

September 19, 2019

INTERCHU  
**Datathon Omop**

**CHU, Paris;  
DREES, Paris;  
CHU, Bordeaux;  
CHU, Clermont-Ferrand  
CHU, Grenoble;  
CHU, Lille;  
CHU, Lyon;  
CHU, Rennes;  
CHU, Toulouse;**

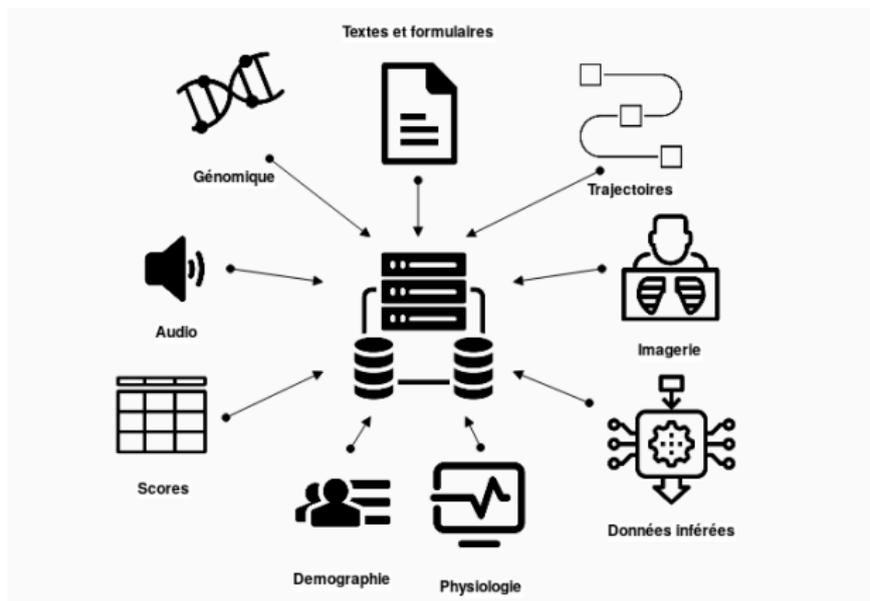


Les Données Hospitalières

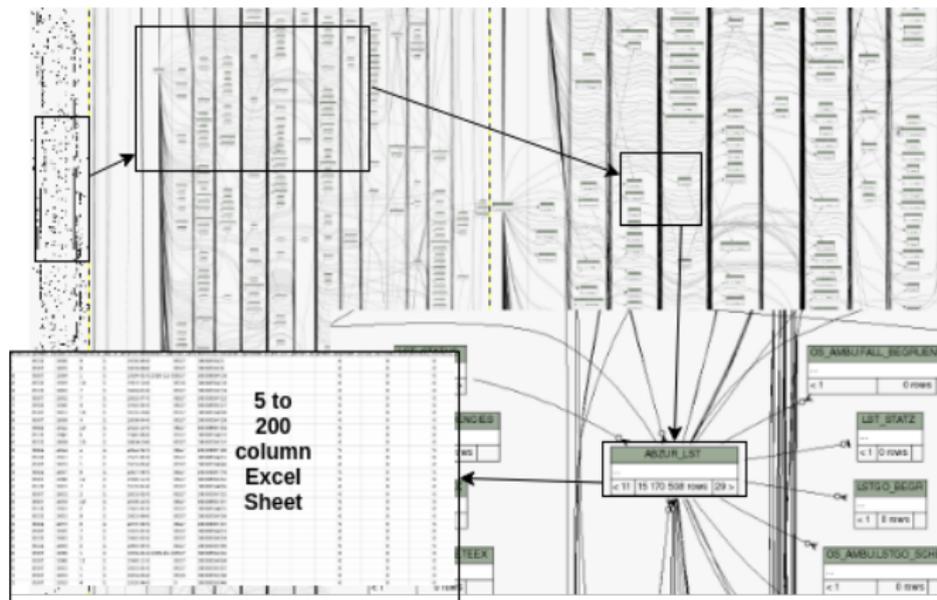
Modèle Commun de Données OMOP

Les apports d'OMOP

Datathon



Informations consignées / recoupées / vérifiées par des professionnels de santé.





