

Documento técnico OSARA — borrador

No revisado por pares. Borrador v5 (2026). El contenido puede cambiar. © José Luis D'Andrade. **Todos los derechos reservados.** Cita breve con atribución permitida. **Queda prohibida** la reproducción o reutilización, total o parcial, sin **autorización previa por escrito** del titular.

OSARA — Integración con el estándar abierto de agentes soberanos

Documento de evaluación de integración. Define qué es OSARA, por qué importa para la visión de servir a **instituciones públicas**, y una hoja de ruta priorizada para alinear la plataforma con el estándar.

Fuente externa: <https://github.com/elkordym/osara-open-standard> (OSARA v0.4-draft.1). **Estado del documento:** propuesta de planeación.

La versión por defecto (inglés) de este documento está en `OSARA-Integration.md`.

1. Por qué este documento existe

La plataforma tiene una visión declarada: servir también a **instituciones públicas**. En producción ya se despliegan organizaciones como entidades de primera clase: la federación multi-host modela ministerios, bancos y direcciones con departamentos, topología y políticas de confianza (*trust policies*).

En paralelo, un esfuerzo en GitHub —**OSARA, Open Sovereign Agent Reference Architecture**, publicado por **Open Source United, una Community of Practice de la ONU**— está fijando el estándar abierto para agentes de IA *antes* de que se imponga un default propietario. Su modelo de tiers culmina en el **G-Agent (agente de gobierno)**: precisamente el comprador al que apunta la visión de la plataforma.

La conclusión es simple: **si las instituciones públicas adoptan OSARA como criterio de compra o certificación, alinearse deja de ser opcional**. Este documento prepara ese camino sin comprometer trabajo prematuro contra un estándar todavía en draft.

2. OSARA en cinco ideas

1. **Tesis:** *"La soberanía no puede concederse por términos de servicio — debe imponerla la criptografía."* El dueño tiene la llave privada, no la plataforma.
2. **Cuatro pilares:** Soberanía · Portabilidad · Interoperabilidad · Rendición de cuentas.
3. **Tres tiers:** P-Agent (personal), E-Agent (empresa), G-Agent (gobierno) — mismos requisitos de identidad, auditoría e inspección; distinto privilegio y anclaje de identidad.

4. **Identidad:** un **GIAC** (certificado X.509 v3 con extensiones OSARA), con la clave privada en un **dispositivo del dueño (OAD)** y nunca en logs ni mensajes.
5. **Ocho componentes obligatorios:** AIM (identidad), ECL (comunicación cifrada), ALE (auditoría encadenada), DAE (delegación), HSB (beacon de salud), IG (gateway de interoperabilidad), PMS (memoria portable), POA (autorización física del dueño). Más IMA (integridad por TPM), estados de bloqueo, certificado anual de salud y badges C/A/S.

Detalle normativo completo (puertos, esquemas GIAC/ALE/Skill Inventory, criptografía) en el reporte de brechas que acompaña a este documento.

3. Punto de partida: lo que ya tenemos

La plataforma **no parte de cero**. Dos cosas juegan a favor:

(a) Encaje conceptual con el modelo de identidad persistente. La arquitectura ya adopta una posición explícita sobre identidad, cambio y verdad que anticipa tres de los cuatro pilares OSARA:

- **Continuidad de la identidad** ≈ Portabilidad / *Person-and-Agent Model*: los agentes se archivan, no se borran; un **archivo de identidad persistente** ancla al agente a través de cambios de modelo; la derivación de agentes es **linaje, no copia**. La memoria sale con su dueño.
- **Validación colectiva de la verdad** ≈ Rendición de cuentas: nada es verdad por declaración unilateral; hay *provenance*, supersesión y *peer review* sobre una **base de conocimiento compartida**.
- **Aprendizaje por uso gobernado** ≈ Soberanía gobernada: una **escalera de gobernanza por niveles** con *compuertas* humanas donde hay riesgo.

