AOS-20 - existing GI material review and design questions

Document: aos-20-gi-material-review-and-design-questions-v0-1 · Source revision or record: v0.2

AOS Specifications v1.0 · Edition dated 3 November 2026

Original architecture and specifications: Mark Sweeny, Governance Intelligence. © 2026 Mark Sweeny.

AOS-authored material is licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).
https://creativecommons.org/licenses/by/4.0/

Preserve third-party credits. Referenced external works remain subject to their own terms.

Canonical collection: https://www.governanceintelligence.co.uk/aos

Publication copy with editorial notices, public references and disclosure edits where applicable.

Version: 0.2
Date: 6 October 2026
Status: author review; Q01-Q07 directions approved by Mark on 6 October, Q04 supplemented by agentic research for adaptation of GI itself. Exact requirement wording remains to be drafted; AOS-20 not yet drafted. No implementation or independent assurance claim.

1. Review conclusion

The existing material supplies a substantial foundation for the GI charter: cross-domain synthesis, a governed corpus, interpretation and option preparation, faithful presentation, and an authorising/obligating charter. RP v6 already incorporates the working paper's precedent-drift and protected-escalation principles. These are governing constraints, not fresh optional proposals.

The current AOS-00/Map v0.18 and kernel v0.6 demarcation governs reconciliation: GI interprets and prepares; humans adjudicate Mode 2; AOS-14 handles external exchange; shared controls enforce permissions. AOS-20 defines the responsibilities of GI as a whole, while 21-25 and 30 supply detailed component workflows. This review does not reopen approved kernel decisions.

Seven choices were submitted at charter/workflow level and are now approved under DEC-079-085 (see §6). Organisational appointments, particular recipient identities, numerical thresholds, review deadlines, implementation services and model evaluation evidence remain later commissioning dependencies. The questions below concern reusable architecture, not a single repair process or a single-organisation structure.

2. Sources consulted and standing

Table row 1
Source: AOS-00 and Specification Map v0.18
Relevant anchors: Architecture §§2.2/3.1-3.3/10; Map §§5-8; DEC-005/008/018/024-025; current kernel adoption
Standing / reconciliation: Governing recorded decisions; use as current baseline.

Table row 2
Source: AOS Reasoning Principles v6
Relevant anchors: Domains 1-3, 6-9; human accountability, boardroom exclusion, precedent integrity, uncertainty, protected escalation and use limits

Standing / reconciliation: Governing source. Annual dual-authored independent reconciliation is already specified; no proposal to remove it.

Table row 3
Source: Book 4
Relevant anchors: Governance Intelligence; Governance Hub; AI, Governance and Ethics; epistemic brake
Standing / reconciliation: Explains purpose/institutional design. Metaphorical conscience/membrane/adjudication language must respect the approved software/human boundary.

Table row 4
Source: Governance Intelligence - Internal Architecture - Working Paper
Relevant anchors: Five elements; corpus curation; charter; governance of drift; validation regime
Standing / reconciliation: Working proposal originally checked against RP v5; compare with actual v6, not the note's old principle counts. Corpus steward powers and charter amendment mechanics remain underspecified.

Table row 5
Source: GIL Technical Specification v2, May 2026
Relevant anchors: §§1-3, governance instruments, assessments, §§6-9, human review and UDL events
Standing / reconciliation: Reuse principally in AOS-21/23. Older six-peer-domain, sole-agent-per-type, numeric confidence, stack/schema and timing assumptions are not adopted cross-AOS defaults.

Table row 6
Source: GIL Technical Specification - Forward Tasks
Relevant anchors: RP v6 recheck; reciprocal agent/oversight references; capability-calibration proposal
Standing / reconciliation: Ordinary evaluation/uncertainty/monitoring remains in scope. Exploratory self-calibration/reciprocal oversight remains deferred by DEC-018.

Table row 7
Source: Petra note
Relevant anchors: Presentation/persona characteristics
Standing / reconciliation: Tone reference only; fictional feelings, conscience and natural authority create no software powers or claims of interiority.

Table row 8
Source: Trade-Off Divergence Tool working draft / CAP-01
Relevant anchors: §§2-4/6-7; shared synthesis/corpus, no ranking or recommendations between persona readings; independent configuration; output-routing question
Standing / reconciliation: Included capability under DEC-017, detailed terms proposed. Preserve its specific prohibition; do not silently generalise it into a prohibition on all GI analytical recommendations. Routing is unresolved; see Q07.

Table row 9
Source: Current Core/Shared v0.4, Exchange Register v0.3, Control Coordination v0.2, Protected Reporting and Recovery v0.1
Relevant anchors: Canonical analysis, acknowledgement/application/responsibility distinctions, separate permissions, current-state and failed-route duties
Standing / reconciliation: Proposed logical definitions under approved directions. Adopt reciprocally in later GI/Handoff drafting; protected notification cannot depend on optional executive release.

