

Modèle d'intégration de données



OMOP CDM

Mercredi 5 mai 2021

**Module Interopérabilité et Standards de Données Médicales
(ISDM) - Majeure Santé EPITA**

Dr Damien Leprovost – AP-HP / Limics

Licence Creative Commons BY-SA 4.0

Ce support de cours est distribué sous licence
Attribution - Partage dans les Mêmes Conditions 4.0 International (CC BY-SA 4.0)

Vous êtes autorisé à :

- **Partager** — copier, distribuer et communiquer le matériel par tous moyens et sous tous formats
- **Adapter** — remixer, transformer et créer à partir du matériel pour toute utilisation, y compris commerciale.

Selon les conditions suivantes :

-  **Attribution** — Vous devez [créditer](#) l'œuvre, intégrer un lien vers la licence et [indiquer](#) si des modifications ont été effectuées à l'œuvre. Vous devez indiquer ces informations par tous les moyens raisonnables, sans toutefois suggérer que l'Offrant vous soutient ou soutient la façon dont vous avez utilisé son œuvre.
-  **Partage dans les Mêmes Conditions** — Dans le cas où vous effectuez un remix, que vous transformez, ou créez à partir du matériel composant l'œuvre originale, vous devez diffuser l'œuvre modifiée dans les même conditions, c'est à dire avec [la même licence](#) avec laquelle l'œuvre originale a été diffusée.



Objectifs du cours

- Savoir ce que sont **OMOP** et **OHDSI**
- Se familiariser avec **OMOP Common Data Model** (CDM)
- Comprendre le principe des **vocabulaires standardisés**
- Être en capacité de trouver des **codes** et des **concepts**
- Pouvoir **utiliser OMOP CDM**
- Naviguer dans la **hiérarchie des concepts**

Plan du cours

I. Introduction

Historique, motivations et enjeux

II. OMOP CDM

Le modèle de données et ses principes

III. Terminologies

Implémentation et motivations

IV. Les tables du modèles

Tables principales et secondaires par domaine

V. Outils OHDSI

Issus de la communauté

VI. Conclusion

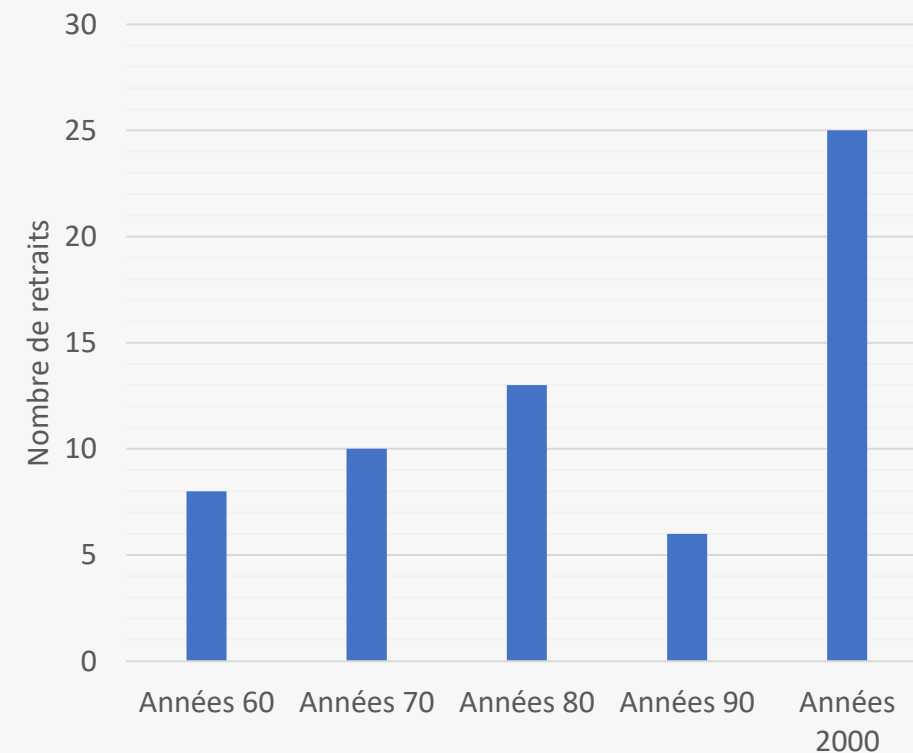


I. Introduction

- Les origines historiques
- L'expérimentation OMOP
- La création d'OHDSI

Aux origines – Un peu d'histoire

- Forte augmentation des retraits après mise sur le marché
 - Informatisation croissante des systèmes de santé
- Scandale du Rofécoxib (Vioxx[®] Merck) en 2005
- Intervention du congrès américain
 - Limites des études initiales
 - Besoin d'études systématiques après mise sur le marché



Retraits de médicaments par la FDA par décennie (USA)

Food and Drug Administration Amendments Act (FDAAA) 2007

SEC. 905. ACTIVE POSTMARKET RISK IDENTIFICATION AND ANALYSIS.

(a) IN GENERAL.—Subsection (k) of section 505 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 355) is amended by adding at the end the following:

“(3) ACTIVE POSTMARKET RISK IDENTIFICATION.—

“(A) DEFINITION.—In this paragraph, the term ‘data’ refers to information with respect to a drug approved under this section or under section 351 of the Public Health Service Act, including claims data, patient survey data, standardized analytic files that allow for the pooling and analysis of data from disparate data environments, and any other data deemed appropriate by the Secretary.

“(B) DEVELOPMENT OF POSTMARKET RISK IDENTIFICATION AND ANALYSIS METHODS.—The Secretary shall, not later than 2 years after the date of the enactment of the Food and Drug Administration Amendments Act of 2007, in collaboration with public, academic, and private entities—

“(i) develop methods to obtain access to disparate data sources including the data sources specified in subparagraph (C);

“(ii) develop validated methods for the establishment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources, with the goals of including, in aggregate—

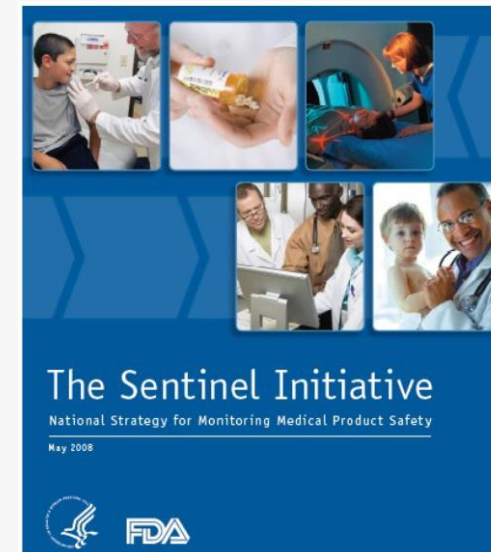
“(I) at least 25,000,000 patients by July 1, 2010; and

“(II) at least 100,000,000 patients by July 1, 2012; and

“(iii) convene a committee of experts, including individuals who are recognized in the field of protecting data privacy and security, to make recommendations to the Secretary on the development of tools and methods for the ethical and scientific uses for, and communication of, postmarketing data specified under subparagraph (C), including recommendations on the development of effective research methods for the study of drug safety questions.

“(C) ESTABLISHMENT OF THE POSTMARKET RISK IDENTIFICATION AND ANALYSIS SYSTEM.—

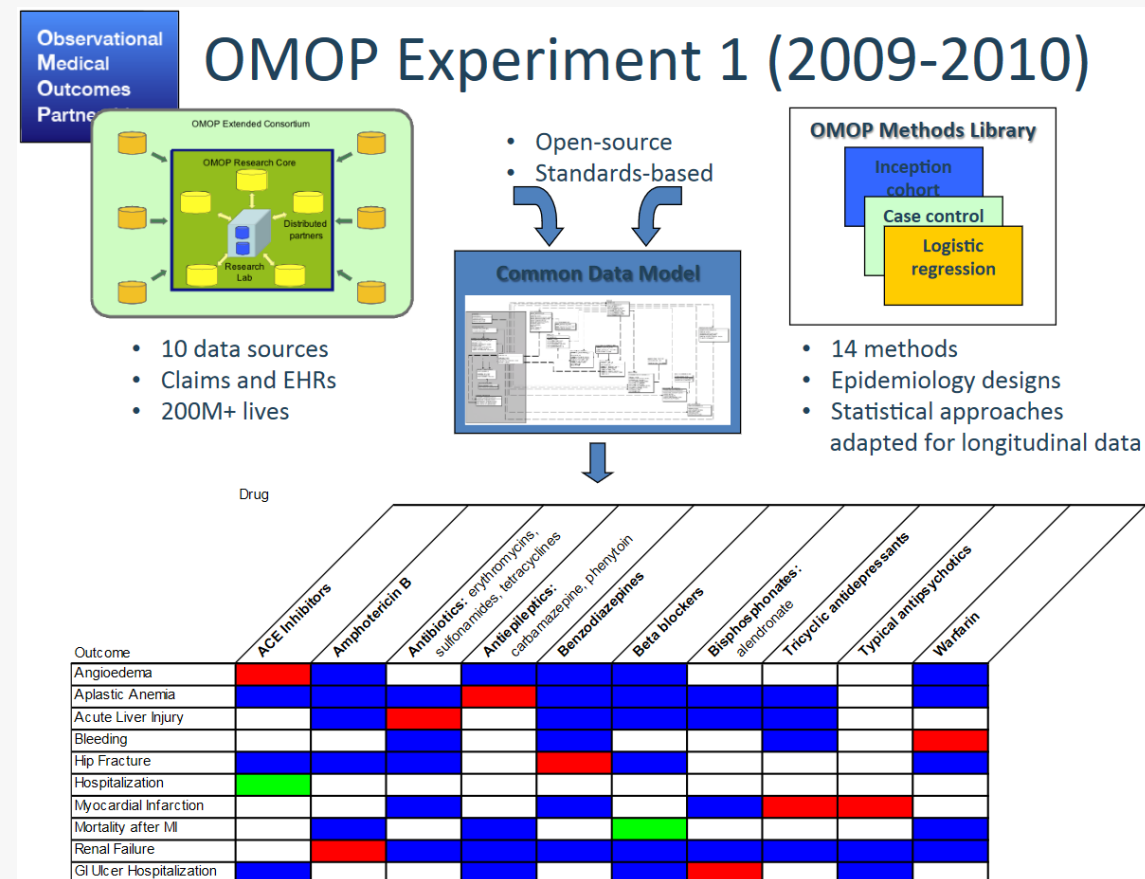
- The Sentinel Initiative
 - processus systématique et reproductible
 - générer des preuves, caractériser des effets potentiels
 - via un réseau de sources de données d'observation disparates



- Quelles méthodes ?
- Sur quelles données ?

OMOP (Observational Medical Outcomes Partnership)

- Partenariat industrie pharmaceutique – FDA
- Reproductibilité des faits par sélection des sources et des méthodes
- Phase 1 (2009-2010)
 - ~10k scripts SAS par base de données et par cas
- Phase 2 (2011-2012)
 - Etudes en profondeur, par médicaments
- Expérimentation européenne (eu-adr)



Résultats de l'expérimentation OMOP

- Forte sensibilité du choix de base de données
 - à cibles et méthodes identiques par ailleurs
- Forte sensibilité des choix d'analyses
- Faible sensibilité des critères d'inclusion
- $AUC > 0.7$ sur les méthodes optimales de l'expérimentation pour distinguer les contrôles positifs et négatifs
- Meilleures performances des méthodes autocontrôlées, sur les cohortes et les méthodes de disproportionnalité

ODHSI (Observational Health Data Sciences and Informatics)

- Fondé par les investigateurs du projet OMOP
- Reprend les conclusions de l'expérimentation
- Réseau de collaboration pour l'analyse à grand échelle des données de santé
- Hébergé par l'université de Columbia
 - Public et ouvert
 - Indépendant de l'industrie pharmaceutique
 - International

[illegible]

Motivation

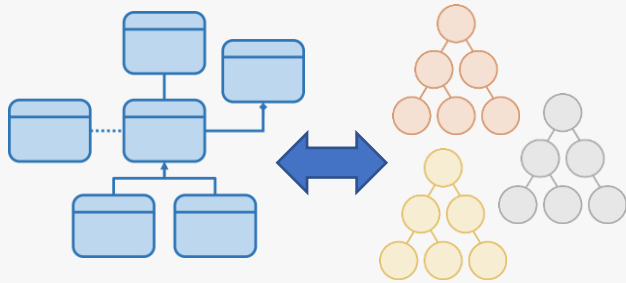
- Standardisation des données pour la recherche systématique
- Issu du besoin d'interopérabilité de l'expérimentation OMOP
- Rendre applicable toute méthode future à n'importe quelle source de données
 - Sans a priori sur les formats sources
 - Sans besoin d'accès aux données
- Standardisation d'un modèle commun de données (**CDM**) et de sa méthodologie
- Développement communautaire d'outils sur ces standards



