COMPLIANCE-022 — Archive Review (HIPAA/GDPR/CQC)

MC #100025 | Published 2026-05-08 | Status: Approved (Pattern 3 — Skybound) / Compliance gate pending (Dr. Sarah Chen M3+M5 blockers)

Related: ADR-022 • SPEC-022

⚠ **PRE-EMPTIVE BLOCKERS** — Pattern 3 cannot ship to production with EU personal data until:

- **(M3)** Azure VM disk encryption verified
- **(M5)** GDPR Art. 28(4) sub-processor DPA chain documented in Bilko Terms + Privacy Notice

See section 9 for full MUST list.

COMPLIANCE-022: Healthcare & Privacy Compliance Review

Bilko Document Archive — Pattern 3 (Blob Queue) ADR-022 / SPEC-022

Reviewer: Dr. Sarah Chen, Healthcare IT Systems Architect **Date:** 2026-05-08 **MC:** #100025 Subtask 3 of 5 **Status:** Final — sign-off conditions in §10

---

1. Scope — Applicable Regulations

Jurisdiction and context

Bilko is a Balkan accounting SaaS (Serbia, BiH, Croatia), EU residency claimed (GCP europe-north1), operated by ALAI Holding AS (Norway). The doc types named in ADR-022 §Context include care plans and incident reports. Those two types trigger healthcare regulatory scope even in an accounting product.

Regulations evaluated

Regulation	Trigger	Applies?
---	---	---
GDPR / EU GDPR (Regulation 2016/679)	EU residency, Balkan clients in EU data space, special category Art. 9 data possible in care plans	YES — primary
HITECH Act (US)	Only if Bilko serves US-based covered entities or their BAs. No US presence confirmed in BUILD-BLUEPRINT.	NOT YET — but architecture must not preclude compliance if US expansion occurs
HIPAA Privacy + Security Rules	Same trigger as HITECH.	NOT YET — apply when US expansion scoped
CQC / Health and Social Care Act 2008	Only if Bilko serves UK-registered domiciliary care agencies. Not confirmed.	NOT YET — same comment
NIS2 Directive (EU 2022/2555)	ALAI Holding AS as digital infrastructure provider processing health data above medium-enterprise threshold. Likely not in scope at current scale but architecture must support NIS2-compliant incident response by design.	MONITOR — review at 50+ orgs
Norway Bokføringsloven §13	Invoices, financial records, 7-year retention	YES — invoices
Serbia Zakon o računovodstvu / Croatia equivalents	Same financial retention	YES — domain packages
GDPR Art. 17 (Right to Erasure)	Active for all EU data subjects	YES — open gap in SPEC-022 §10.4
GDPR Art. 28 (Sub-processor chain)	ALAI Azure VM Paperless is a sub-processor of Bilko	YES — gap in both documents

Legal basis assumed

For care plans and incident reports: **GDPR Art. 6(1)(b)** (contract performance) as primary basis; **Art. 9(2)(h)** (health/social care purposes) as special-category basis. This must be reflected in Bilko's Privacy Notice and any DPA issued to tenants.

---

2. Data Classification

Document Type	GDPR Classification	Special Category (Art. 9)?	Financial Record?	Recommended Paperless Tag
---	---	---	---	---
Invoice	Personal data (contact name, address, VAT ID)	No	Yes (Bokføringsloven, 7y)	`data-class:financial`
Contract	Personal data (signatories, company data)	No	Quasi-financial (5y post-expiry)	`data-class:legal`
Care plan	**Special category health data**	YES — diagnosis, medication, functional status	No	`data-class:health` `sensitivity:high`
Incident report	**Special category health/social data**	YES — if describes injury, clinical event	Potentially	`data-class:health` `sensitivity:high`
Onboarding document	Personal data (identity verification, scanned ID)	No (unless medical screen)	No	`data-class:identity`

Tag strategy amendment

SPEC-022 §5.2 defines four tag types: `org:`, `doc-type:*`, `bilko-source`, `bilko-source-uuid:`.

