

# DATA CATALOGUE & DATA DICTIONARY

## Projet MDM - Clinique Privée

---

### 1. MDM\_PATIENT

#### DATA CATALOGUE

**Nom de la table :** `mdm.mdm_patient`

**Description :** Référentiel maître unique des patients de la clinique. Consolide les informations patients provenant des systèmes de rendez-vous, ERP (dossiers médicaux, laboratoire) et facturation. Élimine les doublons et garantit une vue 360° fiable de chaque patient pour améliorer la qualité des soins et la sécurité médicale.

**Propriétaire métier :** Dr. Sophie Laroche - Directrice Médicale

**Propriétaire technique :** Équipe Data Governance - Service IT Clinique

**Fréquence de mise à jour :** Temps réel (streaming via Kafka)

**Volumétrie :**

- Patients actifs : ~80 (après dédoublonnage)
- Enregistrements sources totaux : ~230 (avant consolidation MDM)
- Croissance estimée : +150 nouveaux patients/an

**Sensibilité :**

- **TRÈS ÉLEVÉE** - Contient des données de santé à caractère personnel (RGPD Art. 9)
- Données Personnelles Identifiables (PII)
- Données sensibles de santé (HADS - Health and Social Care Data)

**Sources de données :**

1. **Système Rendez-vous** (source\_rdv\_patients.csv)
  - Colonnes : nom, prénom, sexe, téléphone, adresse
  - Qualité : Doublons présents (~20%), incohérences de format
2. **Système ERP** (source\_erp\_patients.csv + source\_erp\_dossiers\_medicaux.csv)
  - Colonnes : nom, prénom, date\_naissance, email, num\_dossier, historique\_médical

- Qualité : IDs différents du système RDV, format dates DD/MM/YYYY
- 3. **Système Facturation** (source\_fact\_patients.csv)
  - Colonnes : nom\_famille, prénoms, adresse\_facturation
  - Qualité : Encore d'autres IDs, noms de colonnes différents

**Services consommateurs :**

- **Dossier Patient Partagé** (Application Web Médecins) - Accès lecture/écriture
- **Dashboards Médicaux** (Direction & Qualité) - Accès lecture, analyses statistiques
- **Système d'Alertes Médicales** (Notifications automatiques) - Accès lecture temps réel
- **Outils d'Analyse Statistique** (Études épidémiologiques) - Accès lecture agrégée
- **Module Facturation** (Lien vers factures patients) - Accès lecture
- **Admissions & Secrétariat** (Prise de RDV, coordination soins) - Accès lecture

**Relations MDM :**

- **Lié à MDM\_Médecin** : Un patient peut être suivi par plusieurs médecins (relation gérée dans systèmes opérationnels)
- **Lié à MDM\_Service** : Les services (labo, imagerie) produisent des données liées aux patients (relation indirecte)

## DATA DICTIONARY - MDM\_PATIENT

Colonne	Type	Description métier	Contraintes	Sensibilité	Source(s)	Règle de qualité
<b>Patient_ID</b>	UUID	Identifiant unique universel du patient dans le référentiel MDM. Généré automatiquement et immuable à vie. Utilisé comme clé de référence dans tous les systèmes consommateurs.	PRIMARY KEY, NOT NULL, UNIQUE, DEFAULT gen_random_uuid()	Faible	Généré par PostgreSQL (MDM)	Complétude = 100% (auto-généré)
<b>Nom</b>	VARCHAR(100)	Nom de famille du patient. Normalisé en MAJUSCULES sans accents pour faciliter recherches et matching. En cas	NOT NULL, Longueur 2-100 caractères	Très Élevée (PII)	RDV (nom), ERP (nom), FACT (nom_famille)	Complétude = 100%, Normalisation UPPER + sans accents

		de doublon, prend la valeur la plus fréquente ou la plus récente.				
<b>Prénom</b>	VARCHAR(100)	Prénom du patient. Conserve la casse d'origine (majuscule initiale). En cas de doublon, prend la valeur la plus fréquente.	NOT NULL, Longueur 2-100 caractères	Très Élevée (PII)	RDV (prenom), ERP (prenom), FACT (prenoms)	Complétude = 100%, Capitalisation standardisée
<b>Date_Naissance</b>	DATE	Date de naissance du patient au format YYYY-MM-DD. Champ critique pour le dédoublement et le calcul d'âge.	NOT NULL, Date < TODAY, Date > '1900-01-01'	Très Élevée (PII)	ERP (date_naissance format DD/MM/YYYY)	Complétude = 100%, Format YYYY-MM-DD, Cohérence temporelle

