

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

Objet : définition d'un socle de données commun aux entrepôts de données de santé hospitaliers

**L'utilisation secondaire des données collectées dans le cadre des soins au sein des établissements de santé à des fins de recherche, d'études, d'évaluations ou encore de pilotage se développe rapidement en France. Cette tendance est notablement renforcée grâce à la mise en place d'entrepôts de données de santé (EDS). En harmonisant les efforts de collecte, de mise en qualité et de standardisation des données hospitalières vers un objectif commun, il devient possible de massifier la collecte de données d'intérêt et de concrétiser la consolidation d'un patrimoine national des données de santé. Cela favorise également le rapprochement de ces données de santé avec la base principale du Système National des Données de Santé.**

**Le groupe de travail mandaté en janvier 2023 par le comité stratégique des données de santé, constitué auprès des ministres chargés de la santé et de la recherche, propose à terme que l'ensemble des EDS hospitaliers du territoire national adopte un socle commun de 51 items. Ce socle permettra de caractériser l'ensemble des séjours et des patients, quels que soient les motifs de leur recours aux soins. Le socle est composé principalement des données les plus couramment collectées par les établissements dans le cadre des soins, notamment les données cliniques, médicamenteuses, les données socio-démographiques, ainsi que les résultats biologiques. Ces données sont généralement collectées de manière structurée et fiable au sein des systèmes d'information hospitaliers.**

**Le groupe de travail formule quatre recommandations spécifiques pour la mise en œuvre du socle :**

- i) établir les standards appropriés pour agréger les données du socle commun ;**
- ii) clarifier le cadre réglementaire de rapprochement entre bases de données, en lien étroit avec la CNIL ;**
- iii) renforcer la formation des acteurs hospitaliers concernant la collecte et la réutilisation de ces données ;**
- iv) clarifier rapidement le financement à long terme des EDS hospitaliers.**

### **I. Le socle de données commun aux entrepôts de données de santé hospitaliers désigne un périmètre prioritaire pour regrouper des données d'intérêt général, quelle que soit l'échelle territoriale**

La loi Organisation et Transformation du Système de Santé, promulguée en juillet 2019, a élargi le périmètre du Système National des Données de Santé à toutes les données de santé financées directement ou indirectement par des fonds publics, notamment les données cliniques collectées dans le cadre du parcours de soins, englobant les activités de prévention, de diagnostic, de soins et de prise en charge sociale ou médico-sociale. La majorité des bases de données hospitalières font ainsi, par définition, partie du SNDS et doivent respecter des règles similaires en termes d'accès, de protection des droits des

personnes, et de respect des exigences de sécurité. Afin de définir une stratégie nationale pour les entrepôts de données de santé (EDS) et d'accompagner le déploiement d'un réseau d'entrepôts hospitaliers sur l'ensemble du territoire français, plusieurs travaux ont été lancés depuis 2022 par le comité stratégique des données de santé, chargé par la loi d'émettre des recommandations afin d'améliorer la collecte et de favoriser le partage des données relevant du SNDS élargi. Issue d'un travail collaboratif, la stratégie nationale des EDS a pour objectif de soutenir une collecte pérenne des données issues du parcours de soins hospitalier et d'en favoriser le partage et la réutilisation. Ce soutien à la construction d'un patrimoine des données de santé hospitalier s'est concrétisé par le lancement en 2022 par le Ministère de la Santé et de la Prévention, dans le cadre de la « Stratégie nationale d'Accélération Santé Numérique » et du plan France 2030, d'un appel à projets (AAP) de 75M€ permettant d'accompagner la construction et la consolidation d'entrepôts de données de santé hospitaliers sur l'ensemble du territoire français. Les lauréats de cet appel à projet sont des acteurs clés de la construction de la stratégie nationale des EDS.

Le comité stratégique des données de santé (constitué auprès des ministres chargés de la santé et de la recherche) a missionné un groupe de travail chargé d'identifier les axes prioritaires d'alimentation des entrepôts de données de santé hospitaliers à l'échelle nationale. En effet, converger vers un plan d'intégration harmonisé de données prioritaires garantira l'établissement d'un socle commun de données, pouvant servir de cible à atteindre pour l'ensemble des EDS hospitaliers. Opérationnellement, définir un socle de données commun permet de dresser les priorités pour la collecte, la mise en qualité, la standardisation et la documentation des données, tout en assurant une meilleure disponibilité des données au sein des EDS hospitaliers. En tant que pierre angulaire de la constitution d'un réseau national interopérable d'EDS, le socle favorisera la mutualisation des données entre établissements et leur collecte à grande échelle pour encourager leur réutilisation dans des projets nouveaux. Il facilitera également le rapprochement des données hospitalières avec d'autres grandes bases de données nationales, telle que la base principale du SNDS, grâce à son inclusion progressive au catalogue du SNDS.