Missing: data classification tags. The `PAPERLESS_TAG_IDS_MAP` env var must include entries for `data-class:health`, `data-class:financial`, `data-class:legal`, `data-class:identity`, and `sensitivity:high`. These are required for:

- Retention policy enforcement (different rules per class)
- Access control (human admins in Paperless must not see `sensitivity:high` docs without justification)
- Incident response scoping (breach = all `data-class:health` docs in affected org)

---

3. Audit Trail Requirements

3.1 What SPEC-022 §9.3 provides

`archive_audit_log` records: who queued the archive (implicit — ArchiveService called in user request context), R2 object key, sha256, status transitions, timestamps, Paperless doc ID, retry count, errors.

This covers the **archival pipeline itself** adequately.

3.2 Critical gap — per-access logging for archived documents

SPEC-022 contains **no provision** for logging human access to archived documents in Paperless.

When a CEO-level user or ALAI admin opens a care plan or incident report in the Paperless UI at `archive.alai.no`, there is no audit record in any Bilko system.

GDPR Art. 5(1)(f) (integrity and confidentiality) and, when US healthcare clients are added, **HIPAA §164.312(b)** (audit controls) require that every access to records containing personal or health data is logged with:

- Viewer identity (Paperless username or service account)
- Document ID and document type
- Timestamp (UTC)
- Source IP address
- Access outcome (viewed, downloaded, printed)

Paperless-ngx does not natively emit per-document access logs to an external SIEM. It maintains an internal Django audit trail (`auditlog` tables), but that trail lives on the Azure VM and is not exported to GCP Cloud Logging where Bilko's other audit records live.

Gap: no tamper-evident export of Paperless access logs.

3.3 Retention of access logs

GDPR Article 5 + Recital 39 require demonstrability — logs must be retained long enough to respond to a subject access request or supervisory authority inquiry. Minimum: same retention as the documents they describe. For care plans (25 years per SPEC-022 §6.2), access logs must survive 25 years. For invoices, 7 years.

SPEC-022 §6.2 is silent on access log retention.

3.4 Tamper evidence

The `archive_audit_log` table in Bilko DB is defined in SPEC-022 §9.3. It has no tamper-evidence mechanism (no hash chaining, no write-once constraint beyond application code). PostgreSQL row-level updates are possible for any user with DB access.

Minimum required: ensure `archive_audit_log` has no application-level UPDATE path for `created_at` and core fields (`sha256`, `organization_id`, `bilko_document_id`). A DB-level check constraint or trigger preventing modification of those columns after insert provides tamper-resistance without requiring a separate append-only log infrastructure.

---

4. Access Control Deltas

4.1 Worker process — least privilege (SPEC-022 §4)

SPEC-022 §4.2 recommends a separate R2 API token for the worker scoped to the two archive buckets. SPEC-022 §4.3 recommends a separate Paperless API token. Both are correct. No gap here from an access control standpoint.

4.2 Human admin access to Paperless — unaddressed

Neither ADR-022 nor SPEC-022 defines who may log into `archive.alai.no` as a human user and what they can access. Currently, the only documented credential is `alembasic` (Bitwarden item referenced in ADR-022 §Context). That is a single superuser account.

For multi-tenant data containing health records:

- Superuser access = unrestricted cross-tenant document access
- No audit of which documents the superuser viewed
- No segregation between financial docs (low sensitivity) and care plan / incident docs (high sensitivity)

Required additions:

- Create a Paperless `bilko-ops` service account for operational tasks (queue monitoring, DLQ triage)

with access scoped to `bilko-source` tagged documents only, no `sensitivity:high` filter bypass.

- The CEO (`alembasic`) personal account must not be used for routine Paperless access once healthcare

doc types are live. A named admin account with restricted permissions per document type is required.

- Paperless does not natively enforce per-tag ACLs. This means cross-tenant isolation in the Paperless

UI relies entirely on the discipline of human users to filter by `org:` tag. This is not adequate for healthcare data.