II. OMOP Common Data Model (CDM)

1. Les principes centraux du modèle
2. Vue générale des tables
3. Les conventions de nommage
4. Les aspects terminologiques

Principes centraux



- Modèle et vocabulaires imbriqués
- Modèle extensible
- Vocabulaires évolutifs



Centré sur le patient

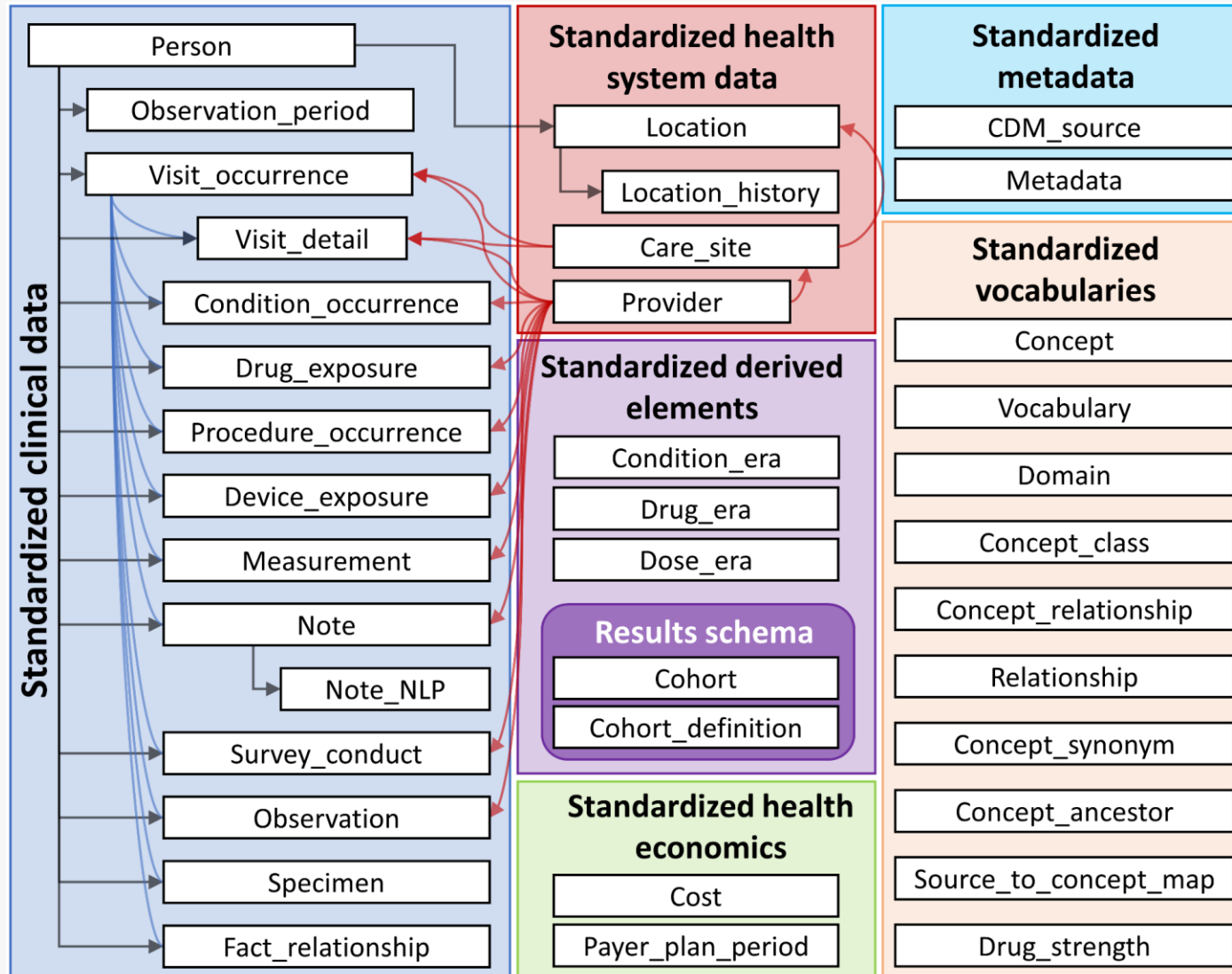


- Multi-sources
- Préserve les données sources



**Indépendant des plateformes
de base de données**

OMOP CDM v6.0

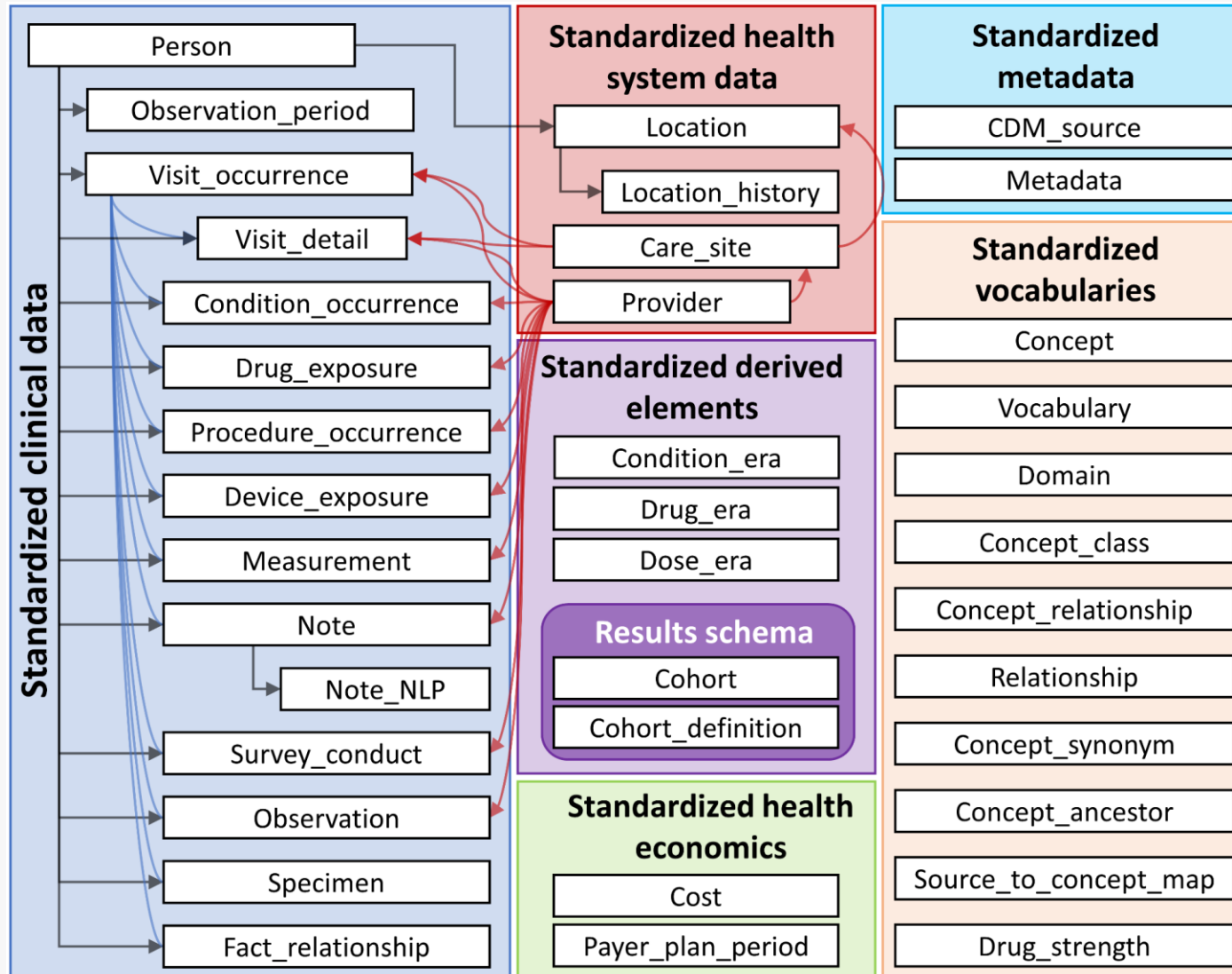


Ce schéma ne représente pas toutes les relations entre les tables

38 tables :

- **Données cliniques** (15 tables)
- **Données terminologiques** (10 tables)
- **Données sur le système de prise en charge** (4 tables)
- **Données dérivées** (5 tables)
- **Données économiques** (2 tables)
- **Métadonnées** (2 tables)

OMOP CDM v6.0



Ce schéma ne représente pas toutes les relations entre les tables

- Tables fondamentales (pivots):
 - Person
 - Concept
- Toute donnée est liée à un patient
 - Les données du **système de prise en charge** ne constituent pas une base de données médico-administrative
- Toute donnée est liée à un concept
 - Tout concept est liée à une terminologie
 - Tout concept vise à être aligné sur un standard

OMOP CDM v6.0 : Conventions

- ANSI/ISO SQL
 - Types limités à VARCHAR, INTEGER, FLOAT, DATE, DATETIME, CLOB
- Limitation stricte des tables aux domaines
 - Non garanti par les terminologies
- Conventions de nommage des champs :

Notation	Description	Exemple
[event]_SOURCE_VALUE	Verbatim des données sources, généralement utilisé par les ETL pour mapper à CONCEPT_ID. <i>Ne doit être utilisé par aucune analyse standard.</i>	CONDITION_SOURCE_VALUE contient le code d'état tel que généré par l'application source.
[event]_ID	Identifiants uniques pour les entités clés. Peuvent servir de clés étrangères pour les relations entre les entités.	PERSON_ID identifie de manière unique chaque individu.
[event]_CONCEPT_ID	Clé étrangère vers la table CONCEPT dans les vocabulaires standardisés, qui sert de base pour toutes les analyses standardisées.	CONDITION_CONCEPT_ID = 31967 contient la valeur de référence pour le concept SNOMED de « nausée ».
[event]_SOURCE_CONCEPT_ID	Clé étrangère vers la table CONCEPT, représentant le concept et la terminologie utilisés dans les données source, le cas échéant.	CONDITION_SOURCE_CONCEPT_ID = 45431665 peut désigner le concept de « nausée » dans la terminologie source.
[event]_TYPE_CONCEPT_ID	Clé étrangère vers la table CONCEPT, utilisé pour délimiter l'origine de l'information source, le cas échéant. normalisée dans les vocabulaires normalisés.	DRUG_TYPE_CONCEPT_ID peut permettre de faire la distinction entre « distribution en pharmacie » et « prescription écrite ».

OMOP CDM v6.0 : Différences entre concepts et valeurs sources

- De nombreuses tables contiennent des informations équivalentes à plusieurs endroits

La valeur source

- Représentation originale dans les données sources
- Il peut s'agir de codes provenant de :
 - systèmes de codage publics (ICD, NDC,...)
 - systèmes de codage propriétaires (SNOMED, MedDRA, ...)
- vocabulaires contrôlés utilisés uniquement dans les données sources (F pour femme et M pour homme)
- textes libres ni standardisés ni contrôlés

stocké dans :

[Event]_SOURCE_VALUE

Le concept source

- Représentation du concept utilisé par la source
- Utilisé uniquement quand la source utilise un codage public ou propriétaire
- La valeur du champ est une ID (integer), clé étrangère sur CONCEPT
- En l'absence de concept, la valeur 0 est utilisée

stocké dans :

[Event]_SOURCE_CONCEPT_ID

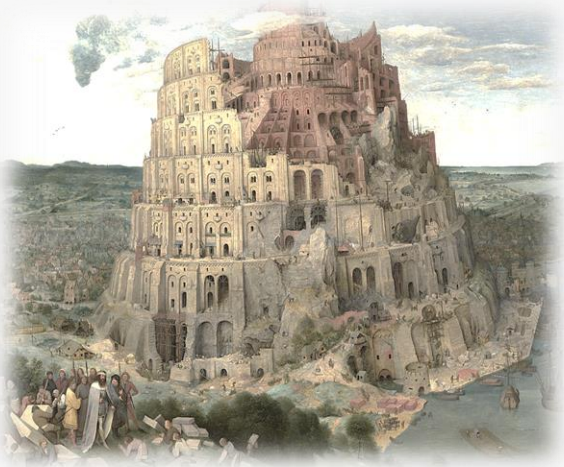
Le concept standard

- Représentation standardisée du concept :
 - unifiée dans toutes les bases OMOP
 - indépendamment des sources
 - définie par OMOP, généralement tirés de systèmes de codage publics ou propriétaires
- Si le concept source est standard, son code est réutilisé ici
- Si le concept source n'est pas standard, l'utilisation d'un concept standard est conditionnée à l'existence d'un mapping avec le vocabulaire standard
- La valeur du champ est une ID (integer), clé étrangère sur CONCEPT
- En l'absence de concept, la valeur 0 est utilisée

stocké dans :

