

**AI-BOK TOOLKIT v1.2**

# Filled Example Templates

Companion to the AI Body of Knowledge v1.2

Jan Willem van Veen · ArchiXL · ai-bok.nl

# Worked-Filled Templates

---

Five worked-filled instances of the templates (Authority Register entry, Mandate Record, Article 4 evidence file, Failure-mode coverage assessment, Cognition-plane ADR) against the composite Stadsdienst Veenwoorden case.

## 1. Authority Register entry - `chra-v2` (CHRA case-routing agent)

---

```
authority_register_entry:
  agent_id: nhi:chra-v2:prod
  agent_kind: agentic
  system: chra-v2
  mandate_ref: mandate://veenwoorden/chra-v2#2026-Q2
  mandate_issued: 2026-04-01
  mandate_expires: 2026-07-01
  mandate_status: active
  chain_of_custody:
    - issued_by: mandate://veenwoorden/ai-board#2026-Q2-001
  issued_on: 2026-04-01
  signature: ed25519:...
  scope_summary:
    - classify_permit (low, full autonomy)
    - route_permit (low, full autonomy)
    - route_permit (medium, HITL)
    - escalate_to_supervisor (high, full autonomy)
  exclusivity: per-case-id
  freshness_budget: 24h
  operating_envelope:
    knowledge_source_trust_floor: high
    policy_engine_freshness: 7d
    sample_rate: n/a (centre tier)
    hardware_health: n/a (centre tier)
  current_kpis:
    actions_last_24h: 412
    escalations_last_24h: 18
    escalation_rate: 4.4%
  mandate_cache_age_max: 18h (within budget)
  fm_coverage_at_runtime: 22/22
  next_review: 2026-06-30
  revocation_status: none
  revocation_endpoint: https://authority.veenwoorden.nl/api/v1/revoke
```

## 2. Mandate Record - `chra-v2` 2026-Q2

---

```
mandate:
  id: mandate://veenwoorden/chra-v2#2026-Q2
  agent: nhi:chra-v2:prod
  version: 2026-Q2
  issued: 2026-04-01
  expires: 2026-07-01
  signed_by:
    - role: AI Board chair
      signature_alg: ed25519
      signature:...
    - role: AI Risk Manager
      signature_alg: ed25519
      signature:...
  freshness_budget: 24h
  exclusivity: per-case-id
  scope:
    - action: classify_permit
      autonomy: full
      risk_band: low
    preconditions:
      - knowledge_sources_trust_grade ≥ high
      - confidence ≥ 0.85
  escalation_on:
    confidence < 0.85
    - action: route_permit
      autonomy: full
      risk_band: low
    - action: route_permit
      autonomy: human_in_the_loop
      risk_band: medium
    hitl_role: case_supervisor
    hitl_max_response: 15min
    fallback_on_timeout: queue_for_supervisor
    - action: escalate_to_supervisor
      autonomy: full
      risk_band: high
    - action: communicate_with_citizen
      autonomy: prohibited
  reason: KA7 reserves direct citizen communication to humans
  operating_envelope:
    knowledge_source_trust_floor: high
    policy_engine_freshness: 7d
    failure_mode_coverage_assessment_ref:
      chra-v2-fm-coverage-2026-Q2
  superseded_by: null
  retirement_plan: linked to portfolio review 2026-Q3
```

### 3. Article 4 evidence file - operator population for `chra-v2`

---

```
article_4_evidence_file:
  system: chra-v2
  reporting_period: 2026-Q2
  competency_profiles:
    - role_group: case_handler_operator
  version: 2026.1
  required_modules:
    - chra-v2-supervision-essentials
    - prompt-injection-recognition
    - escalation-handling
  effective_from: 2026-03-10
    - role_group: case_supervisor
  version: 2026.1
  required_modules:
    - agentic-escalation-handling
    - audit-trail-interpretation
    - bias-recognition
  effective_from: 2026-03-12
  literacy_register:
  operators_in_scope: 24
  operators_current: 24
  operators_current_pct: 100%
  supervisors_in_scope: 6
  supervisors_current: 6
  supervisors_current_pct: 100%
  register_last_attested: 2026-06-01
  drills_conducted:
    - drill: prompt-injection-recognition
  date: 2026-03-15
  participants: 24
  pass_rate: 92%
    - drill: agentic-escalation-tabletop
  date: 2026-04-20
  participants: 6
  outcome: 1 procedure update arising from drill (queue handover protocol)
  measurement:
  competency_assessments_completed: 30
  pass_rate: 100%
  escalations_resolved_correctly_pct: 89% # target 90%; sub-threshold flagged
  incidents_with_literacy_root_cause: 0
  next_refresh_due: 2026-09-10 (operators), 2026-09-12 (supervisors)
  next_full_review: 2026-09-30
  auditor_access: read-only via dpo@veenwoorden.nl
```