4.3 Cross-tenant containment — tag-based vs. physical separation

SPEC-022 §5.3 states: "Cross-tenant queries are impossible if the caller only has access to their own `org_tag_id`. Isolation is enforced by the Bilko application layer controlling which `org_tag_id` each user can query."

This is **correct for machine-to-machine access** (worker reads, API queries). It is **not sufficient for human access** to the Paperless UI, where all documents from all tenants are visible to any logged-in user. Until Paperless supports per-tag or per-user-group ACLs (which it does not as of v2.x), physical separation — one Paperless instance per tenant — is the only way to enforce tenant isolation for human UI access.

Recommendation (SHOULD — not an immediate ship blocker provided care plans are not in scope for MVP): Before enabling care plan or incident report archival through this pipeline, deploy per-tenant Paperless instances or ensure the Paperless UI is not accessible to any human user other than a designated compliance officer who has executed an appropriate access agreement and whose access is logged separately.

4.4 Break-glass access procedure

Neither document defines a break-glass procedure: how does ALAI access a specific tenant's archived documents if the Bilko DB `org_paperless_cache` is corrupted or unavailable?

Required: Document and test a break-glass procedure: (a) query Paperless directly by `org:` tag using the bilko-ops service account, (b) log the access reason and approver, (c) notify the affected tenant within 72 hours if the access was to health data.

---

5. Sub-Processor Analysis

5.1 The data flow chain

```
Bilko tenant (data subject's data)
  → Bilko Cloud Run API (controller / data processor acting on behalf of tenant)
    → Cloudflare R2 (sub-processor #1 — staging queue)
      → archiver-worker Cloud Run (internal processor — ALAI infrastructure)
        → ALAI Azure VM / Paperless-ngx (sub-processor #2 — long-term storage)
```

5.2 Gap: no GDPR Art. 28 chain documented

GDPR Art. 28(4) requires that where a processor engages a sub-processor, the same data protection obligations as set out in the controller-processor contract are imposed on the sub-processor.

ADR-022 notes "Paperless-ngx at archive.alai.no = ALAI Azure VM (separate org from Bilko tenants). Cross-org data flow = sub-processor relationship; needs DPA articulation" — and then defers to this review. SPEC-022 does not address it at all.

Minimum required DPA chain articulation:

1. Bilko's Terms of Service / DPA with each tenant must list: - Cloudflare (R2) as a sub-processor - ALAI Holding AS hosting (Azure VM, Paperless) as a sub-processor

2. The existing ALAI AI Services Legal Pack (BookStack shelf `https://docs.alai.no/shelves/ai-services-legal-pack`, TOMs published) provides a DPA template. That template must be extended with a Schedule listing sub-processors and their processing purposes. For the archive pipeline: Purpose = "Long-term document retention for audit and compliance purposes"; Location = EU (Azure westeurope); Retention per §6.2 of SPEC-022.

3. ALAI must have a DPA with Microsoft Azure (for the VM hosting Paperless). Standard Microsoft Online Services DPA covers this if the Azure subscription is enrolled — verify this is in place.

4. Bilko tenants uploading care plans or incident reports must be explicitly informed (Privacy Notice update) that health data is stored in Paperless on an ALAI-operated EU server.

5.3 Cloudflare R2 sub-processor status

Cloudflare R2 is covered by Cloudflare's standard Data Processing Addendum. ALAI should confirm it is signed as part of the Cloudflare account setup. The R2 bucket must be configured to a confirmed EU jurisdiction (Cloudflare R2 location hint `WEUR` or `EEUR`).

---

6. Encryption Requirements

6.1 At rest — R2 (staging queue)

Cloudflare R2 provides AES-256 encryption at rest by default for all objects. No customer-managed key option was selected per SPEC-022 §4. For current Bilko document types (invoices, contracts), platform-managed encryption is adequate. For care plans and incident reports (special category health data), consider whether tenant-controlled encryption keys are a contractual requirement with any healthcare clients before that doc type goes live.