La collecte, la mise en qualité et la standardisation d'un périmètre de données commun à l'ensemble des EDS hospitaliers ouvrent de nouvelles perspectives d'utilisation de ces données. De l'échelon local à une vision globale du système de santé hospitalier, ce socle permettra la réalisation d'études, d'évaluations ou de recherches jusqu'ici inaccessibles. En tant que référence généraliste, ce socle soutiendra également la gestion des établissements de santé, l'évaluation des parcours et des pratiques de soins en offrant une vision commune des patients et de leur prise en charge à l'hôpital. La disponibilité de données cliniques, biologiques et thérapeutiques à différentes échelles, revêt également un intérêt majeur pour la surveillance et la sécurité sanitaire. A long terme, ce socle pourra servir d'interface avec d'autres initiatives d'EDS, comme l'Entrepôt National des Données de Biologie médicale, les entrepôts de médecine de ville, ainsi qu'avec la base principale du SNDS, afin de construire une description complète du patient et de son parcours, intégrée entre la ville et l'hôpital, quelle que soit la raison de son recours aux soins. Enfin, le déploiement opérationnel du socle marque la constitution d'un réseau national d'EDS, une ressource propice au développement de la recherche et de l'innovation en santé, contribuant ainsi à faire progresser la France vers son ambition de devenir un "leader" international en Santé Numérique.

Composé de variables usuellement collectées dans le cadre des soins hospitaliers et couramment recueillies dans les EDS pour une utilisation secondaire, ce socle commun généraliste complète d'autres socles proposés pour certaines maladies ou populations par

des acteurs spécialisés<sup>1</sup>. Ce socle commun généraliste est destiné à évoluer en fonction avec une maturité croissante des EDS hospitaliers et l'émergence de nouveaux cas d'usage prioritaires, intégrant de nouvelles caractéristiques essentielles des parcours de soins et de prévention.

La première version du socle commun aux EDS hospitaliers, présentée en **annexe 1**, résulte d'un travail collaboratif mené sur une période de six mois depuis le début de l'année 2023. Les modalités de fonctionnement du groupe de travail sont détaillées en **annexe 2**. Ces travaux ont débuté par une consultation de l'écosystème élargi des dépositaires et des utilisateurs de données de santé afin de recueillir leurs besoins et leurs attentes en matière de données prioritaires à collecter au sein des EDS hospitaliers. Forts des conclusions de cette consultation, les membres du groupe de travail ont élaboré un socle initial dont le périmètre et la faisabilité de déploiement opérationnel ont été discutés avec les six premiers lauréats de l'AAP EDS, qui ont été conviés à rejoindre le groupe de travail dès leur désignation en avril 2023.

## **II. Le groupe de travail propose un socle commun de 51 items répondant aux enjeux stratégiques de développement du réseau d'EDS hospitaliers et adapté aux limites de collecte actuelles pour permettre une mise en œuvre efficiente dans chaque établissement**

Dans le but d'engager efficacement les travaux d'intégration du socle au sein des EDS hospitaliers, celui-ci se divise en deux lots de données.

Le **premier lot** est composé d'items dont l'intégration est estimée comme prioritaire. Il s'agit de données essentiellement recueillies dans le cadre des soins, et dont la collecte au sein des systèmes d'information des hôpitaux est généralement structurée et fiable. Ce premier segment du socle comprend :

- Des données du PMSI<sup>2</sup> collectées localement par les hôpitaux. Bien qu'elles soient déjà disponibles et consolidées à l'échelle nationale *via* la base principale du SNDS, le groupe de travail propose de capitaliser sur la collecte et les mécanismes de mise en qualité du PMSI déjà en place au sein des établissements. L'objectif est d'obtenir une caractérisation globale des patients et de leur séjour à l'hôpital au sein du socle. Le groupe de travail recommande donc l'inclusion d'un socle minimal de 9 items génériques collectés par les recueils PMSI dans les quatre "champs de soins de santé"<sup>3</sup>.