## Causes

- ▶ DPI: données fiables
- ▶ Outil simple / immédiat pour l'utilisateur
- ▶ Adaptabilité aux pratiques
- ▶ Données à caractère sensible

## Conséquences

- ▶ Modèle OLTP (Codd)
- ▶ Données manquantes, doublons, erreurs
- ▶ Données non standardisées
- ▶ Accès règlementés

→ Comment obtenir OLAP, qualité, standardisé et sécurisé?



## Approche groupée autour d'un modèle commun de données

- ▶ Partage des terminologies
- ▶ Partage des retro-ingénieries
- ▶ Partage des développements



Les Données Hospitalières

**Modèle Commun de Données OMOP**

Les apports d'OMOP

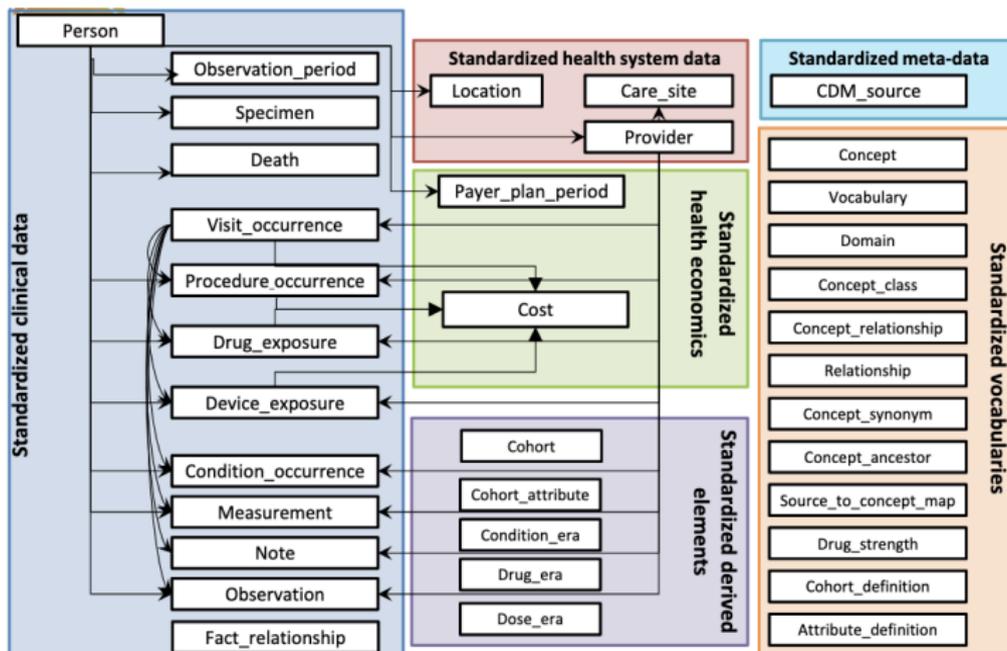
Datathon

# Alignement structurel

un formalisme pour le schéma des tables



- ▶ 13 tables cliniques, 3 tables sur le système de soin, 2 sur les aspects économiques



# Alignement structurel

exemple : la table des visites



<b>visit_occurrence_id</b>
person_id
visit_start_date
visit_start_datetime
visit_end_date
visit_end_datetime
visit_type_concept_id
provider_id
care_site_id

visit_concept_id
admitted_from_concept_id
discharge_to_concept_id
visit_source_value
visit_source_concept_id
admitted_from_source_value
discharge_to_source_value

