
ENKIDU

MIOVSKI GROUP

undefined

Glosario

- Datos maestros** Los « sustantivos » compartidos y de cambio lento sobre los que opera la organización — partes, activos, ubicaciones, productos, cuentas.
- Entidad maestra** El sustantivo concreto que un despliegue domina primero; el ejemplo recurrente es una parte genérica.
- Registro maestro (golden record)** El registro compuesto autoritativo de una entidad, ensamblado atributo por atributo desde la mejor fuente.
- Sistema contribuyente** Todo sistema operativo que posee su versión de la entidad y alimenta el hub (sistemas de entrada, de registro, de interacción, de operaciones).
- Coexistencia** El estilo MDM adoptado: el hub domina y depura la entidad y se mantiene sincronizado de forma bidireccional con las fuentes.
- Bronce / Plata / Oro** Las tres zonas duraderas del pipeline: bruta e inmutable (Bronce); de trabajo y depuración (Plata); fiable y terminada (Oro).
- Supervivencia (survivorship)** Las reglas que seleccionan el mejor valor por atributo cuando las fuentes discrepan.
- Resolución de identidad** Reconocer que varios registros de origen describen la misma entidad del mundo real.
- Procedencia / Cadena de custodia** De dónde vino un valor, y la secuencia ordenada y registrada de decisiones automáticas y humanas que lo produjeron.
- HITL** Humano en el bucle — un administrador decide donde la automatización no debe.
- XREF** La referencia cruzada que asocia el ID maestro con cada clave de origen.
- Almacén de datos maestros (DWH)** Parte 2 — la capa de consolidación analítica que conforma los datos de los contribuyentes a los golden records.
- Dimensión conformada** Una entidad maestra usada de forma coherente en el almacén para que los hechos compartan una única vista.
- Zona de retención** Donde los registros DWH de una entidad no reconocida se aparcen hasta que el MDM produce un golden record.

Resumen

Toda organización de escala opera varios sistemas que poseen cada uno su versión de las mismas entidades centrales — el mismo cliente, ciudadano, paciente, soldado, activo o proveedor, con identificadores, ortografías y estados distintos. Esa fragmentación produce duplicados, informes poco fiables y exposición al cumplimiento. El problema es universal; solo cambia el nombre de la entidad según el sector.

Este documento describe la arquitectura objetivo de una plataforma de datos maestros en dos partes, independiente del sector. La Parte 1 (MDM) mantiene datos maestros autoritativos; la Parte 2 (DWH) consolida los datos de los mismos contribuyentes contra esos golden records y los sirve desnormalizados a la analítica y la IA. Ambas siguen la disciplina Bronce -> Plata -> Oro.

No se puede integrar lo que no se puede identificar.

Esta plataforma hace que las entidades centrales sean identificables, fiables y compartibles — sea

cual sea el nombre que les dé el sector.

1. Contexto y planteamiento del problema

1.1 El problema universal

Una organización opera varios sistemas que poseen cada uno su versión de las mismas entidades, con identificadores, ortografías y estados distintos. El hub MDM produce y mantiene la representación autoritativa de cada entidad — el golden record — para que las fuentes se integren contra un único registro fiable en lugar de unas contra otras.

1.2 Un modelo neutral respecto al sector

Los sistemas contribuyentes se describen por su rol, no por categoría de producto. Todo despliegue real asigna sus sistemas a estos roles:

Sistema de entrada

Donde la entidad se crea y se posee. Autoridad sobre la existencia y la identidad básica.

Sistema de registro

Donde residen los datos operativos canónicos. Autoridad sobre conjuntos de atributos concretos.

Sistema de interacción

Donde se interactúa con la entidad a diario. Aporta contacto, actividad y estado.

Sistema de operaciones / planificación

Donde la entidad se despliega o planifica. Aporta identidad y estado operativos.

Consumidor posterior

Analítica, informes, IA/AA e integraciones. Lee el golden record; no lo domina.