Initialement recueillies à des fins médico-administratives, la qualité des données du PMSI en fait désormais une source d'information largement utilisée pour la gestion des établissements, l'évaluation des parcours de soins, ainsi que pour la recherche. Par conséquent, la collecte des données du PMSI doit être durable et indépendante du modèle de financement des hôpitaux.

- Des données socio-démographiques. Ces données complètent la caractérisation générale des patients par des données administratives qui ne sont pas collectées dans le cadre du PMSI. Parmi ces données, figurent les identifiants nationaux tels que le NIR, l'INS ou l'adresse géocodée. En tant que clés d'appariement direct, elles

---

<sup>1</sup> L'audition de l'INCa par le groupe de travail a souligné les efforts en cours sur la structuration du dossier communicant de cancérologie dans le cadre des soins et dont les données permettent d'alimenter un socle minimum de données de vie réelle OSIRIS-RWD. D'autres initiatives comme le *minimum set* de la BNMDR pour le suivi des maladies rares ou la liste exhaustive de variables dédiées au suivi du VIH/SIDA sont également à souligner.

<sup>2</sup> PMSI : Programme de médicalisation des systèmes d'information

<sup>3</sup> Le PMSI se compose de 4 champs : médecine, chirurgie, obstétrique et odontologie (MCO) ; soins de suite ou de réadaptation (SSR) ; hospitalisation à domicile (HAD) ; psychiatrie (PSY).

revêtent une importance primordiale pour rendre plus fiable et exhaustif le croisement des données cliniques collectées à l'hôpital avec d'autres bases de données nationales, notamment la base principale du SNDS, ainsi que des bases environnementales ou sociales (bases de données de la statistique publique). La mise en relation de ces données permettrait de promouvoir l'analyse des déterminants sociaux ou environnementaux de la santé, mesurés au moins à l'échelon collectif, sinon à un échelon individuel.

Actuellement, le traitement de ces données sensibles à des fins d'appariement suscite des interrogations légitimes pour les EDS conformes au référentiel sur les entrepôts de données de santé établi par la CNIL qui ne les autorise, ni ne les interdit explicitement.

- Les données de prescription et d'administration permettant de caractériser l'exposition médicamenteuse des patients. Ces données sont essentielles pour décrire les maladies et pour évaluer le rôle des médicaments comme déterminant de santé. Ces données répondent à un besoin exprimé des utilisateurs pour améliorer la qualité des études en vie réelle en complétant les données médico-administratives de prescription et de délivrance en ville, disponibles dans la base principale du SNDS. Disposer d'une vision globale de l'exposition médicamenteuse des patients sur l'ensemble de leur parcours de soins vise à améliorer l'évaluation des prises en charge des patients.

Le socle inclut les données relatives aux prescriptions et à l'administration des médicaments dans le milieu hospitalier quel que soit le type de traitement. A terme, le socle commun consolidera l'ensemble des informations concernant l'exposition médicamenteuse individuelle des patients à l'hôpital, ce qui constitue une évolution très favorable pour cette source d'information.

- Des résultats d'examens de biologie médicale. Élément majeur de la prise en charge des patients, contribuant au diagnostic, à apprécier la gravité ou le pronostic d'une maladie ou encore à évaluer la tolérance des produits de santé, les résultats des examens de biologie médicale ont un intérêt scientifique majeur. Les résultats de certains examens de biologie médicale réalisés à l'hôpital complètent les données relatives aux actes de dépistage et de diagnostic disponibles dans le PMSI. Étant donné la grande diversité des examens réalisables et le volume de données généré, le groupe de travail propose un socle minimal de 23 items, correspondant aux examens de routine, offrant une vision globale de l'état de santé du patient. Ces items incluent l'hémogramme, une caractérisation des fonctions rénale et hépatique et du bilan glycémique.
- Des données issues de l'examen clinique des patients. Des pratiques hétérogènes de collecte, privilégiant notamment une collecte de l'information au sein des comptes-rendus plutôt que dans les champs structurés alloués dans les logiciels métier, complique la collecte standardisée et leur utilisation secondaire. Le groupe de travail a néanmoins souhaité soutenir la collecte d'un nombre limité et prioritaire de données cliniques (taille, poids, pression artérielle) qui ne sont disponibles par ailleurs dans aucune grande base de données nationale en dehors d'enquêtes ponctuelles en population générale.