<b>Sexe</b>	VARCH AR(10)	Sexe biologiqu e du patient. Valeurs standardi sées : 'Masculin 'Féminin'. Utilisé pour statistiqu es épidémiol ogiques et protocole s de soins spécifiqu es.	CHECK (Sexe IN ('Masculin', 'Féminin')), NULL autorisé	Élevée (PII)	RDV (sexe)	Complé tude cible ≥90%, Valeurs contrôl ées
<b>Adresse</b>	VARCH AR(255)	Adresse résidenti elle complète du patient. Utilisée pour correspo ndance postale et localisati on géograph ique (études de santé publique) .	NULL autorisé, Longueur max 255	Élevée (PII)	RDV (adresse), ERP (adresse), FACT (adresse_factura tion)	Complé tude cible ≥85%, Consoli dation meilleur e source

<b>Téléphone</b>	VARCHAR(20)	Numéro de téléphone principal du patient. Normalisé à 10 chiffres sans espaces ni caractères spéciaux. Utilisé pour rappels RDV et urgences.	NULL autorisé, Format 10 chiffres, Regex <code>^[0-9]{10}\$</code>	Élevée (PII)	RDV (telephone), FACT (tel_contact)	Complétude cible $\geq 90\%$ , Validation format 10 chiffres
<b>Email</b>	VARCHAR(255)	Adresse email du patient. Validée par regex. Utilisée pour communications non-urgentes et accès portail patient.	NULL autorisé, Format email valide, Regex RFC 5322	Élevée (PII)	ERP (email), FACT (email_contact)	Complétude cible $\geq 70\%$ , Validation regex email
<b>Num_Dossier</b>	VARCHAR(50)	Identifiant unique du dossier médical patient dans le système	NULL autorisé, UNIQUE si renseigné	Très Élevée (Santé)	ERP (num_dossier format DOSS-XXXXXX)	Complétude cible $\geq 95\%$ , Unicité = 100%

		ERP. Permet la traçabilité vers l'historique médical complet.				
<b>Historique_Médical</b>	TEXT	Antécédents médicaux, allergies et pathologies chroniques du patient. Champ libre enrichi au fil du temps. CRITIQUE pour décisions médicales et prévention erreurs médicamenteuses.	NULL autorisé, Longueur max 10000 caractères	CRITIQUE (Donnée de santé)	ERP (source_erp_dossiers_medicaux : historique_médical + allergies)	Complétude cible ≥70%, Enrichissement continu
<b>Date_Création</b>	TIMESTAMP	Date et heure de création de l'enregistrement MDM. Utilisée pour audit et traçabilité.	NOT NULL, DEFAULT CURRENT_TIMESTAMP	Faible (Métadonnée)	Généré automatiquement par MDM	Complétude = 100% (auto)

## 2. MDM\_MEDECIN

### DATA CATALOGUE

**Nom de la table :** `mdm.mdm_medecin`

**Description :** Référentiel maître unique des médecins et praticiens exerçant dans la clinique. Consolide les informations professionnelles provenant du système RH/ERP et du système de rendez-vous. Garantit l'identification unique de chaque praticien et facilite la gestion des plannings, affectations de services et traçabilité des prescriptions.

