
ENKIDU

MIOVSKI GROUP

undefined

Glossaire

Données maîtres	Les « noms » partagés et à évolution lente sur lesquels l'organisation opère — parties, actifs, lieux, produits, comptes.
Entité maître	Le nom précis qu'un déploiement maîtrise en premier ; l'exemple courant est une partie générique.
Enregistrement de référence (golden record)	L'enregistrement composite faisant autorité pour une entité, assemblé attribut par attribut à partir de la meilleure source.
Système contributeur	Tout système opérationnel détenant sa propre version de l'entité et alimentant le hub (systèmes d'entrée, de référence, d'engagement, d'exploitation).
Coexistence	Le style MDM adopté : le hub maîtrise et nettoie l'entité et reste synchronisé de façon bidirectionnelle avec les sources.
Bronze / Argent / Or	Les trois zones durables du pipeline : brute et immuable (Bronze) ; de travail et de nettoyage (Argent) ; fiable et finie (Or).
Survivance	Les règles qui sélectionnent la meilleure valeur par attribut lorsque les sources divergent.
Résolution d'identité	Reconnaître que plusieurs enregistrements sources décrivent la même entité du monde réel.
Provenance / Chaîne de possession	D'où vient une valeur, et la séquence ordonnée et enregistrée des décisions automatiques et humaines qui l'ont produite.
HITL	Humain dans la boucle — un intendant décide là où l'automatisation ne doit pas.
XREF	La référence croisée associant l'identifiant de référence à chaque clé source.
Entrepôt de données maîtres (DWH)	Partie 2 — la couche de consolidation analytique qui conforme les données des contributeurs aux golden records.
Dimension conforme	Une entité de référence utilisée de façon cohérente dans l'entrepôt pour que les faits partagent une seule vue.
Zone d'attente	Où les enregistrements DWH d'une entité inconnue sont garés jusqu'à ce que le MDM produise un golden record.

Résumé

Toute organisation d'envergure exploite plusieurs systèmes qui détiennent chacun leur version des mêmes entités centrales — le même client, citoyen, patient, soldat, actif ou fournisseur, sous des identifiants, orthographes et statuts différents. Cette fragmentation produit des doublons, des rapports peu fiables et une exposition à la conformité. Le problème est universel ; seul le nom de l'entité change selon le secteur.

Ce document décrit l'architecture cible d'une plateforme de données maîtres en deux parties, indépendante du secteur. La Partie 1 (MDM) maintient des données maîtres faisant autorité ; la Partie 2 (DWH) consolide les données des mêmes contributeurs contre ces golden records et les

sert dénormalisées à l'analytique et à l'IA. Les deux suivent la discipline Bronze -> Argent -> Or.

On ne peut intégrer ce qu'on ne peut identifier.

Cette plateforme rend les entités centrales identifiables, fiables et partageables — quel que soit le nom que leur donne le secteur.

1. Contexte et énoncé du problème

1.1 Le problème universel

Une organisation exploite plusieurs systèmes détenant chacun leur version des mêmes entités, avec des identifiants, orthographes et statuts différents. Le hub MDM produit et maintient la représentation faisant autorité de chaque entité — le golden record — pour que les sources s'intègrent contre un seul enregistrement fiable plutôt que les unes contre les autres.

1.2 Un modèle neutre par rapport au secteur

Les systèmes contributeurs sont décrits par leur rôle, non par catégorie de produit. Tout déploiement réel associe ses systèmes à ces rôles :

Système d'entrée

Où l'entité est créée et possédée. Autorité sur l'existence et l'identité de base.

Système de référence

Où résident les données opérationnelles canoniques. Autorité sur des ensembles d'attributs précis.

Système d'engagement

Où l'on interagit avec l'entité au quotidien. Contribue coordonnées, activité et statut.

Système d'exploitation / planification

Où l'entité est déployée ou planifiée. Contribue l'identité et l'état opérationnels.

Consommateur en aval

Analytique, rapports, IA/AA et intégrations. Lit le golden record ; ne le maîtrise pas.

L'exemple courant est une partie générique. L'architecture s'applique inchangée à un client, un citoyen, un patient, un fournisseur ou une personne sous mandat de défense — et aux entités non-parties (actifs, lieux, produits, comptes).

1.3 Le même modèle entre secteurs

La structure est constante ; seule l'instanciation change. Un déploiement nomme son entité, assigne chaque système à un rôle et fixe la précedence de survivance — rien dans l'architecture ne bouge.

Commercial / entreprise

Entité : client, fournisseur, produit. Rôles : CRM (engagement), ERP (référence), facturation, e-commerce (entrée).

