

Note à l'attention du Comité Stratégique du Système National des Données de Santé

Objet : Conclusion et perspectives du groupe de travail "Standardisation et interopérabilité du socle commun"

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'examen 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.

Le groupe de travail a également initié une réflexion sur l'implémentation du socle commun au format FHIR.

Cette première proposition de standardisation pour les données du socle va pouvoir être implémentée et testée dans des conditions réelles au sein des EDSH.

I. La standardisation, une étape nécessaire au déploiement du socle commun au sein des entrepôts de données de santé hospitaliers

Depuis sa création en 2021, le comité stratégique des données de santé construit une stratégie nationale pour améliorer la collecte et favoriser le partage des données relevant du SNDS. Pour mettre en œuvre cette stratégie, le comité a lancé plusieurs groupes de travail. L'un d'entre eux a élaboré en 2023 un socle de données commun aux entrepôts de données de santé hospitaliers dans le but d'harmoniser les démarches, de proposer des priorités dans l'ingestion des données au sein des entrepôts et de permettre de ce fait d'augmenter le potentiel de mutualisation entre les centres.

En lien avec les recommandations formulées lors de la conclusion de ces travaux, un nouveau groupe a été lancé en janvier 2024, pour formuler des orientations en matière de mise en qualité et de standardisation des données du socle. En effet, si une partie des données présentes dans les établissements de santé et leurs EDS est déjà standardisée ou transformée, il n'y a pas aujourd'hui de pratiques communes clairement identifiées et chacun opère selon sa culture interne et les finalités poursuivies. Pour accompagner les nouveaux entrants et poursuivre la dynamique d'harmonisation des démarches dans l'optique notamment de mutualiser les efforts mais également les données, ce groupe de travail constituait donc une étape essentielle.

Il s'agissait notamment de formuler des orientations concernant :

Les modèles de données : ils permettent d'harmoniser les tables, variables, organisation et nommage des données, facilitant ainsi leur manipulation à des fins de recherches, en particulier pour les données de vie réelle dont ce n'est pas l'usage premier. Ces modèles facilitent notamment la réutilisation et le partage de scripts entre établissements et équipes de recherche.

L'alignement des terminologies : il représente également un enjeu majeur pour le partage des données de santé, assurant qu'une variable donnée va être codée de la même manière entre différents établissements. L'existence de nombreux standards sémantiques complexifie l'exercice d'alignement et la mise en place d'un standard international pourrait contribuer à faciliter les choix tout en ouvrant la possibilité d'études transnationales.

C'est grâce aux travaux réalisés sur les alignements que des études multicentriques peuvent être réalisées, aussi bien dans un modèle fédéré que centralisé, à une échelle nationale ou internationale (selon le standard cible retenu).

Étaient exclues les questions relatives à la standardisation du stockage des données au sein des EDS plus spécifiquement.

Constitué de représentants des projets lauréats de l'AAP EDS et d'institutions nationales liées aux données de santé et à l'interopérabilité, le groupe a eu à cœur de produire des recommandations fruit d'un travail collégial fondé sur l'expérience du terrain.

II. Le groupe de travail propose une standardisation des cinq thématiques du socle au format OMOP-CDM et initie une réflexion vers une surcouche de spécifications FHIR

Le groupe constitué a conduit ses travaux en trois temps :

1. Réalisation d'un état des lieux des standards et terminologies utilisés par les établissements partenaires sur le périmètre du socle commun
2. Identification et au choix d'un modèle standard de données
3. Choix des standards et terminologies cibles pour chaque variable du socle

Le choix du modèle standard de données a constitué un premier jalon important des travaux. L'état des lieux a fait émerger deux modèles principaux plébiscités par les membres :

- OMOP-CDM, modèle permettant la standardisation de la structure et de la sémantique, pensé pour les bases de données relationnelles, avec un cadre très structuré et orienté vers les requêtes.
- FHIR, modèle orienté vers le stockage et la distribution des données, centré sur le patient, orienté vers la flexibilité et le partage des données et incluant une démarche très complète de spécification et de documentation.

