

Spécifications techniques et fonctionnelles

Standardisation et interopérabilité du socle commun

Introduction	2
1. Rappel du contexte du groupe de travail	2
2. Consignes de lecture et mode d'emploi	3
3. Références et documentation utiles	3
Éléments de standardisation des données de biologie du socle commun	1
1. Extrait des items du socle	1
2. Standardisation des items	1

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 des 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 préconisation 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 de biologie du socle commun

1. Extrait des items du socle

Groupe d'items	Intérêt du recueil	Conditions	Item	Définition de l'item	Valeur attendue	Métadonnées
3.a. Fonction rénale	Les données de biologie contribuent au diagnostic, à l'évaluation de la sévérité, de la tolérance des produits de santé, au suivi des patients, etc., et constituent en tant que tels des indicateurs d'intérêt scientifique		Dosage de l'urée dans le sang	Dosage sanguin de l'urée qui permet d'évaluer la fonction rénale	Valeur numérique	Métadonnées LOINC Grandeur Temporalité Milieu Echelle Technique Autres métadonnées Date et heure du prélèvement Unité de mesure Statut de validation Borne inférieure de normalité du résultat Borne supérieure de normalité du résultat
			Créatininémie	Dosage sanguin de la créatinine qui permet d'évaluer la fonction rénale	Valeur numérique	
			Débit de filtration glomérulaire (DFG)	Le débit de filtration par les glomérules rénaux permet d'évaluer la fonction rénale en dehors des épisodes aigus	Valeur numérique	
3.b. Hémogramme	Données cliniques non disponibles dans la base principale du SNDS Les variables de biologie proposées constituent un socle minimal pour fournir une vision de l'état		Leucocytes	Dosage sanguin des globules blancs	Valeur numérique	
			Hémoglobine	Dosage sanguin de l'hémoglobine permettant le diagnostic de l'anémie	Valeur numérique	
			Hématocrite	Volume occupé par les globules rouges	Valeur numérique	

	général du patient			dans le sang, par rapport au volume sanguin total. Utile pour évaluer l'état d'hydratation du patient et poser de nombreux diagnostics		
			Globules rouges	Dosage sanguin des globules rouges	Valeur numérique	
			Volume Globulaire Moyen (VGM)	Volume moyen des globules rouges, qui est une valeur centrale pour le diagnostic des anémies	Valeur numérique	
			Plaquettes	Dosage sanguin des plaquettes, utile pour toute pathologie faisant intervenir le système immunitaire / hématologique	Valeur numérique	
			Neutrophiles	Dosage sanguin des globules blancs de type polynucléaires neutrophiles, utile pour toute pathologie faisant intervenir le système immunitaire / hématologique	Valeur numérique	

			Lymphocytes	Dosage sanguin des globules blancs de type lymphocytes, utile pour toute pathologie faisant intervenir le système immunitaire / hématologique	Valeur numérique	
			Eosinophiles	Dosage sanguin des globules blancs de type éosinophiles, utile pour toute pathologie faisant intervenir le système immunitaire / hématologique	Valeur numérique	
			Monocytes	Dosage sanguin des globules blancs de type monocytes, utile pour toute pathologie faisant intervenir le système immunitaire / hématologique	Valeur numérique	
			Taux de prothrombine (TP)	Le taux de prothrombine permet d'évaluer la coagulation du sang	Valeur numérique	
			Temps de céphaline activée (TCA)	Le temps de céphaline activée permet d'évaluer la coagulation du sang	Valeur numérique	

3.c. Bilan hépatique			Aspartate aminotransférase (AST)	Dosage sanguin d'une enzyme permettant d'évaluer la fonction hépatique	Valeur numérique	
			Alanine aminotransférase (ALT)	Dosage sanguin d'une enzyme permettant d'évaluer la fonction hépatique	Valeur numérique	
			Gamma-glutamyl transférase (GGT)	Dosage sanguin d'une enzyme permettant d'évaluer la fonction hépatique	Valeur numérique	
			Phosphatases alcalines (PAL)	Dosage sanguin d'une enzyme permettant d'évaluer la fonction hépatique ou certaines maladies osseuses ou métaboliques	Valeur numérique	
			Bilirubine totale	Dosage sanguin de la bilirubine, produit de dégradation de l'hémoglobine, permettant d'évaluer la fonction hépatique	Valeur numérique	
			Bilirubine conjuguée	Dosage sanguin de la bilirubine conjuguée, produit de dégradation de l'hémoglobine, permettant d'évaluer la fonction hépatique	Valeur numérique	

3.d. Autres			Glycémie à jeun	Dosage sanguin pour diagnostic du diabète	Valeur numérique	
			Hémoglobine glyquée	Dosage sanguin de l'hémoglobine glyquée, qui est la part de l'hémoglobine qui capte le sucre dans le sang et est un reflet cumulé de la glycémie sur les derniers mois. Permet le diagnostic et le suivi du diabète	Valeur numérique	