Secteur public / gouvernement

Entité : citoyen, dossier, prestataire. Rôles : inscription (entrée), gestion de cas (référence), portails (engagement).

Santé

Entité : patient, prestataire, consultation. Rôles : admission/ADT (entrée), dossier clinique (référence), planification (exploitation).

Défense / sécurité publique

Entité : personne, unité, plateforme, actif. Rôles : système du personnel (entrée), C2 opérationnel (exploitation), logistique (référence).

Main-d'œuvre / RH

Entité : employé, poste, unité. Rôles : SIRH (entrée), paie (référence), planification (exploitation), annuaire (engagement).

Le modèle étant agnostique à l'entité, on peut maîtriser d'abord l'entité la plus problématique puis en ajouter d'autres sans ré-architecturer.

1.4 Portée

Dans la portée (Phase 1) : une entité maître et ses attributs de base, sa hiérarchie et la référence croisée d'identité. Hors portée (phases ultérieures) : d'autres domaines de données maîtres, l'intégration transactionnelle complète et la mise hors service des interfaces héritées.

2. Principes et décisions d'architecture

L'architecture est façonnée par quatre décisions critiques, plus un choix de style MDM.

Décision 1 • Ingestion multi-étapes

Les données passent par trois zones nommées — Bronze (brut), Argent (nettoyage et résolution), Or (golden records finis). Chaque zone est une frontière durable et gouvernée.

Décision 2 • Sous-systèmes explicites

Chaque capacité du hub est nommée, bornée et dotée d'une responsabilité.

Décision 3 • Provenance et chaîne de possession

Pour chaque attribut de référence, on enregistre son origine et la séquence des décisions — de quoi reconstituer et défendre tout enregistrement.

Décision 4 • Curation automatisée + humain dans la boucle

Le moteur de curation résout automatiquement quand la confiance est élevée et route l'ambigu, l'incertain ou le sensible vers un intendant. La séparation est régie par des seuils et des politiques.

2.1 Style choisi — Coexistence

Parmi les quatre styles MDM classiques, ce modèle adopte la Coexistence : le hub maîtrise l'enregistrement et reste synchronisé de façon bidirectionnelle avec les sources, qui restent les systèmes d'entrée de leurs propriétaires.

Registre — rejeté

Index mince renvoyant aux sources ; golden record virtuel à la lecture. Trop peu de contrôle.

Consolidation — partiel

Copies centralisées surtout pour les rapports. Bon pour l'analytique, ne renvoie pas de données fiables aux opérations.

Coexistence — retenu

Le hub maîtrise, nettoie et synchronise bidirectionnellement. Équilibre l'autorité et un changement acceptable des sources.

Centralisé — option ultérieure

Le hub devient l'unique système d'entrée. Cohérence maximale, mais la plus grande perturbation organisationnelle.

2.2 Principes de soutien

- Programme, pas projet : gouvernance et intendance sont de premier plan et continues.
- Les données brutes sont immuables : le Bronze est le registre rejouable de tout ce qui est reçu.
- Chaque décision est attribuable : aucune valeur n'atteint l'Or sans source et justification enregistrées.
- Source de données fiables, pas seulement une source unique de vérité.

3. Architecture de référence de haut niveau

L'architecture se lit de haut en bas comme un flux et de gauche à droite par zone : les sources alimentent l'intégration ; les données atterrissent en Bronze, sont curées en Argent, maîtrisées en Or ; les golden records sont publiés. Deux plans transversaux couvrent chaque couche : gouvernance et intendance ; sécurité, traçabilité et audit.

4. Catalogue des sous-systèmes

Chaque sous-système a une responsabilité claire, couvrant tout le cycle de vie d'une entité maître.

SUB-01 • Intégration et connecteurs

Capte chaque source (CDC, API, lot/fichier, événements), alimente le Bronze et syndique les golden records.

SUB-02 • Modèle et schéma d'entité

Définit l'entité maître, ses attributs, relations et hiérarchies.

SUB-03 • Moteur de qualité

Profile, standardise, valide et enrichit selon les six dimensions de qualité.

SUB-04 • Appariement et dédoublonnage

Résolution d'identité par appariement exact, flou, probabiliste et assisté par IA.

SUB-05 • Survivance et résolution de conflits

Fusionne le meilleur attribut de chaque source et résout les valeurs en conflit.

SUB-06 · Flux d'intendance (HITL)

Revue humaine, gestion des exceptions et approbation des changements incertains.

SUB-07 · Golden record et XREF

Assemble et stocke l'enregistrement faisant autorité et la référence croisée.

SUB-08 · Hiérarchies et relations

Maintient les regroupements parent/enfant et les lignes hiérarchiques.