Un **second lot** de données comprend quatre items relatifs à des comportements à risque pour la santé, mesurés à l'échelle individuelle : consommation de tabac, consommation d'alcool, autres addictions, activité physique. Faisant partie des bonnes pratiques de soins et essentielles à la conduite d'études et d'évaluations dans le domaine de la santé, le recueil de ces données au sein des EDS reste aujourd'hui insuffisamment structuré pour une utilisation secondaire. Par ailleurs, selon la finalité de leur utilisation, les informations concernant ces comportements à risque pour la santé peuvent être collectées de différentes manières, que

ce soit à l'échelon des unités médicales, des établissements de santé ou à l'échelle nationale. Cela rend l'interprétation de ces informations fortement dépendante du contexte de leur collecte. Leur collecte structurée et pérenne nécessite donc des travaux complémentaires, notamment en s'appuyant sur des définitions utilisées dans de grandes enquêtes européenne (par exemple, EHIS) ou nationale (par exemple, baromètre de Santé publique France), et en promouvant les travaux de comparaison, voire de généralisation des résultats issus des EDS hospitaliers, à la population française. Ce lot ne pourra donc pas être mis en œuvre simultanément avec la première partie du socle.

Enfin, le socle commun se compose d'informations généralement structurées à la source, mais les difficultés d'extraction de données depuis les systèmes d'information hospitaliers devront être considérées. En effet, certains éditeurs, sur la base de considérations liées aux droits de copyright et aux investissements à réaliser, choisissent de restreindre l'accès aux données, ce qui peut rendre l'extraction particulièrement complexe ou coûteuse pour les établissements, tout en limitant leur réutilisation au sein des EDS.

### **III. Les établissements de santé devront être accompagnés dans la mise en place du socle, le maintien et l'évolution du socle afin de consolider dans la durée ce patrimoine national des données de santé : quatre recommandations du groupe de travail**

Les travaux menés au cours du premier semestre 2023, notamment grâce au retour d'expérience partagé par plusieurs acteurs internationaux<sup>4</sup>, ont souligné l'importance de collecter un ensemble de données identiques tout en définissant des terminologies communes pour restituer l'information, et en déployant un système robuste de contrôle et d'amélioration continue de la qualité des données. S'accorder sur une standardisation des données et sur des méthodes communes de contrôle de leur qualité renforcera l'interopérabilité des informations et leur consolidation à différentes échelles.

Implémenter le socle commun au sein des établissements est donc un travail conséquent qui nécessitera un accompagnement approfondi et pérenne des acteurs de terrain. Dans le cadre de l'Appel à Projets (AAP) EDS, le Health Data Hub accompagnera les établissements lauréats pour élaborer, au cours des prochaines semaines, une *feuille de route* d'implémentation du socle et de son partage au catalogue du SNDS, adaptée à la maturité et au projet d'EDS porté par chaque établissement.

Pour garantir la réussite du dispositif proposé et, de manière générale, le déploiement du socle au sein de l'ensemble des EDS selon la *feuille de route* proposée, le groupe de travail formule les recommandations suivantes :

- Établir les standards appropriés pour les données du socle et définir les modalités de contrôle qualité des données collectées afin de garantir leur agrégation à différentes échelles. Le groupe de travail recommande donc la réalisation de travaux complémentaires visant à préciser les spécifications techniques des formats de données et des terminologies essentielles à la mise en œuvre pratique d'un standard commun. Ces travaux pourront utilement être accompagnés d'une *phase pilote*, opérationnelle visant à qualifier l'état initial d'implémentation de ce standard au sein des EDS lauréats de l'AAP EDS. Cette démarche permettra d'améliorer la cartographie initiale du socle en évaluant la qualité des données collectées, en définissant les

---

<sup>4</sup> Le groupe de travail a auditionné en juillet 2023 la Danish Health Data Authority et le réseau d'EDS hospitaliers japonais MID-NET.

besoins généraux des EDS et en favorisant le partage d'expérience entre les parties prenantes ;