# Alignement conceptuel

## un formalisme pour les terminologies



09/24



# Alignement conceptuel

exemple : une ligne dans la table concept



CONCEPT_ID	4108832
CONCEPT_NAME	Atrial fibrillation and flutter
DOMAIN_ID	condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	195080001
VALID_START_DATE	1970-01-01
VALID_END_DATE	2099-12-31
INVALID_REASON	

Table: Concept standard

CONCEPT_ID	2002201945
CONCEPT_NAME	fibrillation et flutter auriculaires
DOMAIN_ID	condition
VOCABULARY_ID	SNDS - cim10
CONCEPT_CLASS_ID	EMPTY
STANDARD_CONCEPT	
CONCEPT_CODE	l48
VALID_START_DATE	2019-06-06
VALID_END_DATE	2099-01-01
INVALID_REASON	

Table: Concept local

# OMOP: quelles limites ?



- ▶ Un modèle **opensource** centré sur la SNOMED-CT avec **license payante**
- ▶ Coût d'entrée élevé
- ▶ Des spécificités françaises

## Solutions : Travail commun pour un alignement français

The screenshot shows the 'Concepts Mapper App' interface. On the left, there is a sidebar with a search bar and a list of filters under 'Filtres spécifiques à omop-mapper'. The main area displays a table of concepts with columns for 'A mapper', 'id', 'Nom', 'Vocabulaire', and 'Domain'. The table contains 20 rows of data, including medical terms like 'Medical and Surgical @ Heart and Great Vessels @ Supplement @ Atrial Septum @ Pericardium @ Adipose Tissue Substitute' and 'Métazapine 45 MG Oral Tablet [MERTAZAPINE SANDOZ] Box of 100 by Novartis'. The interface also shows a search bar at the top right with the ID '11037629' and a 'Recherche' field.

A mapper	id	Nom	Vocabulaire	Domain
<input type="checkbox"/>	2962016	Medical and Surgical @ Heart and Great Vessels @ Supplement @ Atrial Septum @ Pericardium @ Adipose Tissue Substitute	ICD10PCS	Procedure
<input type="checkbox"/>	21370532	Ativan class 2 (23-32mmHg) arm/sleeve with elbow pocket lymphoedema garment regular size 1 Black (Ativan L1) 1 device	dm+d	Device
<input type="checkbox"/>	44660360	FERRUM SIDERIUM ORAL SOLUTION [FERRIUM SIDERIUM]	DPD	Drug
<input type="checkbox"/>	41103217	Métazapine 45 MG Oral Tablet [MERTAZAPINE SANDOZ] Box of 100 by Novartis	Rohim Extension	Drug
<input type="checkbox"/>	2007475704	Likoria peanoriana	ndi	measurement
<input type="checkbox"/>	40672416	Procaine 10 MG/mL / Thiamine 0.24 MG/mL injectable Solution [Procain Loge]	Rohim Extension	Drug
<input type="checkbox"/>	2003056009	Rbdymena sp. 2 DWF-2013	ndi	observation
<input type="checkbox"/>	4095461	Anticagulant control test	SNOMED	Measurement
<input type="checkbox"/>	2003694957	Genetta decolorata	ndi	observation
<input type="checkbox"/>	19077826	Guafenesin 300 MG / Phenylephrine 20 MG Extended Release Oral Tablet	Rohim	Drug
<input type="checkbox"/>	30003297	Staphylococcus aureus enterotoxin D [IgE Ab RAST class [Presence] in Serum	LOINC	Measurement
<input type="checkbox"/>	4162000	Medication administered prior to percutaneous coronary intervention	SNOMED	Observation
<input type="checkbox"/>	2005607875	lungal sp. aure1143	ndi	observation
<input type="checkbox"/>	40504978	Sore throat symptom	Read	Condition
<input type="checkbox"/>	2002106114	Status respiratory [ Patient ] -	JOV_complet	measurement