6.2 At rest — Paperless on Azure VM

SPEC-022 does not confirm disk encryption on the Azure VM hosting Paperless. Azure VM OS disks are **not encrypted by default** — Azure Disk Encryption (ADE using BitLocker/DM-Crypt) or server-side encryption with customer-managed keys must be explicitly enabled. This must be verified by FlowForge before any healthcare document type is archived.

Required (MUST): Confirm Azure VM hosting Paperless has disk encryption enabled. Run `az vm encryption show --name  --resource-group` and include output in the ship checklist evidence.

6.3 In transit — GCP Cloud Run to R2

Cloudflare R2 S3-compatible API enforces TLS 1.2+ on all endpoints. Confirmed adequate.

6.4 In transit — archiver-worker to Paperless (archive.alai.no)

ADR-022 §Context: Paperless is "behind Cloudflare Access (service token required)". Cloudflare Access enforces HTTPS on all traffic to the origin. The origin-to-Cloudflare tunnel should use Cloudflare Tunnel (cloudflared) or an authenticated origin pull — confirm this is configured so the Azure VM does not expose port 443 directly to the internet.

If the Azure VM is exposed directly (no cloudflared), a misconfigured security group could allow direct HTTP access bypassing CF Access entirely. FlowForge must confirm the network path.

6.5 Field-level encryption

Field-level encryption of PDF content is not feasible within this architecture and is not required at this stage. The PDF is the record. Encryption at transport and at rest is the appropriate control. If any structured extracted fields from care plans are ever stored in Bilko DB as queryable columns,

those columns must be treated as special category data and assessed for column-level encryption.

---

7. Erasure / Right to Be Forgotten (GDPR Art. 17)

7.1 Current state

SPEC-022 §10.4 acknowledges erasure as an open question: "a separate erasure worker, out of scope for this implementation phase."

ADR-022 §Phase 4 (Q3) provides a three-step Paperless erasure process (query by `org:` tag, delete documents, delete correspondent). This is architecturally sound.

7.2 Interim recommendation (required before care plans go live, SHOULD before MVP)

GDPR Art. 17(1) requires that erasure be executed "without undue delay." For a SaaS with a documented erasure process, "without undue delay" means the capability must exist and be operable when a valid erasure request arrives — it does not require automatic self-service.

For MVP (invoices, contracts, onboarding — not health data): an operationally-documented manual erasure procedure is acceptable interim. The procedure must be documented in the RUNBOOK.md and tested before any EU data subject data reaches production.

Manual erasure procedure (document in RUNBOOK.md before MVP ship):

1. Receive verified erasure request (tenant admin or data subject via support ticket). 2. Confirm no legal hold applies (Bokføringsloven 7-year financial record exception — invoices cannot be erased within retention period under legitimate interest override). 3. Delete Bilko DB records for the org (existing DB delete cascade paths — confirm with CodeCraft). 4. Query R2 for any pending queue objects: `aws s3 ls s3://bilko-archive-queue/org//` and delete all. 5. Query Paperless: `GET /api/documents/?tags__id__in=`, delete all results. 6. Delete Paperless correspondent: `DELETE /api/correspondents/`. 7. Delete `org_paperless_cache` row for the org. 8. Log erasure completion with timestamp and executor identity.

Before care plans or incident reports are archived: an automated erasure worker is required (child MC, FlowForge). Manual erasure for health records under Art. 17 is too slow and too error-prone.

7.3 Financial record exception

Invoices subject to Bokføringsloven §13 (Norway) or equivalent (Serbia, Croatia, BiH) cannot be erased within the mandatory retention period even on Art. 17 request. The Privacy Notice must inform data subjects of this limitation. The erasure procedure must check document type and skip financial records with a logged exception.

