

Biomedical Data Mediator

Development of a mediator system for
biomedical data

Projet GMD 2020



- ① Objectifs et introduction du projet.
- ② Bases de données
 - Schéma global inter-bases.(explications sur le schéma,.)
 - Mapping vers notre schéma.
- ③ Les fonctionnalités réalisées.
- ④ Les technologies utilisées.



* Ce projet consiste à créer un système médiateur pour des données diverses sur les symptômes, maladies et médicaments,

* Ces données sont issues de plusieurs bases de données biomédicales hétérogènes.

ORPHADATA - CouchDB

- Vues -> JSON
- Références

SIDER - MySQL

- Meddra et meddra_all_indications
- Meddra_all_se et meddra_freq

HPO_annotation - SQLite

- Association Orpha+OMIM

HPO.obo - Index

- Recherche de synonymes

OMIM.txt & OMIM_onto.csv - Index

- Maladies et liens vers SIDER (CUI)

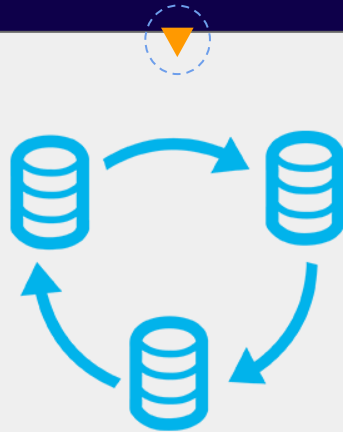
STITCH et ATC - Index

- Lien des composants chimiques vers le label des médicaments

DrugBank.xml - Index

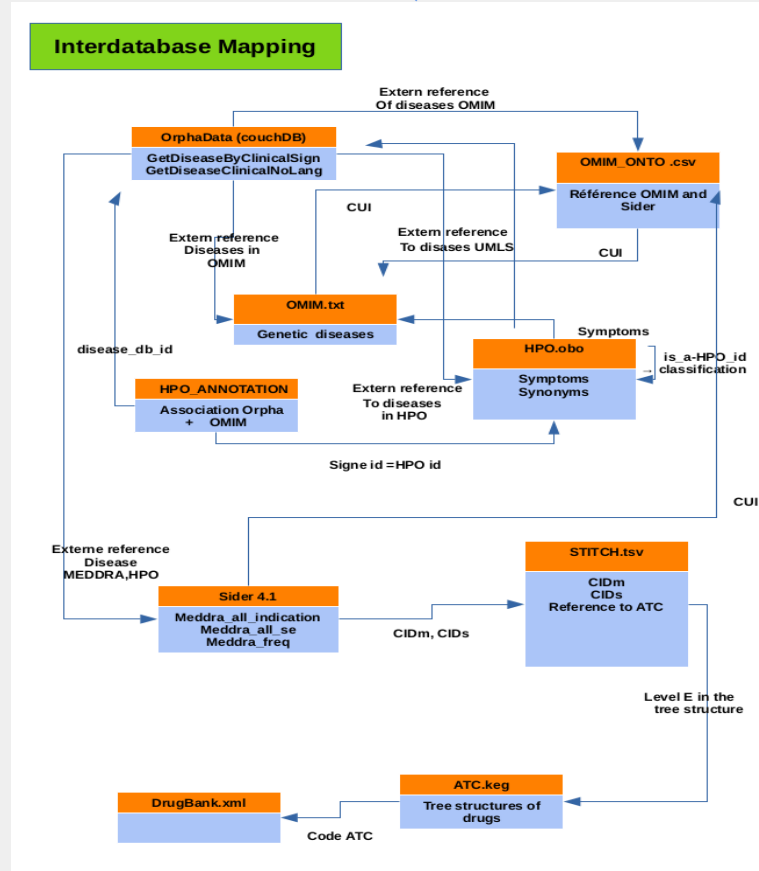
- Indication des médicaments et effets secondaires