SUB-09 · Métadonnées et données de référence

Définitions, propriété, règles de qualité et listes de codes contrôlées.

SUB-10 · Sécurité, traçabilité et audit

Contrôle d'accès plus provenance et chaîne de possession pour chaque valeur et décision.

SUB-11 · Console de gouvernance et d'intendance

La surface opérationnelle pour définitions, droits de décision, politiques et exceptions.

SUB-12 · Publication et syndication

Pousse les golden records aux consommateurs opérationnels (resync) et à l'analytique.

5. Le pipeline d'ingestion — Bronze · Argent · Or

Un pipeline médaillon à trois zones. La promotion n'a lieu qu'au franchissement du critère suivant ; rien n'est supprimé du Bronze, donc tout est rejouable.

5.1 Bronze — zone d'atterrissage

Reçoit les données telles qu'envoyées, sans les modifier ; chaque enregistrement est estampillé. C'est le registre rejouable : l'Argent peut être reconstruit sans réextraire des sources.

5.2 Argent — zone de travail

Où se font le nettoyage et la résolution : profiler et standardiser, valider et enrichir, apparier et dédoublonner, puis appliquer survivance et résolution de conflits. L'incertain est escaladé à l'intendance.

5.3 Or — zone fiable

Ne détient que des golden records finis. La promotion attribue un identifiant de référence stable, met à jour la XREF, rafraîchit les hiérarchies et publie ; en coexistence, les valeurs corrigées sont resynchronisées vers les sources.

PROPRIÉTÉS DES ZONES

Contenu Bronze : brut ; Argent : grappes en cours ; Or : golden records finis.

Mutabilité Bronze : immuable ; Argent : re-dérivable ; Or : versionné, contrôlé.

Accès Bronze : pipeline et audit ; Argent : moteur et intendants ; Or : consommateurs (lecture), resync.

6. Moteur de curation — automatisé et humain dans la boucle

Le moteur de curation décide de l'identité et résout les conflits, avec une séparation délibérée entre une voie automatisée (décisions confiantes) et une voie HITL (ambigu, incertain, sensible), régie par des seuils et des politiques.

6.1 La voie automatisée

Au-dessus du seuil de fusion automatique et sans attribut sensible en conflit, le moteur agit seul : il fusionne, applique la survivance, enregistre la règle et sa justification, et avance vers l'Or. C'est la voie à fort volume.

6.2 La voie humain dans la boucle

L'incertain ou le sensible est routé à un intendant qui voit la grappe, les valeurs concurrentes, les scores et la traçabilité, puis décide. Chaque action est capturée avec l'identité et un motif, et sert de signal d'ajustement.

6.3 Où un humain doit intervenir

Appariement médian — HITL

Score entre les seuils de fusion et de non-appariement — identité réellement ambiguë.

Conflit entre sources autoritaires — HITL

Deux sources fiables divergent et la survivance ne tranche pas.

Changement d'attribut sensible — HITL

Les champs réglementés (identifiant national, données protégées) exigent une approbation explicite.

Fusion / scission douteuse — HITL

Une fusion combinerait deux entités réelles, ou une sur-fusion doit être annulée.

Exception de validation — HITL

Un enregistrement échoue à une règle non auto-corrigeable (date impossible, boucle de hiérarchie).

Attribut nouveau / non mappé — HITL

Une source envoie un champ sans mappage ; un intendant décide de son traitement.

Appariement clair à forte confiance — AUTO

Aucun humain requis — appliqué automatiquement avec justification enregistrée.

7. Sécurité, traçabilité et auditabilité

Cette capacité répond, pour toute valeur : qui peut la voir ou la modifier (sécurité), d'où elle vient (provenance), et quelle séquence de décisions l'a produite (chaîne de possession). Elle capture des preuves à chaque étape.

7.1 Sécurité et contrôle d'accès

Accès par rôle et attribut sur zones et champs ; chiffrement et masquage des attributs sensibles ; moindre privilège par zone — le Bronze n'est jamais édité.

7.2 Traçabilité et provenance

Pour chaque attribut de référence, le hub enregistre la source gagnante, les valeurs concurrentes et la règle ou décision qui a tranché. La provenance est au niveau attribut, non seulement enregistrement.

7.3 Audit et chaîne de possession

Chaque transformation et décision — automatique ou humaine — est écrite dans un journal en ajout seul, suffisant pour reconstituer et défendre tout enregistrement. Les régimes de vie privée (RGPD, LPRPDE) exigent de savoir où vivent les données ; la provenance au niveau attribut rend cela requêtable.

