

Spécifications techniques et fonctionnelles

Standardisation et interopérabilité du socle commun

Introduction	1
1. Rappel du contexte du groupe de travail	1
2. Consignes de lecture et mode d'emploi	2
3. Références et documentation utiles	2
Éléments de standardisation des données cliniques et du style de vie	3
1. Items du socle	3
1. Standardisation des items "style de vie"	3
2. Standardisation des items "dossier de soins"	4

Introduction

1. Rappel du contexte du groupe de travail

Dans la continuité du groupe de travail (GT) "Ensemble commun de données", qui a établi en 2023 un socle commun aux entrepôts de données hospitaliers (EDSH), le comité stratégique des données de santé a mandaté début 2024 un nouveau groupe de travail pour proposer des standards pour les 51 items qui constituent le socle. Cette étape est essentielle pour permettre son utilisation harmonisée et sa mutualisation.

Les représentants des EDS hospitaliers lauréats de l'appel à projet (AAP) France 2030 soutenant le déploiement des EDSH et d'institutions nationales liées aux données de santé et à l'interopérabilité se sont penchés sur les cinq familles de variables du socle (données démographiques, données du PMSI, données biologiques, données du médicament et données d'examens cliniques) pour identifier les éléments de standardisation les plus pertinents au sein du modèle OMOP et consigner ces choix, arbitrages et préconisations au sein de livrables.

Pour plus d'informations sur le contexte du groupe de travail, consultez la note de synthèse des travaux.

2. Consignes de lecture et mode d'emploi

Méthodologie du groupe de travail "Standardisation et interopérabilité du socle commun aux EDSH"

Le groupe de travail s'est organisé en sous-groupes de réflexion, répartis sur chacune des cinq thématiques du socle commun (identité, biologie, médicaments, PMSI, données cliniques et style de vie).

Chaque sous-groupe s'est d'abord penché sur la standardisation de son ensemble de données, puis les 5 sous-groupes ont restitué leurs propositions à l'ensemble du GT pour validation et pour s'accorder sur les préconisations adéquates compte-tenu des contraintes réelles des EDSH.

Ce livrable final a pour objectif de restituer les choix préconisés par le GT, afin de permettre aux entrepôts de données de santé de mettre en oeuvre le socle commun au sein de leur structure et de le partager de manière standardisée

Il s'accompagne également d'un fichier Excel avec les tables OMOP utilisées pour la standardisation et les variables associées.

Travaux et décisions du GT

Le groupe de travail a choisi le standard OMOP-CDM comme référence afin de spécifier les données du socle commun.

En effet, OMOP est maîtrisé par la majorité des membres du GT et permet de répondre aux objectifs du socle commun :

- Créer des bases standardisées pour réaliser des études communes sur le socle de données,
- Pouvoir interroger toutes les données de la même façon,
- Pouvoir interpréter les résultats de la même manière, peu importe leur EDS d'origine, grâce à une sémantique commune et représentation homogène des données.

En complément des travaux sur le modèle OMOP, un sous-groupe dédié à réfléchir à une manière homogène et opérationnelle de documenter les solutions retenues au format FHIR. Ces travaux étant encore au stade expérimental, ils n'ont pas été intégrés dans ces spécifications, mais sont mis à disposition sur un Github public : [Lien vers le répertoire Github contenant les travaux FHIR..](#)

Notes des rédacteurs et rédactrices

Les livrables du GT "Standardisation et interopérabilité du socle commun aux EDSH" contiennent les choix du groupe de travail quant à la standardisation des données du socle commun aux EDSH uniquement. Ils n'ont pas vocation à couvrir l'ensemble des données des EDSH et pourront évoluer avec le temps. Les porteurs d'EDS restent libres des choix de standard au sein de leurs structures.

Pour toute question ou remarques sur ces documents veuillez vous adresser à l'adresse suivante : contact@health-data-hub.fr.