**Propriétaire métier :** Marie Dubois - Directrice des Ressources Humaines Médicales

**Propriétaire technique :** Équipe Data Governance - Service IT Clinique

**Fréquence de mise à jour :** Quotidienne (batch nocturne à 02h00) + Temps réel pour modifications urgentes

**Volumétrie :**

- Médecins actifs : 6
- Médecins historiques (retraités/partis) : Conservés pour traçabilité
- Croissance estimée : +2-3 nouveaux praticiens/an

**Sensibilité :**

- **MOYENNE** - Données professionnelles, non publiques
- Informations de contact professionnel
- Numéro de licence (RPPS) = donnée sensible professionnelle

**Sources de données :**

1. **Système RH/ERP** (source\_erp\_medecins.csv) - SOURCE PRINCIPALE
  - Colonnes : nom, prénom, spécialité, num\_licence, email\_pro, téléphone\_pro, service\_affecté, disponibilités
  - Qualité : Données complètes et validées RH
2. **Système Rendez-vous** (source\_rdv\_medecins.csv) - SOURCE SECONDAIRE
  - Colonnes : nom\_complet (format "Dr. Prénom Nom"), spécialité, téléphone, email
  - Qualité : Format inconsistant, données partielles

**Services consommateurs :**

- **Module Planning** (Gestion agenda médecins) - Accès lecture/écriture
- **Système Rendez-vous** (Affectation RDV patients) - Accès lecture
- **Dossier Patient** (Identification prescripteur/médecin traitant) - Accès lecture
- **Facturation** (Lien actes médicaux → médecin) - Accès lecture
- **Direction Médicale** (Tableaux de bord activité) - Accès lecture agrégée



- **Analyses Statistiques** (Études de performance, charge de travail) - Accès lecture

**Relations MDM :**

- **Lié à MDM\_Service** : Un médecin appartient à un service hospitalier (FK Service\_ID)
  - **Relation inverse** : Un médecin peut être responsable d'un service (lien via MDM\_Service.Responsable\_ID)
-

## DATA DICTIONARY - MDM\_MEDECIN

Colonne	Type	Description métier	Contraintes	Sensibilité	Source(s)	Règle de qualité
<b>Medecin_ID</b>	UUID	Identifiant unique universel du médecin dans le référentiel MDM. Généré automatiquement et immuable. Utilisé comme clé de référence dans tous les systèmes.	PRIMARY KEY, NOT NULL, UNIQUE, DEFAULT gen_random_uuid()	Faible	Généré par PostgreSQL (MDM)	Complétude = 100% (auto-généré)
<b>Nom</b>	VARCHAR (100)	Nom de famille du médecin. Normalisé en MAJUSCULES pour cohérence. Utilisé pour recherche et affichage dans interfaces.	NOT NULL, Longueur 2-100 caractères	Moyenne (Professionnel)	ERP (nom), RDV (parsing de nom_complet)	Complétude = 100%, Normalisation UPPER

<b>Prénom</b>	VARCHAR (100)	Prénom du médecin. Conserve la casse d'origine (majuscule initiale). Utilisé pour communication professionnelle.	NOT NULL, Longueur 2-100 caractères	Moyenne (Professionnel)	ERP (prenom), RDV (parsing de nom_complet)	Complétude = 100%, Capitalisation standard
<b>Spécialité</b>	VARCHAR (100)	Spécialité médicale principale du praticien (Cardiologie, Pédiatrie, etc.). Valeurs contrôlées par table de référence. Utilisée pour affectation RDV et analyses statistiques.	NOT NULL, Valeurs contrôlées (table ref_specialites), ENUM recommandé	Faible	ERP (specialite), RDV (specialite)	Complétude = 100%, Valeurs contrôlées via FK
<b>Service_ID</b>	UUID	Identifiant du service hospitalier d'affectation du médecin. Clé étrangère vers MDM_Service.	FOREIGN KEY (MDM_Service), NULL autorisé	Faible	ERP (service_affecte - matching par nom service)	Complétude cible ≥95%, Intégrité référentielle

		Définit le rattachement organisationnel.				
<b>Num_Licence</b>	VARCHAR (11)	Numéro RPPS (Répertoire Partagé des Professionnels de Santé). Identifiant national unique du praticien. Format : 11 chiffres. CRITIQUE pour validation légale d'exercice.	UNIQUE, NOT NULL, Format '[0-9]{11}\$', Validation via API RPPS	Élevée (ID professionnel)	ERP (num_licence)	Complétude = 100%, Unicité = 100%, Validation API RPPS
<b>Disponibilité</b>	JSON	Horaires de disponibilité du médecin par jour de la semaine. Format JSON : {"lundi": "08:00-18:00", "mardi": "08:00-18:00", ...}. Utilisé pour planification RDV.	NULL autorisé, Format JSON valide, Schema validation	Faible	ERP (disponibilité_lundi, _mardi, etc. - agrégation)	Complétude cible ≥80%, Validation schema JSON