Ces modèles ont fait l'objet d'une comparaison (voir schéma en annexe), au regard des finalités du socle mais également du niveau de maturité de l'écosystème les concernant.

Après discussions, les membres du GT ont choisi de privilégier le modèle OMOP-CDM comme cadre de référence, car il a été conçu pour réaliser des études épidémiologiques. C'est aussi un standard reconnu à l'échelle européenne et qui est ou sera utilisé dans l'ensemble des projets lauréats de l'AAP EDS. Il apparaît également que le standard OMOP est plus répandu dans les établissements de santé du groupe de travail que le standard FHIR.

Le modèle FHIR présentant également un intérêt certain, le GT a souhaité instruire son utilisation dans le contexte de la standardisation du socle. Plus souple, ce modèle pourrait à terme s'avérer intéressant pour multiplier les usages du socle. Un sous-groupe spécifique a été mis en place pour créer un guide d'implémentation FHIR pour les variables du socle, afin de proposer une méthode et une démarche à suivre. Les travaux réalisés sont mis à

disposition sur un répertoire Gitlab dédié hébergé par l'ANS, afin de que les EDS qui le souhaitent puissent en prendre connaissance.

[Lien vers le répertoire Github contenant les travaux FHIR](#)

Pour formuler des choix de standards et terminologie par item du socle, les membres du groupe se sont répartis en sous-groupes selon les 5 familles de données répertoriées dans le socle : données démographiques, données du PMSI, données biologiques, données du médicament et données d'examens cliniques.

Menés par deux experts de la thématique et de la standardisation, chaque sous-groupe a travaillé en chambre et en parallèle pour fournir au groupe plénier des propositions de standardisation des items.

Les membres du GT devaient aboutir à deux livrables permettant à la fois de déterminer la cible la plus pertinente pour chaque variable et disposer de l'ensemble des spécifications pour implémenter le socle :

- un document de spécifications incluant une description précise des items pour permettre leur bonne identification par les établissements, la liste des tables et champs OMOP sélectionnés pour correspondre à chaque item, les terminologies attendues et le détail des réflexions, problématiques et arbitrages ayant conduit les membres à formuler ces choix.
- un tableau excel, permettant de replacer les variables du socle commun au sein des tables et variables telles qu'elles sont organisées et présentées au sein du modèle OMOP-CDM. Ce tableau reprend la description des tables fournie par OHDSI (Observational Health Data Sciences and Informatics) sur son [répertoire Github](#).

III. La standardisation du socle et sa mise en qualité devront se poursuivre à travers une implémentation en conditions réelles et des travaux complémentaires

S'agissant du modèle de données cible, les travaux du GT se sont majoritairement consacrés au modèle OMOP mais les membres du groupe maintiennent leur intérêt pour le standard FHIR, qui dispose d'un grand nombre de ressources, comme le [guide d'implémentation Fr Core](#), qui pourront nourrir les réflexions sur l'implémentation du socle et permettra d'élargir ses possibilités d'usage. Le standard OMOP est notamment particulièrement adapté pour le multiculturel et l'international.

Toutefois, les travaux menés par le sous-groupe FHIR ont confirmé la nécessité de disposer de connaissances approfondies pour proposer adopter ce modèle. Il s'avère que l'écosystème des EDS et les membres du GT ne maîtrisaient pas encore suffisamment le sujet pour finaliser les spécifications FHIR pour l'ensemble socle dans le temps imparti.

Convaincu de l'intérêt du standard FHIR pour le futur du socle commun et des EDS, le GT réfléchit toutefois à la poursuite de ce travail dans un cadre plus large.

Dans l'ensemble, les standards OMOP et FHIR peuvent coexister. Les établissements de santé peuvent adopter le standard OMOP ou FHIR en fonction de leurs compétences internes. En cas d'adoption du standard OMOP, les établissements sont vivement encouragés à s'appuyer sur les spécifications des livrables GT Standardisation.

S'agissant des standards cibles pour les items, le groupe a pu mener à terme la quasi-totalité de ses travaux et proposer des éléments de standardisation qui ont fait consensus parmi les membres pour les 5 familles de données présentes dans le socle.