[Event]_CONCEPT_ID

OMOP CDM v6.0 : Importance du codage standardisé



- Les valeurs sources n'existent dans le modèle qu'à des fins de commodité et d'assurance qualité
- Elles peuvent contenir des informations qui n'ont de sens que dans le contexte d'une source de données spécifique
- L'utilisation des valeurs sources et des concepts sources est facultative
 - Elle est fortement recommandée si les données sources utilisent des systèmes de codage
- Les concepts standards sont obligatoires
 - Cette utilisation obligatoire des concepts standards permet à toutes les instances d'OMOP CDM de parler la même langue

OMOP CDM v6.0 : Point d'attention sur des dates

- Utilisation du champ de type **datetime** obligatoire sur toute donnée temporellement annotée
- Présence de champs de type date, facultatifs
 - Compatibilité historique avec OMOP CDM v5.3
- Convention de complétion dépendant de la table et du champs
- Priorités générales de résolution:
 1. Règles spécifiques au champ
 - *Exemples:*
 - les naissances n'ayant qu'une année et sans élément daté pouvant être rattaché sont fixées au 15 juin de l'année fournie
 - une date de fin manquante d'exposition à un médicament est inférée de sa date de début et de sa prescription
 - une date de fin d'exposition à un dispositif médical non renseignée ne doit pas être inférée et laissée vide
 2. Inférence depuis un élément daté relatif et pertinent
 3. Cas général d'une date vers un datetime : heure définie à 00:00:00



III. Terminologies

1. Définition introductive
2. (quelques) Terminologies en santé
3. Approche et motivations
4. Fonctions implémentées

Qu'est-ce qu'une terminologie ?

- Système de codage de l'information
- Représentation de la connaissance
- Référenciel normatif spécialisé
- Notion fondamentale de **concept**



- Organisation hiérarchique
- Ensembles connexes :
 - Thésaurus
 - Ontologies

(quelques) Terminologies en santé

Nom	Distribution	Domaine	Type	Codes (approx)
CIM-10 (ICD-10)	Publique (OMS)	Diagnostics, co-morbidités	Classification	12 000
SNOMED-CT	Propriétaire	Diagnostics, actes, résultats, substances, dispositifs, ...	Terminologie	311 000
LOINC	Gratuite	Observations, mesures, documents, ...	Terminologie	95 000
RxNorm	Publique* (US NLM)	Médicaments (US)	Classification	250 000
MeSH	Publique (US NLM)	Indexation bibliographique	Thésaurus	41 000
CPT	Propriétaire	Diagnostics, actes	Classification	14 000
CCAM	Publique (FR HAS)	Actes	Classification	7 600
ATC	Publique (OMS)	Substances	Classification	6 300

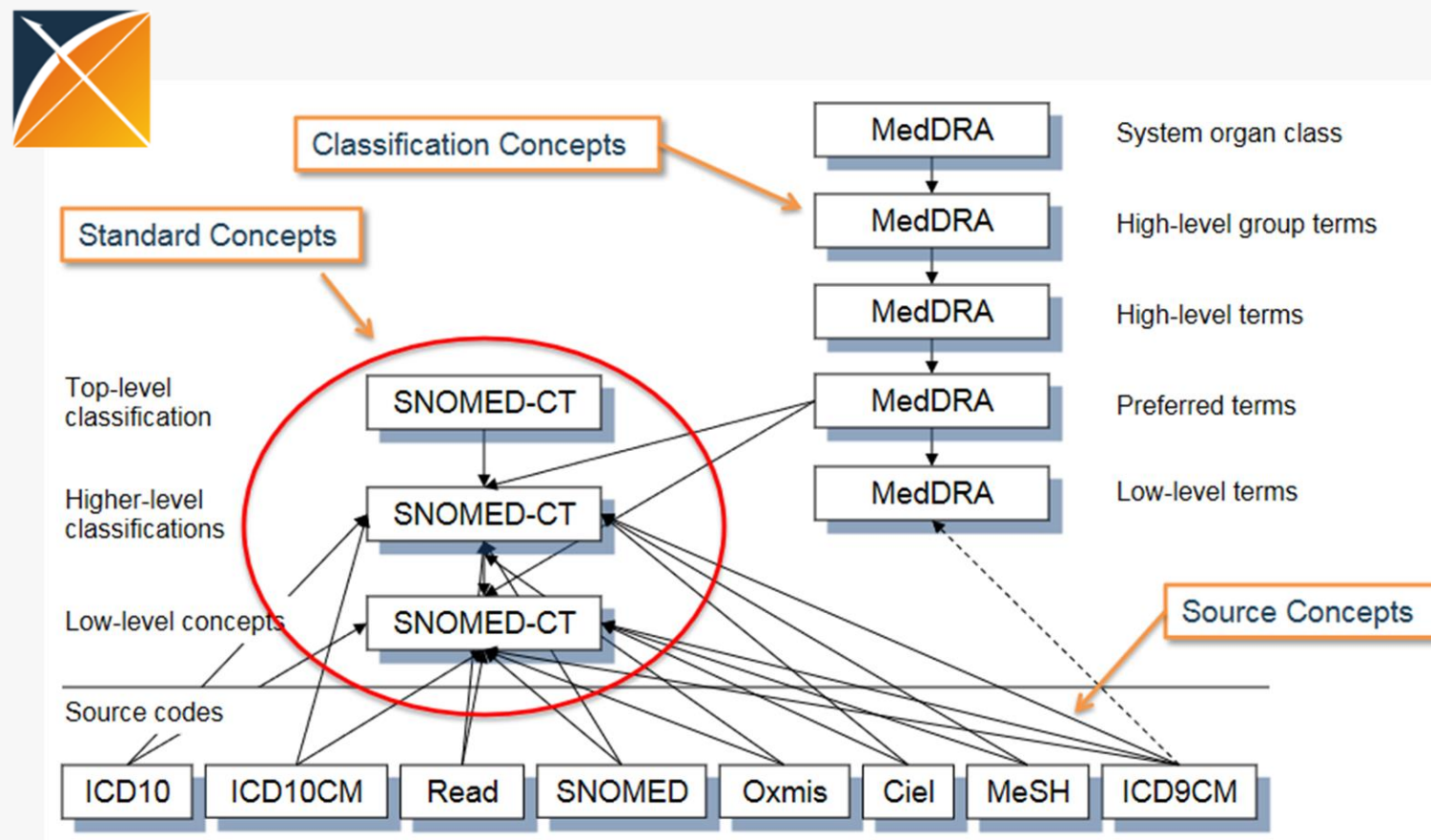
Approche OMOP CDM

- Raison d'être d'OMOP: Création d'un **réseau**
 - **x** systèmes nécessitent jusqu'à **x²** efforts de traductions **et de maintenance**

	A	B	C	D	E
A		A->B	A->C	A->D	A->E
B	B->A		B->C	B->D	B->E
C	C->A	C->B		C->D	C->E
D	D->A	D->B	D->C		D->E
E	E->A	E->B	E->C	E->D	

- Définition d'un vocabulaire standard
 - Identifie explicitement tout concept utilisable dans le CDM
- Mapping des terminologies sources vers le vocabulaire standard
 - Traduction ascendante uniquement

Philosophie de mapping

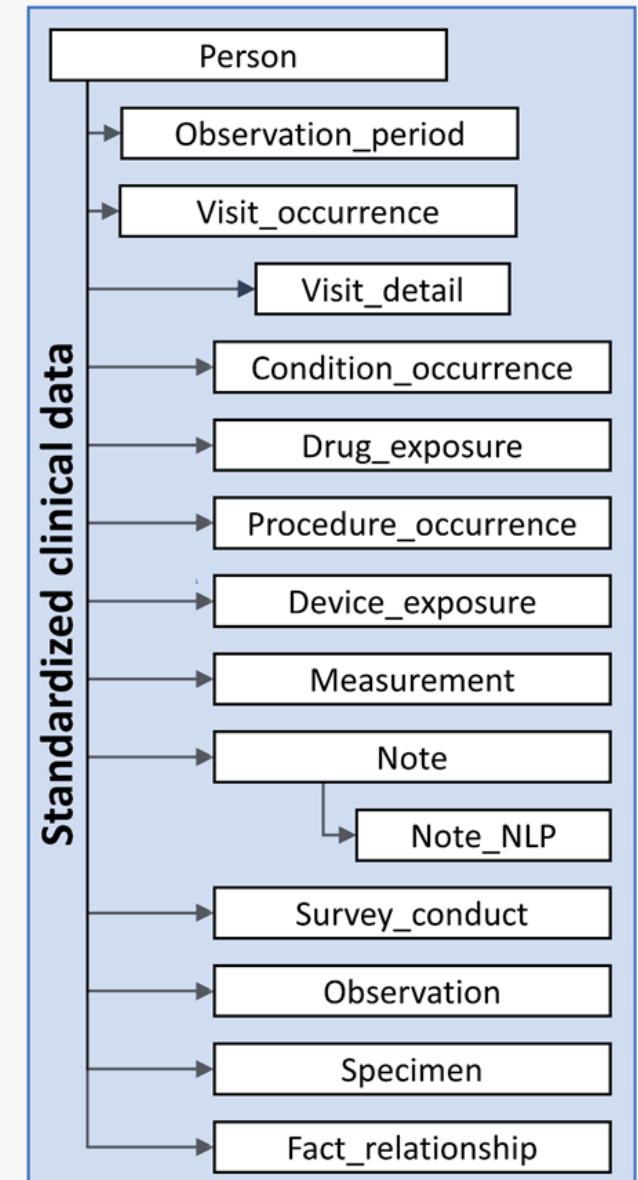


IV. Les tables d'OMOP CDM

1. Données cliniques
2. Données terminologiques
3. Système de prise en charge
4. Données dérivées
5. Données économiques
6. Métadonnées

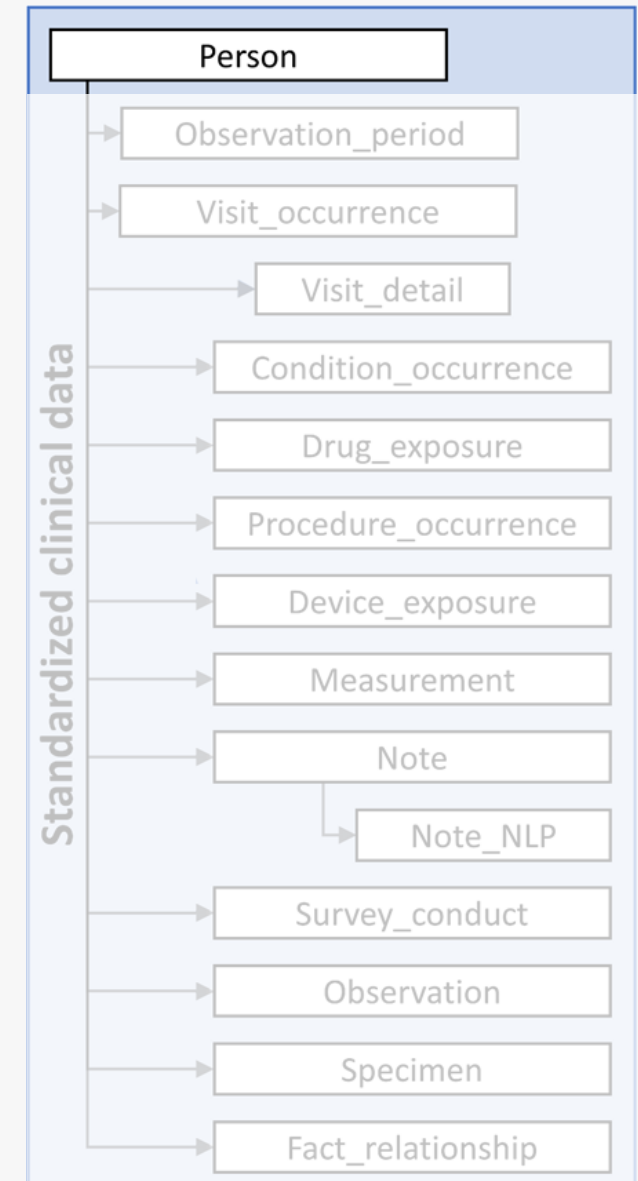
OMOP CDM v6.0 : Données cliniques

- Données relatives au patient et à son **parcours clinique**
 - Données individuelles
 - Prises en charges
 - Etats médicaux
 - Traitements
 - etc.
- Structurées autour de la table Person
 - Toute donnée est rattachée à un person_id