---

8. Incident Response / Breach Notification

8.1 Breach scenarios

Scenario	Severity	GDPR notification	Responsible party
---	---	---	---
Bilko Cloud Run API compromise (R2 staging queue exposed)	HIGH if health data in queue	72h to supervisory authority (Datatilsynet, Norway; or relevant Balkan DPA)	ALAI (as Bilko operator)
Azure VM compromise (Paperless data exposed)	HIGH	72h — triggers sub-processor notification chain: ALAI Azure → ALAI Bilko team → tenant notification	ALAI (as sub-processor); tenant notifies their data subjects
Worker credential leak (CF Access + Paperless API token)	MEDIUM-HIGH (allows read of all archived docs across all tenants)	72h if PHI/health data accessible	ALAI
Cross-tenant Paperless UI access (human error)	MEDIUM	72h if health data accessed	ALAI

8.2 Notification chain (required in RUNBOOK.md)

Neither ADR-022 nor SPEC-022 defines a breach notification chain. The following must be documented:

1. **Detection:** Cloud Monitoring alert (unauthorised 401/403 spike, DLQ depth spike, anomalous ListObjectsV2 calls from unexpected IP) fires to `dev@alai.no`. 2. **Triage:** Within 1 hour — ALAI ops determines whether PHI/PII was exposed. 3. **Internal declaration:** ALAI Compliance (Alem Basic as DPO for current scale) declares breach. 4. **Supervisory authority notification:** Within 72 hours of awareness — notify Datatilsynet (Norway) via `https://www.datatilsynet.no/en/about-privacy/notification-of-a-data-breach/`. If Serbian or Croatian data subjects affected: notify relevant authority (POVP, Serbia; AZOP, Croatia) simultaneously. 5. **Tenant notification:** Within 72 hours — notify affected tenant(s) via documented contact (tenant owner email on record in Bilko DB). 6. **Data subject notification:** If "likely to result in a high risk to rights and freedoms" (Art. 34), notify data subjects directly. Care plan or incident report exposure = high risk threshold met automatically.

---

9. Recommended Changes

MUST — compliance blockers (must fix before production ship)

ID	Document	Section	Required change
---	----------	----------	----------
M1	SPEC-022	§5.2	Add `data-class:health`, `data-class:financial`, `data-class:legal`, `data-class:identity`, and `sensitivity:high` to `PAPERLESS_TAG_IDS_MAP`. Worker must apply `data-class` and (for care plans / incident reports) `sensitivity:high` tag on every archive call.

M2	SPEC-022	§9.3	Add DB-level protection on `archive_audit_log`: a Postgres trigger or check constraint must prevent UPDATE of `organization_id`, `sha256`, `bilko_document_id`, and `created_at` after row insert. Append-only semantics enforced at DB layer, not only application layer.
M3	ADR-022 + SPEC-022	§4 / §Context	Document and verify Azure VM disk encryption is enabled before care plans or incident reports are archived. Add to ship checklist: `az vm encryption show` output as evidence.
M4	SPEC-022	§10.4	Document manual erasure procedure in RUNBOOK.md (see §7.2 of this review) before MVP ship. Must include: financial record exception logic, Paperless deletion steps, audit log of erasure.
M5	ADR-022	§Consequences	Update Bilko Terms of Service / Privacy Notice and sub-processor DPA to list Cloudflare R2 and ALAI Azure VM (Paperless) as sub-processors per GDPR Art. 28(4). This must exist before any EU personal data flows through the archive pipeline. Must reference ALAI AI Services Legal Pack DPA template on BookStack.
M6	SPEC-022	§4 / §9	Paperless access log export: configure Paperless Django audit log export (or Cloudflare Access request logging for `archive.alai.no`) to ship access events to Cloud Logging. Access log entries must contain: user/service account identity, document ID, document type, timestamp, source IP. Retain per document class retention period.