The reviewed sources are the preserved material under V0.1/Source Documents, V0.1/GIL_Technical_Specification_V2.docx, current V0.1 architecture/map and 02 Architectural Contracts. Book chapters and relevant GIL sections were reviewed for charter implications; this is not a line-by-line reconciliation of all 121 artefact schemas or a current legal/regulatory assessment.

3. Corrections required without reopening settled principles

- Replace the older GIL description of GI as a sixth peer domain with GI as the whole AOS-20-25 layer above the five sensing domains.

- Translate adjudication/conscience/membrane language into specified software duties and accountable human judgement. GI cannot confer authority, adjudicate Mode 2 or govern by changing the kernel's controls.
- Preserve synthesis/interpretation versus rendering, and charter versus executable function. AOS-23's artefact assessment is an explicit capability within GI; no additional peer intelligence is introduced.
- Preserve permanent human Mode 2, boardroom exclusion and authoritative human-approved post-meeting records. Conversational confirmation cannot be an indirect meeting-capture channel.
- Preserve source hierarchy, current authority/use, protected escalation, acknowledged handoff, human case closure and conservative recovery rules. No protected finding is held indefinitely for a discretionary management sign-off.
- Reconcile the GIL's blanket review-before-UDL wording with its own urgent-alert UDL publication and the canonical-analysis review-state contract. Separate storage of provisional evidence, authorised challenge access and formal reliance; Q06 resolves the remaining workflow choice.
- Retain CAP-01's shared evidence base and its no-ranking/no-recommendation constraint between persona readings. Personas cannot create separate competing factual realities.
- Do not adopt historic numerical confidence thresholds, particular products, 2015 regulatory references or claims of solved technical capability as present validated defaults. Current regulatory or standard-specific obligations need primary-source verification when used in the specifications.

4. Submitted design questions and recommendations - approved disposition in §6

Q01 - charter amendment powers

Question: should GI's core permissions, prohibitions and mandatory duties be a protected board-approved charter, with operational profiles maintained separately?

Recommendation: yes. The board approves and amends the charter's constitutional core; technical administrators and ordinary delegates cannot weaken it. Organisation-specific routes, recipients, timing and analytical settings sit in separately approved operational profiles under AOS-01 change powers. A profile change cannot dilute a core duty or broaden authority. Applicable organisational constitutional instruments determine the precise process; do not universally prescribe a special resolution. This resolves the working paper's “same bar as telos” proposal without treating every parameter change as a telos amendment.

Q02 - corpus steward's delegated powers

Question: what may the human corpus steward approve, and what must go to another empowered authority?

Recommendation: establish a human Governance Knowledge Steward function, often located with the company secretary/governance team. Delegate routine source admission, provenance, version/currency maintenance and faithful capture of already approved decisions within defined policy. Separate verification of source status from organisational adoption of a new obligation. Disputed authority, materially changed meaning, new normative interpretations and conflicts go to the appropriate authorised decision-maker; the steward cannot invent constitutional authority. Oversight needs an independent reading of the evidence, including independent-of-GI preparation of the annual report. Role names and organisational location may vary; powers and separation remain explicit.

Q03 - confirmation and precedent

Question: when does a human confirmation become an authoritative interpretation or precedent?

Recommendation: distinguish conversational understanding, reviewed analysis, an empowered human decision and approved precedent. “Yes, that is what I meant” in a Petra exchange does not itself amend the corpus. Approval binds exact

wording/revision, decision-maker's powers, rationale, context, limitations and applicability; ordinary interpretations may be approved within an explicit human delegation, while constitutional disputes/amendments follow protected routes. GI can propose and challenge precedent but cannot approve its own interpretation. Preserve differing rulings and review their cumulative drift against the founding statement; do not treat the latest ruling as universal truth.

Q04 - proactive GI review

Question: may GI initiate cross-domain review without waiting for a human request?

Recommendation: yes, within explicitly approved purposes, sources and event/cycle triggers. GI should notice inconsistency, emerging patterns, stale governance material and purpose/precedent drift and initiate internal review. No additional sensing or private resident/worker access follows from this authority. Automated noticing and reassessment do not authorise external contact, commitments, operational effects or Mode 2 resolution. Persistent/combined patterns are assessed alongside individual events so that thresholds do not hide accumulation.

Q05 - required escalation classes and routing

Question: which findings must be protected from suppression or discretionary delay?