El ejemplo recurrente es una parte genérica. La arquitectura se aplica sin cambios a un cliente, ciudadano, paciente, proveedor o persona bajo un mandato de defensa — y a entidades no-parte (activos, ubicaciones, productos, cuentas).

1.3 El mismo modelo entre sectores

La estructura es constante; solo cambia la instanciación. Un despliegue nombra su entidad, asigna cada sistema a un rol y fija la precedencia de supervivencia — nada en la arquitectura se mueve.

Comercial / empresa

Entidad: cliente, proveedor, producto. Roles: CRM (interacción), ERP (registro), facturación, e-commerce (entrada).

Sector público / gobierno

Entidad: ciudadano, caso, beneficiario. Roles: registro (entrada), gestión de casos (registro), portales (interacción).

Salud

Entidad: paciente, proveedor, encuentro. Roles: admisión/ADT (entrada), historia clínica (registro), planificación (operaciones).

Defensa / seguridad pública

Entidad: persona, unidad, plataforma, activo. Roles: sistema de personal (entrada), C2 operativo (operaciones), logística (registro).

Fuerza laboral / RR. HH.

Entidad: trabajador, puesto, unidad. Roles: SIRH (entrada), nómina (registro), planificación (operaciones), directorio (interacción).

Como el modelo es agnóstico a la entidad, se puede dominar primero la entidad más problemática y añadir otras después sin rediseñar.

1.4 Alcance

En alcance (Fase 1): una entidad maestra y sus atributos básicos, su jerarquía y la referencia cruzada de identidad. Fuera de alcance (fases posteriores): otros dominios de datos maestros, la integración transaccional completa y la baja de interfaces heredadas.

2. Principios y decisiones de arquitectura

La arquitectura se moldea con cuatro decisiones críticas, más una elección de estilo MDM.

Decisión 1 · Ingestión multietapa

Los datos pasan por tres zonas — Bronce (bruto), Plata (depuración y resolución), Oro (golden records terminados). Cada zona es una frontera duradera y gobernada.

Decisión 2 · Subsistemas explícitos

Cada capacidad del hub tiene nombre, límites y una responsabilidad.

Decisión 3 · Procedencia y cadena de custodia

Para cada atributo maestro se registra su origen y la secuencia de decisiones — suficiente para reconstruir y defender cualquier registro.

Decisión 4 · Curación automatizada + humano en el bucle

El motor de curación resuelve automáticamente cuando la confianza es alta y enruta lo ambiguo, incierto o sensible a un administrador. La división se rige por umbrales y políticas.

2.1 Estilo elegido — Coexistencia

De los cuatro estilos MDM clásicos, este diseño adopta la Coexistencia: el hub domina el registro y se mantiene sincronizado de forma bidireccional con las fuentes, que siguen siendo los sistemas de entrada de sus dueños.

Registro (Registry) — rechazado

Índice fino que apunta a las fuentes; golden record virtual en lectura. Muy poco control.

Consolidación — parcial

Copias centralizadas sobre todo para informes. Buena para analítica, no devuelve datos fiables a operaciones.

Coexistencia — elegido

El hub domina, depura y sincroniza de forma bidireccional. Equilibra autoridad y cambio aceptable de las fuentes.

Centralizado — opción posterior

El hub se convierte en el único sistema de entrada. Máxima coherencia, pero la mayor disrupción organizativa.

2.2 Principios de apoyo

- Programa, no proyecto: gobernanza y administración de primer nivel y continuas.
- Los datos brutos son inmutables: el Bronce es el registro reproducible de todo lo recibido.
- Cada decisión es atribuible: ningún valor llega a Oro sin origen y justificación registrados.
- Fuente de datos fiables, no solo una única fuente de verdad.