2. Standardisation des items

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. Description

Les items de biologie retenus dans le socle commun de données et traités dans le cadre de ce groupe de travail sont liés au bilan hépatique, rénal, hématologique, et glycémique. Le périmètre de ces items se limite aux résultats de biologie (mesure, unité et date de prélèvement) et ne comprend pas les demandes d'examens de biologie.

c. Stratégie de standardisation OMOP générale des items de biologies

Standards

- **Concepts de biologie** : LOINC (et leur équivalent dans OMOP). Le choix des codes à utiliser a été réalisé sur la base de la documentation LOINC, BIOLOINC et ATHENA.
- **Unités** : normalisation des unités en UCUM (*Unified Code for Units of Measure*) et leur équivalent¹ dans OMOP

¹ Pour chaque code issu d'une terminologie standard, un code interne OMOP est disponible et doit, en réalité, être utilisé.

Tables OMOP utilisées :

- Measurement
- Concept

Exemple de l'[Hemoglobin in Blood](#) : le code interne LOINC correspondant à l'hémoglobine en masse par volume est "718-7" (concept code : code issu de la terminologie intégrée dans OMOP). Dans OMOP, le code utilisé pour représenter l'hémoglobine en masse par volume est "3000963" (concept ID : code intégré d'une terminologie standard dans OMOP).

Champs OMOP utilisés :

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	3000963 (718-7 Hemoglobin [Mass/volume] in Blood ***)	Concept STANDARD dans OMOP décrivant le résultat de biologie
Measurement	measurement_date	date	01/01/2024	Date du prélèvement
Measurement	measurement_datetime	datetime	01/01/2024 12:00	Date et heure du prélèvement
Measurement	measurement_time	varchar(10)	12:00:00	Heure du prélèvement
Measurement	measurement_type_concept_id	integer	32817 (EHR)	Source de laquelle est issue le résultat de biologie
Measurement	operator_concept_id	integer	4172703 (=)	En cas de résultat numérique, nature de l'opérateur mathématique (=, >, <...)
Measurement	value_as_number	float	125	Valeur numérique associée au résultat
Measurement	unit_concept_id	integer	8636 (g/L gram per liter)	En cas de résultat numérique, CONCEPT

				STANDARD représentant l'unité
Measurement	range_low	float	120	Borne de normalité inférieure
Measurement	range_high	float	180	Borne de normalité supérieure
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:HbSg	CONCEPT représentant le résultat de biologie, tel que présent dans la source
Measurement	measurement_source_concept_id	integer	3052169862	ID du CONCEPT associé au résultat dans la source, tel qu'intégré dans la table CONCEPT
Measurement	unit_source_value	varchar(50)	g/L	Verbatim de l'unité tel que définie dans la source
Measurement	unit_source_concept_id	integer	3000004087	ID du CONCEPT de l'unité de la source, tel qu'intégré dans la table CONCEPT
Measurement	value_source_value	varchar(50)	125 g/L	Verbatim du résultat de biologie, tel que présent dans la source

Validité des données :

Le groupe de travail recommande d'inclure dans le socle des données de biologie uniquement les résultats jugés comme valides localement par les équipes des EDS.

Le groupe de travail recommande de documenter dans les métadonnées associées au jeu de données si/comment ont été filtrés les résultats le cas échéant. Le format des métadonnées n'a pas été spécifié par le groupe et reste au choix de l'EDS.

i. Éléments de qualité

L'évaluation de la qualité des données de biologie standardisée serait principalement descriptive, et inclurait des informations sur la complétude des données :

- Description des concepts utilisés :
 - Concepts standards disponible dans les données du socle
 - Concepts locaux alignés aux concepts standards du socle
- Description des données intégrées dans le socle :
 - Nombre d'observations au cours du temps
 - Ratio du nombre d'observations / nombre de venues au cours du temps (normalisation par rapport à l'activité)
- Autres éléments de qualité pour chaque item :
 - Description stratifiée par concepts standards d'OMOP, par concept source et par unités
 - Représentation graphique des paramètres centraux au cours du temps

Le groupe de travail met à disposition des acteurs un script d'analyse des données intégrées au langage R pour faciliter et standardiser cette étape (données stockées dans une base de données relationnelle). Ce script permettra de produire des résultats agrégés éclairant l'utilisateur et permettant la comparaison de la qualité des données inter-établissements dans une perspective de partage des données.

Le groupe de travail recommande de mettre à disposition des utilisateurs les éléments liés à la qualité des items de biologie.

ii. Description des concepts de biologie des items du socle :

Pour chaque élément du socle, un élément LOINC a été identifié. Le détail des concepts standards OMOP retenus par le groupe de travail pour chaque item du socle et pour son unité de mesure est détaillé dans le tableau Excel en annexe.