8. Golden record et publication

Le golden record est la représentation faisant autorité, assemblée attribut par attribut. Ce n'est pas la copie d'un système ; c'est le composite convenu plus la référence croisée.

Identifiant de référence et XREF

Chaque entité reçoit un identifiant stable ; la XREF l'associe à chaque clé locale des sources.

Hierarchies et relations

Un changement de relation est résolu une fois, dans le hub, et publié partout.

Publication et resynchronisation

Publié par événements et API ; en coexistence, les valeurs corrigées reviennent aux sources.

9. Gouvernance et intendance

Le MDM réussit ou échoue sur la gouvernance, non l'outillage. Ces rôles opèrent la console et fixent les politiques de la séparation auto/HITL.

Propriétaire des données

Autorité métier du domaine ; approuve définitions, survivance et précédence par attribut.

Intendant des données

Gardien quotidien de la qualité ; traite la file HITL — fusions, annulations, conflits, exceptions.

Dépositaire des données

Côté TI ; exploite et sécurise le hub, les magasins et la plomberie.

Conseil de gouvernance

Fixe les normes, arbitre les différends et approuve seuils et politiques.

10. Préoccupations transversales

- Évolutivité et rejeu : le Bronze immuable permet de reconstruire Argent et Or par rejeu.
- Idempotence : réingérer le même enregistrement donne le même résultat.
- Modes de latence : événement/CDC pour le quasi temps réel ; lots pour les charges en masse.
- Versionnage : golden records, règles et modèle d'entité versionnés.
- Surveillance de la qualité : les six dimensions mesurées en continu.
- Déploiement progressif : commencer par le domaine le plus problématique.

11. Entrepôt de données maîtres — aperçu et relation avec le MDM

Le DWH est la couche de consolidation analytique. Là où le MDM résout l'identité, le DWH intègre les données opérationnelles et transactionnelles plus larges, les conforme aux entités de référence, puis les reformate pour la requête et la modélisation.

Division nette de l'autorité : le MDM (Partie 1) est l'autorité des dimensions ; le DWH (Partie 2) rencontre ces dimensions conformes — il ne résout jamais l'identité lui-même, il la consulte auprès du MDM. Les deux étagent Bronze -> Argent -> Or, mais l'Argent du DWH est le modèle intégré normalisé et l'Or la forme dénormalisée prête pour l'analytique.

12. Le pipeline d'ingestion du DWH — Bronze · Argent · Or

Un flux médaillon à trois zones, ajusté pour la consolidation analytique : le Bronze reçoit les extraits bruts ; l'Argent consolide et normalise contre les données MDM ; l'Or dénormalise pour la consommation.

12.1 Bronze — zone d'atterrissage

Reçoit les extraits bruts tels quels, les estampille et les persiste de façon immuable ; le pipeline reste rejouable.

12.2 Argent — zone de consolidation (normalisée)

Consolidation contre le MDM : normaliser et nettoyer, résoudre les clés en identifiants de référence via la XREF, joindre les codes de référence et attacher les attributs de référence comme dimensions conformes. C'est ici que l'entité inconnue déclenche la zone d'attente.

12.3 Or — zone de service (dénormalisée)

Reforme le modèle en formes dénormalisées (schémas en étoile, tables larges, agrégats) et publie aux data marts, outils BI et magasin de caractéristiques. L'identité étant déjà conforme, l'Or peut dénormaliser librement.

PROPRIÉTÉS DES ZONES DU DWH

Contenu Bronze : extraits bruts ; Argent : modèle intégré consolidé ; Or : tables dimensionnelles / larges.

Forme Bronze : tel que reçu ; Argent : normalisé (3FN) ; Or : dénormalisé.

Dépendance au MDM Bronze : aucune ; Argent : identifiant de référence, XREF, données de référence ; Or : héritée de l'Argent.

Références d'identité Bronze : clés locales ; Argent : résolues en identifiants de référence ; Or :

identifiants de référence partout.

Consommateur Bronze : pipeline et audit ; Argent : pipeline (interne) ; Or : BI, IA/AA, ad hoc.

13. Catalogue des composants du pipeline

Chaque composant du pipeline DWH porte une responsabilité, implantant le flux automatisé et la voie d'exception de la zone d'attente.

DWH-01 • Extracteurs / connecteurs sources

Tirent les données de chaque source (CDC, lot, API, fichier) vers le Bronze.

DWH-02 • Magasin brut Bronze

Atterrissage durable et immuable des extraits ; le point de rejeu du pipeline.

DWH-03 • Normalisation et nettoyage

Conforme types et formats, corrige la structure, retire les doublons intra-source.

