EPIC 4 — Stockage IA & données utilisateur (RGPD)

1. Objectif de l'EPIC

Définir et mettre en œuvre une **architecture de stockage des données IA et utilisateur conforme au RGPD**, adaptée aux métiers à données sensibles (comptabilité, santé, juridique), garantissant : - la **minimisation des données stockées**, - l'**isolation stricte par tenant et par utilisateur**, - la **confidentialité, intégrité et traçabilité** des données, - une base technique compatible avec l'exercice des droits RGPD (EPIC 5).

📌**Cadre RGPD** : cet EPIC met en œuvre les obligations relatives à la **structuration, à la conservation et à la sécurisation des données personnelles** (articles 5, 25 et 32 du RGPD). L'exercice effectif des droits (accès, effacement, portabilité) est traité dans **EPIC 5**.

---

2. Positionnement dans la roadmap

- EPIC 1 — Socle applicatif sécurisé
- EPIC 2 — Durcissement serveur & réseau
- EPIC 3 — Validation technique IA locale (POC contrôlé)
- **EPIC 4 — Stockage IA & données utilisateur (ce document)**
- EPIC 5 — Pipeline RGPD (droits des personnes)
- EPIC 6 — Stack IA Docker RGPD-ready (industrialisation)
- EPIC 7 — Kit conformité & audit

---

3. Périmètre couvert

Inclus

- Stockage des données utilisateur liées aux usages IA
- Stockage des entrées / sorties IA (prompts, réponses) lorsqu'elles sont conservées
- Stockage des métadonnées (journaux fonctionnels, références d'événements)
- Chiffrement des données au repos
- Politique de rétention et de classification des données

Exclus

- Sécurité infra bas niveau (EPIC 2)
- Gouvernance des accès applicatifs (EPIC 1)
- Exécution des droits RGPD (EPIC 5)
- Orchestration et runtime IA (EPIC 6)

---

4. Principes directeurs (non négociables)

1. **Minimisation des données stockées** (aucune conservation par défaut)
2. **Isolation stricte tenant / utilisateur**
3. **Chiffrement systématique au repos**
4. **Traçabilité sans exposition de contenu sensible**
5. **Durées de conservation explicites et documentées**

---

5. Typologie des données stockées

Données utilisateur

• Identifiants techniques (user_id, tenant_id)
• Préférences fonctionnelles
• Paramètres de confidentialité

Données IA

• Prompts et réponses **uniquement si nécessaires**
• Données pseudonymisées ou redactionnées
• Embeddings vectoriels

Métadonnées

• Horodatage
• Type d'opération
• Références anonymisées

---

6. Architecture de stockage cible

• Base relationnelle sécurisée (données structurées)
• Stockage chiffré pour données temporaires
• Base vectorielle pour embeddings
• Séparation logique par tenant
• Clés de chiffrement segmentées

---

7. User Stories & Exigences

US-1 — Isolation stricte du stockage par tenant

Je veux garantir qu'aucune donnée ne soit accessible entre tenants

Exigences - Partitionnement logique par `tenant_id` - Vérification systématique côté API - Tests d'isolation automatisés

---

US-2 — Minimisation et opt-in de conservation

Je veux ne stocker que les données strictement nécessaires

Exigences - Conservation désactivée par défaut - Opt-in explicite par tenant / utilisateur - Justification fonctionnelle documentée

---

US-3 — Chiffrement des données au repos

Je veux protéger les données stockées

Exigences - Chiffrement fort au repos - Clés distinctes par tenant - Rotation des clés

---

US-4 — Classification et rétention des données

Je veux maîtriser la durée de conservation

Exigences - Classification par type de donnée - Durée de rétention configurable - Suppression automatique en fin de cycle

---

US-5 — Traçabilité sans exposition de contenu

Je veux tracer les usages sans stocker de données sensibles

Exigences - Logs orientés événements - Références anonymisées - Aucun contenu métier en clair

---

US-6 — Stockage compatible avec les droits RGPD

Je veux permettre l'exercice futur des droits RGPD

Exigences - Données indexées par utilisateur - Mécanismes de suppression logique - Compatibilité avec export structuré

---

8. Dépendances

- Dépend de : EPIC 1, EPIC 2
- Alimenté par : EPIC 3 (contraintes techniques)
- Prérequis pour : EPIC 5, EPIC 6, EPIC 7

---

9. Livrables

- Schéma de stockage documenté
- Politique de rétention validée

- Chiffrement opérationnel
- Isolation tenant vérifiée
- Documentation des types de données

---

10. Risques & points de vigilance

- Surstockage inutile → non-conformité
- Mauvaise isolation → fuite inter-tenant
- Rétention trop longue → risque RGPD

---

11. Definition of Done (DoD)

- Conservation désactivée par défaut
- Chiffrement actif pour toutes les données
- Isolation tenant testée
- Rétention automatisée
- Documentation RGPD prête

Fin EPIC 4