OMOP CDM v6.0 : Table PERSON

- Table de gestion centralisée des identités
- Identification unique de chaque patient
 - Tout enregistrement est un patient indépendant
- Toutes les personnes d'une base de données source doivent avoir un enregistrement
 - Sauf si elles ne respectent pas les exigences de qualité des données spécifiées dans l'ETL
 - Les personnes sans événements devraient néanmoins avoir un dossier
- Si plusieurs sources de données contribuent à des événements dans la base de données, un seul enregistrement doit être créé par personne
 - Rapprochement/réconciliation par l'ETL

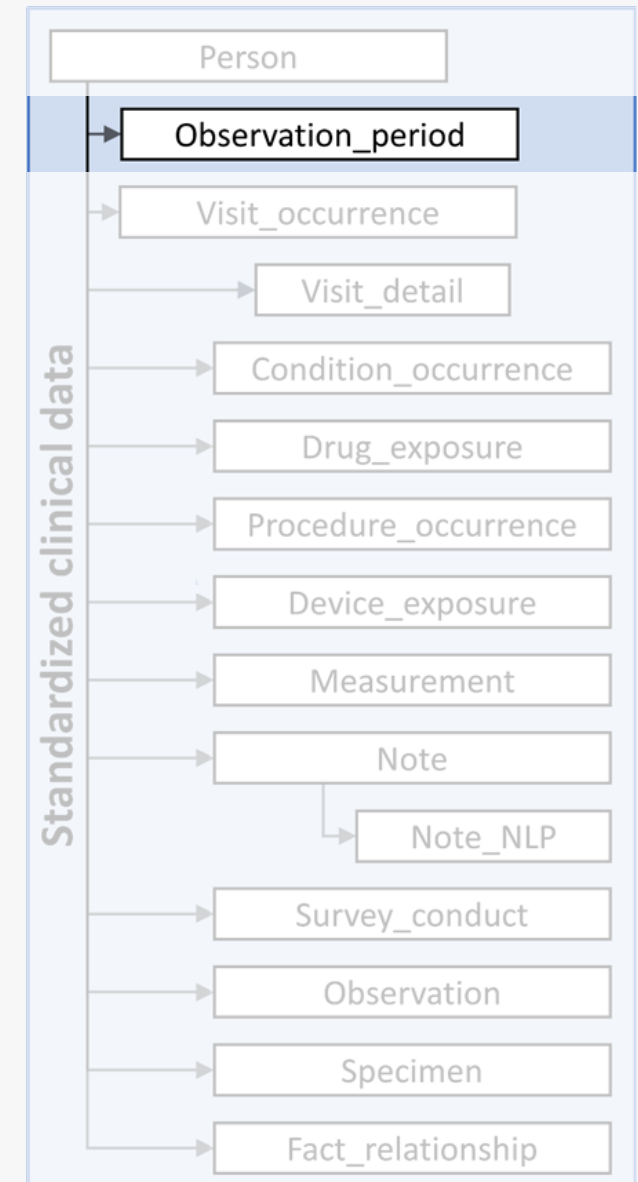


OMOP CDM v6.0 : Table PERSON

Champ CDM	Guide d'utilisation	Type de données	Champ requis	Clé étrangère
person_id	Chaque identifiant unique différent doit être une personne unique et être traitée indépendamment	integer	oui	
gender_concept_id	Relatif au sexe biologique à la naissance, sans notion de genre (sic)	integer		CONCEPT
year_of_birth	Année de naissance sur 4 chiffres	integer	oui	
month_of_birth	Mois de naissance (1-12)	integer		
day_of_birth	Jour de naissance (1-31)	integer		
birth_datetime	Non obligatoire mais fortement encouragé. Règles spécifiques de remplissage.	datetime		
death_datetime	Date et heure de décès (si applicable)	datetime		
race_concept_id	<i>Non utilisé en dehors des USA</i>	integer	oui	CONCEPT
ethnicity_concept_id	<i>Non utilisé en dehors des USA</i>	integer	oui	CONCEPT
location_id	Dernière adresse postale connue	integer		LOCATION
provider_id	Dernier médecin généraliste connu	integer		PROVIDER
care_site_id	Référence le CARE_SITE lié au provider_id	integer		CARE_SITE
person_source_value	Référence du patient dans les données sources. Recommandé pour permettre la vérification des données.	varchar		
gender_source_value	Valeur depuis les données sources. Utilisé uniquement à titre de référence et de contrôle.	varchar		
gender_source_concept_id	Généralement 0, sauf si les données sources codent le sexe biologique dans un vocabulaire non standard.	integer	oui	CONCEPT
race_source_value	<i>Non utilisé en dehors des USA</i>	varchar		
race_source_concept_id	<i>Non utilisé en dehors des USA</i>	integer	oui	CONCEPT
ethnicity_source_value	<i>Non utilisé en dehors des USA</i>	varchar		
ethnicity_source_concept_id	<i>Non utilisé en dehors des USA</i>	integer	oui	CONCEPT

OMOP CDM v6.0 : Table OBSERVATION_PERIOD

- Table des périodes d'observations liées à un patient
- Désigne une période d'attention clinique
 - Assure que tout évènement enregistré dans cette période a eu lieu
 - Assure que tout évènement non-enregistré dans cette période n'a pas eu lieu
- Des évènements peuvent être enregistrés en dehors des périodes d'OBSERVATION_PERIOD
 - Pas de garantie qu'un évènement non-enregistré n'ait pas eu lieu en dehors des périodes
- Les périodes sont exclusives entre elles pour un patient
 - Les périodes se chevauchant doivent être fusionnées
 - Même recommandation pour les périodes adjacentes

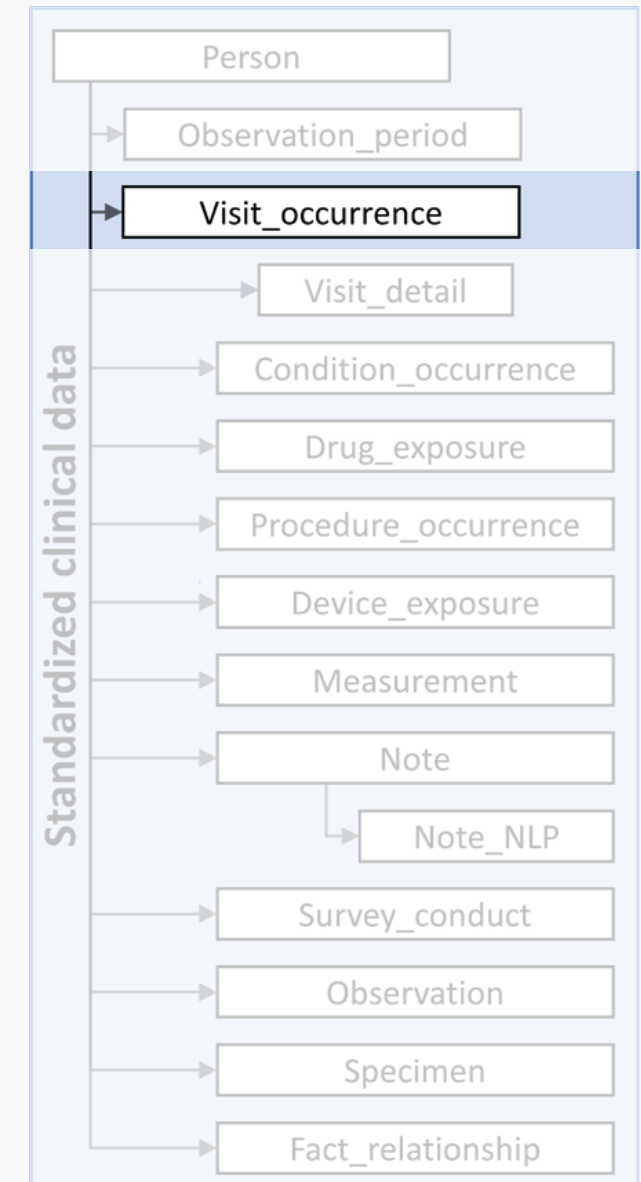


OMOP CDM v6.0 : Table OBSERVATION_PERIOD

Champ CDM	Guide d'utilisation	Type de données	Champ requis	Clé étrangère
observation_period_id	Identifiant unique de la période	integer	oui	
person_id	Patient concerné	integer	oui	PERSON
observation_period_start_date	En l'absence de date explicite, la date du premier évènement est utilisée	date	oui	
observation_period_end_date	Date à laquelle on présume tous les événements liés comme enregistrés. Peut être inférée du dernier enregistrement concerné	date	oui	
period_type_concept_id	Définit la provenance des observations (dossier patient, consultation, etc)	integer	oui	CONCEPT

OMOP CDM v6.0 : Table VISIT_OCCURENCE

- Table des interactions du patient avec le système de santé
- Concerne les hospitalisations comme les consultations
- Si la date de sortie est identique à la date d'entrée, la durée de la visite stockée est de 0
- Les visites ne peuvent se chevaucher
 - Mais elles peuvent se juxtaposer (date d'entrée identique à date de sortie de la précédente)

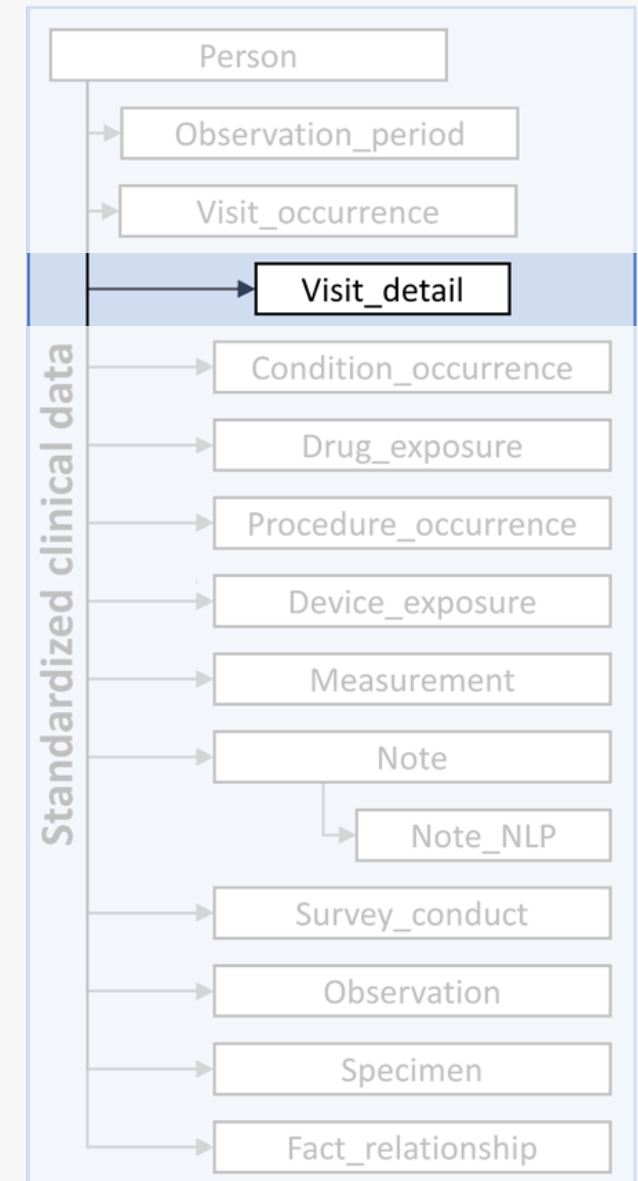


OMOP CDM v6.0 : Table VISIT_OCCURRENCE

Champ CDM	Guide d'utilisation	Type de données	Champ requis	Clé étrangère
<u>visit_occurrence_id</u>	Identifiant unique, utilisé comme clé étrangère dans les autres tables pour les événements liés à cette visite	integer	oui	
person_id	Patient concerné	integer	oui	PERSON
visit_concept_id	Concept standardisé du type de visite (hospitalisé, ambulatoire, etc)	integer	oui	CONCEPT
visit_start_date	Pour un patient hospitalisé, la date de début correspond à la date d'admission. En ambulatoire, il s'agit de la date de consultation	date		
visit_start_datetime		datetime	oui	
visit_end_date	Date de fin de la visite	date		
visit_end_datetime		datetime	oui	
visit_type_concept_id	Origine de la visite ou de l'enregistrement. Code notable : visit_type_concept_id = 32220 : "Still patient"	integer	oui	CONCEPT
provider_id	Prestataire lié à la visite. En cas de prestataires additionnels, ils peuvent être stockés dans VISIT_DETAIL	integer		PROVIDER
care_site_id	Où se déroule la visite	integer		CARE_SITE
visit_source_value	La valeur du type de visite telle que représentée dans les données sources.	varchar		
visit_source_concept_id	Si le type de visite est standardisé, son code apparaît ici (sinon, 0)	integer	oui	CONCEPT
admitted_from_source_value	Origine de l'admission depuis données sources. Cela ne s'applique généralement qu'aux patients hospitalisés ou en prise en charge de longue durée	varchar		
admitted_from_concept_id	Origine de l'admission depuis les terminologies standards, si existant	integer	oui	CONCEPT
discharge_to_source_value	Destination de sortie dans les données sources. Usuellement, il s'agit d'une unité de soin ou d'une sortie	varchar		
discharge_to_concept_id	Destination depuis les terminologies standards, si existant	integer	oui	CONCEPT
preceding_visit_occurrence_id	Référence vers la dernière visite du patient	integer		VISIT_OCCURRENCE