3. Références et documentation utiles

- [Socle commun défini dans le cadre du groupe de travail « Ensemble des variables communes à tous les entrepôts de données de santé »](#)
- Note de synthèse du groupe de travail « standards et interopérabilité »
- Ressources OMOP-CDM :
 - [Standardized Data: The OMOP Common Data Model - OHDSI](#)
 - [OMOP Common Data Model - Github](#)
 - [Athena – OHDSI Vocabularies Repository](#)
-

Éléments de standardisation des données cliniques et du style de vie

1. Items du socle

Groupe d'items	Intérêt du recueil	Item	Définition de l'item	Valeur attendue	Métadonnées
Dossier de soins	Caractérisation générale du patient et facteurs de risque. Données cliniques non disponibles dans la base principale du SNDS	Taille	Taille du patient telle que relevée à un instant donné lors de sa venue à l'hôpital	Valeur numérique	Date de la mesure Unité de mesure
		Poids	Poids du patient tel que relevé à un instant donné lors de sa venue à l'hôpital	Valeur numérique	Date de la mesure Unité de mesure
		Pression artérielle systolique	Pression artérielle systolique mesurée à un instant donné de la venue du patient à l'hôpital	Valeur numérique	Date et heure de la mesure Unité de mesure
		Pression artérielle diastolique	Pression artérielle diastolique mesurée à un instant donné de la venue du patient à l'hôpital	Valeur numérique	Date et heure de la mesure Unité de mesure
Style de vie		Consommation de tabac	Consommation de tabac par le patient	A définir dans des travaux complémentaires	A définir dans des travaux complémentaires
		Consommation d'alcool	Consommation d'alcool par le patient	A définir dans des travaux complémentaires	A définir dans des travaux complémentaires
		Consommation d'autres drogues	Nom des autres drogues consommées régulièrement le cas échéant	A définir dans des travaux complémentaires	A définir dans des travaux complémentaires
		Activité physique	Activité physique du patient	A définir dans des travaux complémentaires	A définir dans des travaux complémentaires

1. Standardisation des items "style de vie"

Le groupe de travail a restreint ses travaux aux données du dossier de soin, sans inclure d'éléments de standardisation liés aux données de style de vie. En effet, les données de style de vie nécessitent des travaux préalables auprès de sociétés savantes et agences sanitaires afin de définir les données et les terminologies attendues pour ces items du socle.

2. Standardisation des items "dossier de soins"

NB : pour l'ensemble des items, une saisie de texte libre est nécessaire pour compléter / préciser les données remontées.

Pour ces données issues de texte libre il s'agira de demander aux établissements de documenter leurs variables, leurs algorithmes de remontée de valeurs, la méthodologie d'extraction utilisée, l'évaluation de l'algorithme.

Il s'agit entre autres de permettre ainsi la mutualisation des algorithmes et encourager à terme la mise en place d'un algorithme commun.

L'enjeu pour ces items réside notamment dans le rattachement des concepts à la date, aussi le GT préconise-t-il que la date de saisie (date retranscrite dans les données) soit la date de recueil, i.e. "la valeur recueillie (mesurée ou estimée) sur le lieu des soins par un professionnel de santé et reportée en temps réel dans le dossier".

a. Recommandations générales de liaison des informations au séjour

Le GT recommande de **relier les différents items du socle à la notion de "visit"** de la table "visit_occurrence", comprise comme un contact pour un patient avec un établissement de santé :

- Une visit_occurrence commence lors de la venue du patient dans l'établissement de santé (cela exclut donc la planification de sa venue) pour une modalité de prise en charge spécifique (consultation, acte externe, hospitalisation complète, passage aux urgences...).
- Une visit_occurrence se termine lors de la sortie du patient de l'établissement de santé pour la même modalité de prise en charge.

La définition d'établissement de santé reste à la discrétion de chaque structure (FINESS juridique ou géographique).

