

TAXONOMIE DES DONNÉES

Classification et Gouvernance des Données Sensibles

Projet Federateur- ENSIAS

Plateforme de Gouvernance de Données
et Référentiel de Métadonnées

Apache Atlas & Apache Ranger

Realise par :

BAZZAOUI Younes

ELGARCH Youssef

IBNOU-KADY Nisrine

TOUZANI Yousef

Encadre par :

Professor. K.BAINA

Dr. GASMI Manal

Année Universitaire 2024-2025

Table des Matières

1. Introduction à la Taxonomie

Contexte et Objectifs

La taxonomie des données constitue la pierre angulaire d'une gouvernance efficace des données. Dans le cadre de notre projet de plateforme de gouvernance utilisant Apache Atlas et Apache Ranger, nous avons développé une classification complète des données sensibles adaptée au contexte marocain.

Importance de la Classification

- Conformité réglementaire (Loi 09-08, GDPR)
- Protection des données personnelles et sensibles
- Sécurisation des accès par politiques basées sur les tags
- Traçabilité et auditabilité complètes
- Facilitation de la découverte et du partage sécurisé

1.1 Niveaux de Classification

Niveau	Score	Description
PUBLIC	0	Données publiquement disponibles sans restriction
INTERNAL	25	Usage interne organisation uniquement
CONFIDENTIAL	50	PII - Données personnelles nécessitant protection (Email, Téléphone, Adresse)
RESTRICTED	75	SPI - Données hautement sensibles (CIN, IBAN, Dossier médical)
CRITICAL	100	Données critiques - Impact catastrophique si compromises (CVV, Mots de passe)

2. Classification par Domaine Métier

2.1 Domaine : Identité & Documents Officiels

Propriétaire du Domaine : Direction RH / Service État Civil

Description : Documents officiels d'identification émis par les autorités marocaines permettant l'identification unique d'une personne.

Type de Donnée	Caté g.	Risq ue	Description & Conformité
CIN Marocaine	SPI	Critiq ue	Format : 2 lettres + 5-6 chiffres (ex: AB123456) Conformité : Loi 09-08, GDPR Art.9 Sanction : 10K-600K MAD + prison Sécurité : Chiffrement AES-256, MFA
Passeport	SPI	Critiq ue	Format : 2 lettres + 6-7 chiffres (ex: AA1234567) Usage : Document voyage international Rétention : 10 ans après fin relation
Permis de conduire	PII	Élev é	Usage : Document identification secondaire Masking : Partiel recommandé

2.2 Domaine : Coordonnées & Contact

Propriétaire du Domaine : Service Marketing / CRM

Description : Informations permettant de contacter directement une personne.

Type de Donnée	Caté g.	Risq ue	Description & Conformité
Téléphone	PII	Moye n	Format : +212 6XX XXX XXX ou 06XX XXX XXX Risque : Spam, phishing, 2FA bypass Masking : +212 6XX XXX 789 (derniers chiffres)
Email	PII	Moye n	Risque : Phishing, reset password, usurpation GDPR : Droit à l'effacement applicable Masking : u***r@example.com
Adresse postale	PII	Moye n	Risque : Localisation physique, intrusion domicile Usage : Livraison, correspondance officielle

2.3 Domaine : Données Financières

Propriétaire du Domaine : Direction Financière / Trésorerie

Description : Informations bancaires et financières ultra-sensibles. Standard de conformité : PCI-DSS

Type de Donnée	Catég.	Risque	Description & Conformité
IBAN/RIB	SPI	Critique	Format : MA + 2 chiffres + 20 caractères Impact : Prélèvement frauduleux, perte financière directe Sécurité : AES-256, accès Finance uniquement Rétention : 7 ans après clôture compte
Carte bancaire	SPI	Critique	Standard : PCI-DSS Level 1 Règle absolue : JAMAIS stocker en clair Tokenization : Obligatoire pour stockage
CVV	SPI	Critique	Code 3 chiffres au dos de carte INTERDICTION TOTALE de stockage sous toute forme
Salaire	PII	Élevé	Sensibilité : Secret professionnel, discrimination possible Accès : RH + Paie uniquement

2.4 Domaine : Données Professionnelles

Propriétaire du Domaine : Direction RH

Description : Informations liées à l'emploi et la relation de travail.