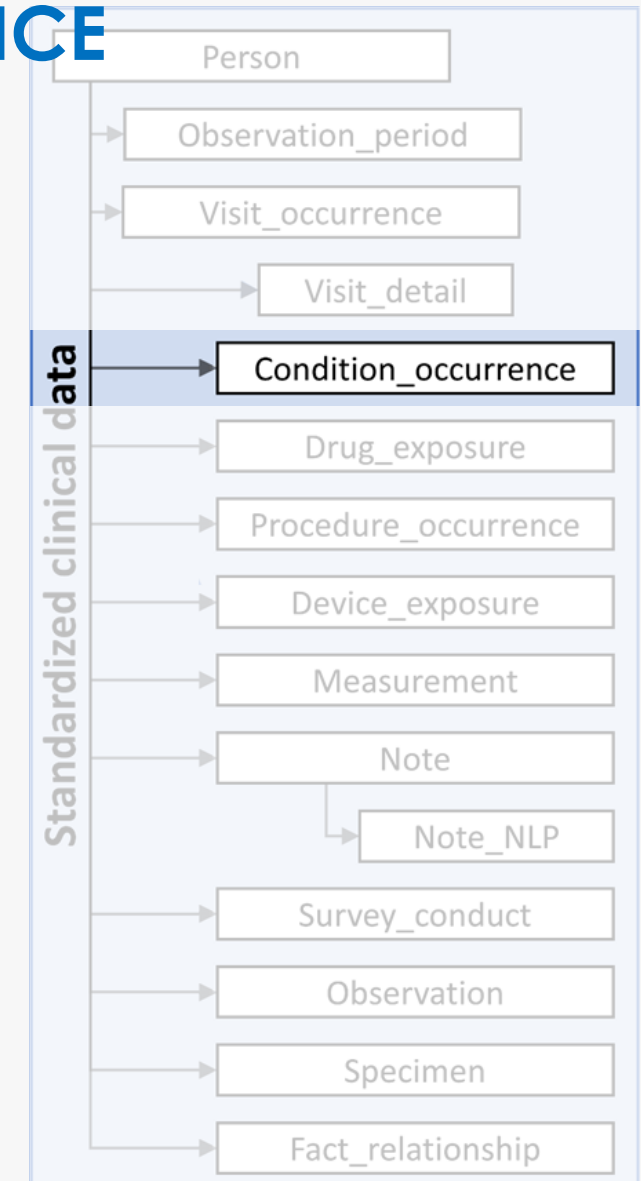
OMOP CDM v6.0 : Table VISIT_DETAIL

- Table optionnelle pour détailler [VISIT_OCCURENCE](#)
- Chaque enregistrement dans [VISIT_OCCURENCE](#) peut être lié à 0 ou plus enregistrements dans [VISIT_DETAIL](#)
 - Peut être utilisé pour décrire la granularité d'une visite (rencontre avec différents praticiens, etc)
- Schéma identique à [VISIT_OCCURENCE](#) plus deux champs:
 - Référence à la [VISIT_OCCURENCE](#) mère
 - Référence au précédent [VISIT_DETAIL](#) de la visite, si existant



OMOP CDM v6.0 : Table **CONDITION_OCCURENCE**

- Tables des conditions médicales, maladie ou état, qu'il s'agisse d'un diagnostic, d'un symptôme
 - observé par un soignant ou rapporté par le patient
- Durée : existence d'une date de début et de fin
 - Fin rarement connue en raison de la nature de la donnée
 - Manquement traité dans **CONDITION_ERA**
 - Enregistré en tant qu'élément unitaire : date de début = date de fin
- Ne concerne que les éléments « actuels »
 - Antécédents personnels et familiaux sont stockés dans **OBSERVATION**

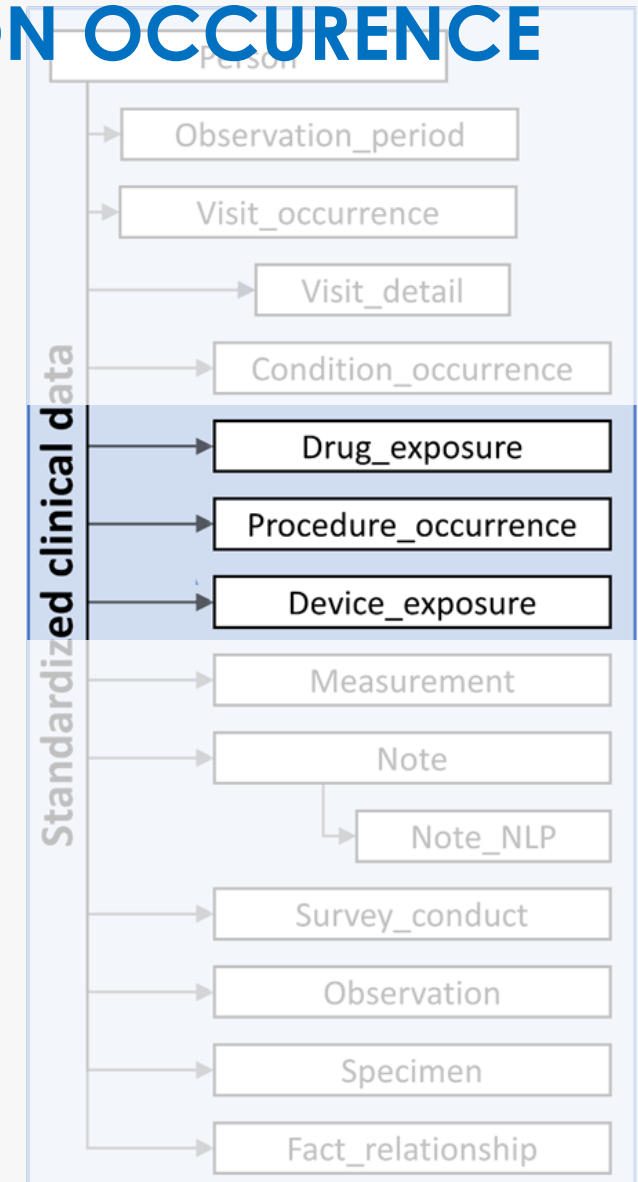


OMOP CDM v6.0 : Table CONDITION_OCCURENCE

Champ CDM	Guide d'utilisation	Type de données	Champ requis	Clé étrangère
condition_occurrence_id	Identifiant unique de la condition	integer	oui	
person_id	Patient concerné	integer	oui	PERSON
condition_concept_id	Concept standardisé de la condition	integer	oui	CONCEPT
condition_start_date	Date de début de la condition. À défaut, peut être inféré de la prise en charge	date		
condition_start_datetime		datetime		
condition_end_date	Peut être laissé vide si non connu, peut également utiliser la date de début comme date de fin	date		
condition_end_datetime		datetime	oui	
condition_type_concept_id	Concept standardisé du type de la condition	integer	oui	CONCEPT
condition_status_concept_id	Concept représentant le moment où le diagnostic a été posé (admission, diagnostic final), diagnostic déterminé en fonction des résultats de laboratoire, diagnostic exclusion, diagnostic préliminaire, etc.	integer	oui	CONCEPT
stop_reason	Raison pour laquelle l'enregistrement n'est plus valide. Peut être vide, peut ne pas être implémenté. Peut contenir des données et rester valide	varchar		
provider_id	Qui a réalisé le diagnostic ou procédé à l'enregistrement (choix d'implémentation de l'ETL)	integer		PROVIDER
visit_occurrence_id	Visite durant laquelle la condition est observée, si applicable	integer		VISIT_OCCURRENCE
visit_detail_id	Instance de VISIT_DETAIL lié à l'enregistrement de cette condition, si applicable	integer		VISIT_DETAIL
condition_source_value	Valeur de la condition dans les données sources (par exemple, un code CIM10)	varchar		
condition_source_concept_id	Code de la condition dans les données sources	integer	oui	CONCEPT
condition_status_source_value	Valeur issue des données sources caractérisant condition_status	varchar		

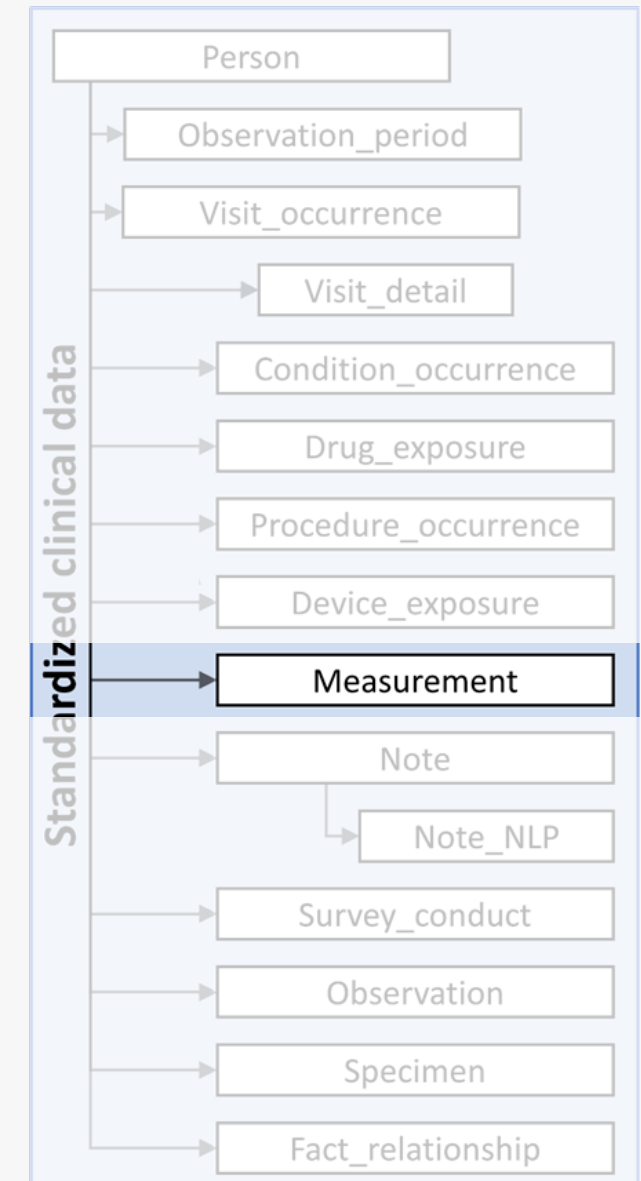
OMOP CDM v6.0 : Tables similaires à CONDITION OCCURENCE

- Table DRUG_EXPOSURE
 - Prise / administration de substances médicamenteuses ou assimilées
 - Définies par principe actif et quantités
- Table PROCEDURE_OCCURENCE
 - Actes médicaux réalisés
 - Date unique
 - Une procédure sur plusieurs jours doit être enregistrée sous plusieurs occurrences
- Table DEVICE_EXPOSURE
 - Exposition à un appareil physique ou instrument
 - Tant pour des raisons thérapeutiques que de diagnostic
 - Domaine vaste: pacemakers, plâtres, points de suture, défibrillateurs, etc.
 - Attention à la distinction entre *procedure* et *device*



OMOP CDM v6.0 : Table MEASUREMENT

- Table des mesures (généralement biologiques) réalisées
- Peuvent être numériques ou catégorielles
 - Numérique: si la mesure à une unité, elle est référencée par un concept
 - Catégorielle: les catégories possibles sont référencées par des concepts
- Peuvent être mathématiquement discrètes (par défaut)
 - Ou supérieure ou inférieure à une valeur (opérateurs <, >, <=, >=)
- Mesures unitaires
 - Les mesures multi-dimensionnelles doivent utiliser plusieurs enregistrements (ex : pression artérielle)