SHOULD — best practice (not immediate ship blockers)

ID	Document	Section	Recommended change
---	--------	--------	----------
S1	SPEC-022	§5.3	Before enabling care plan or incident report doc types, assess whether tag-based isolation in the shared Paperless instance is sufficient or whether a dedicated per-healthcare-tenant Paperless instance is required. Tag isolation is adequate for machine queries but not for human Paperless UI access.
S2	SPEC-022	§4.3	Replace `INTERNAL_API_KEY` shared secret for worker-to-api callback with GCP Cloud Run service-to-service OIDC auth (already in SPEC-022 §10.5 as Phase 2 item). Shared secret is a credential management risk. This is already flagged; confirm it is a Phase 2 child MC, not indefinitely deferred.
S3	ADR-022	§Phase 4	Create child MC for automated erasure worker before enabling care plan archival. Manual erasure is not appropriate for health data under GDPR Art. 17.

S4	SPEC-022	§6.2	Add care plan retention to 25 years in Paperless Workflow rule (SPEC-022 already notes this as out of scope). File the child MC before health doc types go live. 25-year retention is a CQC/NHS standard; for Balkan jurisdiction equivalents, confirm with local counsel (no equivalent statutory period confirmed for Serbia/BiH/Croatia).
S5	SPEC-022	§10	Add Breach Notification Runbook to RUNBOOK.md (§8.2 of this review) as child MC. Required before any production data flows through the pipeline.
S6	ADR-022	§Context	Verify Cloudflare R2 bucket `bilko-archive-queue` location hint is set to `WEUR` or `EEUR` to maintain EU data residency. Not confirmed in either document.

---

10. Sign-Off Conditions

The following must be true before Pattern 3 ships to production. Each item maps to a MUST above.

1. **[M5] Sub-processor DPA chain published.** Bilko ToS / Privacy Notice lists Cloudflare R2 and ALAI Azure VM as sub-processors. Bilko tenant DPA template updated. Evidence: BookStack page with published DPA addendum (reference ALAI AI Services Legal Pack shelf).

2. **[M1] Data classification tags deployed in Paperless.** `data-class:*` and `sensitivity:high` tags exist in Paperless, IDs populated in `PAPERLESS_TAG_IDS_MAP`, worker applies them. Evidence: Proveo test showing a care plan doc archived with `data-class:health` + `sensitivity:high` tags visible in Paperless.

3. **[M3] Azure VM disk encryption verified.** FlowForge provides `az vm encryption show` output confirming encryption enabled on the VM hosting Paperless. Evidence: output attached to ship checklist.

4. **[M2] Archive audit log tamper-protection deployed.** Flyway migration adds DB-level constraint on `archive_audit_log`. Evidence: Proveo attempts direct SQL UPDATE on `created_at` and

`sha256` columns and confirms rejection.

5. **[M6] Paperless access log export live.** Cloudflare Access request logs for `archive.alai.no` (or Paperless Django auditlog export) flowing to Cloud Logging. Evidence: Cloud Logging query showing access log entries from a test document retrieval.

6. **[M4] RUNBOOK.md updated with manual erasure procedure.** Procedure includes financial record exception, Paperless deletion steps, confirmation of `org_paperless_cache` cleanup. Evidence: Proveo executes erasure procedure end-to-end in staging and documents result.

Pre-emption clause: Items M3 (Azure disk encryption) and M5 (DPA chain) are pre-emptive — they must be resolved before any personal data of any kind is archived in Paperless production. They are not "ship before care plans go live" items; they are "ship before any data flows" items. If either is unresolved at production launch, the pipeline must be restricted to internal test data only via a feature flag.

Reviewed against: ADR-022 (all sections), SPEC-022 (all sections §1–§10), BUILD-BLUEPRINT.md (multi-tenancy model, GCP deployment, R2 config). GDPR 2016/679 Arts. 5, 6, 9, 17, 28, 34; HIPAA §164.312 (noted for future US expansion); CQC Key Lines of Enquiry Safe domain (noted for future UK healthcare expansion); NIS2 Directive 2022/2555 (monitor threshold).

Revision #4
Created 2026-05-08 19:29:47 UTC by John
Updated 2026-07-19 20:00:57 UTC by John