Recommendation: require protected internal escalation for credible material indications of: telos/constitutional conflict or drift; Mode 2 encroachment/authority bypass; attempted suppression or control tampering; failure to meet a required reporting/compliance obligation; material safety/welfare implications; and evidence/assurance failure that makes an important governance conclusion unsafe to rely on. These are indications for human assessment, not AI verdicts of wrongdoing. Organisation-approved criteria add context, urgency, accountable recipients/cover and response times; no invented universal thresholds. If the normal recipient is implicated or fails to respond, route through a separately authorised alternative to the appropriate independent governance authority. Apply approved permission, minimal payload, acknowledgement/escalation and failed-route rules from the kernel. Protected reporting remains owned even when normal operation is stopped.

Q06 - review before publication or reliance

Question: must every GI output await human review before anyone can see it, or should review depend on the intended use?

Recommendation: use explicit graduated review states. Permitted provisional analyses can be stored, interrogated and used to alert authorised reviewers, visibly labelled with uncertainty, source coverage and review limits. Formal governance assessments and material decision-support conclusions need the designated human validation before being represented as reviewed or placed in the applicable formal decision workflow. Action authority remains separate in every state. Urgent protected escalation does not wait for the executive review it may challenge. A raw model confidence score cannot substitute for source assurance or human review. Preserve one identified canonical analysis revision and review status across permitted presentations.

Q07 - independent challenge access

Question: should GI's challenge output always be shared with management immediately, or may NEDs/the relevant independent governance body have a protected preparation window?

Recommendation: allow an explicitly governed independent preparation window for authorised NED/committee challenge work, including CAP-01 divergence analysis. The executive whose paper is being tested cannot tune or suppress that challenge profile. This is a timing/access rule over the same canonical evidence and analysis, not a second secret factual corpus. Relevant management receives the permitted findings/questions in time to correct facts and respond, under an approved workflow. The window cannot delay protected mandatory reporting or conceal material governance evidence from its required audience; no AI enters live deliberation. Board-approved scope, permitted recipients and exceptions belong in AOS-03/24/30 and the CAP-01

profile. Do not adopt an always-secret or always-shared default for every GI use.

5. Next step

Present Q01-Q07 to Mark. Record only explicit answers as approved directions. Then draft AOS-20 against FP v4/RP v6/GT, AOS-00/Map and the common kernel contracts, with source/requirement/verification links and organisational settings left explicit. AOS-30 follows for detailed human decisions/handoffs. The design directions do not complete implementation, appoint operational people or establish independent assurance. Research pilot remains a separate open workstream.

6. Approval disposition and Q04 addition - 6 October 2026

Mark approved all seven recommendations and added: “The GI function will be supported by agentic AI processes, similarly to our boundary intelligence agent, researching for changes that may imply adaption of the GI function.” AOS-00/Map v0.19 records DEC-079-085. Original §4 text is preserved as the submitted rationale; this section is its current disposition.

Table row 1
Question: Q01
Decision: DEC-079
Current disposition: Approved protected charter/separate operational profiles.

Table row 2
Question: Q02
Decision: DEC-080
Current disposition: Approved human Governance Knowledge Steward powers and referral boundaries.

Table row 3
Question: Q03
Decision: DEC-081
Current disposition: Approved distinct confirmation/analysis/decision/precedent approval.

Table row 4
Question: Q04
Decision: DEC-082
Current disposition: Approved proactive organisational analysis, supplemented by agentic research into developments requiring adaptation of GI itself.

Table row 5
Question: Q05
Decision: DEC-083
Current disposition: Approved protected escalation classes and separately authorised alternative routing.

Table row 6
Question: Q06
Decision: DEC-084
Current disposition: Approved graduated review states, preserving urgent protected escalation.

Table row 7
Question: Q07
Decision: DEC-085
Current disposition: Approved independently governed challenge preparation window and permitted response workflow.

Q04 has two related but distinct functions: (a) review the organisation against current governed purposes/evidence; (b) research whether current GI design/function remains adequate in light of external developments. Research observes, analyses, challenges

and proposes; governed human decisions and controlled change determine adaptation. Relevant findings must identify primary sources/date/coverage, current GI baseline and affected element, evidence and uncertainty, implications/options and cross-component impacts. Accepted/rejected/deferred/unresolved dispositions retain accountable owners; urgent protected findings use the approved routes. A single agent or multiple processes are implementation choices, not numerical defaults.

AOS-20 defines the research remit and powers; AOS-25 develops adequacy/drift/proposal tracking; AOS-14 controls external retrieval/exchange; AOS-04 supplies general agent controls; AOS-01/03/05 enforce authority/use/runtime limits; AOS-02 preserves evidence. Research may recommend changes but cannot amend specifications, widen access or give its own proposals approval. Deferred exploratory self-calibration is not reopened. Protected charter review and adaptation remain human governed.

Next: draft AOS-20 with all seven approved directions and the amended research remit. Actual source access, cadence/coverage, operational recipients/cover, numerical settings, briefing and delivery proof remain setup dependencies. No new research automation has been created or scheduled; AOS-20 is not yet drafted. Exact requirements, conformance/readiness and independent review remain open.