(b) Análogos funcionales en producción de varios componentes obligatorios:

Componente OSARA	¿Lo tenemos?	Análogo ya existente
HSB (beacon de salud)	✓	Servicio de telemetría: <i>heartbeats</i> vivos por agente/runtime
IG (gateway interop)	✓	Gateway de agentes + servidores MCP (memoria, skills, artefactos, telemetría)
DAE (delegación)	●	Aprobaciones de gobernanza + presupuestos por objetivo + sobres (<i>envelopes</i>) cross-host estructurados y auditados, NO firmados , con políticas de confianza. La confianza entre hosts descansa en la política más las fronteras de red, no en criptografía ([PA] §4.4)
ALE (auditoría)	✓	Log de auditoría en la base relacional + traza causal en telemetría + tablero de tareas
PMS (memoria portable)	✓	Memoria vectorial por-agente (RAG), memoria del runtime, memoria episódica, almacén de artefactos
Cross-tier rules	✓	Políticas de confianza + topología (flat / hub-spoke / hierarchical)

La identidad por-agente **ya existe y es rica**: cada agente tiene su propio archivo de identidad persistente, memoria, modelo y base de conocimiento, y toda la tesis de diseño de la plataforma es la continuidad de identidad. Lo que **aún no se añade** es la **atestación criptográfica** de esa identidad — un certificado por-agente (GIAC), una llave en manos del dueño, y logs firmados y encadenados por hash. Eso está diferido **por diseño**: construir criptografía pesada contra un estándar todavía en draft sería prematuro (ver §4). Es la capa de atestación lo que convertiría una arquitectura de identidad-y-gobernanza ya sólida en una *certificable* — no un cimientito que falte hoy.

4. Estrategia: tres horizontes, no un big-bang

OSARA está en **v0.4-draft**. Implementar criptografía pesada contra un blanco móvil sería caro y prematuro. La estrategia es escalonada:

Horizonte 1 — OSARA-aware (postura + vocabulario + quick wins)

Bajo costo, alto valor narrativo frente a compradores públicos. No requiere criptografía nueva.

Horizonte 2 — OSARA-ready (contratos y firmas)

Introduce identidad criptográfica por-agente y encadenado de auditoría. Disparador: un cliente público que lo exija contractualmente, o OSARA llegando a rc/v1.0.

Horizonte 3 — OSARA-certified (badges C → A → S)

Inspección, certificado anual de salud y el ecosistema de certificación. Solo tiene sentido cuando exista demanda real y un emisor de GIAC operando en la jurisdicción objetivo.

5. Hoja de ruta priorizada (track o)

Propuesta de track OSARA mapeado a los componentes funcionales de la plataforma. Tamaño: B (bajo) · M (medio) · A (alto). **Nada de esto está comprometido**; es el menú del que se prioriza.

Horizonte 1 — OSARA-aware

ID	Trabajo	Componente	Tamaño
O0	Elegir licencia OSI para los componentes de identidad/auditoría/cifrado. Prerrequisito de cualquier conformidad — y buena higiene de todos modos.	raíz	B
O1	Mapeo de tiers P/E/G sobre el modelo de federación: declarar en cada <i>spec</i> de institución y host su tier OSARA (<code>osara.tier</code>) y jurisdicción (ISO 3166-1). Una institución pública = G-Agent; un departamento = sub-agente.	Fleet manager, <i>specs</i> de organización	B

ID	Trabajo	Componente	Tamaño
O2	Vocabulario en sobres y políticas de confianza: mapear la semántica cross-tier OSARA sobre las <i>trust policies</i> existentes ($P \rightarrow G$ prohibido directo, $G \leftrightarrow G$ por tratado, etc.) como validación opt-in.	Fleet manager	M
O3	Skill Inventory legible estilo OSARA (JSON-LD: id, versión, método de adquisición, atribución, licencia, estado) como extensión del marketplace de skills. Sin firma todavía.	Marketplace de skills	M
O4	Conformance Statement (documento): declarar tier, clase de conformidad parcial y versión de esquema. Honesto sobre qué se cumple y qué no.	Documentación	B