Les Données Hospitalières

Modèle Commun de Données OMOP

Les apports d'OMOP

Datathon

# L'approche multicentre

exemple: le Federated Learning

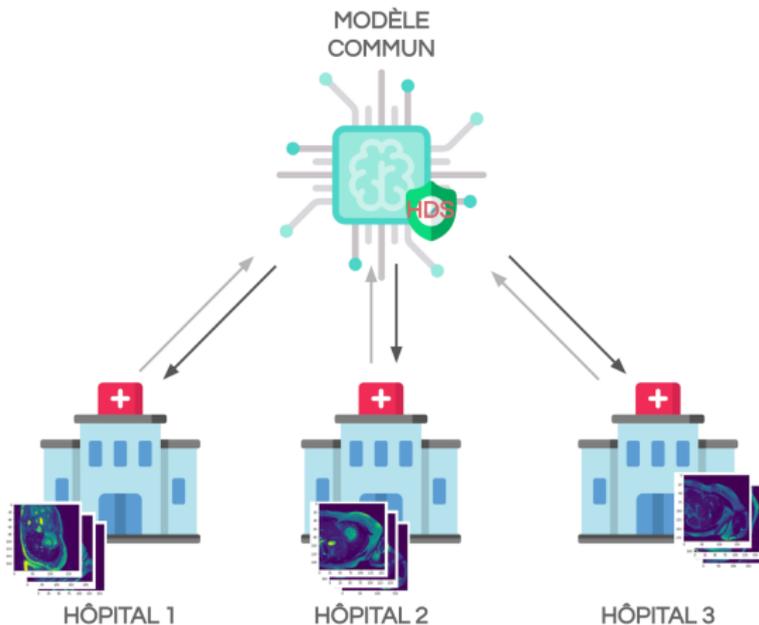


Figure: Apprentissage fédéré d'un modèle d'apprentissage

# Des outils partageables

Visualisation, cohortes, préparation des données

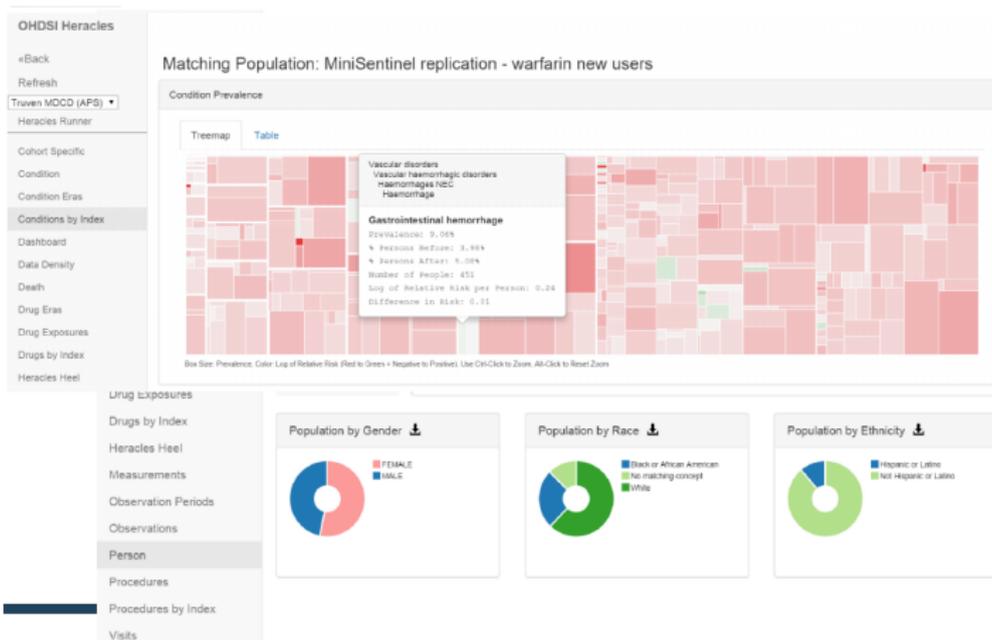


Figure: Achille, construction d'indicateurs agrégés et visualisations