Mapping



Le schéma global inter-bases

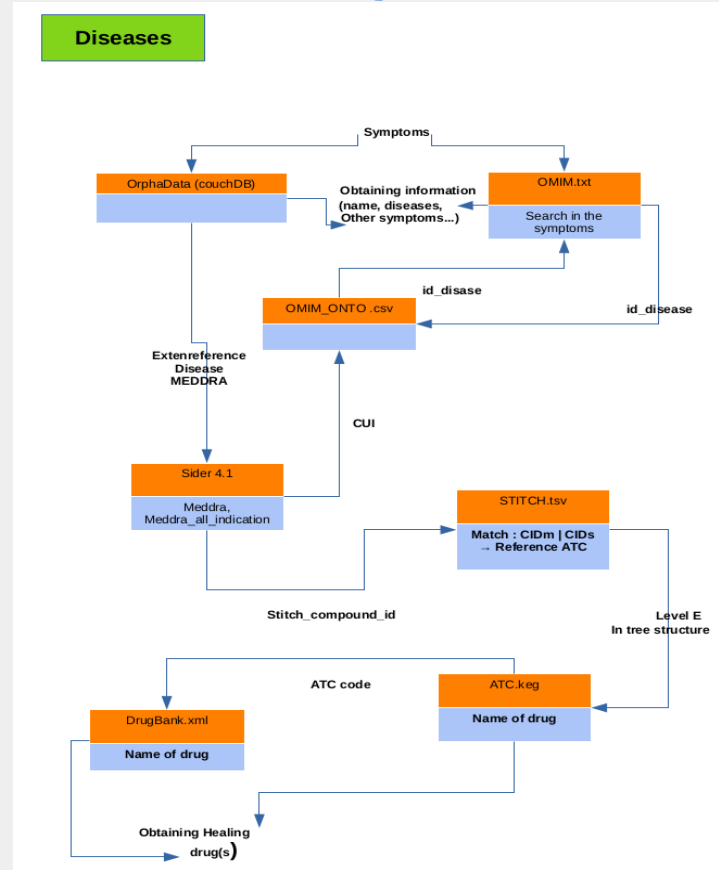
5



Le mapping des maladies associées

6

MALADIES



Les mappings des médicaments

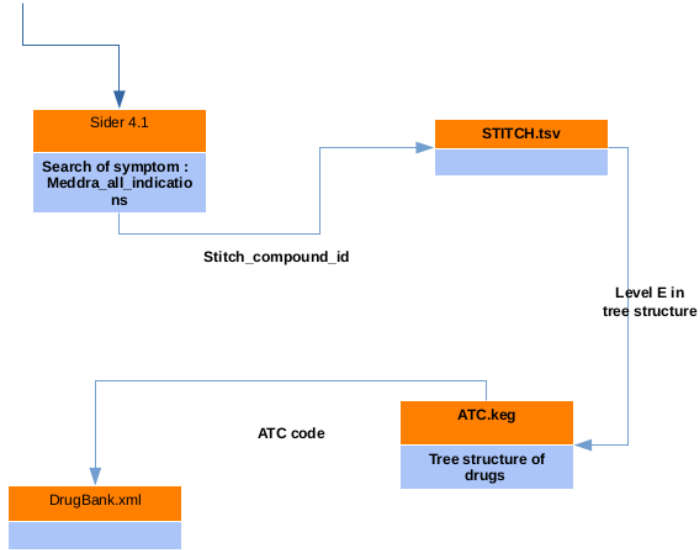
7



MÉDICAMENTS TRAITANTS

DRUGS INDICATION

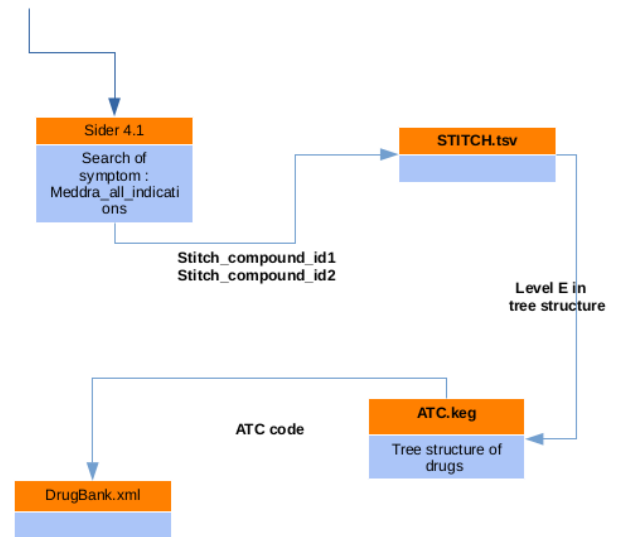
Symptoms entered by the user



MÉDICAMENTS CAUSANTS

DRUGS SIDE EFFECTS

Symptoms entered by the user





FONCTIONNALITÉS DE BASE

- Système médiateur ✓
- Indexation *full-text* ✓
- Recherche de symptômes
 - Maladies associées ✓
 - Médicaments causants ✓
 - Médicaments traitants ✓
- Requêtes avec AND ✓

FONCTIONNALITÉS BONUS

- Utilisation de jokers dans la requête ✓
- Requêtes avec des OR et des AND ✓
- Utilisation des synonymes ✓
- Qualité des mappings ✓
- Tri des résultats ✓
- Provenance des données ✓
- Visualisation des données ✓
- Interface graphique ✓



- Le langage de programmation java.
- Apache Lucene 4.0.0 pour l'indexation de certaines bases de données.
- Qt-Jambi pour la réalisation de l'interface graphique.



- Utiliser un langage plus intuitif comme python.
- Utiliser des technologies de parallélisation comme (Spark,Hadoop ...), pour réduire le temps de recherche dans les bases de données.