- En lien étroit avec la CNIL, clarifier le cadre réglementaire relatif au rapprochement direct au niveau individuel, entre les bases de données de santé et d'autres bases nationales ou locales, notamment sociales ou environnementales. Cette clarification vise à faciliter la réalisation d'études et de recherches intégrant un grand nombre de déterminants de santé (position sociale, conditions de vie, de logement, expositions physico-chimique, etc.) ;
- Renforcer la formation de l'ensemble des acteurs hospitaliers à l'importance de la qualité des données de santé, et du recueil de ces données (exactitude, précision, complétude), notamment pour améliorer la pratique des soins, et pour leur utilisation secondaire. Une réflexion sur le choix et la mise en place d'outils ergonomiques à l'interface humain-machine pour la collecte simplifiée et efficiente des données doit être envisagée, de même que pour la réutilisation de ces données. La réflexion sur ces enjeux doit également couvrir le périmètre de la formation initiale et continue de ces professionnels afin de favoriser l'émergence de profils multidisciplinaires disposant de compétences pour la mise en place et l'exploitation des données d'un réseau pérenne d'EDS hospitaliers ;
- Clarifier rapidement le financement pérenne des EDS hospitaliers. Ce financement doit couvrir les coûts d'implémentation, de maintenance, d'évolution et de partage des données du socle commun sur la durée. En effet, la conduite de ces opérations dans la durée ne saurait reposer exclusivement sur le financement obtenu suite à l'AAP EDS.

## Annexe 1 - Socle commun des EDS hospitaliers

Le tableur présenté ci-dessous détaille le contenu du socle commun des EDS hospitaliers, composé de 51 items d'intérêt. Un travail complémentaire devra être proposé afin d'assurer l'interopérabilité du socle.

Groupe d'items	Intérêt du recueil	Lot	N° de l'item	Points d'attention sur la faisabilité du recueil	Condition(s)	Item	Définition de l'item	Valeur attendue	Métadonnées
1. Données socio-démographiques	1.a. Identité patient	1	1.a.1		Au moins un identifiant attendu	Numéro d'inscription au Répertoire (NIR)	Numéro unique attribué à chaque personne à sa naissance sur la base d'éléments d'état civil transmis par les mairies à l'INSEE	Chaîne de caractères	N/A
		1	1.a.2			Identité Nationale de Santé (INS)	Numéro d'identité unique, pérenne, partagée par l'ensemble des professionnels du monde de la santé	Chaîne de caractères	N/A
		1	1.a.3			Date de naissance	Date de naissance des papiers d'identité utilisés pour la production de l'INS	Date	N/A
		1	1.a.4			Date de décès	Date de décès à l'hôpital, ou date de décès collectée par chaînage avec une base externe comme l'INSEE ou le CepiDc	Date	Source de la donnée
		1	1.a.5			Rang gémellaire du bénéficiaire	Pour le régime général, il permet de distinguer les naissances gémellaires de même sexe	Chaîne de caractères	N/A
	1.b. Environnement	Clé d'appariement avec les bases de données environnementales	1	1.b.1			Adresse géocodée	Coordonnées géographiques (latitude et longitude) de l'adresse du patient	Coordonnées géographiques
1			1.b.2		Si adresse géocodée (1.b.1) non disponible	IRIS de l'adresse	Codage de l'adresse du patient selon la méthode d'îlots Regroupés pour l'Information Statistique de l'INSEE	Chaîne de caractères	Date du recueil de l'information

		et d'études liées aux déterminants de santé							
2. PMSI	Description générale du patient et de la venue à l'hôpital, dans les différents PMSI (MCO, SSR, HAD, Psy)	1	2.1		Si date de naissance (1.a.3) ou date de début du séjour (2.8) non disponibles	Age	Age lors de la venue à l'hôpital tel que défini dans le PMSI	Valeur numérique	Date du recueil de l'information
		1	2.2			Sexe	Sexe du patient figurant sur les papiers d'identité utilisés pour la production de l'INS	Chaîne de caractères	N/A
		1	2.3			Code géographique de résidence	Code de la commune de résidence du patient selon la classification PMSI	Chaîne de caractères	Date du recueil de l'information
		1	2.4			Tous les diagnostics selon la classification CIM-10 à l'échelle du séjour	L'ensemble des diagnostics (qu'il s'agisse de diagnostics principaux, diagnostics reliés, ou diagnostics associés significatifs) selon la classification CIM-10, à l'échelle du séjour	Chaîne de caractères	N/A
		1	2.5			Actes	Ensemble des actes médicotechniques réalisés par les médecins, selon la classification utilisée dans chaque PMSI	Chaîne de caractères	Date du recueil de l'information
		1	2.6			Date de début du séjour	Date de début du séjour	Date	N/A
		1	2.7			Date de fin du séjour	Date de fin du séjour	Date	N/A
		1	2.8			Mode d'entrée du séjour	Mode d'entrée pendant le séjour ; dans le cas d'un séjour avec passage dans plusieurs unités médicales, il s'agit du mode d'entrée dans la première unité médicale dans l'ordre chronologique pendant le séjour	Valeur numérique	Libellé
		1	2.9			Mode de sortie du séjour	Mode de sortie pendant le séjour ; dans le cas d'un séjour avec passage dans plusieurs unités médicales, il s'agit du mode de sortie de la dernière unité médicale dans l'ordre chronologique pendant le séjour	Valeur numérique	Libellé