Horizonte 2 — OSARA-ready

ID	Trabajo	Componente	Tamaño
O5	AIM mínimo — identidad criptográfica por-agente: par de llaves (Ed25519) por agente, certificado tipo-GIAC con extensiones OSARA (<code>identityAnchorHash</code> , <code>ownerDID</code> , <code>scope</code> , <code>jurisdiction</code>). Define dónde vive la llave del "dueño" (institución $\rightarrow$ HSM, no SaaS). Desbloquea POA, ECL, partes de DAE/IG.	Nuevo servicio de identidad	A
O6	ALE encadenado: añadir hash-chain (SHA-3-256/BLAKE3) y política de retención (16 meses) al log de auditoría y a los eventos de grafo; export JSON-LD/CBOR sin contenido de mensajes.	Telemetría, gobernanza	M
O7	HSB firmado: firmar el <i>heartbeat</i> de telemetría y normalizar formato/cadencia hacia el contrato OSARA (beacon $\geq$ cada 60 s). Reusa el sustrato existente — no reescribir.	Telemetría	B-M
O8	DAE con tokens scoped + doble firma: elevar los sobres/aprobaciones a permisos <i>time-bound</i> con doble firma; alinear con el contrato de tokens de delegación de OSARA.	Gobernanza, fleet manager	M
O9	PMS con export garantizado: cifrado-at-rest (AES-256-GCM) de la memoria vectorial y artefactos + manifiesto firmado + garantía de export en ≤ 30 días (continuidad de identidad hecha criptografía).	Almacén de artefactos, memoria vectorial	A
O10	Firma de skills (SKL-01/02): firmar cada skill en su adquisición con la llave AIM; una skill firmada por A no corre en B sin grant explícito.	Marketplace de skills	M

Horizonte 3 — OSARA-certified

ID	Trabajo	Tamaño
O11	IMA (medición de integridad por TPM 2.0) + máquina de estados de bloqueo YELLOW/ORANGE/RED sobre la cuarentena y los <i>kill-switch</i> existentes.	A
O12	Endpoint de inspección (puerto 4406, POA-gated) + flujo de Certificado Anual de Salud (AHC).	A
O13	Badges $C \rightarrow A \rightarrow S$: suite de conformidad automatizada + auditoría manual + Compliance Attestation Document (Verifiable Credential).	A

Camino crítico: O0 (licencia) → O1/O2 (tiers+vocabulario, baratos y de alto valor) → O5 (AIM, la llave que desbloquea casi todo) → el resto del Horizonte 2 en paralelo.

6. Recomendación

1. **Hacer ya (barato, sin esperar a nadie):** O0 (licencia OSI) y O1 (declarar tiers P/E/G y jurisdicción en los specs de institución). Da postura "OSARA-aware" creíble frente a un comprador público con días de trabajo, no meses.
 2. **Diferir hasta tener señal de demanda:** todo el Horizonte 2 (criptografía). Disparador explícito: **primer cliente público que exija OSARA, o OSARA llegando a rc/v1.0**. No construir contra un draft que aún puede cambiar.
 3. **No reescribir lo nativo.** HSB, IG y partes de PMS ya tienen análogo; el delta legítimo es *envolver y firmar*, no levantar un runtime paralelo.
 4. **Vigilar el modelo OAD/POA.** La llave del dueño en hardware choca con el despliegue on-prem actual; para un G-Agent, "el dueño" es la institución y el OAD plausible es un HSM. Es decisión de arquitectura, conviene pensarla antes de O5.
-

7. Decisiones abiertas

- ¿Se prioriza el track OSARA ahora, o se mantiene como inteligencia de mercado a la espera de un cliente público concreto?
 - ¿Qué licencia OSI para los componentes obligatorios (O0)? (Apache-2.0 y MPL-2.0 son candidatos típicos para infraestructura con intención comercial.)
 - Para un G-Agent institucional, ¿quién es "el dueño" y qué hace de OAD (HSM gestionado por la institución, smartcard de un funcionario responsable, ...)?
 - ¿Conviene abrir contacto con Open Source United / el repo OSARA para seguir el draft y, en su caso, contribuir desde la experiencia de despliegues institucionales reales de la plataforma?
-

Referencias

- **Evidencia acompañante:** `OSARA-Gap-Analysis-es.md` — análisis de brechas y matriz de conformidad.
- **OSARA Open Standard:** <https://github.com/elkordym/osara-open-standard>
- OSARA Specification, `specs/osara/current.md` (v0.4-draft.1)
- OSARA Overview & Concepts, `specs/osara-overview/current.md` (v0.4-draft.1)