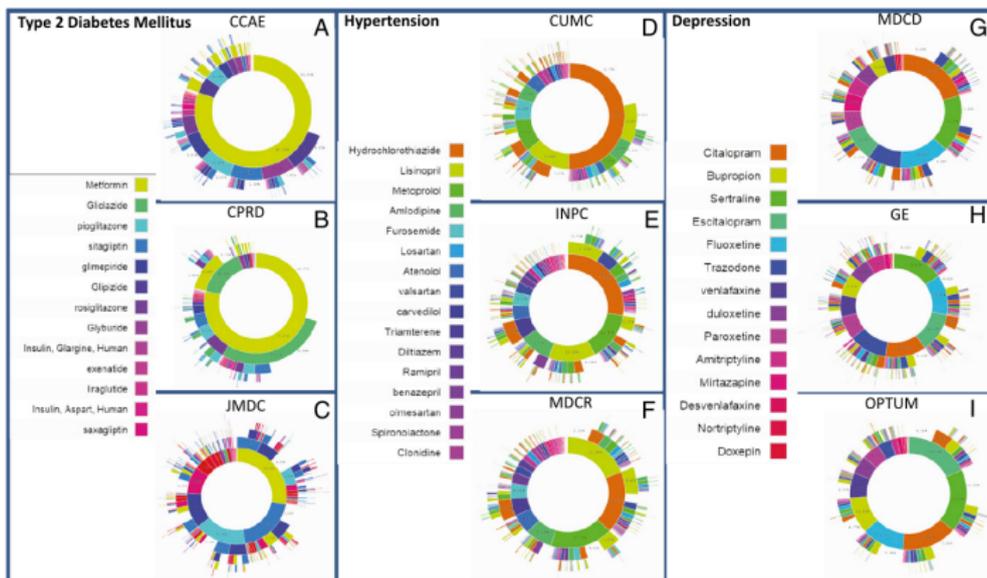
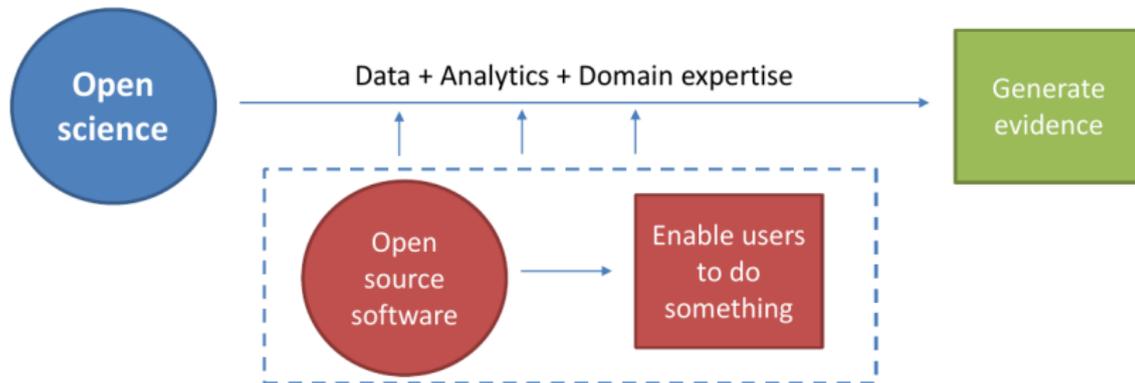
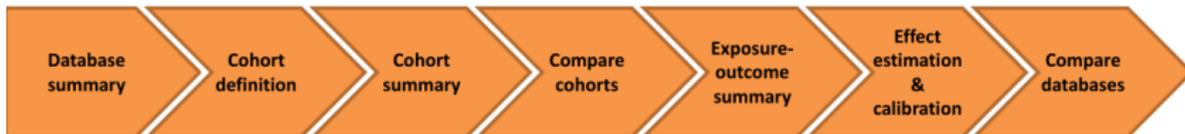


Fig. 3. For each disease, diabetes (A–C), hypertension (D–F), and depression (G–I), the inner circle shows the first relevant medication that the patient took, the second circle shows the second medication, and so forth. Three data sources are shown for each disease; the data source abbreviations are defined in Table 2.