Ces modalités de définition de la notion de "visit" induisent les conséquences suivantes :

- Plusieurs résumés PMSI peuvent donner une seule visit_occurrence (Résumé d'Unité Médicale, Résumés Hebdomadaires Standardisés, etc)
- Une même venue avec des modalités de prise en charge différentes produisent plusieurs visit_occurrence. Par exemple, un patient venant en consultation puis hospitalisé à la suite de la consultation entraînera la création de 2 "visit_occurrence".

Pour les séjours, le GT recommande aux établissements d'alimenter également la table "visit_detail" afin de **repérer les unités médicales particulières traversées par les patients** (par exemple, réanimation ou psychiatrie), mais **cette alimentation resterait à la discrétion des établissements**, ainsi que la mise en lien avec les tables regroupant les actes, les diagnostics, etc. qui serait laissée à l'initiative des établissements.

La description de la standardisation des items liés au séjour est disponible dans le livrable 02A_GT Standardisation - spécifications - données PMSI.

b. Éléments de standardisation communs à tous les items "dossier de soins"

Valeurs cibles : Pour l'ensemble des items "dossier de soins", le groupe de travail précise que les valeurs retranscrites doivent être **les valeurs recueillies (mesurées ou estimées) sur le lieu des soins par un professionnel de santé et reportées en**

temps réel dans le dossier. Ainsi, la **date de saisie** (date retranscrite dans les données) **doit être la date de recueil.**

A la suite de ce document, sont proposées des sources de données possibles pour les items "dossier de soins". Le groupe de travail souligne que **le respect de la définition des valeurs cibles ci-dessus doit éclairer les entrepôts de données de santé dans leur choix d'inclure ou pas les données issues d'une source ou une autre.**

Le groupe de travail relève également la **priorité des données structurées par rapport aux données issues du texte libre** par traitement automatique du langage naturel. Le texte libre n'est envisagé comme source de données par le groupe de travail que dans des situations limitées détaillées ci-dessous. Afin d'améliorer la performance des algorithmes de traitement automatique du langage et d'accroître leur utilisabilité, le groupe de travail est très favorable aux perspectives de mutualisation et de mise à disposition de ces algorithmes entre les acteurs et les établissements.

Gestion des données invalides : Le groupe de travail souligne le rôle des entrepôts de données de santé dans le filtrage en amont des données jugées invalides, à la lumière des particularités locales d'un entrepôt de données de santé. Ainsi, **si l'entrepôt de données de santé juge probable que l'utilisateur ne soit pas capable d'identifier qu'une donnée est aberrante ou invalide via une exploration de la qualité des données,** alors **cette donnée doit être corrigée ou exclue par l'entrepôt de données de santé en amont.** Un exemple serait une donnée issue d'un formulaire structuré où des champs pour le poids et la taille sont à proximité et mènent à de très nombreuses erreurs par inversion : cette situation ne sera pas connue par l'utilisateur final, donc si la qualité de la donnée est jugée insuffisante par l'entrepôt de données de santé, il est légitime qu'elle soit exclue ou corrigée.

Documentation : Le groupe de travail réitère l'importance pour un entrepôt de données de santé de **fournir une documentation la plus claire possible des items partagés,** afin d'éclairer l'analyse de l'utilisateur. En effet, cela est fondamental lors de l'exploration qualité d'un projet utilisateur, en facilitant l'identification de certains problèmes liés à la qualité des données qui n'auraient pas pu être identifiées en amont par l'entrepôt de données de santé.

c. Items "poids" et "poids de naissance"

i. Description

Poids : poids du patient tel que relevé à un instant donné lors de sa venue à l'hôpital.