3. Arquitectura de referencia de alto nivel

La arquitectura se lee de arriba abajo como un flujo y de izquierda a derecha por zona: las fuentes alimentan la integración; los datos aterrizan en Bronce, se curan en Plata y se dominan en Oro; los golden records se publican. Dos planos transversales abarcan cada capa: gobernanza y administración; seguridad, linaje y auditoría.

4. Catálogo de subsistemas

Cada subsistema tiene una responsabilidad clara, cubriendo todo el ciclo de vida de una entidad maestra.

SUB-01 · Integración y conectores

Capta cada fuente (CDC, API, lote/archivo, eventos), alimenta el Bronce y syndica los golden records.

SUB-02 · Modelo y esquema de entidad

Define la entidad maestra, sus atributos, relaciones y jerarquías.

SUB-03 · Motor de calidad

Perfila, estandariza, valida y enriquece según las seis dimensiones de calidad.

SUB-04 · Emparejamiento y deduplicación

Resolución de identidad por emparejamiento exacto, difuso, probabilístico y asistido por IA.

SUB-05 · Supervivencia y resolución de conflictos

Fusiona el mejor atributo de cada fuente y resuelve valores en conflicto.

SUB-06 · Flujo de administración (HITL)

Revisión humana, gestión de excepciones y aprobación de cambios inciertos.

SUB-07 · Golden record y XREF

Ensambla y almacena el registro autoritativo y la referencia cruzada.

SUB-08 · Jerarquías y relaciones

Mantiene agrupaciones padre/hijo y líneas jerárquicas.

SUB-09 · Metadatos y datos de referencia

Definiciones, propiedad, reglas de calidad y listas de códigos controladas.

SUB-10 · Seguridad, linaje y auditabilidad

Control de acceso más procedencia y cadena de custodia para cada valor y decisión.

SUB-11 · Consola de gobernanza y administración

La superficie operativa para definiciones, derechos de decisión, políticas y excepciones.

SUB-12 · Publicación y sindicación

Envía los golden records a consumidores operativos (resincronización) y a la analítica.

5. El pipeline de ingestión — Bronce · Plata · Oro

Un pipeline medallón de tres zonas. La promoción solo ocurre al cumplir el criterio siguiente; nada se borra del Bronce, por lo que todo es reproducible.

5.1 Bronce — zona de aterrizaje

Recibe los datos tal como los envió la fuente, sin modificarlos; cada registro se sella. Es el registro reproducible: la Plata puede reconstruirse sin volver a extraer de las fuentes.

5.2 Plata — zona de trabajo

Donde ocurre la depuración y resolución: perfilar y estandarizar, validar y enriquecer, emparejar y deduplicar, y luego aplicar supervivencia y resolución de conflictos. Lo incierto se escala a la administración.

5.3 Oro — zona fiable

Solo contiene golden records terminados. La promoción asigna un ID maestro estable, actualiza la XREF, refresca jerarquías y publica; en coexistencia, los valores corregidos se resincronizan a las fuentes.

PROPIEDADES DE LAS ZONAS

Contenido Bronce: bruto; Plata: cúmulos en curso; Oro: golden records terminados.

Mutabilidad Bronce: inmutable; Plata: rederivable; Oro: versionado, controlado.

Acceso Bronce: pipeline y auditoría; Plata: motor y administradores; Oro: consumidores (lectura), resincronización.

6. Motor de curación — automatizado y humano en el bucle

El motor de curación decide la identidad y resuelve conflictos, con una división deliberada entre una vía automatizada (decisiones seguras) y una vía HITL (ambiguo, incierto, sensible), regida por

umbrales y políticas.

6.1 La vía automatizada

Por encima del umbral de fusión automática y sin atributo sensible en conflicto, el motor actúa solo: fusiona, aplica supervivencia, registra la regla y su justificación, y avanza hacia Oro. Es la vía de alto volumen.