Figure: G. Hripcsak et al., "Characterizing treatment pathways at scale using the OHDSI network", Proc Natl Acad Sci U S A, vol. 113, n 27, p. 73297336, juill. 2016



## Standardized, transparent workflows





Les Données Hospitalières

Modèle Commun de Données OMOP

Les apports d'OMOP

**Datathon**



## Principe Général

- ▶ proposer un challenge
- ▶ publier un jeu de données synthétique
- ▶ mettre au point des jeux de train/test
- ▶ lancer le code des participants
- ▶ renvoyer les résultats

:-) Données réelles, multicentrique

:-o Sécurisé, scalable

:-O Dématerialisé / sur place

:-( Boite noire, debugging

# Datathon Framework Organisation



**ZULIP** datathon - teams stream:datathon++teams

Tous les messages  
Messages privés  
Mentions  
Messages favoris

CANAUX

- datathon - teams
  - Team FCD
  - challenge task
  - OMOP-CDM presentation
  - welcome
  - salut
  - plus de sujets
- FHIR
- general

**datathon - teams** welcome

**Notification Bot**  
Adrien PARROT renamed stream **Datathon - teams** to **datathon - teams**

JUL 05 JUL 21

**datathon - teams** OMOP-CDM presentation

**Jim Tyhurst**  
Le **README.md** du projet **datathon\_dataset** dit qu'il y aura une «OMOP-CDM presentation and challenge presentation: 22/07/2019». Comment peut-on assister à ces présentations? Elles seront à quelle heure?

JUL 21 JUL 22

**Adrien PARROT**  
Bonjour.lim

**RAMP**

- RAPID challenge**
  - RAPID 2018 challenge, number of participants = **22**, number of submissions = **10**, click here for score vs time plot
  - RAPID 2016 challenge, number of participants = **2**, number of submissions = **1**, click here for score vs time plot
- Mechanics classification**
  - Saclay M2 Data Camp 2018/19, number of participants = **1.148**, click here for score vs time plot
- Solar wind classification**
  - Saclay M2 Data Camp 2018/19, number of participants = **0.17**, click here for score vs time plot
  - Solar wind hackathon, number of participants = **55**, number of submissions = **10**, click here for score vs time plot
- Storm intensity forecast**
  - CI2018 hackaton, number of participants = **47**, number of submissions = **10**, click here for score vs time plot
  - initial hackaton, number of participants = **14**, number of submissions = **10**, click here for score vs time plot
  - Geilo Winter School 2019 challenge, number of participants = **10**, number of submissions = **10**, click here for score vs time plot

File Notebook Editor Terminal Console Help

Files

Name	Last Modified
step	2 minutes ago
JDBC Tutor.ipynb	8 days ago
Tutorial_Omop.ipynb	9 days ago
Untitled.ipynb	10 days ago

Running

Console

Python 3 PySpark 3.0 R SparkR 2.12

id	titre
200204370	Pilote de cours d'analyse relationnelle interactive distribuée - Big Data
200204142	Autre cours de formation de cours (AIR2) - A transformation de langage et de
200204307	CRD - Recherche de cours d'analyse relationnelle interactive distribuée et
200204463	Autre analyse relationnelle interactive distribuée distribuée distribuée
200117816	Chaque fois que vous avez une session de cours (AIR2) - A transformation de
200140186	Pilote de cours d'analyse relationnelle interactive distribuée distribuée
200103004	Autre analyse relationnelle interactive distribuée distribuée distribuée
200201561	STREPTO - GROUPE 2 - CHAISE DE HALL - CLASSEES CHAISES D'UNIVERSITE



## Jeux de données OMOP publics

- ▶ MIMIC-OMOP
- ▶ Synpuf (synthétiques)

## Jeux de données OMOP

PERSON, VISIT\_OCCURRENCE, CONDITION\_OCCURRENCE,  
PROCEDURE\_OCCURRENCE, DRUG\_EXPOSURE ...