- Source des données : **pancarte** (recueil systématisé par les soignants médicaux ou paramédicaux dans un logiciel du système d'information hospitalier)

Poids de naissance : poids du nouveau-né tel que relevé à la naissance

- Source des données : en priorité, **PMSI ou données structurées issues de questionnaires ou formulaires.** En l'absence de données structurées pour le poids de naissance, il est possible d'extraire ces données dans du texte libre (consultations, formulaires non structurés, etc.) via un algorithme de traitement automatique du langage naturel, en documentant la méthodologie d'extraction et son évaluation en matière de complétude et de validité des données.

ii. Standardisation OMOP

Table OMOP utilisée : Measurement

Champs OMOP utilisés : unité de mesure, date de mesure, source des données

Table	Champ	Format	Exemple	Informations complémentaires
Measurement	measurement_id	Integer	1	ID du résultat de biologie (numéro auto-incrément)
Measurement	person_id	Integer	1000000	ID du patient, tel que défini dans la table PERSON
Measurement	measurement_concept_id	Integer	3025315 (29463-7 Body weight)	Concept standard OMOP de l'item mesuré
Measurement	measurement_date	Date	01/01/2024	Date de mesure
Measurement	measurement_datetime	Datetime	01/01/2024 12:00	Date et heure de mesure
Measurement	measurement_time	Varchar(10)	12:00:00	Heure de mesure
Measurement	measurement_type_concept_id	Integer	32836 (EHR physical examination)	Source de laquelle est issue la mesure
Measurement	operator_concept_id	integer	4172703 (=)	En cas de résultat numérique, nature de l'opérateur mathématique (=, >, <...)
Measurement	value_as_number	float	90	Valeur numérique associée au résultat
Measurement	unit_concept_id	integer	9529 (kg)	En cas de résultat numérique,

				CONCEPT STANDARD représentant l'unité
Measurement	visit_occurrence_id	integer	2000000	ID de la venue, tel que défini dans la table VISIT_OCCURRENCE
Measurement	measurement_source_value	varchar(50)	CHU:Pds	CONCEPT représentant le résultat de la mesure, tel que présent dans la source
Measurement	measurement_source_concept_id	integer	3052969862	ID du CONCEPT associé au résultat de la mesure dans la source, tel qu'intégré dans la table CONCEPT
Measurement	unit_source_value	varchar(50)	kilo	Verbatim de l'unité tel que définie dans la source
Measurement	unit_source_concept_id	integer	2000004087	ID du CONCEPT de l'unité de la source, tel qu'intégré dans la table CONCEPT
Measurement	value_source_value	varchar(50)	90 kilos	Verbatim du résultat de la mesure, tel que présent dans la source

Précisions :

- Le groupe de travail propose d'utiliser les concepts standards OMOP de la terminologie LOINC pour le poids et le poids de naissance.
- Le poids peut être rattaché aux concepts standards OMOP de la terminologie LOINC 3013762 - "body weight measured" (dans le cas où le poids est mesuré), 3026600 -

"body weight estimated" (dans le cas où le poids est estimé), ou 3025315 - "body weight" (en l'absence d'information sur le caractère estimé ou mesuré du poids).

- Le poids de naissance peut être rattaché au concept standards OMOP de la terminologie LOINC 3011043 - "birth weight measured".
- Le poids et le poids de naissance peuvent être exprimés en g et/ou kg
- L'ensemble des concepts standards OMOP permettant de saisir le poids et le poids de naissance sont précisés dans l'onglet "Measurement CONCEPT_ID" du document Excel attaché, ainsi que les concepts standards OMOP rattachés à l'unité de mesure.
- Le champ "**measurement_type_concept_id**" permet de retracer la source des données. Il peut prendre plusieurs possibilités (référence) :

Source des données	Measurement_type_concept_id
Pancarte ou équivalent	32836 - EHR physical examination
PMSI	32810 - Claim
Formulaire ou questionnaire (données structurées)	32851 - Healthcare professional filled survey
Texte libre	32858 - NLP

d. Items "taille" et "taille de naissance"

i. Description

Taille : taille du patient telle que relevée à un instant donné lors de sa venue à l'hôpital

- Source des données : **pancarte** (recueil systématisé par les soignants médicaux ou paramédicaux dans un logiciel du système d'information hospitalier)

Taille de naissance : taille du nouveau-né telle que relevée à l'instant de sa naissance

- Source des données : **en priorité, ou données structurées issues de questionnaires ou formulaires**. En l'absence de données structurées pour le poids de naissance, il est possible d'**extraire ces données dans du texte libre** (consultations, formulaires non structurés, etc.) via un algorithme de traitement automatique du langage naturel, **en documentant la méthodologie d'extraction et son évaluation** en matière de performance et de qualité des données.

