

DOCUMENT DE CADRAGE

1. Contexte général du projet

(À compléter / corriger si information inexacte ou manquante)

Le projet vise à améliorer l'interopérabilité des données de santé afin de faciliter leur partage, leur standardisation et leur exploitation dans différents environnements internes et externes au CHU.

L'objectif principal est d'établir une correspondance fiable entre les données internes de l'EDSaN et les ressources du standard FHIR, largement utilisé dans de nombreux systèmes d'information en santé.

Les données utilisées dans le cadre du projet doivent être entièrement fictives afin de respecter les contraintes légales et éthiques liées aux données de santé réelles.

Le fichier de métadonnées fourni par les commanditaires constitue la base de référence décrivant la structure des modules, des colonnes et des identifiants de l'EDSaN.

2. Structure des données – EDSaN

(À compléter / corriger si information inexacte ou manquante)

L'EDSaN est structuré en plusieurs modules représentant différents domaines fonctionnels :

DOCEDS (documents), PMSI, BIOL (biologie), VIRO (virologie), MOUVEMENTS (gestion des séjours), BACTERIO (bactériologie), etc.

Chaque module possède ses propres colonnes mais partage certains identifiants transversaux tels que :

- **PatID** (patient)
- **EvtID** (séjour/événement)
- **EltID** (selon les modules)

Les données sont manipulées sous forme de DataFrames et stockées au format Parquet, ce qui permet un traitement performant et adapté aux gros volumes.

Les schémas de colonnes fournis dans le fichier de métadonnées permettent de reconstruire un EDS fictif cohérent et représentatif de la structure réelle utilisée au CHU.

3. Données fictives et génération

(À compléter / corriger si information inexacte ou manquante)

Aucune donnée réelle n'est mise à disposition : toutes les données utilisées dans le cadre du projet doivent être générées de manière entièrement fictive.

L'outil **Synthea** sera mobilisé pour générer des données FHIR fictives et réalistes qui serviront de base de travail.

Les données fictives produites devront être revues et validées avec les commanditaires du CHU pour garantir qu'elles respectent les structures, types et cohérences attendues.

4. Objectifs techniques du projet

(À compléter / corriger si information inexacte ou manquante)

- Concevoir un pipeline permettant de transformer un EDS fictif en un ensemble de ressources conformes au standard FHIR.
- Mettre en place un pipeline inverse pour reconstruire un EDS fictif cohérent à partir de données FHIR.
- Définir et maintenir les mappings **EDS → FHIR** et **FHIR → EDS** dans des fichiers de configuration (YAML ou JSON).
- Analyser et utiliser correctement la structure des principales ressources FHIR pertinentes pour le projet.
- Concevoir une architecture modulaire, documentée et évolutive afin de faciliter la compréhension et la maintenance du projet.

5. Environnement technique et outils

(À compléter / corriger si information inexacte ou manquante)

- Le développement sera réalisé en **Python**, en cohérence avec les pratiques de l'EDSaN.
- La manipulation des DataFrames pourra s'appuyer sur **Polars**, notamment pour ses performances et sa compatibilité avec Parquet.
- Les données fictives seront stockées en **Parquet**, garantissant une bonne intégration avec l'écosystème existant.
- L'utilisation d'API (ex. **FastAPI**) est pressentie pour la partie automatisation.
- La conteneurisation (ex. **Podman**) a été évoquée, mais ses modalités exactes restent à confirmer avec les commanditaires.
- L'environnement technique du CHU repose principalement sur des systèmes **Linux Debian**, ce qui pourra orienter certains choix.

7. Compléments d'information

(Section volontairement laissée libre pour toute précision ou directive complémentaire.)