6.2 La vía humano en el bucle

Lo incierto o sensible se enruta a un administrador que ve el cúmulo, los valores en competencia, las puntuaciones y el linaje, y decide. Cada acción se captura con identidad y motivo, y sirve de señal de ajuste.

6.3 Dónde debe intervenir un humano

Coincidencia de banda media — HITL

La puntuación cae entre los umbrales de fusión y de no coincidencia — identidad genuinamente ambigua.

Conflicto entre fuentes autoritativas — HITL

Dos fuentes fiables discrepan y la supervivencia no desempata.

Cambio de atributo sensible — HITL

Los campos regulados (ID nacional, datos protegidos) requieren aprobación explícita.

Fusión / división dudosa — HITL

Una fusión combinaría dos entidades reales, o una sobrefusión debe deshacerse.

Excepción de validación — HITL

Un registro falla una regla no autocorregible (fecha imposible, bucle de jerarquía).

Atributo nuevo / no mapeado — HITL

Una fuente envía un campo sin mapeo; un administrador decide cómo tratarlo.

Coincidencia clara de alta confianza — AUTO

Sin humano — aplicada automáticamente con justificación registrada.

7. Seguridad, linaje y auditabilidad

Esta capacidad responde, para cualquier valor: quién puede verlo o cambiarlo (seguridad), de dónde vino (procedencia), y qué secuencia de decisiones lo produjo (cadena de custodia). Captura evidencia en cada paso.

7.1 Seguridad y control de acceso

Acceso por rol y atributo sobre zonas y campos; cifrado y enmascaramiento de atributos sensibles; mínimo privilegio por zona — el Bronce no se edita.

7.2 Linaje y procedencia

Para cada atributo maestro, el hub registra la fuente ganadora, los valores en competencia y la regla o decisión que lo eligió. La procedencia es a nivel de atributo, no solo de registro.

7.3 Auditoría y cadena de custodia

Cada transformación y decisión — automática o humana — se escribe en un registro de solo anexo, suficiente para reconstruir y defender cualquier registro. Los regímenes de privacidad (RGPD, PIPEDA) exigen saber dónde viven los datos; la procedencia a nivel de atributo lo convierte en una consulta.

8. Golden record y publicación

El golden record es la representación autoritativa, ensamblada atributo por atributo. No es la copia de un sistema; es el compuesto acordado más la referencia cruzada.

ID maestro y referencia cruzada

Cada entidad recibe un ID estable; la XREF lo asocia a cada clave local de las fuentes.

Jerarquías y relaciones

Un cambio de relación se resuelve una vez, en el hub, y se publica en todas partes.

Publicación y resincronización

Publicado por eventos y API; en coexistencia, los valores corregidos vuelven a las fuentes.

9. Gobernanza y administración

El MDM triunfa o fracasa por la gobernanza, no las herramientas. Estos roles operan la consola y fijan las políticas de la división auto/HITL.

Propietario de datos

Autoridad de negocio del dominio; aprueba definiciones, supervivencia y precedencia por atributo.

Administrador de datos

Guardián diario de la calidad; trabaja la cola HITL — fusiones, deshacer fusiones, conflictos, excepciones.

Custodio de datos

Lado de TI; opera y asegura el hub, los almacenes y la plomería.

Consejo de gobernanza

Fija normas, arbitra disputas y aprueba umbrales y políticas.

10. Preocupaciones transversales

- Escalabilidad y reproceso: el Bronce inmutable permite reconstruir Plata y Oro por reproceso.
- Idempotencia: reingresar el mismo registro produce el mismo resultado.

- Modos de latencia: evento/CDC para casi tiempo real; lotes para cargas masivas.
- Versionado: golden records, reglas y modelo de entidad versionados.
- Monitoreo de calidad: las seis dimensiones medidas de forma continua.
- Despliegue por fases: empezar por el dominio donde los datos malos más dañan.