ii. Stratégie de standardisation OMOP

Table OMOP utilisée : measurement

Champs OMOP utilisés : unité de mesure, date de mesure, source des données

Table	Champ	Format	Exemple	Informations complémentaires
Measurement	measurement_id	Integer	1	ID du résultat de biologie (numéro auto-incrément)
Measurement	person_id	Integer	1000000	ID du patient, tel que défini dans la table PERSON
Measurement	measurement_concept_id	Integer	3036277 (<i>8302-2 Body height</i>)	Concept standard OMOP de l'item mesuré
Measurement	measurement_date	Date	01/01/2024	Date de mesure
Measurement	measurement_datetime	Datetime	01/01/2024 12:00	Date et heure de mesure
Measurement	measurement_time	Varchar(10)	12:00:00	Heure de mesure
Measurement	measurement_type_concept_id	integer	32836 (<i>EHR physical examination</i>)	Source de laquelle est issue la mesure
Measurement	operator_concept_id	integer	4172703 (=)	En cas de résultat numérique, nature de l'opérateur mathématique (=, >, <...)
Measurement	value_as_number	float	1.86	Valeur numérique associée au résultat
Measurement	unit_concept_id	integer	9546 (<i>m</i>)	En cas de résultat numérique, CONCEPT STANDARD représentant l'unité

Measurement	visit_occurrence_id	integer	2000000	ID de la venue, tel que défini dans la table VISIT_OCCURRENCE
Measurement	measurement_source_value	varchar(50)	CHU:Taille	CONCEPT représentant le résultat de la mesure, tel que présent dans la source
Measurement	measurement_source_concept_id	integer	3952969862	ID du CONCEPT associé au résultat de la mesure dans la source, tel qu'intégré dans la table CONCEPT
Measurement	unit_source_value	varchar(50)	m	Verbatim de l'unité tel que définie dans la source
Measurement	unit_source_concept_id	integer	2000005087	ID du CONCEPT de l'unité de la source, tel qu'intégré dans la table CONCEPT
Measurement	value_source_value	varchar(50)	1,86 m	Verbatim du résultat de la mesure, tel que présent dans la source

Précisions :

- Le groupe de travail propose d'utiliser les concepts standards OMOP de la terminologie LOINC pour la taille et la taille de naissance.
- La taille peut être rattachée aux concepts standards OMOP de la terminologie LOINC 3023540 - "body height measured" (dans le cas où la taille est mesurée), 3035463 - "body height estimated" (dans le cas où la taille est estimée), ou 3036277 - "body height" (en l'absence d'information sur le caractère estimé ou mesuré de la taille).
- La taille de naissance peut être rattachée au concept standard OMOP de la terminologie LOINC 36304231 - "Body height Measured --at birth".
- La taille et la taille de naissance peuvent être exprimés en m et/ou cm

- L'ensemble des concepts standards OMOP permettant de saisir la taille et la date de naissance sont précisés dans l'onglet "Measurement CONCEPT_ID" du document Excel attaché, ainsi que les concepts standards OMOP rattachés à l'unité de mesure.
- Le champ "**measurement_type_concept_id**" est la variable permettant de retracer la source des données. Il peut prendre plusieurs possibilités (référence) :

Source des données	Measurement_type_concept_idFormat / terminologies
Pancarte ou équivalent	32836 - EHR physical examination
Formulaire ou questionnaire (données structurées)	32851 - Healthcare professional filled survey
Texte libre	32858 - NLP

e. Items "pression artérielle systolique" et "pression artérielle diastolique"

i. Description

Pression artérielle systolique : pression artérielle systolique mesurée à un instant donné de la venue du patient à l'hôpital, sans préjuger des modalités de prise (type d'appareil, position du patient, etc).