<b>Email</b>	VARCHAR (255)	Adresse email professionnelle du médecin (domaine @clinique.fr). Utilisée pour communications internes et notifications système.	NULL autorisé, Format email valide, Domaine @clinique.fr	Moyenne (Contact pro)	ERP (email_pro), RDV (email)	Complétude cible ≥95%, Validation domaine + regex
<b>Téléphone</b>	VARCHAR (20)	Numéro de téléphone professionnel du médecin. Normalisé à 10 chiffres. Utilisé pour contact urgences et coordination soins.	NULL autorisé, Format 10 chiffres	Moyenne (Contact pro)	ERP (telephone_pro), RDV (telephone)	Complétude cible ≥90%, Validation format

---

## 3. MDM\_SERVICE

### DATA CATALOGUE

**Nom de la table :** `mdm.mdm_service`

**Description :** Référentiel maître unique des services et départements de la clinique. Consolide la structure organisationnelle incluant services médicaux (Cardiologie, Pédiatrie, etc.), services techniques (Laboratoire, Imagerie) et services administratifs. Facilite la gestion hiérarchique, la planification et la traçabilité des activités par service.

**Propriétaire métier :** Jean Moreau - Directeur des Opérations

**Propriétaire technique :** Équipe Data Governance - Service IT Clinique

**Fréquence de mise à jour :** Hebdomadaire (batch dimanche 23h00) + Ponctuel lors de réorganisations

**Volumétrie :**

- Services actifs : 6 (Cardiologie, Pédiatrie, Radiologie, Chirurgie, Gynécologie, Urgences)
- Services historiques : Conservés pour traçabilité
- Croissance estimée : +1 nouveau service tous les 2-3 ans

**Sensibilité :**

- **FAIBLE** - Données organisationnelles publiques en interne
- Informations de contact publiques (standard téléphonique)

**Sources de données :**

1. **Système ERP** (source\_erp\_services.csv) - SOURCE UNIQUE
  - Colonnes : service\_id\_erp, nom\_service, description, responsable\_nom, localisation, horaires, téléphone, email
  - Qualité : Données complètes et validées Direction

**Services consommateurs :**

- **Système Rendez-vous** (Affichage services disponibles pour RDV) - Accès lecture
- **Dossier Patient** (Traçabilité services consultés) - Accès lecture
- **Module Planning** (Affectation médecins aux services) - Accès lecture
- **Facturation** (Imputation coûts par service) - Accès lecture
- **Direction** (Tableaux de bord par service, KPIs) - Accès lecture agrégée

- **Patients & Secrétariat** (Informations pratiques : localisation, horaires) - Accès lecture public

**Relations MDM :**

- **Relation avec MDM\_Médecin** : Un service a plusieurs médecins affectés (FK depuis Médecin.Service\_ID)
- **Relation hiérarchique** : Un service peut avoir un médecin comme responsable (Responsable\_ID → Medecin\_ID)
- **Relation indirecte avec MDM\_Patient** : Services (labo, imagerie) produisent des données liées aux patients

---

## DATA DICTIONARY - MDM\_SERVICE

Colonne	Type	Descripti on métier	Contraintes	Sensib ilité	Source(s)	Règle de qualit é
<b>Service_ID</b>	UUID	Identifiant unique universel du service dans le référentiel MDM. Généré automatiquement et immuable. Utilisé comme clé de référence dans tous les systèmes.	PRIMARY KEY, NOT NULL, UNIQUE, DEFAULT gen_random_uuid()	Faible	Généré par PostgreSQL (MDM)	Complétude = 100% (auto-généré)
<b>Nom_Service</b>	VARCHAR(100)	Nom officiel du service hospitalier (ex: "Cardiologie", "Laboratoire Biochimie"). Valeur unique servant de référence dans toute la clinique.	NOT NULL, UNIQUE, Longueur 3-100 caractères	Faible	ERP (nom_service)	Complétude = 100%, Unicité = 100%



