALAI-Tech-D00, BasicConsulting, client

Storage Paths (21)

Folder hierarhija po kompaniji + funkciji:

```
/ALAI/legal/{created_year}/{title}
/ALAI/research/kuran-19/{title}
/ALAI/research/general/{created_year}/{title}
/ALAI/partnerships/sintef/{title}
/ALAI/partnerships/intesa/{title}
/ALAI/partnerships/pbz/{title}
/ALAI/regulators/finanstilsynet/{created_year}/{title}
/ALAI/regulators/skatteetaten/{created_year}/{title}
/ALAI/regulators/brønnoysund/{created_year}/{title}
/ALAI/contacts/{title}
/Drop/legal/{created_year}/{title}
/Drop/contracts/{title}
/Bilko/legal/{created_year}/{title}
/Bilko/contracts/{title}
/Tok/legal/{created_year}/{title}
/Lobby/legal/{created_year}/{title}
/LumisCare/legal/{created_year}/{title}
/Plock/legal/{created_year}/{title}
/ALAI-Tech-D00/legal/{created_year}/{title}
/BasicConsulting/{created_year}/{title}
/clients/Entur/{created_year}/{title}
```

Initial Correspondents (11 seeded, currently 25 active, auto-expand)

SINTEF, Finanstilsynet, Skatteetaten, Brønnøysundregistrene, PBZ Zagreb, Intesa Sanpaolo, Anthropic, Cloudflare, Tryg, Fiken AS, Entur AS — auto-create on classify match.

Upload workflow

Manual single file

```
source ~/.config/alai/paperless-token.env
curl -s --interface "$PAPERLESS_BIND_INTERFACE" \
  -H "Authorization: Token $PAPERLESS_TOKEN" \
  -F "title=My Document" \
  -F "storage_path=1" \
  -F "tags=30" -F "tags=17" \
  -F "document=@/path/to/file.pdf" \
  -X POST "$PAPERLESS_BASE/api/documents/post_document/"
```

Returns task UUID. Verify success via:

```
curl ... "$PAPERLESS_BASE/api/tasks/?task_id=<UUID>"
```

Batch upload sa klasifikacijom

Skripta: `/tmp/paperless-classify-v2.py` (commit u repo-u TBD)

```
python3 /tmp/paperless-classify-v2.py --dry --all # dry-run all ~/ALAI/*
python3 /tmp/paperless-classify-v2.py --all # actual upload
python3 /tmp/paperless-classify-v2.py FILE [FILE...] # specific files
```

Klasifikator mapira path → (storage_path, correspondent, document_type, tags) prema rules engine-u. Pre-upload dedup po normalized title; Paperless takođe ima vlastiti content-hash dedup (rejects file ako mu je sadržaj već prisutan).

Operations cheat sheet

```
# Document count
curl ... "$BASE/api/documents/?page_size=1" | jq '.count'

# Latest 10 docs
curl ... "$BASE/api/documents/?ordering=-created&page_size=10" | jq
'.results[]|{id,title,created}'

# Search by tag
curl ... "$BASE/api/documents/?tags__id=17" | jq '.results[].title'
```

```
# Search by storage path
curl ... "$BASE/api/documents/?storage_path__id=1"

# Full-text search (OCR'd content)
curl ... "$BASE/api/documents/?query=finanstilsynet"

# Task queue status
curl ... "$BASE/api/tasks/?page_size=200" | jq
'group_by(.status)|map({status:.[0].status,count:length})'

# Failed tasks (often = content duplicates)
curl ... "$BASE/api/tasks/" | jq
'[.[]|select(.status=="FAILURE")|{file:.task_file_name,reason:.result}]'
```

Architecture

```
[ALAI LAN egress 92.221.168.61]
  |
  ▼
[Cloudflare]
  ├── IP Access Rule: bypass WAF (Layer 1)
  └── CF Access policy: bypass Zero Trust (Layer 2)
    |
    ▼
[Caddy on Azure VM 4.223.110.181]
  archive.alai.no → paperless-ngx:8000
    |
    ▼
[alai-paperless-1 container]
  ├── alai-paperless-redis-1 (queue)
  ├── alai-paperless-gutenberg-1 (PDF preview)
  └── alai-paperless-tika-1 (text extraction)
    |
    ▼
[Postgres + media volume on Azure VM]
```

Web login

CEO `alembasic` superuser created 2026-04-28. Initial password rotirana — koristi BW item ili lični password.

Pristup sa Mac Air (LAN egress 92.221.168.61, u CF Access bypass) → direktno na `https://archive.alai.no` bez CF SSO challenge. Login Paperless web UI sa username + password. Promijeni password kroz Profile → Change Password.

Iz Mac Studio (VPN aktivan) — backend dostupan ali samo via API sa bind interface, ne web browser (browser ne prima `--interface` flag).

Outstanding (TODO)

- **MC #9982** — DR backup automation: `pg_dump` cron + media volume snapshot + B2/R2 upload + 30-day retention
- **Bitwarden token storage** — `bw create item` blocked by node 25 incompat (`Invalid version` error). Manually add via Vaultwarden web UI ako traje
- **Token rotation policy** — currently no expiry; consider 90-day rotation za admin token
- **Per-user tokens** — kreiraj user-specific tokens za audit trail (admin token shared = no per-user audit)

Related

- [CF IP Access Rules — ALAI LAN Bypass](#) — both layers documented
- [DEPLOY-MAP — System Infrastructure](#) — CF Access policies + Paperless API entry
- [ZAKON NETWORK EGRESS](#) — VPN exit vs ISP egress
- **Incident origin:** 2026-04-28 ALAI legal docs upload task — discovered Paperless instance had 58 pre-existing docs; after dedup-aware bulk upload, 99 docs total

Revision #3

Created 2026-04-28 13:12:48 UTC by John

Updated 2026-05-31 20:06:47 UTC by John