- ▶ APHP : 13M de patients
- ▶ Rennes : 1.3M de patients
- ▶ Drees : SNDS échantillon, 4 millions de personnes sur 10 ans

→ 20k patients pour le train/test dans ces centres

# OMOP Terminologies

L'avancement du groupe interChu



Centres	Vocabulaires	Total	Mappés (%)
Bordeaux	Cim10, ccam	20 467	9 653 (47%)
OSIRIS	ATC, ICD03, medra, urls, loinc, fhir, uicc	8 622	1 336 (15%)
Lille	cim10, ccam	37 113	12 927 (34%)
Paris	Anabio, antibio, ccam, cim10, ghm	83 246	35 921 (43%)
Rennes	Adicap, ccam, cim10, laboratoire	72 994	20 421 (27%)
Toulouse	ccam, cim10, médicaments	31 995	2 106 (6%)
SNDS	cim10, cip13, ccam, lpp, nabm, ald, pmsi	48 884	16 996 (34%)

# Une première tâche proposée

## Déduplication des Patients

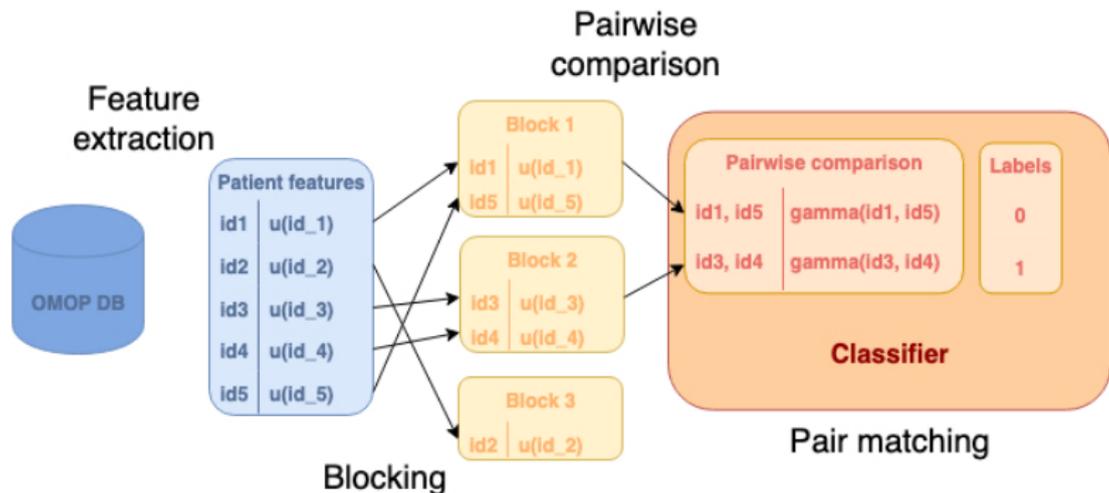


Figure: *Entity Resolution* en 4 étapes

# Comment participer ?

Plusieurs niveaux d'implications



- ▶ Proposer des challenges ou y participer
- ▶ Faire partie des centres
- ▶ Mapper des concepts



- ▶ Le forum <https://interchu.zulipchat.com>
- ▶ Le git <https://framagit.org/interchu>
- ▶ Le challenge <https://interchu.frama.io/website/>
- ▶ La communauté OMOP <https://forums.ohdsi.org/>



Questions ?

[interchu@riseup.net](mailto:interchu@riseup.net)