3. Résultats d'examen biologiques	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	1	3.a.1			Dosage de l'urée dans le sang	Dosage sanguin de l'urée qui permet d'évaluer la fonction rénale	Valeur numérique	<b>Métadonnées LOINC</b> Grandeur Temporalité Milieu Echelle Technique  <b>Autres métadonnées</b> Date et heure du prélèvement Unité de mesure Statut de validation Borne inférieure de normalité du résultat Borne supérieure de normalité du résultat
			1	3.a.2			Créatininémie	Dosage sanguin de la créatinine qui permet d'évaluer la fonction rénale	Valeur numérique	
			1	3.a.3			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	1	3.b.1			Leucocytes	Dosage sanguin des globules blancs	Valeur numérique	
			1	3.b.2			Hémoglobine	Dosage sanguin de l'hémoglobine permettant le diagnostic de l'anémie	Valeur numérique	
			1	3.b.3			Hématocrite	Volume occupé par les globules rouges dans le sang, par rapport au volume sanguin total. Utile pour évaluer l'état d'hydratation du patient et poser de nombreux diagnostics	Valeur numérique	
			1	3.b.4			Globules rouges	Dosage sanguin des globules rouges	Valeur numérique	
			1	3.b.5			Volume Globulaire Moyen (VGM)	Volume moyen des globules rouges, qui est une valeur centrale pour le diagnostic des anémies	Valeur numérique	
			1	3.b.6			Plaquettes	Dosage sanguin des plaquettes, utile pour toute pathologie faisant intervenir le système immunitaire / hématologique	Valeur numérique	
			1	3.b.7			Neutrophiles	Dosage sanguin des globules blancs de type polynucléaires neutrophiles, utile pour toute pathologie faisant intervenir le système immunitaire / hématologique	Valeur numérique	
1	3.b.8			Lymphocytes	Dosage sanguin des globules blancs de type lymphocytes, utile pour toute pathologie faisant intervenir le système immunitaire / hématologique	Valeur numérique				
1	3.b.9			Eosinophiles	Dosage sanguin des globules blancs de type éosinophiles, utile pour toute pathologie faisant intervenir le système immunitaire / hématologique	Valeur numérique				
1	3.b.10			Monocytes	Dosage sanguin des globules blancs de type monocytes, utile pour toute pathologie faisant intervenir le système immunitaire / hématologique	Valeur numérique				

3.c. Bilan hépatique	général du patient	1	3.b.11			Taux de prothrombine (TP)	Le taux de prothrombine permet d'évaluer la coagulation du sang	Valeur numérique	
		1	3.b.12			Temps de céphaline activée (TCA)	Le temps de céphaline activée permet d'évaluer la coagulation du sang	Valeur numérique	
		1	3.c.1			Aspartate aminotransférase (AST)	Dosage sanguin d'une enzyme permettant d'évaluer la fonction hépatique	Valeur numérique	
		1	3.c.2			Alanine aminotransférase (ALT)	Dosage sanguin d'une enzyme permettant d'évaluer la fonction hépatique	Valeur numérique	
		1	3.c.3			Gamma-glutamyl transférase (GGT)	Dosage sanguin d'une enzyme permettant d'évaluer la fonction hépatique	Valeur numérique	
		1	3.c.4			Phosphatases alcalines (PAL)	Dosage sanguin d'une enzyme permettant d'évaluer la fonction hépatique ou certaines maladies osseuses ou métaboliques	Valeur numérique	
		1	3.c.5			Bilirubine totale	Dosage sanguin de la bilirubine, produit de dégradation de l'hémoglobine, permettant d'évaluer la fonction hépatique	Valeur numérique	
		1	3.c.6			Bilirubine conjuguée	Dosage sanguin de la bilirubine conjuguée, produit de dégradation de l'hémoglobine, permettant d'évaluer la fonction hépatique	Valeur numérique	
		1	3.d.1			Glycémie à jeun	Dosage sanguin pour diagnostic du diabète	Valeur numérique	
		1	3.d.2			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	
4. Exposition médicale	L'exposition médicale est essentielle pour la caractérisation et l'étude des pathologies médicales et de leurs	1	4.1			Médicament prescrit	Libellé du médicament prescrit	Chaîne de caractères	Libellé UCD Libellé DCI Libellé commercial Code UCD Code ATC Code CIP 13 Voie d'administration