DWH-04 • Service de résolution de clés

Mappe chaque clé métier à son identifiant de référence via la XREF. La porte qui détecte l'inconnu.

DWH-05 • Jointure des données de référence

Résout les valeurs codées en descriptions standardisées via les données MDM.

DWH-06 • Enrichissement par données maîtres

Attache les attributs du golden record comme dimensions conformes.

DWH-07 • Modèle intégré / normalisé

Assemble le modèle intégré 3FN en Argent, toutes références pointant des identifiants de référence.

DWH-08 • Gestionnaire de la zone d'attente

Gare les enregistrements non résolus, soumet l'entité au MDM et retraite à la résolution.

DWH-09 • Dénormalisation / modélisateur dimensionnel

Transforme le modèle intégré en schémas en étoile et tables larges pour l'Or.

DWH-10 • Agrégation et pré-calcul

Matérialise résumés, métriques et regroupements pour des requêtes rapides.

DWH-11 • Couche de service / data marts

Expose l'Or à la BI, à l'analytique ad hoc et au magasin de caractéristiques.

DWH-12 • Orchestration et planification

Pilote le flux automatisé de bout en bout — dépendances, déclencheurs, reprises et rappel de la zone d'attente.

DWH-13 • Qualité et surveillance

Vérifie fraîcheur, exhaustivité et rapprochement entre zones ; lève des alertes.

DWH-14 · Traçabilité et audit (partagé)

Étend la traçabilité de la Partie 1 aux transformations du DWH.

14. Flux automatisé et zone d'attente

Le pipeline DWH s'exécute entièrement automatisé. Une seule exception, elle aussi automatique : une donnée référant une entité sans golden record est mise de côté plutôt que chargée avec une identité devinée.

14.1 La voie d'exception

Quand DWH-04 ne trouve pas d'identifiant, DWH-08 gare l'enregistrement, soumet l'entité au pipeline MDM et attend. Une fois le golden record publié et la XREF mise à jour, les enregistrements retenus sont réévalués automatiquement et poursuivent vers l'Or.

14.2 Pourquoi garer plutôt que deviner

Inventer une identité est l'autorité du MDM, non de l'entrepôt. Garer préserve la cohérence interne et le principe d'autorité unique : le DWH ne résout jamais l'identité, il défère au MDM.

14.3 Où les humains apparaissent (ou non)

La voie DWH est entièrement automatisée. Un humain n'intervient que dans la Partie 1 : si le moteur de curation du MDM ne résout pas l'entité, un intendant s'en charge là. Vu de l'entrepôt, c'est invisible.

ÉTATS DE LA ZONE D'ATTENTE

Garé	Enregistrement retenu ; entité soumise au MDM ; en attente d'un golden record. (automatique)
Déclenché	Le MDM a publié le golden record et mis à jour la XREF. (automatique)
Retraitement	Enregistrements retenus réévalués ; les clés résolvent désormais. (vers l'Or)
Libéré	Enregistrements conformés et promus ; entrée de la zone d'attente effacée. (terminé)
Escaladé	Le MDM n'a pu résoudre après reprises / délai. (exception)

15. De bout en bout : Partie 1 et Partie 2

Ensemble, les deux parties forment une seule plateforme avec une seule autorité d'identité. Le MDM résout et publie ; le DWH consolide le reste contre elles. La boucle de la zone d'attente est la jonction qui les garde synchronisés.

15.1 Points clés

- Les données maîtres sont les noms partagés de l'entreprise ; le MDM produit un golden record et est un programme, non un projet.
- L'architecture est indépendante du secteur : nommer l'entité, assigner les rôles, fixer la survivance — la structure ne change pas.
- Le pipeline Bronze -> Argent -> Or garde le brut immuable et rejouable et ne promeut que le fini.
- La curation est explicitement séparée : automatiser le confiant, escalader l'ambigu et le sensible.

- Chaque valeur porte sa provenance et chaque décision sa chaîne de possession — base de la confiance et de la conformité.
- La coexistence équilibre autorité et changement acceptable ; ne se centraliser que par choix délibéré.
- Le DWH consolide les données contre les golden records : Argent normalisé, Or dénormalisé et prêt pour l'IA.
- Le pipeline DWH est entièrement automatisé ; l'unique exception (entité sans golden record) est gérée par la mise en attente et le retraitement.
- Une seule autorité d'identité : l'entrepôt ne résout jamais les entités ; le MDM le fait, et l'entrepôt attend.

Références

- [1] MIOVSKI GROUP / Enkidu Enterprises (2026). Trust Assurance Overview (Origine), 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 (2e éd.).

FIN DU DOCUMENT