Cependant, une partie des données cliniques portait sur le "style de vie" des patients, c'est-à-dire leur consommation d'alcool, de tabac et d'autres drogues. Le GT n'a pas pu

mener à bien la proposition de standards pour ces items dans la mesure où ils s'avéraient trop vagues à ce stade. Pour faire l'objet d'une proposition de standard, ces variables nécessiteront des travaux complémentaires de spécifications, en concertation avec les sociétés savantes et les organismes de santé publique.

Au-delà de la question des standards cibles, la mutualisation des données et leur utilisation requiert de porter une attention particulière à leur mise en qualité. Les membres du GT ont proposé une démarche de mise en qualité pour les données de biologie : il s'agit dans un premier temps d'exécuter sur les bases de données locales des scripts permettant de mesurer l'exhaustivité et la qualité (description des concepts standards et locaux, cohérence du nombre d'observations recueillies, cohérence des données au cours du temps) des données intégrées puis de comparer cette analyse locale avec celle d'autres établissements ayant implémenté le socle. Les porteurs d'EDS pourront ainsi identifier et corriger de potentiels écarts et problématiques, et tendre vers une harmonisation de la collecte et du traitement des données au sein des EDS.

Il serait intéressant d'étendre cette démarche aux autres familles de données et de prendre en compte les travaux par ailleurs réalisés à l'échelle européenne, tels que le projet Quantum qui œuvre à la construction d'un label de qualité des données. Le GT préconise donc la poursuite de cette réflexion dans un groupe de travail dédié.

Enfin, il a semblé aux membres du groupe que pour assurer la pérennité de la mise en œuvre du socle commun aux EDS, il était essentiel de désigner une gouvernance nationale qui serait chargée de maintenir le socle et sa standardisation. Une instance nationale devra établir puis maintenir la cohérence avec les terminologies françaises de référence utilisées au sein du socle mais aussi faire le lien entre le socle et les travaux réalisés à l'échelle européenne et internationale.

Les livrables du GT sont mis à disposition de l'écosystème des entrepôts de données de santé pour permettre aux porteurs d'EDS d'implémenter le socle, mais aussi aux futurs utilisateurs des données du socle commun de mieux comprendre les données qui seront mises à leur disposition. La proposition de standards formulée par le groupe pour les variables du socle commun aux EDS est le fruit d'un important travail collaboratif entre les acteurs de cet écosystème mais doit tout de même se confronter à la mise en œuvre pratique, ce un retour d'expérience étant nécessaire pour confirmer les choix faits par le GT et les faire évoluer si nécessaire. Le Comité Stratégique des Données de Santé et le Health Data Hub fourniront un support opérationnel et proposeront une feuille de route aux établissements qui initieront l'implémentation du socle pour permettre une coordination et une mutualisation des efforts et outils.

Annexe 1 - Livrables du groupe de travail

- Données démographiques
 - Spécifications technico-fonctionnelles
 - Tableau OMOP-CDM
- Données du PMSI
 - Spécifications technico-fonctionnelles
 - Tableau OMOP-CDM
- Données du médicament
 - Spécifications technico-fonctionnelles
 - Tableau OMOP-CDM
- Données de biologie
 - Spécifications technico-fonctionnelles
 - Tableau OMOP-CDM
- Données d'examens cliniques
 - Spécifications technico-fonctionnelles
 - Tableau OMOP-CDM

Annexe 2 - Organisation du groupe de travail

Le groupe s'est réuni à 14 reprises dans le cadre de séances plénières d'une heure trente entre les mois de janvier et septembre 2024 pour déterminer un modèle standard de référence, échanger sur les travaux et arbitrages proposés par les sous-groupes thématiques et travailler à l'élaboration des livrables.

Il est piloté par Vianney Jouhet, Responsable de l'entrepôt de données du CHU de Bordeaux, et Olivier Steichen, Professeur de médecine interne à l'AP-HP, avec le soutien du Health Data Hub et est composé de représentants d'entrepôts de données de santé et d'institutionnels.