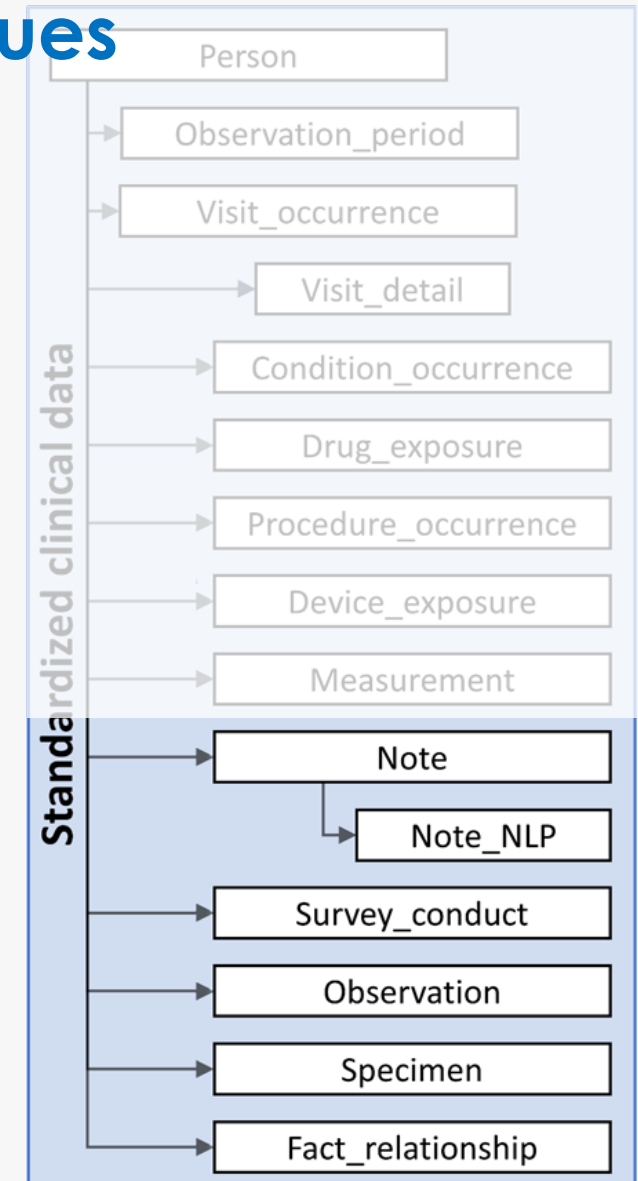


OMOP CDM v6.0 : Table MEASUREMENT

Champ CDM	Guide d'utilisation	Type de données	Champ requis	Clé étrangère
measurement_id	Identifiant unique de la mesure	integer	oui	
person_id	Patient concerné	integer	oui	PERSON
measurement_concept_id	Référence du concept définissant la mesure	integer	oui	CONCEPT
measurement_date	Date de la mesure.	date		
measurement_datetime		datetime	oui	
measurement_time	Déprécié	varchar		
measurement_type_concept_id	Référence du concept définissant la provenance de la mesure (dossier médical, réclamation d'assurance, registre, autre source)	integer	oui	CONCEPT
operator_concept_id	Référence à l'opérateur logique (< , <= , = , >= , >) entre la mesure et sa valeur. Par commodité, le non-remplissage de ce champ doit être interprété comme utilisant le code 4172703 pour =	integer		CONCEPT
value_as_number	Valeur numérique de la mesure	float		
value_as_concept_id	Référence du concept de la valeur de la mesure, si celle-ci est catégorielle	integer		CONCEPT
unit_concept_id	Référence du concept de l'unité de la mesure, si celle-ci en a une	integer		CONCEPT
range_low	Limite basse de la normale, si fournie. NULL en l'absence de normale	float		
range_high	Limite haute de la normale, si fournie. NULL en l'absence de normale	float		
provider_id	Le prestataire de soins lié à la mesure (celui qui l'a ordonné ou enregistré, en fonction des choix d'implémentation de l'ETL)	integer		PROVIDER
visit_occurrence_id	Référence à la visite durant laquelle la mesure a été réalisée, si application	integer		VISIT_OCCURRENCE
visit_detail_id	Si la mesure est liée à un enregistrement dans VISIT_DETAIL	integer		VISIT_DETAIL
measurement_source_value	Valeur supposée catégorielle de la mesure dans les données sources	varchar		
measurement_source_concept_id	Concept source de la mesure	integer	oui	CONCEPT
unit_source_value	Copie du champ "unité" des données sources, si existant	varchar		
value_source_value	Valeur supposée numérique de la mesure dans les données sources, sans présumer du type	varchar		

OMOP CDM v6.0 : Autres tables Données cliniques

- Table NOTE
 - Informations non structurées pour un patient à un temps donné.
 - Les notes sont typées par des concepts, leur contenu est en texte libre.
- Table NOTE_NLP
 - Stocke les résultats d'analyses NLP sur le contenu de la table [NOTE](#)
- Table SURVERY_CONDUCT
 - Enregistrements de remplissage de questionnaires et métadonnées liées
- Table OBSERVATION
 - Tout fait médical collecté sur le patient durant son séjour (antécédents, situation familiale, circonstances sociales, choix de vie, ...)
- Table SPECIMEN
 - Echantillons biologiques liés à un patient
- Table FACT_RELATIONSHIP
 - Relations entre des entités stockées dans n'importe quelles tables du modèle (entre [PERSON](#) : parent-enfant ; entre [CARE_SITE](#) : structure hiérarchique, ...)



OMOP CDM v6.0 : Données terminologiques

- Données relatives aux références terminologiques utilisées dans le CDM
- Structurées autour de la table Concept et Vocabulary
 - La table Concept contient tous les **concepts sémantiques** utilisés dans le CDM
 - La table Vocabulary référence les **terminologies** dont sont issus les concepts
 - Les autres tables définissent les **relations** entre les concepts

Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_relationship
Relationship
Concept_synonym
Concept_ancestor
Source_to_concept_map
Drug_strength

OMOP CDM v6.0 : Table CONCEPT

- Table listant tous les concepts maintenus par OHDSI
 - Terminologies standards OMOP CDM
 - Terminologies maintenues (90+ actuellement)
- Table de référence
 - Toutes terminologies confondues
 - Tous domaines
 - Utilisée pour les clés étrangères de toutes les tables du modèle
- Date de validité
 - Date de début au 1^{er} janvier 1970 si pas de date de début connue
 - Date de fin pour les concepts dépréciés
 - Définie au 31 décembre 2099 pour les concepts en cours de validité
- Extensions possibles
 - concept_id supérieur à 2 000 000 000

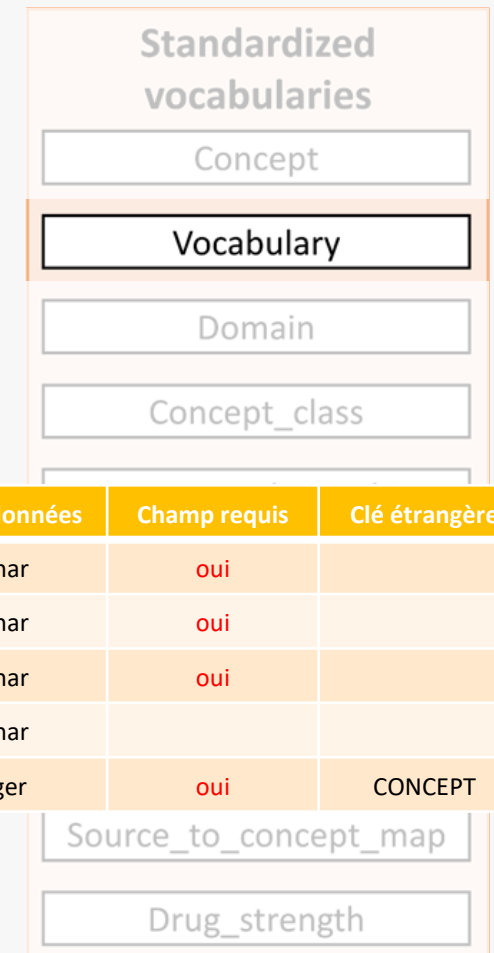
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_relationship
Relationship
Concept_synonym
Concept_ancestor
Source_to_concept_map
Drug_strength

OMOP CDM v6.0 : Table CONCEPT

Champ CDM	Guide d'utilisation	Type de données	Champ requis	Clé étrangère
concept_id	Identifiant unique du concept	integer	oui	
concept_name	Un nom sans ambiguïté, significatif et descriptif pour le concept	varchar	oui	
domain_id	Le domaine terminologique auquel le concept appartient	varchar	oui	DOMAIN
vocabulary_id	Le vocabulaire dont le concept est issu	varchar	oui	VOCABULARY
concept_class_id	La classe du concept au sein de sa terminologie	varchar	oui	CONCEPT_CLASS
standard_concept	Flag pour déterminer si le concept est un concept standard (champ à 'S') ou un concept de classification (champ à 'C'). Sinon, le champ est à NULL.	varchar		
concept_code	Le code du concept dans sa propre terminologie	varchar	oui	
valid_start_date	Date de début de validité. 1970-01-01 par défaut	date	oui	
valid_end_date	Date de dépréciation. 2099-12-31 pour les concepts valides	date	oui	
invalid_reason	Flag de motif de dépréciation. 'D' si le concept a été supprimé ou 'U' s'il est remplacé par une mise à jour (et donc un nouveau code). NULL si le concept est toujours valide	varchar		

OMOP CDM v6.0 : Table VOCABULARY

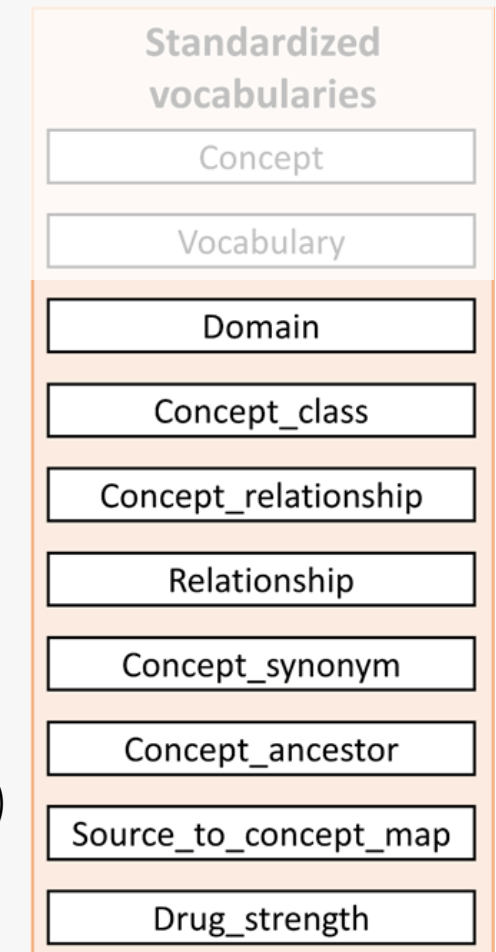
- Table listant tous les vocabulaires dont le contenu est chargé dans la table **CONCEPT**
- Inclut quelques terminologies ad hoc OMOP CDM
 - Exemples : domaines de concepts, vocabulaire des terminologies



Champ CDM	Guide d'utilisation	Type de données	Champ requis	Clé étrangère
<u>vocabulary_id</u>	Identifiant unique de la terminologie	varchar	oui	
vocabulary_name	Nom complet de la terminologie	varchar	oui	
vocabulary_reference	Référence externe (généralement une URL) vers la documentation de la terminologie	varchar	oui	
vocabulary_version	Version de la terminologie	varchar		
vocabulary_concept_id	Clé étrangère du concept représentant le vocabulaire. 0 si non applicable	integer	oui	CONCEPT

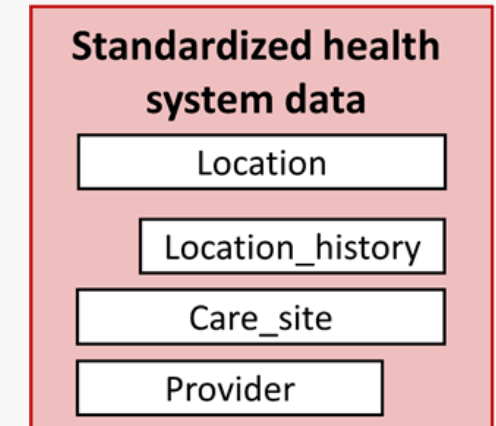
OMOP CDM v6.0 : Autres tables terminologiques

- Table DOMAIN
 - Domaines définis par OMOP pour qualifier les concepts et leur utilisation
- Table CONCEPT_CLASS
 - Référence les classifications internes aux terminologies
- Table CONCEPT_RELATIONSHIP
 - Définit les liens entre les concepts
- Table RELATIONSHIP
 - Référence les relations possibles entre les concepts (hiérarchiques, équivalences, remplacements, ...)
- Table CONCEPT_SYNONYM
 - Stocke les noms alternatifs pour un même concept
- Table CONCEPT_ANCESTOR
 - Liste les ascendants hiérarchiques des concepts (facilite la généralisation des requêtes)
- Table SOURCE_TO_CONCEPT_MAP
 - Table de compatibilité pour lier manuellement des concepts locaux aux standards
- Table DRUG_STRENGTH
 - Explicite les doses et compositions des concepts médicamenteux.