11. Almacén de datos maestros — panorama y relación con el MDM

El DWH es la capa de consolidación analítica. Donde el MDM resuelve identidad, el DWH integra los datos operativos y transaccionales más amplios, los conforma a esas entidades maestras y los reformatea para consulta y modelado.

División limpia de autoridad: el MDM (Parte 1) es la autoridad de dimensiones; el DWH (Parte 2) es donde los datos encuentran esas dimensiones conformadas — nunca resuelve identidad, la consulta al MDM. Ambos escalonan Bronce -> Plata -> Oro, pero la Plata del DWH es el modelo integrado normalizado y el Oro la forma desnormalizada lista para la analítica.

12. El pipeline de ingestión del DWH — Bronce · Plata · Oro

Un flujo medallón de tres zonas, ajustado para la consolidación analítica: el Bronce recibe los extractos brutos; la Plata consolida y normaliza contra los datos MDM; el Oro desnormaliza para el consumo.

12.1 Bronce — zona de aterrizaje

Recibe los extractos brutos tal cual, los sella y los persiste de forma inmutable; el pipeline sigue siendo reproducible.

12.2 Plata — zona de consolidación (normalizada)

Consolidación contra el MDM: normalizar y depurar, resolver las claves a ID maestros vía la XREF, unir códigos de referencia y adjuntar atributos maestros como dimensiones conformadas. Aquí la entidad no reconocida dispara la zona de retención.

12.3 Oro — zona de servicio (desnormalizada)

Reforma el modelo en formas desnormalizadas (esquemas en estrella, tablas anchas, agregados) y publica a data marts, herramientas BI y el feature store. Como la identidad ya se conformó, el Oro puede desnormalizar libremente.

PROPIEDADES DE LAS ZONAS DEL DWH

Contenido Bronce: extractos brutos; Plata: modelo integrado consolidado; Oro: tablas dimensionales / anchas.

Forma Bronce: tal como se recibe; Plata: normalizado (3FN); Oro: desnormalizado.

Dependencia del MDM Bronce: ninguna; Plata: ID maestro, XREF, datos de referencia; Oro: heredada de Plata.

Referencias de identidad Bronce: claves locales; Plata: resueltas a ID maestros; Oro: ID maestros en todo.

Consumidor Bronce: pipeline y auditoría; Plata: pipeline (interno); Oro: BI, IA/AA, ad hoc.

13. Catálogo de componentes del pipeline

Cada componente del pipeline DWH lleva una responsabilidad, implementando el flujo automatizado y la vía de excepción de la zona de retención.

DWH-01 · Extractores / conectores de origen

Extraen datos de cada fuente (CDC, lote, API, archivo) hacia el Bronce.

DWH-02 · Almacén bruto Bronce

Aterrizaje duradero e inmutable de los extractos; el punto de reproceso del pipeline.

DWH-03 · Normalización y depuración

Conforma tipos y formatos, corrige la estructura y elimina duplicados intrafuente.

DWH-04 · Servicio de resolución de claves

Mapea cada clave de negocio a su ID maestro vía la XREF. La puerta que detecta lo desconocido.

DWH-05 · Unión de datos de referencia

Resuelve valores codificados a descripciones estandarizadas con datos MDM.

DWH-06 · Enriquecimiento con datos maestros

Adjunta atributos del golden record como dimensiones conformadas.

DWH-07 · Modelo integrado / normalizado

Ensambla el modelo integrado 3FN en Plata, con todas las referencias apuntando a ID maestros.

DWH-08 · Gestor de la zona de retención

Aparca los registros no resueltos, envía la entidad al MDM y reprocesa al resolverse.

DWH-09 · Desnormalización / modelador dimensional

Transforma el modelo integrado en esquemas en estrella y tablas anchas para el Oro.

DWH-10 · Agregación y precálculo

Materializa resúmenes, métricas y agrupaciones para consultas rápidas.

DWH-11 · Capa de servicio / data marts

Expone el Oro a BI, analítica ad hoc y el feature store de IA/AA.