	facteurs de risque, l'évaluation et l'amélioration de leur prise en charge, ainsi que pour les études (notamment de pharmacovigilance) en vie réelle.  Données cliniques non disponibles dans la base principale du SNDS	1	4.2	Terminologies de recueil de l'information en cours de définition par l'ANS (livraison prévue en 2023)		Posologie	Posologie quantitative de la prescription	Chaîne de caractères	Métadonnées en cours de définition par l'ANS (livraison fin 2023) Date de début de la prescription Date de fin de la prescription	
		1	4.3	Si le recueil n'est pas prévu dans les systèmes d'information source : l'information peut-être collectée par un intermédiaire comme la dispensation		Médicament administré	Libellé du médicament administré	Chaîne de caractères	Libellé UCD Libellé DCI Libellé commercial Code UCD Code ATC Code CIP 13 Voie d'administration	
		1	4.4	Si le recueil n'est pas prévu dans les systèmes d'information source : l'information peut-être collectée par un intermédiaire comme la dispensation		Dosage	Dosage du médicament administré par prise	Valeur numérique	Quantité Unité Date et heure de début de l'administration Date et heure de fin d'administration	
5. Examen clinique	5.a Dossier de soins	Caractérisation générale du patient et facteurs de risque. Données cliniques non disponibles dans la base principale du SNDS	1	6.a.1	Telle que collectée dans les systèmes d'information source		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
			1	6.a.2	Tel que collecté dans les systèmes d'information source		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
			1	6.a.3	Telle que collectée dans les systèmes d'information source		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
			1	6.a.4	Telle que collectée dans les systèmes d'information source		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

5.b Style de vie	2	6.b.1	Informations dont les pratiques de collecte sont actuellement très hétérogènes et nécessitant des travaux complémentaires notamment de définition.		Consommation de tabac	Consommation de tabac par le patient	A définir dans des travaux complémentaires	A définir dans des travaux complémentaires
	2	6.b.2	Informations dont les pratiques de collecte sont actuellement très hétérogènes et nécessitant des travaux complémentaires notamment de définition.		Consommation d'alcool	Consommation d'alcool par le patient	A définir dans des travaux complémentaires	A définir dans des travaux complémentaires
	2	6.b.3	Informations dont les pratiques de collecte sont actuellement très hétérogènes et nécessitant des travaux complémentaires notamment de définition.		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
	2	6.b.4	Informations dont les pratiques de collecte sont actuellement très hétérogènes et nécessitant des travaux complémentaires notamment de définition.		Activité physique	Activité physique du patient	A définir dans des travaux complémentaires	A définir dans des travaux complémentaires

## Annexe 2 - Modalités de fonctionnement du groupe de travail

Le groupe de travail s'est réuni à 27 reprises dans le cadre de séances d'une heure et demie entre les mois de janvier et septembre 2023 afin de proposer un socle de données commun à l'ensemble des EDS hospitaliers. Il est piloté par le Pr Geneviève Chêne (CHU et Université de Bordeaux, INSERM) et le Dr Cyrille Delpierre (INSERM). Le groupe de travail est composé de membres aux expertises variées et complémentaires sur les données de santé et de représentants des projets lauréats de la première vague de l'appel à projet d' "accompagnement et soutien à la constitution d'entrepôts de données de santé hospitaliers" lancé par le Ministère de la Santé et de la Prévention en 2022. La composition du groupe de travail est détaillée ci-après.

### **Organisation des travaux**

Les travaux se sont agencés autour de 2 phases principales :

1. Consultation de l'écosystème afin d'établir un état des lieux du recueil des données de santé hospitalières et des attentes des utilisateurs de données de santé
2. Construction itérative du socle commun avec les acteurs de l'écosystème et notamment avec les projets lauréats de la première vague de l'AAP EDS

### **Auditions réalisées**

L'état des lieux initial et la construction itérative du socle commun ont mené le groupe de travail à conduire un ensemble d'auditions avec les acteurs de l'écosystème des données de santé :