<b>Descripti on</b>	TEXT	Descriptio n détaillée du service : missions, types de soins, équipeme nts dispo nibles. Util isée pour document ation in terne et in formatio n patients.	NULL autorisé, Longueur max 1000 caractères	Faible	ERP (description)	Compl étude cible ≥90%
<b>Respons able_ID</b>	UUID	Identifiant du médecin responsa ble/chef de service. Clé étrangère vers MDM_Mé decin. Définit le leadershi p médical du service.	FOREIGN KEY (MDM_Medecin ) , NULL autorisé	Faible	ERP (responsable_n om - matching vers MDM_Medecin )	Compl étude cible ≥80%, Intégrit é référe ntielle
<b>Localisat ion</b>	VARCHAR(200)	Localisati on physique du service dans la clinique. Format standardi sé : "Bâtiment	NULL autorisé, Format "Bâtiment X, Étage Y"	Faible (Public interne)	ERP (localisation)	Compl étude cible ≥95%, Validat ion format

		[A-Z], Étage [1-9]". Utilisée pour orientatio n patients et staff.				
<b>Horaires</b>	JSON	Horaires d'ouvertur e du service par jour de la semaine. Format JSON : { "lundi": "08:00-18 :00", ... }. Utilisé pour informatio n patients et gestion des accès.	NULL autorisé, Format JSON valide, Schema validation	Faible (Public interne)	ERP (horaires Ouver ture_lundi, _mardi, etc. - agrégation)	Compl étude cible ≥90%, Validat ion schem a JSON
<b>Date_Cré ation</b>	TIMESTA MP	Date et heure de création de l'enregistr ement MDM. Utilisée pour audit et traçabilité organisati onnelle.	NOT NULL, DEFAULT CURRENT_TI MESTAMP	Faible (Métad onnée)	Généré automatique ment par MDM	Compl étude = 100% (auto)

# ANNEXES

## A. Glossaire des Termes Métier

**MDM (Master Data Management)** : Processus de gestion des données de référence permettant d'établir une source unique de vérité (Single Source of Truth - SSOT) pour les données critiques de l'entreprise.

**Golden Record** : Enregistrement de référence unique consolidant les meilleures données provenant de multiples sources, après dédoublonnage et enrichissement.

**PII (Personally Identifiable Information)** : Données personnelles permettant d'identifier directement ou indirectement une personne physique (nom, prénom, date de naissance, etc.).

**HADS (Health and Social Care Data)** : Données de santé à caractère personnel, protégées par réglementation stricte (RGPD Art. 9).

**RPPS (Répertoire Partagé des Professionnels de Santé)** : Identifiant national unique attribué à chaque professionnel de santé en France (11 chiffres).

**Dédoublonnage** : Processus d'identification et de fusion des enregistrements en doublon représentant la même entité réelle.

**Score de Qualité** : Indicateur numérique (0-100) mesurant la fiabilité d'un enregistrement basé sur complétude, exactitude, cohérence et fraîcheur.

**UUID (Universally Unique Identifier)** : Identifiant unique universel de 128 bits garantissant l'unicité globale sans coordination centrale.

---

## B. Mapping Sources → MDM (Traçabilité)

### MDM\_PATIENT

Colonne MDM	Source RDV	Source ERP	Source FACT
Patient_ID	-	-	- (généré)
Nom	nom	nom	nom_famille
Prénom	prenom	prenom	prenoms
Date_Naissance	-	date_naissance	-
Sexe	sexe	-	-

Adresse	adresse	adresse	adresse_facturation
Téléphone	telephone	-	tel_contact
Email	-	email	email_contact
Num_Dossier	-	num_dossier	-
Historique_Médical	-	historique_medical + allergies	-

## MDM\_MEDECIN

Colonne MDM	Source RDV	Source ERP
Medecin_ID	-	- (génééré)
Nom	nom_complet (parsing)	nom
Prénom	nom_complet (parsing)	prenom
Spécialité	specialite	specialite
Service_ID	-	service_affecte (matching)
Num_Licence	-	num_licence
Disponibilité	-	disponibilite_* (agrégation)
Email	email	email_pro
Téléphone	telephone	telephone_pro