Type de Donnée	Catég.	Risque	Description & Conformité
Numéro CNSS	SPI	Critique	Format : 12 chiffres Caractère : Unique, permanent, lie à tous droits sociaux Usage : Déclarations sociales, retraite, assurance Organisme : CNSS (Caisse Nationale Sécurité Sociale)
Matricule employé	PII	Moyen	Usage : Identifiant interne dans systèmes RH Sensibilité : Révèle appartenance à l'organisation

2.5 Domaine : Données Médicales

Propriétaire du Domaine : Service Médical / DPO

Protection : Secret médical absolu - Code pénal art. 446

Type de Donnée	Catég.	Risque	Description & Conformité
Dossier médical	SPI	Critique	Contenu : Historique santé complet du patient Protection légale : Secret médical (Code pénal art. 446) Impact violation : Poursuites pénales du médecin Rétention : 20 ans minimum
N° Assurance AMO	SPI	Critique	Organisme : CNOPS, CNSS Usage : Remboursement frais médicaux
Données biométriques	SPI	Critique	Types : Empreintes, reconnaissance faciale, iris Caractère : IMMuable - ne peut pas être changé Consentement : Explicite obligatoire

2.6 Domaine : Données Techniques

Propriétaire du Domaine : DSI / Service IT

Description : Données techniques et métadonnées IT moins sensibles mais réglementées.

Type de Donnée	Catégorie	Risque	Description & Conformité
Adresse IP	PII	Faible	GDPR : Considérée comme PII Usage : Géolocalisation approximative, tracking Rétention : Maximum 12 mois
Cookie ID	PII	Faible	Usage : Tracking comportement, publicité ciblée Obligation : Bannière de consentement

3. Matrice de Conformité Réglementaire

Cette section présente les principales réglementations applicables à notre contexte de gouvernance des données au Maroc.

Régulation	Scope	Autorité	Sanctions	Applicabilité
Loi 09-08	Maroc - Toutes données personnelles	CNDP	10K-600K MAD + prison	✓ Applicable
GDPR	UE - Données citoyens UE	CEPD	20M EUR ou 4% CA	⚠ Si clients UE
ISO 27001	International - Volontaire	ISO	Perte certification	Recommandé
PCI-DSS	Cartes bancaires	PCI SSC	5K-100K USD/mois	⚠ Si paiements

4. Intégration avec Apache Atlas & Ranger

4.1 Architecture de Classification

Notre taxonomie s'intègre directement dans l'écosystème Apache Atlas via plusieurs mécanismes :

- **Business Glossary** : Définition des termes métier par domaine
- **Classifications** : Tags PII, SPI, CONFIDENTIAL, etc.
- **Propagation** : Héritage automatique via lineage
- **Labels** : Étiquettes pour recherche et découverte

4.2 Politiques Apache Ranger

Les classifications Atlas sont utilisées pour créer des politiques d'accès dynamiques dans Ranger :

Groupe	Tag	Permissions	Masking
DPO_Team	SPI, CIN_MAROC	READ, WRITE	Aucun
RH_Managers	SPI, CIN_MAROC	READ	Partiel (3 derniers chiffres)
Analysts	SPI, CIN_MAROC	READ	Total (Hash)
Public	SPI, CIN_MAROC	DENY	N/A

5. Conclusion et Recommandations

5.1 Points Clés

Cette taxonomie constitue la base fondamentale pour une gouvernance efficace des données dans notre plateforme. Elle permet de :

- Classifier automatiquement les données selon leur sensibilité
- Appliquer les bonnes politiques de sécurité via Ranger
- Assurer la conformité avec la Loi 09-08 et GDPR
- Faciliter les audits et la traçabilité complète
- Protéger les données personnelles des citoyens marocains

5.2 Prochaines Étapes

Implémentation Technique :

- Import de la taxonomie dans Atlas via REST API
- Configuration des classifications et glossaires
- Mise en place des politiques Ranger tag-based
- Tests de propagation via lineage

Formation et Documentation :

- Formation des data stewards sur la classification
- Documentation des procédures de gouvernance
- Guides d'utilisation Atlas et Ranger

--- Fin du Rapport ---