



LES ENTREPRISES DE LA
RECHERCHE CLINIQUE

ÉTUDES OBSERVATIONNELLES AVEC DES DONNÉES STANDARDISÉES : LE FORMAT OMOP



Synthèse du petit-déjeuner
du 25 avril 2024
organisé par des membres
du Groupe de travail RWD
de l'AFCROs

07/2023

Ce petit-déjeuner s'est tenu le 25 avril 2024.

Il était animé par **Manon Belhassen**, représentante de l'AFCROs et Présidente de PELyon

avec pour intervenants :

- **Gaëlle Collumeau**, Data engineer au HDH
- **Laureen Majed**, Responsable Medical Evidence BioPharma de Astra Zeneca
- **Alexandre Malouvier**, représentant de l'AFCROs et Senior Director Scientific Affairs & Digital Innovation de ICON
- **Gilles Paubert**, représentant de l'AFCROs et Managing Director, Global head Cegedim Health Data de GERS
- **Marie Verdoux**, Data Scientist à l'AP-HP (URC)

Pour accroître leur potentiel d'utilisation, les bases de données peuvent être standardisées vers des modèles reconnus à l'échelle internationale et largement utilisés, comme l'OMOP-CDM (Observational Medical Outcomes Partnership -Common Data Model), les rendant ainsi interopérables avec d'autres bases de données car partageant la même structure et utilisant les mêmes terminologies.

En France, plusieurs bases de données ont été ou sont en cours de traduction au format OMOP, que ce soient des dossiers médicaux hospitaliers, des dossiers médicaux de médecine générale, ou le SNDS (Système National des Données de Santé).

Ces transformations ouvrent de nouvelles perspectives de collaborations internationales.

Que veut dire « interopérabilité des données » ?

Dans les industries de la santé, l'interopérabilité est la capacité pour différents systèmes d'information et de logiciels de communiquer, d'échanger des données et d'utiliser ces données échangées. L'utilisation de standards et de modèles d'échange des données permettent à l'information d'être partagées par les établissements de soins, les professionnels de santé, les patients, les pharmacies, les laboratoires, etc., quelle que soit l'application utilisée.

Les standards définissent le format et la structure des données pour faciliter les échanges entre différents systèmes cliniques. On peut citer par exemple HL7 V2.X, HL7 V3 pour les données de démographie et de données cliniques et administratives, DICOM (*Digital Imaging and Communications in Medicine*) pour les données d'imagerie médicale et informations associées, ICD-10 (*International Classification of Diseases*) pour les diagnostics et les procédures cliniques, LOINC (*Logical Observation Identifiers Names & Codes*) pour les examens de laboratoire, les mesures et observations cliniques, et enfin SNOMED CT (*Systematized Nomenclature Of Medicine Clinical Terms*) qui est un système terminologique international permettant de restituer les informations cliniques contenues dans les dossiers électroniques des patients, de façon cohérente et exploitable par des ordinateurs.

Quelle interopérabilité entre sources de données, au niveau international ?

Le modèle commun de donnée OMOP-CDM est un modèle relationnel de bases de données de santé, qui a pour objectif l'interopérabilité entre les différentes bases d'analyse en santé, qu'elles soient cliniques ou médico-administratives. Cette interopérabilité se traduit concrètement par la possibilité de partage d'outils, de méthodes, d'études et de programmes entre les centres disposant d'une base au format OMOP. Cette initiative a été lancée par **OHDSI**, Observational Health Data Sciences and Informatics. OHDSI est une collaboration scientifique ouverte, multidisciplinaire et impliquant de multiples parties prenantes, visant à mettre en valeur les données de santé grâce à des analyses à grande échelle, dans plus de 70 pays (<https://www.ohdsi.org>).

Au niveau européen, **EHDEN** (European Health Data Evidence Network) vise à collaborer avec diverses institutions, sources de données et gardiens de données à travers l'UE, dans le but d'harmoniser les données sources, à grande échelle, selon le modèle de données commun OMOP, au sein d'un réseau fédéré. Grâce aux sept appels organisés, 187 partenaires de données provenant de 29 pays différents mappent leurs données au modèle OMOP (<https://www.ehden.eu>).

En France, quelles utilisations du format de données OMOP ?

La standardisation au format OMOP de la base principale du SNDS est en cours de mise en place par le HDH (Health Data Hub). Les étapes clés de la standardisation ont été :

1. Le design de l'ETL (*Extract Transform Load*), cela signifie établir des règles de correspondance entre les tables et variables sources et le modèle OMOP.
2. L'alignement des codes des nomenclatures sources vers des codes standards : OMOP-CDM utilise des terminologies standards (SNOMED-CT, RxNorm, etc.) pour décrire les événements médicaux. Comme les bases sources utilisent généralement des terminologies différentes (locales ou nationales), il devient nécessaire de les aligner avec celles standards, utilisées par OMOP-CDM, i.e., définir une correspondance entre les valeurs des terminologies locales et celles standards
3. L'implémentation de l'ETL
4. La validation de la qualité de la base standardisée, via la visualisation de statistiques générales sur la population (pyramide des âges, répartition homme/femme), répartition par domaine (nombre de procédures, diagnostics...) et par concept (nombre de personnes concernées en fonction du temps, etc.).

Au niveau de l'EDS (entrepôt de données de santé) de l'APHP, les données des 11 millions de patients répartis dans les 39 hôpitaux sont standardisés au format i2b2 ou OMOP. Cet EDS contient des données administratives, des données médico-économiques (avec le PMSI), ainsi que des données cliniques structurées et non structurées. Cet EDS a par exemple été utilisé pour réaliser l'étude UROCOVID, qui avait pour objectif d'étudier l'association entre l'exposition à l'acide ursodésoxycholique et le risque d'infection au COVID-19.

Enfin, en ce qui concerne les données de soins primaires, la base de données THIN®, qui est une base de données de vie réelle, médicalisée, multicentrique, multi source, est déjà mappée au format OMOP.

En conclusion, le format OMOP est centré sur le chercheur, et non sur le soin, c'est à dire que les concepts et les abstractions en sont les clés. Le mapping permet ainsi de lier les vocabulaires standards et les vocabulaires locaux. Le processus ETL peut cependant entraîner une perte de précision si les champs sources ne correspondent pas au *common data model*.

Même si OMOP présente encore des obstacles structurels, naturels à sa mise en œuvre harmonieuse au sein des bases de données de soins primaires, et que les données mappées sont actuellement surtout des données hospitalières, ce standard est un prérequis fondamental pour la collaboration entre équipes de recherche multi pays, et l'accès facilité à des analyses statistiques de haut niveau.

A propos de l'AFCROs

Les Entreprises de la Recherche Clinique (AFCROs), représentent près de 100 entreprises de conseil et de services dans le domaine de la recherche biomédicale pour les promoteurs d'études cliniques, pour l'industrie pharmaceutique, biotechnologique et du dispositif médical. L'AFCROs s'engage au quotidien pour une recherche éthique, responsable et durable.

www.afcros.com