## MDM\_SERVICE

Colonne MDM	Source ERP
Service_ID	- (génééré)
Nom_Service	nom_service
Description	description
Responsable_ID	responsable_nom (matching)
Localisation	localisation
Horaires	horaires_ouverture_* (agrégation)

---

## C. Règles de Dédoublonnage Patient

Deux enregistrements patients sont considérés comme doublons si **AU MOINS UNE** des conditions suivantes est vraie :

1. **(Nom similaire à 85%+) ET (Prénom similaire à 85%+) ET (Date naissance identique)**
  - Similarité calculée via distance de Levenshtein
2. **Téléphone identique (10 chiffres)**
  - Ignore doublons avec téléphone vide ou invalide
3. **Email identique (format valide)**
  - Comparaison insensible à la casse
4. **Num\_Dossier identique**
  - Si renseigné dans les deux enregistrements

**Fusion des doublons (Golden Record) :**

- Nom : Valeur la plus fréquente ou la plus récente
  - Prénom : Valeur la plus fréquente
  - Date\_Naissance : Valeur la plus ancienne (source ERP prioritaire)
  - Autres champs : Prioriser valeurs non vides, sinon source ERP > RDV > FACT
- 

## D. Calcul du Score de Qualité

**Formule globale :**

$\text{Score\_Qualité} = (\text{Complétude} \times 40\%) + (\text{Exactitude} \times 30\%) + (\text{Cohérence} \times 20\%) + (\text{Fraîcheur} \times 10\%)$

### Dimension 1 : Complétude (40%)

$\text{Complétude} = (\text{Nombre de champs obligatoires remplis} / \text{Nombre total de champs obligatoires}) \times 100$

Champs obligatoires MDM\_Patient : Nom, Prénom, Date\_Naissance, Num\_Dossier  
Champs importants (bonus) : Sexe, Téléphone, Email, Adresse, Historique\_Médical

### Dimension 2 : Exactitude (30%)

$\text{Exactitude} = \text{Moyenne des validations réussies}$

Validations :

- Email : Format regex RFC 5322 valide ? (Oui=100, Non=0)
- Téléphone : Format 10 chiffres ? (Oui=100, Non=0)
- Date\_Naissance : Cohérence temporelle (< aujourd'hui, > 1900) ? (Oui=100, Non=0)

- Num\_Dossier : Format DOSS-XXXXXX ? (Oui=100, Non=0)

### Dimension 3 : Cohérence (20%)

Cohérence = Respect des règles métier

Règles :

- Date\_Naissance < Date\_Création (Oui=100, Non=0)
- Sexe dans valeurs autorisées (Oui=100, Non=0)
- Num\_Dossier unique dans MDM (Oui=100, Non=0)
- Au moins 1 ID source renseigné (RDV, ERP ou FACT) (Oui=100, Non=0)

### Dimension 4 : Fraîcheur (10%)

Fraîcheur = Score basé sur ancienneté dernière mise à jour

Si dernière\_maj < 7 jours : 100

Si 7 jours ≤ dernière\_maj < 30 jours : 80

Si 30 jours ≤ dernière\_maj < 180 jours : 60

Si 180 jours ≤ dernière\_maj < 365 jours : 40

Si dernière\_maj ≥ 365 jours : 20

Seuils d'interprétation :

- **Excellent** : Score ≥ 90
- **Bon** : 70 ≤ Score < 90
- **Moyen** : 50 ≤ Score < 70
- **Faible** : Score < 50 (nécessite action corrective)

---

## E. Standards de Normalisation

### Normalisation des Noms

1. Conversion MAJUSCULES : "dupont" → "DUPONT"
2. Suppression accents : "ÉLÉONORE" → "ELEONORE"
3. Suppression espaces multiples : "MARTIN PAUL" → "MARTIN PAUL"
4. Trim espaces début/fin

Table de remplacement des accents :

Caractère	Remplacement
	t

É, È, Ê, Ë	E
À, Â, Ä	A
Ù, Ú, Ü	U
Ô, Ö	O
Î, Ï	I
Ç	C

## Normalisation des Dates

Formats sources détectés :

- DD/MM/YYYY (ERP) → Conversion vers YYYY-MM-DD
- YYYY-MM-DD (RDV) → Aucune conversion
- YYYY/MM/DD (FACT) → Remplacement "/" par "-"