- Source des données : **pancarte** (recueil systématisé par les soignants médicaux ou paramédicaux dans un logiciel du système d'information hospitalier)

Pression artérielle diastolique : pression artérielle diastolique mesurée à un instant donné de la venue du patient à l'hôpital, sans préjuger des modalités de prise (type d'appareil, position du patient, etc)

- Source des données : **pancarte** (recueil systématisé par les soignants médicaux ou paramédicaux dans un logiciel du système d'information hospitalier)

ii. Stratégie de standardisation OMOP

Table OMOP utilisée : Measurement

Champs OMOP utilisés : unité de mesure, date de mesure, source des données

Table	Champ	Format	Exemple	Informations complémentaires
Measurement	measurement_id	Integer	1	ID du résultat de biologie (numéro auto-incrément)
Measurement	person_id	Integer	1000000	ID du patient, tel que défini

				dans la table PERSON
Measurement	measurement_concept_id	Integer	3036277 (8480-6 Systolic blood pressure)	Concept standard OMOP de l'item mesuré
Measurement	measurement_date	Date	01/01/2024	Date de mesure
Measurement	measurement_datetime	Datetime	01/01/2024 12:00	Date et heure de mesure
Measurement	measurement_time	Varchar(10)	12:00:00	Heure de mesure
Measurement	measurement_type_concept_id	integer	32836 (EHR physical examination)	Source de laquelle est issue la mesure
Measurement	operator_concept_id	integer	4172703 (=)	En cas de résultat numérique, nature de l'opérateur mathématique (=, >, <...)
Measurement	value_as_number	float	134	Valeur numérique associée au résultat
Measurement	unit_concept_id	integer	8876 (mm[Hg])	En cas de résultat numérique, CONCEPT STANDARD représentant l'unité
Measurement	visit_occurrence_id	integer	2000000	ID de la venue, tel que défini dans la table VISIT_OCCURRENCE
Measurement	measurement_source_value	varchar(50)	CHU:TAS	CONCEPT représentant le résultat de la

				mesure, tel que présent dans la source
Measurement	measurement_source_concept_id	integer	3952936862	ID du CONCEPT associé au résultat de la mesure dans la source, tel qu'intégré dans la table CONCEPT
Measurement	unit_source_value	varchar(50)	mmHg	Verbatim de l'unité tel que définie dans la source
Measurement	unit_source_concept_id	integer	20000045087	ID du CONCEPT de l'unité de la source, tel qu'intégré dans la table CONCEPT
Measurement	value_source_value	varchar(50)	134 mmHg	Verbatim du résultat de la mesure, tel que présent dans la source

Précisions :

- Le groupe de travail propose d'utiliser les concepts standards OMOP de la terminologie LOINC pour la pression artérielle systolique et diastolique.
- La pression artérielle systolique peut être rattachée au concept standard OMOP de la terminologie LOINC 3004249 - "Systolic blood pressure".
- La pression artérielle diastolique peut être rattachée au concept standard OMOP de la terminologie LOINC 3012888 - "Diastolic blood pressure".
- Les pressions artérielles sont exprimées en mm[Hg] (millimètre de mercure)
- L'ensemble des concepts standards OMOP permettant de saisir la pression artérielle systolique et diastolique sont précisés dans l'onglet "Measurement CONCEPT_ID" du document Excel attaché, ainsi que les concepts standards OMOP rattachés à l'unité de mesure.
- La source de données de la pression artérielle étant uniquement la pancarte (ou équivalent), le champ "**measurement_type_concept_id**" prendra dans ce cas la valeur 32836 - EHR physical examination.
- Les conditions de mesures de la pression artérielle sont rarement spécifiées dans les applications métier. Une seule variable a donc été définie pour la pression artérielle systolique et une seule pour la diastolique. Il revient aux responsables des EDS de ne pas alimenter ces variables avec des valeurs mesurées dans des conditions très particulières. Il est également important que les utilisateurs prennent en compte

dans leurs analyses et leurs interprétations l'absence d'information sur les conditions de mesure et l'hétérogénéité qui peut en résulter