OMOP CDM v6.0 : Données du système de santé

- Données relatives au système de santé liées au patient et à son parcours
- Table LOCATION
 - Stocke les adresses physiques des patients (table **PERSON**) et des lieux de soins (table **CARE_SITE**)
- Table LOCATION_HISTORY
 - Stocke les relations entre les patients et les lieux de soins au fil du temps.
- Table CARE_SITE
 - Stocke les entités de soins, tant physiques qu'organisationnelles
- Table PROVIDER
 - Stocke les identifiants des « fournisseurs de soins » liés aux actes, médicaments et observations.



OMOP CDM v6.0 : Données économiques

- Données liées à la facturation des soins
- Table COST
 - Stocke les informations de coût liés aux événements enregistrés dans [DRUG_EXPOSURE](#), [PROCEDURE_OCCURRENCE](#), [VISIT_OCCURRENCE](#), [VISIT_DETAIL](#), [DEVICE_OCCURRENCE](#), [OBSERVATION](#) ou [MEASUREMENT](#).
- Table PAYER_PLAN_PERIOD
 - Enregistre les périodes de temps liés à des plans de soins (au sens US du terme).

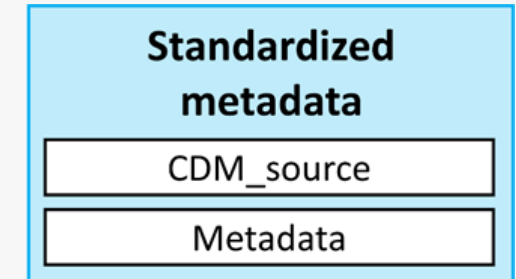
Standardized health economics

Cost

Payer_plan_period

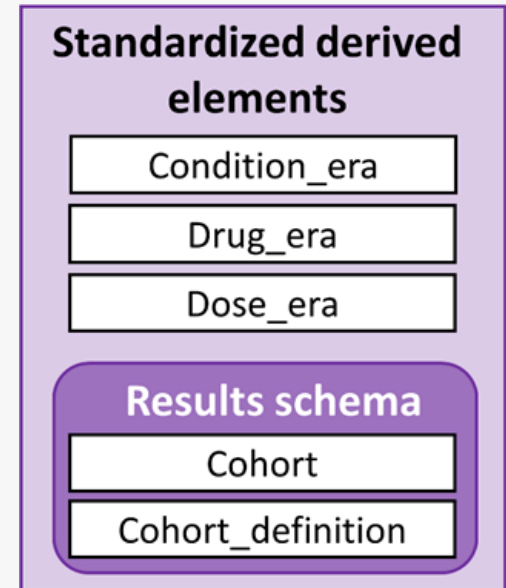
OMOP CDM v6.0 : Métadonnées

- Utilisées pour stocker de l'information sur les données et méthodes utilisées pour peupler le modèle OMOP
- Table CDM_SOURCE
 - Chaque enregistrement décrit une source (base de données) utilisée
 - L'ETL utilisé peut également être référencé
- Table METADATA
 - Stocke les métadonnées liées aux jeux de données sources intégrés dans le modèle



OMOP CDM v6.0 : Données dérivées

- Tables contenant des données secondaires
- Générées à partir du contenu du CDM
 - OHDSI fournit les scripts de génération des données secondaires
- Table DRUG_ERA
 - Période d'exposition d'un patient à un médicament, combinaisons d'instances de [DRUG_EXPOSURE](#)
- Table CONDITION_ERA
 - Même raisonnement sur les instances de [CONDITION_OCCURENCE](#)
- Table DOSE_ERA
 - Stocke les périodes d'exposition à des doses constantes de médicament, à partir des données de [DRUG_EXPOSURE](#) et de [DRUG_STRENGTH](#)
- Table COHORT_DEFINITION
 - Décrit une cohorte et ses métadonnées
- Table COHORT
 - Relie des [PERSON](#) à des cohortes définies dans [COHORT_DEFINITION](#)





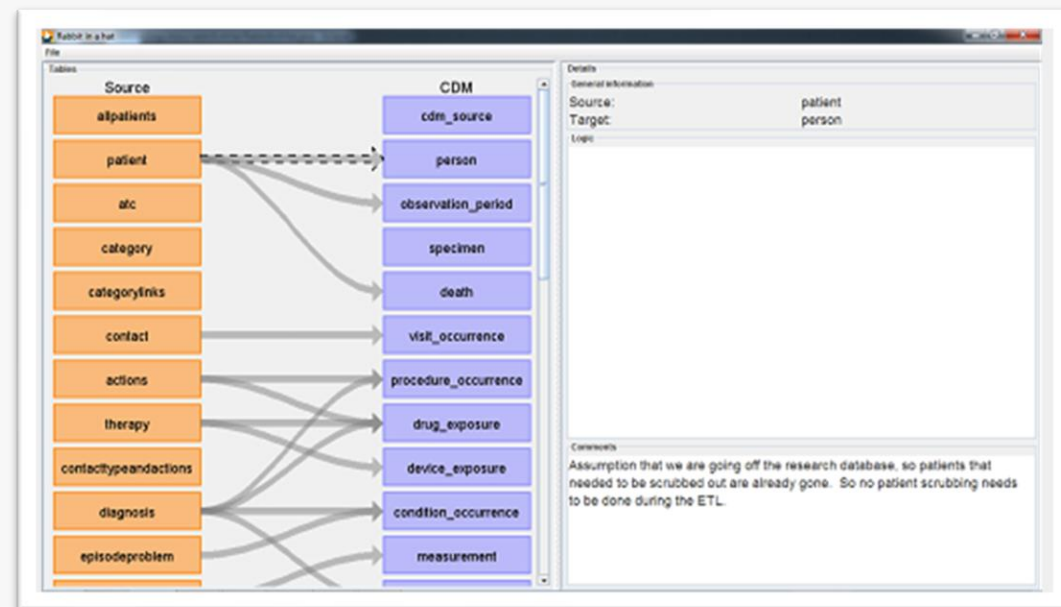
V. Outils OHDSI

1. Aide à la conception
2. Support terminologique
3. Outils d'exploitation
4. Analyses et contrôle qualité

Conception – WhiteRabbit et RabbitInAHat

- WhiteRabbit: application d'analyse de structure et de contenu de base de données
 - en vue de la conception d'un ETL vers OMOP CDM
- RabbitInAHat: application de conception interactive d'ETL
 - Utilise les rapports d'analyse générés par WhiteRabbit
- SQL Server, Oracle, PostgreSQL, MySQL, MS Access, Amazon RedShift, Google BigQuery, fichiers SAS et CSV
- Rapports sur les tables, les champs, les distributions fréquentielles des valeurs
- Seuil de fréquence minimale pour la confidentialité des patients

<https://github.com/OHDSI/whiterabbit>



Conception & Support terminologique – USAGI

- Application pour aider la création de mappages entre les systèmes de codage sources et les concepts standards du vocabulaire
- Pré-mapping basé sur la similarité des termes
- Recherche synonymique
- Revue interactive
- Filtrage par vocabulaire, domaine, sous-ensemble de concepts
- Export au format de la table **SOURCE_TO_CONCEPT_MAP**
- Ré-application de mapping à un jeu de codes mis à jour

Usagi - testlpcCodesMapping.csv

File Edit View Help

Status	Source code	Source term	Frequency	CodeText	Match score	Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard con...	Parents	Children	Comment
Approved	K87.00	Hypertension...	694195	Hypertensie ...	0.81	316866	Hypertensive...	Condition	Clinical Findi...	SNOMED	38341003	S	1	27	
Approved	L99.00	Other diseas...	680422	Andere ziekte...	0.47	0	Unmapped					0	0		Too generic
Approved	D01.00	Abdominal p...	678588	Gegeneralis...	0.61	197988	Generalized ...	Condition	Clinical Findi...	SNOMED	102614006	S	1	0	
Unchecked	S99.00	Skin disease...	675817	Andere ziekte...	0.75	4317258	Disorder of s...	Condition	Clinical Findi...	SNOMED	95320005	S	2	193	
Unchecked	T86.00	Hvnothvroldi...	667283	Hvnothvroldi...	1.00	4113642	Hvnothvroldi...	Condition	Clinical Findi...	SNOMED	286910004	S	1	0	

Source code

Source code	Source term	Frequency	CodeText
S99.00	Skin disease, other	675817	Andere ziekte(n) huid/subcutis

Target concepts

Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard concept	Parents	Children
4317258	Disorder of skin	Condition	Clinical Finding	SNOMED	95320005	S	2	193

Remove concept

Search

Query

☒ Use source term as query

☐ Query:

Filters

☐ Filter by user selected concepts

☒ Filter standard concepts

☒ Include source terms

☐ Filter by concept class: 2-dig nonbill code

☐ Filter by vocabulary: ABMS

☒ Filter by domain: Condition

Results

Score	Term	Concept ID	Concept name	Domain	Concept class	Vocabulary	Concept code	Standard concept	Parents	Children
0.75	Skin disease	4317258	Disorder of skin	Condition	Clinical Finding	SNOMED	95320005	S	2	193
0.65	Skin Disease, Fungal	137213	Dermal mycosis	Condition	Clinical Finding	SNOMED	14560005	S	3	12
0.57	AIDS with skin dise...	4224566	Skin disorder associated with AIDS	Condition	Clinical Finding	SNOMED	421394009	S	2	2
0.56	Chronic skin disease	4134132	Chronic disease of skin	Condition	Clinical Finding	SNOMED	128236002	S	2	26
0.55	Disease, Otologic	378161	Disorder of ear	Condition	Clinical Finding	SNOMED	25906001	S	4	43
0.55	Disease, Hers	4163346	Glycogen storage disease, type VI	Condition	Clinical Finding	SNOMED	29291001	S	2	0
0.55	Other peripheral va...	321052	Peripheral vascular disease	Condition	Clinical Finding	SNOMED	400047006	S	1	44
0.55	Other peripheral va...	4119612	Lower limb ischemia	Condition	Clinical Finding	SNOMED	233961000	S	2	3
0.55	Disease, Ormond	4176725	Retropertoneal fibrosis	Condition	Clinical Finding	SNOMED	49120005	S	1	3
0.54	Pathological fractur...	73571	Pathological fracture	Condition	Clinical Finding	SNOMED	288029009	S	1	21
0.52	Disease, Tooth	4122115	Tooth disorder	Condition	Clinical Finding	SNOMED	234947003	S	3	58
0.52	Disease, Lip	135858	Disorder of lip	Condition	Clinical Finding	SNOMED	90678009	S	3	35
0.61	Disease, Ollier	4113600	Multifocal congenital anosteosis	Condition	Clinical Finding	SNOMED	254014004	S	6	0

Replace concept Add concept

Comment:

Approved / total: 5 / 1037 8.8% of total frequency

Approve

<https://github.com/OHDSI/usagi>

Support terminologique – Athena

- Moteur de recherche terminologique
- Sur les terminologies standards OMOP et les terminologies sources maintenues

- Recherches par:

- Termes
- Codes
- Domaines
- Terminologies
- Classes
- Validité

<https://athena.ohdsi.org>

The screenshot shows the Athena OHDSI search interface. The top navigation bar includes the Athena logo, a search bar with the keyword 'Insomnia', and buttons for SEARCH, DOWNLOAD, LOGIN, and a help icon. Below the search bar, the left sidebar shows filters for 'Insomnia' (selected) and 'Condition'. The main area displays a table of search results with columns: ID, CODE, NAME, CLASS, CONCEPT, VALIDITY, DOMAIN, and VOCAB. The table shows 171 total items, with 15 items displayed per page. The results include various insomnia-related terms from different terminologies like SNOMED and PPI.