Liste des membres

Prénom NOM	Fonction	Organisation
Salam ABBARA	Médecin spécialiste en médecine interne	AP-HP
Jean-Marc ALLIOT	Directeur Scientifique pour les données et l'Intelligence Artificielle	CHU de Toulouse
Guillaume BARTHELEMY	Ingénieur data	GCS ELSAN
Ali BELLAMINE	Médecin de santé publique, Data Scientist	GCS ELSAN
Jean-Christophe BERNADAC	Directeur des Services Numériques	GCS HOURAA - Hospices Civils de Lyon
Olivier BILLUART	Data Manager	Hôpital Saint Joseph Paris

Elsa CECILIA-JOSEPH	Data scientist	CHU de Martinique
Cécile CHEVASSUS-CLEMENT	Chargée de mission Data	GCS G4
Sebastien COSSIN	Praticien hospitalier	GCS NOVA
Thierry DART	Directeur, domaine interopérabilité	ANS
Denis DELAMARRE	Data-ingénieur	GCS HUGO - CHU de Rennes
Catherine DUCLOS	PU-PH de Santé publique	AP-HP
Caroline DUNOYER	Responsable de l'unité Sciences des Données de Santé, SSP	CHU de Montpellier
Alban DUPOUX	Chargé de mission	GCS Grand Est
Hassan FAOUR	Data Architect	Hôpital Saint Joseph Paris
Elisabeth FERY-LEMONNIER	Conseillère médicale	DREES
Laure FEUILLET	Responsable DATA	CHU de Toulouse
Charlotte GEAY-FERRAN	Directrice projet de l'EDS INCLUDE Lille	GCS G4 - CHU de Lille
Pierre-Antoine GOURRAUD	responsable de la clinique des données du CHU de Nantes et membre de la direction scientifique ODH	GCS HUGO
Romain GRIFFIER	Praticien hospitalier	GCS NOVA
Elisabeth HULIER-AMMAR	Directrice de la recherche	Hôpital Foch
Camille JUNG	Cheffe du département hospitalier de territoire DRIIM	CHI Créteil
Benoît LABARTHE	Responsable Accélération Recherche et Transfert de Technologie	Agence de l'Innovation en Santé

Yann LE STRAT	Directeur du département Data Science	Santé Publique France
Kévin LEGUEULT	Département de la Recherche Clinique et de l'Innovation (DRCI)	GIRCI Méditerranée
Pierre-Ephrem MADIOT	Data Analyst	GCS HOURAA - CHU de Grenoble
Laurent MOLINIER	Professeur de santé publique, DIM	CHU de Toulouse
Joris MULLER	PH en Santé Publique au CHU de Strasbourg	CHU de Strasbourg / GCS Grand Est - réseau Paysage
David OUAGNE	Expert interopérabilité, Ingénieur en systèmes d'information et réseaux, PhD informatique médicale et épidémiologie	AP-HP
Samantha PASDELOUP	Directrice Développement et Partenariat	GCS ELSAN
David PELLECUER	Département d'information médicale	CHU de Montpellier
Morgane PIERRE-JEAN	Biostatisticienne	GCS HUGO
Véronique ROBERT	Responsable du Pôle Gestion de la Donnée à la Direction des Data	Unicancer
Mathilde ROUSSEAU	Médecin DIM	CHI Créteil
Christina SADER	Data-manager et Biostatisticienne	CHI Créteil
Sophie SEBILLE	Ingénieur/docteur R&D - Data Scientist	GHU Paris
Audrey SOKOLO-MENAYAMO	Délégué général	GCS HOURAA
Olivier STEICHEN	Professeur de médecine interne	AP-HP
Emmanuelle SYLVESTRE	MCU-PH en Informatique médicale	CHRU Martinique
Pauline TOUCHE	Responsable - Unité Data	Hôpital Foch

Natasha ZANOTTI	Directrice adjointe Stratégie Mutualisation EDS	Unicancer
-----------------	--	-----------

Annexe 3 - Intervention des standards OMOP et FHIR dans le cycle de vie de la donnée

Intervention des standards OMOP et FHIR dans le cycle de vie de la donnée