## 4. Failure-mode coverage assessment - chra-v2

```
fm_coverage_assessment:
  system: chra-v2
  assessment_date: 2026-04-15
```

```
next_assessment: 2026-07-15
assessor: AI Risk Manager (Role 6)
coverage:
FM01 prompt_injection_direct:
control: C-FM01-1 (privilege separation) + C-FM01-2 (instruction hierarchy)
status: operational
evidence_ref: chra-v2-arch-pattern-1, classifier-baseline-2026-Q2
FM02 prompt_injection_indirect:
control: C-FM02-1 (allowlist+signing) + C-FM02-2 (structural separation)
status: operational
evidence_ref: trusted-source-register, rag-pipeline-pattern-1
FM03 tool_misuse:
control: C-FM03-1 (per-tool gating) + C-FM03-2 (pre-flight HITL)
status: operational
evidence_ref: authority-register-policy-v1
FM04 unauthorised_delegation:
control: C-FM04-1 (mandate non-inheritance)
status: operational (no sub-agents currently spawned)
evidence_ref: mandate-mint-log-2026-Q2
FM05 multi_agent_conflict:
control: C-FM05-1 (per-object exclusivity)
status: operational
evidence_ref: exclusivity-conflict-log (empty in Q2)
FM06 memory_poisoning:
control: C-FM06-1 + C-FM06-2 (provenance + revalidation)
status: operational
evidence_ref: knowledge-source-provenance-policy-v1
FM07 context_saturation:
control: C-FM07-1 (input-size guardrails)
status: operational
evidence_ref: input-size-policy
FM08 function_call_abuse:
control: C-FM08-1 (parameter anomaly detection)
status: operational
evidence_ref: anomaly-baseline-2026-Q2
FM09 goal_drift:
control: C-FM09-1 (long-window drift monitor)
status: monitoring (not enough history yet for a stable baseline)
evidence_ref: drift-monitor-onboarding-doc
review_due: 2026-09-15
FM10 sycophancy:
control: C-FM10-1 (hard confidence thresholds)
status: operational
evidence_ref: chra-v2-confidence-thresholds
FM11 jailbreak_persistence:
control: C-FM11-1 (per-NHI baseline + rate limit)
status: operational
evidence_ref: rate-limit-policy
FM12 nhi_credential_leak:
control: C-FM12-1 (short-lived NHI + rotation)
status: operational
evidence_ref: identity-rotation-log-2026-Q2
FM13 sandbox_escape:
```

```

control: C-FM13-1 (tool sandbox)
status: operational (no tool calls outside sandbox in Q2)
evidence_ref: sandbox-policy
FM14 autonomy_creep:
control: C-FM14-1 (mandate-version verification)
status: operational
evidence_ref: policy-consult-log (100% mandate version coverage)
FM15 capability_inheritance:
control: C-FM15-1 (sequence-level policy)
status: monitoring (composite-mandate catalogue v0.3 in build)
evidence_ref: composite-rule-catalogue-v03
review_due: 2026-07-15
FM16 hallucinated_authority:
control: C-FM16-1 (signed decision artefacts)
status: operational
evidence_ref: decision-artefact-schema-v1, verification-rate-2026-Q2 (99.6%)
FM17 sensor_spoofing:
applicability: not applicable (centre tier)
status: documented as N/A
FM18 actuator_hijack:
applicability: not applicable (no actuators)
FM19 closed_loop_drift:
applicability: not applicable (no closed loop)
FM20 sample_rate_mismatch:
applicability: not applicable
FM21 connectivity_induced_policy_bypass:
applicability: not applicable (centre tier, continuous connectivity)
FM22 hardware_degradation:
applicability: not applicable (centre tier)
summary:
applicable_total: 16
operational: 14
monitoring_only: 2 (FM09, FM15)
not_applicable: 6
overall_coverage: 14/16 operational = 87.5%
next_assessment: 2026-07-15

```

## 5. Cognition-plane ADR - chra-v2 mandate-cache freshness budget

```
# ADR-2026-Q2-07 – Mandate-cache freshness budget for CHRA-v2
```

```
**Status:** Accepted, 2026-04-01
```

```
**Decision-makers:** AI Architect (Role 5), AI Agent Governor (Role 18), AI Risk Manager (Role 6)
```

```
**Affected systems:** chra-v2
```

```
## Context
```

```
CHRA-v2 operates centre-tier with permanent connectivity to the Authority Register. However, the Authority Register itself runs in a separate availability zone and can
```

momentarily be unreachable (e.g. during failover, brief network blips, scheduled maintenance). We must decide how stale the mandate cache may be before CHRA-v2 enters degraded mode.

### ## Decision

Mandate-cache freshness budget for CHRA-v2 is **\*\*24 hours\*\***.

### ## Alternatives considered

- **\*\*No cache (always-live consult).\*\*** Rejected: ties CHRA-v2's availability directly to Authority Register's; unacceptable when AR is briefly down.
- **\*\*15 minutes.\*\*** Rejected: tight enough that scheduled AR maintenance disrupts CHRA-v2.
- **\*\*24 hours.\*\*** Selected: long enough that any AR-side outage we currently plan for stays within budget; short enough that mandate revocations propagate within one operational day.
- **\*\*7 days.\*\*** Rejected: too long. A revocation issued today and an undetected outage tomorrow could leave CHRA-v2 running on revoked authority for nearly a week.

### ## Consequences

- CHRA-v2 enters degraded mode (refuse new actions, complete in-flight) if AR has been unreachable for more than 24 hours.
- Operations runbook updated to monitor mandate-cache age.
- The Authority Register's outage budget is implicitly tightened: any planned maintenance > 4 hours requires explicit notice and a revised CHRA-v2 mandate.
- Increases pressure to verify mandate-cache age in every policy consult (already implemented in v2026-Q2 policy engine).

### ## Anchor

- AI-BOK KA3 (cognition plane), cyber-physical addendum §2.3 (freshness budgets), FM21 (connectivity-induced policy bypass).
- Mandate metadata schema (cf. modelling conventions §4).

These five filled templates show the template structure populated with real-feeling values. Each can be lifted into an adopter's environment and edited for their own portfolio.