Format cible : YYYY-MM-DD (ISO 8601)

## Normalisation des Téléphones

1. Suppression caractères non-numériques : "06.12.34.56.78" → "0612345678"
2. Validation longueur = 10
3. Validation préfixe français (01-09)
4. Format stockage : "0612345678" (pas d'espaces ni séparateurs)

## Normalisation des Emails

1. Conversion minuscules : "Jean.DUPONT@example.com" → "jean.dupont@example.com"
2. Trim espaces
3. Validation regex : `^[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$`
4. Vérification domaine @clinique.fr pour emails pro médecins

---

## F. Métriques et KPIs à Suivre

### Métriques de Qualité Données

Métrique	Calcul	Cible	Fréquence mesure
Score qualité moyen patients	AVG(score_qualite)	≥85	Quotidienne

<b>Taux complétude Historique_Médical</b>	COUNT(historique_medical IS NOT NULL) / COUNT(*) × 100	≥70 %	Hebdomadaire
<b>Taux complétude Email</b>	COUNT(email IS NOT NULL AND email VALID) / COUNT(*) × 100	≥70 %	Hebdomadaire
<b>Taux complétude Téléphone</b>	COUNT(telephone IS NOT NULL AND telephone VALID) / COUNT(*) × 100	≥90 %	Hebdomadaire
<b>Taux doublons éliminés</b>	(Nb enreg sources - Nb enreg MDM) / Nb enreg sources × 100	~25 %	Après chaque ETL
<b>Taux validation RPPS médecins</b>	COUNT(num_licence VALID via API) / COUNT(*) × 100	100 %	Hebdomadaire

### Métriques de Performance ETL

Métrique	Cible	Fréquence mesure
<b>Latence temps réel (patients)</b>	<5 secondes	Continue (monitoring)
<b>Taux de succès ETL</b>	≥99%	Après chaque exécution
<b>Durée traitement batch (médecins)</b>	<10 minutes	Quotidienne
<b>Durée traitement batch (services)</b>	<5 minutes	Hebdomadaire
<b>Taux d'erreur transformation</b>	<1%	Après chaque ETL

### Métriques d'Adoption

Métrique	Cible	Fréquence mesure
<b>Nb applications consommatrices MDM</b>	≥5	Mensuelle
<b>Nb requêtes MDM/jour</b>	Croissance continue	Quotidienne
<b>Taux satisfaction utilisateurs</b>	≥80%	Trimestrielle (enquête)

## G. Plan de Gouvernance des Données

### Rôles et Responsabilités



Rôle	Responsable	Responsabilités
<b>Data Owner</b>	Dr. Sophie Laroche (Directrice Médicale)	Validation règles métier, arbitrage qualité, approbation modifications structure
<b>Data Steward</b>	Équipe Data Governance IT	Exécution règles qualité, monitoring métriques, corrections anomalies
<b>Data Custodian</b>	Admin Base de Données	Maintenance technique, backups, sécurité accès
<b>Data Consumer</b>	Médecins, Secrétariat, Direction	Utilisation données, remontée anomalies

## Processus de Modification

### Ajout d'une nouvelle colonne MDM :

1. Demande formelle au Data Owner (justification métier)
2. Validation comité gouvernance (impact sources + consommateurs)
3. Mise à jour Data Catalogue + Data Dictionary
4. Développement ETL
5. Tests qualité
6. Déploiement + communication utilisateurs

### Modification d'une règle métier :

1. Proposition au Data Owner
2. Impact analysis (nb enregistrements affectés)
3. Validation comité
4. Backfill données historiques si nécessaire
5. Documentation + formation

## Comité de Gouvernance

**Fréquence** : Mensuelle (1er lundi du mois, 10h-11h)

### Membres :

- Data Owner (présidence)
- Data Stewards (2 personnes IT)
- Représentant Direction
- Représentant Médecins
- DPO (Data Protection Officer) - pour RGPD

### Ordre du jour type :

1. Revue métriques qualité du mois
2. Incidents/anomalies

3. Demandes d'évolution
4. Conformité RGPD
5. Actions correctives