DWH-12 · Orquestación y planificación

Impulsa el flujo automatizado de extremo a extremo — dependencias, disparadores, reintentos y la devolución de la zona de retención.

DWH-13 · Calidad y monitoreo

Verifica frescura, integridad y conciliación entre zonas; levanta alertas.

DWH-14 · Linaje y auditoría (compartido)

Extiende el linaje de la Parte 1 a las transformaciones del DWH.

14. Flujo automatizado y la zona de retención

El pipeline DWH corre totalmente automatizado. Una sola excepción, también automática: un dato que referencia una entidad sin golden record se aparta en lugar de cargarse con una identidad adivinada.

14.1 La vía de excepción

Cuando DWH-04 no halla ID, DWH-08 aparca el registro, envía la entidad al pipeline MDM y espera. Una vez que el MDM publica el golden record y actualiza la XREF, los registros retenidos se reevalúan automáticamente y continúan a Oro.

14.2 Por qué aparcar en lugar de adivinar

Inventar una identidad es autoridad del MDM, no del almacén. Aparcar preserva la coherencia interna y el principio de autoridad única: el DWH nunca resuelve identidad, defiende al MDM.

14.3 Dónde aparecen (o no) los humanos

La vía DWH es totalmente automatizada. Un humano solo entra en la Parte 1: si el motor de curación del MDM no resuelve la entidad, un administrador lo hace allí. Desde el almacén es invisible.

ESTADOS DE LA ZONA DE RETENCIÓN

Aparcado Registro retenido; entidad enviada al MDM; a la espera de un golden record. (automático)

Disparado El MDM publicó el golden record y actualizó la XREF. (automático)

Reprocesando Registros retenidos reevaluados; las claves ya resuelven. (a Oro)

Liberado Registros conformados y promovidos; entrada de la zona de retención borrada. (hecho)

Escalado El MDM no pudo resolver tras reintentos / tiempo de espera. (excepción)

15. De extremo a extremo: Parte 1 y Parte 2

Juntas, las dos partes forman una plataforma con una única autoridad de identidad. El MDM resuelve y publica; el DWH consolida el resto contra ellas. El bucle de la zona de retención es la juntura que las mantiene sincronizadas.

15.1 Puntos clave

- Los datos maestros son los sustantivos compartidos del negocio; el MDM produce un golden record y es un programa, no un proyecto.
- La arquitectura es independiente del sector: nombrar la entidad, asignar roles, fijar supervivencia — la estructura no cambia.
- El pipeline Bronce -> Plata -> Oro mantiene el bruto inmutable y reproducible y solo promueve lo terminado.
- La curación se divide explícitamente: automatizar lo seguro, escalar lo ambiguo y lo sensible.
- Cada valor lleva su procedencia y cada decisión su cadena de custodia — base de la confianza

y el cumplimiento.

- La coexistencia equilibra autoridad y cambio aceptable; centralizar solo por elección deliberada.
- El DWH consolida los datos contra los golden records: Plata normalizada, Oro desnormalizado y listo para IA.
- El pipeline DWH es totalmente automatizado; la única excepción (entidad sin golden record) se gestiona aparcando y reprocesando.
- Una única autoridad de identidad: el almacén nunca resuelve entidades; el MDM lo hace, y el almacén espera.

Referencias

- [1] MIOVSKI GROUP / Enkidu Enterprises (2026). Trust Assurance Overview (Origi), v1.0.
- [2] MIOVSKI GROUP / Enkidu Enterprises (2026). Aegis Assure — AI Data Governance and Ethics Assessment, v3.0.
- [3] Enkidu Enterprises (2026). Tangram Overview, v1.0.
- [4] The Open Group. ArchiMate 3.2 Specification.
- [5] DAMA International. DAMA-DMBOK: Data Management Body of Knowledge (2.^a ed.).

FIN DEL DOCUMENTO