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
436962	193462001	Insomnia	Clinical Finding	Standard	Valid	Condition	SNOMED
40615648	62577008	Insomnia	Clinical Finding	Non-standard	Invalid	Condition	SNOMED
40355205	270478005	Insomnia	Clinical Finding	Non-standard	Invalid	Condition	SNOMED
4215402	3972004	Primary insomnia	Clinical Finding	Standard	Valid	Condition	SNOMED
43529116	NervousSystemConditions_Insomnia	Insomnia	Answer	Non-standard	Invalid	Condition	PPI
42538607	762348004	Acute insomnia	Clinical Finding	Standard	Valid	Condition	SNOMED
43020938	472819006	Adjustment insomnia	Clinical Finding	Standard	Valid	Condition	SNOMED
37110488	724748004	Chronic insomnia	Clinical Finding	Standard	Valid	Condition	SNOMED
44784625	91411000119105	Drug-induced insomnia	Clinical Finding	Standard	Valid	Condition	SNOMED

Exploitation – ATLAS

- Plateforme web intégrée pour l'exploitation d'OMOP CDM
- Exploration du vocabulaire
- Définition de cohorte
- Examen de cas
- Caractérisation clinique
- Estimation d'incidence
- Prédications

<https://github.com/OHDSI/Atlas>

Serveur de démonstration (partiel):

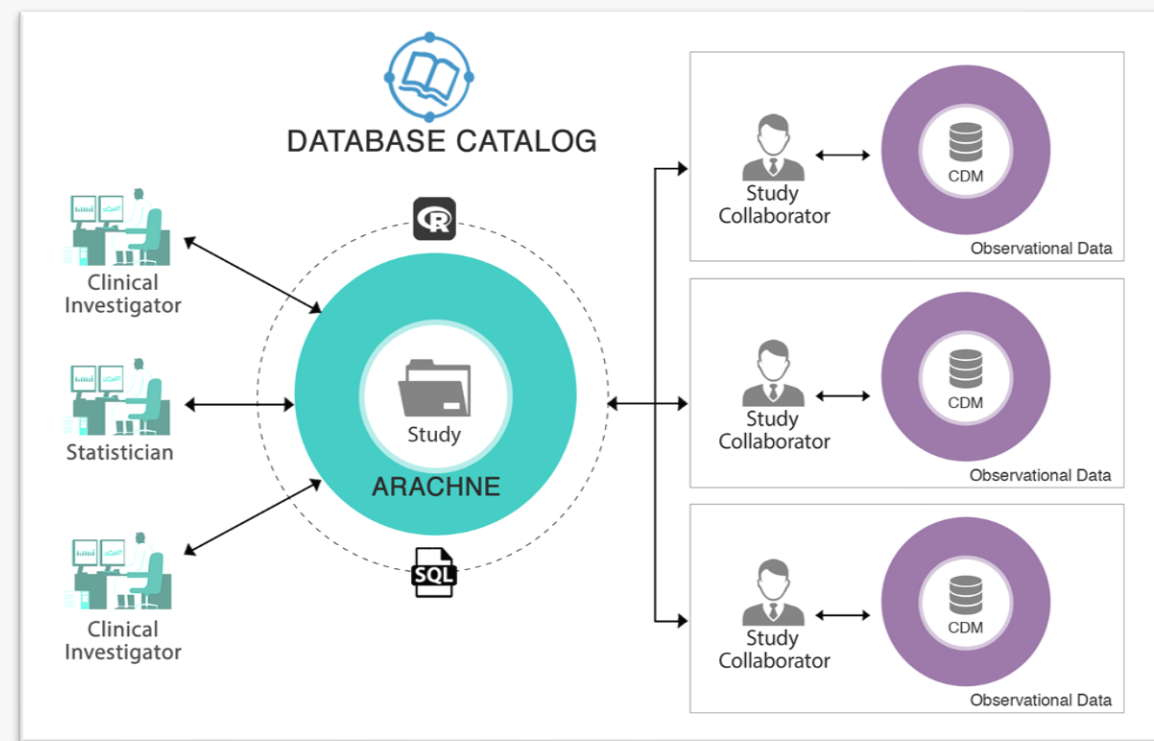
<http://atlas-demo.ohdsi.org>

The screenshot displays the ATLAS web interface for defining a cohort. The left sidebar contains navigation links: Home, Data Sources, Search, Concept Sets, Cohort Definitions (selected), Characterizations, Cohort Pathways, Incidence Rates, Profiles, Estimation, Prediction, Jobs, Configuration, and Feedback. The main content area is titled 'Cohort #1770710' and shows the definition for 'New users of ACE inhibitors as first-line monotherapy for hypertension'. The 'Definition' tab is active, with other tabs like Concept Sets, Generation, Reporting, Export, and Messages visible. A text input field prompts 'enter a cohort definition description here'. Below this, the 'Cohort Entry Events' section shows criteria for events: 'a drug exposure of ACE inhibitors' (selected from a dropdown) and 'for the first time in the person's history' (indicated by a red X). There are buttons for '+ Add attribute...', 'Delete Criteria', and '+ Add Initial Event'. Further down, observation criteria are set to 'at least 365 days before and 0 days after event index date', and initial events are limited to 'earliest event' per person. A 'Restrict initial events' button is present. The 'Inclusion Criteria' section at the bottom lists two criteria: '1. has hypertension diagnosis in 1 yr prior to treatment' and '2. Has no prior antihypertensive drug exposures in medical history'. The footer of the interface mentions 'Apache 2.0 open source software', 'provided by OHDSI', and 'join the journey'.

Exploitation – ARACHNE

- Outil d'infrastructure réseau pour déployer des études collaboratives
- Déploiement d'études unifiées sur plusieurs implémentations OMOP
- Gestion de la confidentialité entre les sources
- Intégration dans ATLAS

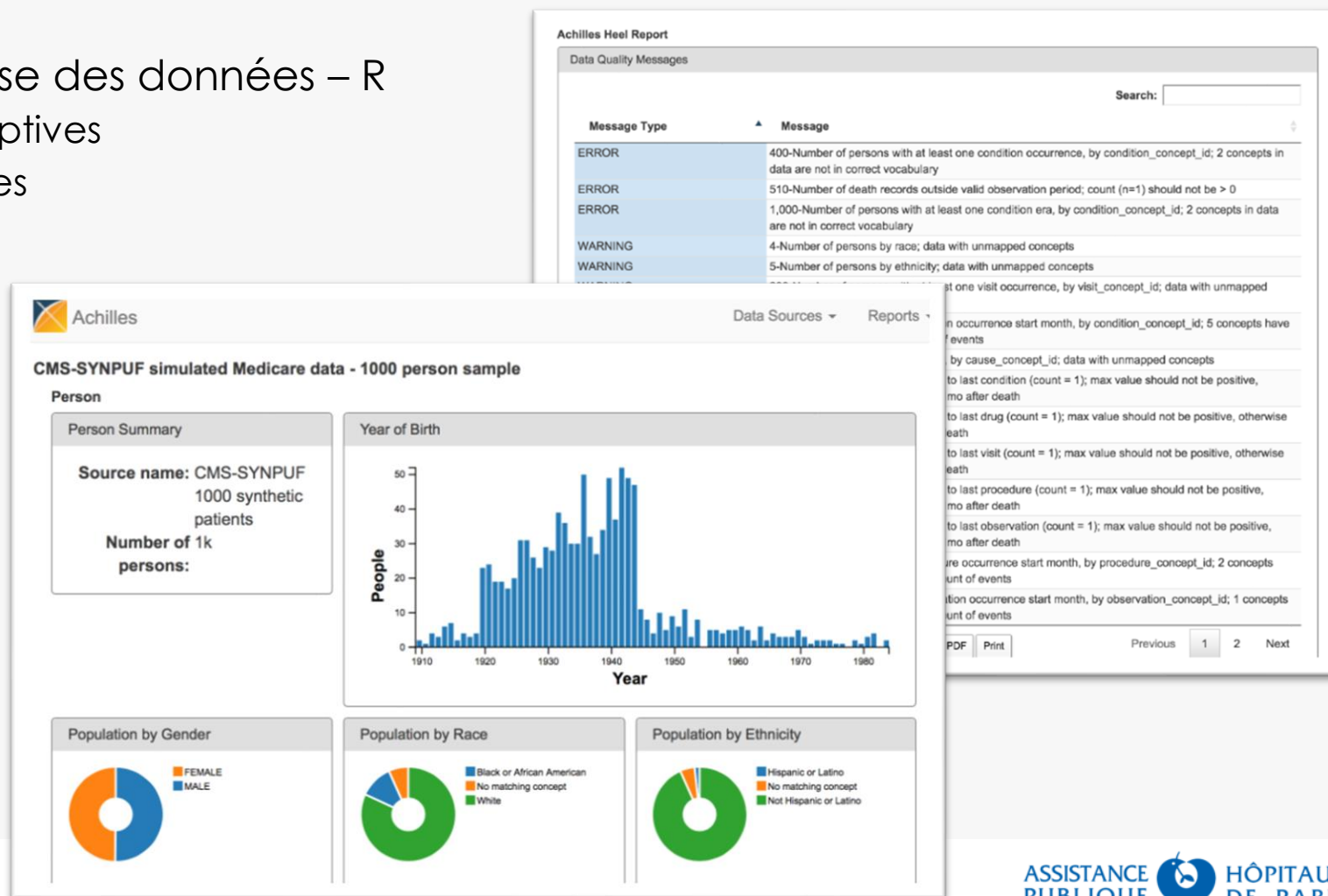
<https://github.com/OHDSI/ArachneUI>



Analyses et Qualité – ACHILLES

- Automated **C**haracterization of **H**ealth **I**nformation at **L**arge-scale **L**ongitudinal **E**vidence **S**ystems (ACHILLES)
- Framework open-source d'analyse des données – R
 - Production de statistiques descriptives
 - Contrôles de qualité des données
- Interface web de visualisation
- Actuellement disponible pour OMOP CDM v5.x uniquement

<http://www.github.com/ohdsi/achilles>



VI. Conclusion

OMOP CDM



- Utilisé pour **standardiser** la structure et les requêtes
- Intégré avec un **vocabulaire contrôlé**
- Consolide des données de sources **hétérogènes**
- Centré sur le patient
- Préserve la provenance des données
- Indépendant des plateformes utilisées

Le vocabulaire OMOP



- Utilisé pour **standardiser** la terminologie
- Regroupe des standards depuis des sources publiques, privées et propres à OMOP
- Permet la coexistence de terminologies multiples
- Conçu pour faciliter des requêtes
- Régulièrement mis à jour, **maintenu** et amélioré

Enjeux des implémentations OMOP



- **Recherche collaborative** sur des sources de données internationales
- Maintenues par une communauté de développeurs et de chercheurs interdisciplinaires
 - Fournisseurs d'une vaste gamme d'outils dédiés
- Spécialisé: reflétant le **domaine clinique**, granulaire et structuré
- Intégré avec un **vocabulaire uniformément structuré** et organisé
- **Extensible**: de nouveaux concepts et attributs peuvent être ajoutés

Sources et références

- [EN] Wiki de documentation d'OHDSI sur OMOP CDM v6.0
 - <https://ohdsi.github.io/CommonDataModel/cdm60.html>
- [EN] Symposium OHDSI 2018
 - <https://www.ohdsi.org/past-events/2018-ohdsi-symposium-materials/>
- [EN] G. Hripcsak et al., Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers, Stud Health Technol Inform, vol. 216, p. 574-578, 2015.
 - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4815923/>
- [FR] Fiche OMOP de la documentation du Health Data Hub
 - <https://documentation-snds.health-data-hub.fr/glossaire/omop.html>

URL du cours et licence

Permalien : <https://www.damien-leprovost.fr/enseignements/OMOP.2021.pdf>

Ce support de cours est distribué sous licence
Attribution - Partage dans les Mêmes Conditions 4.0 International (CC BY-SA 4.0)

Vous êtes autorisé à :

- **Partager** — copier, distribuer et communiquer le matériel par tous moyens et sous tous formats
- **Adapter** — remixer, transformer et créer à partir du matériel pour toute utilisation, y compris commerciale.

Selon les conditions suivantes :

-  **Attribution** — Vous devez [créditer](#) l'œuvre, intégrer un lien vers la licence et [indiquer](#) si des modifications ont été effectuées à l'œuvre. Vous devez indiquer ces informations par tous les moyens raisonnables, sans toutefois suggérer que l'Offrant vous soutient ou soutient la façon dont vous avez utilisé son œuvre.
-  **Partage dans les Mêmes Conditions** — Dans le cas où vous effectuez un remix, que vous transformez, ou créez à partir du matériel composant l'œuvre originale, vous devez diffuser l'œuvre modifiée dans les même conditions, c'est à dire avec [la même licence](#) avec laquelle l'œuvre originale a été diffusée.