- 22 février 2023 : Audition de la Clinique des Données du CHU de Nantes
  - Personne auditionnée :
    - Pr Pierre-Antoine Gourraud, PU-PH et Responsable de la Clinique des Données du CHU de Nantes
- 1er mars 2023 : Audition de l'entrepôt de données de santé du CHU de Bordeaux
  - Personnes auditionnées :
    - Pr Rodolphe Thiébaud, PU-PH et Chef du service d'Information Médicale du CHU de Bordeaux
    - Dr Vianney Jouhet, PH, responsable de l'EDS du CHU de Bordeaux
- 15 mars 2023 : Audition de l'Institut National du Cancer
  - Personne auditionnée :
    - Dr Philippe-Jean Bousquet, Directeur de l'Observation, des Sciences des données et de l'Évaluation de l'INCa
- 22 mars 2023 : Audition du GCS HUGO dans le cadre d'un retour d'expérience sur le projet HUGOSHARE
  - Personnes auditionnées :
    - Pascal Van Hille, Coordinateur Technique du Réseau interrégional des Centres de Données Cliniques du GIRCI Grand Ouest au CHU de Rennes
    - Denis Delamarre, Ingénieur bases de données au CHU de Rennes
- 29 mars 2023 :
  - Audition de la Direction de la recherche, des études, de l'évaluation et des statistiques du Ministère de la Santé et de la Prévention
    - Personne auditionnée
      - Benoît Ourliac, Sous-directeur Observation de la santé et de l'assurance maladie à la DREES

- Audition de Santé publique France
  - Personnes auditionnées
    - Yann Le Strat, Directeur appui, traitements et analyses des données de Santé publique France
    - Sahar Haidar, Chargée d'études scientifiques à Santé publique France
- 5 avril 2023 : Audition de la Haute Autorité de Santé
  - Personnes auditionnées
    - Camille Thomassin, Responsable de la cellule de coordination des données de vie réelle à la HAS
    - Matthieu Doutreligne, chargé de projet data science à la HAS
- 12 avril 2023 : Audition d'EPI-PHARE
  - Personne auditionnée
    - Pr Mahmoud Zureik, Directeur d'EPI-PHARE
- 3 mai 2023 : Audition de l'Institut Thématique de Santé publique de l'INSERM
  - Personne auditionnée
    - Pr Rémy Slama, Directeur de l'Institut Thématique de Santé Publique de l'INSERM
- 24 mai 2023 : Audition du projet *Platform for Data in Primary Care* (P4DP) sur le lien des EDSH avec les EDS de médecine de ville
  - Personnes auditionnées
    - Pr David Darmon, Directeur du département d'enseignement et de recherche en médecine générale de l'Université Côte d'Azur
    - Pr Olivier Saint-Lary, Président du Collège National des Généralistes Enseignants (CNGE)
- 31 mai 2023 : Audition de l'Agence technique de l'information sur l'hospitalisation
  - Personnes auditionnées
    - Sophie Guéant, Responsable du pôle d'information médicale à l'ATIH
    - Etienne Joubert, économiste de la santé au pôle d'information médicale à l'ATIH
- 11 juillet 2023 :
  - Audition du réseau *MID-NET* : réseau d'une vingtaine d'EDS hospitaliers constitué par la *Pharmaceuticals and Medical Devices Agency* japonaise
    - Personnes auditionnées
      - Azumi Takano, M.S., Coordinator, Office of Medical Informatics and Science, Pharmaceuticals and Medical Devices Agency (PMDA)
      - Kazuishi Sekino, Office of Medical Informatics and Science, Pharmaceuticals and Medical Devices Agency (PMDA)
  - Audition de la Danish Health Data Authority
    - Personne auditionnée
      - Henrik Jensen, conseiller à la classification et au recueil des données à la DHDA
- 12 septembre 2023 : Audition d'acteurs économiques
  - Personnes auditionnées
    - Jean-Baptiste Angeloglou, Directeur général du groupe PSIH
    - Isabelle Delarozière, Responsable Données de santé en vie réelle et Partenariats du groupe PSIH
    - Armelle Graciet, Directrice des affaires industrielles du Snitem
    - Yoan Jacquemin, Manager Business Intelligence Advisor du groupe PSIH
    - Magali Lemaitre, Vice-présidente de l'AFCROs

- Nathalie Manaud, Directrice innovation du Leem

### **Composition du groupe de travail**

Les travaux menés sur la constitution du socle commun aux EDS hospitaliers ont débuté en janvier 2023 au sein d'un groupe de travail restreint composé de 8 membres et 4 invités permanents. Le groupe s'est ensuite élargi le 5 avril 2023, suite à l'annonce survenue le 3 avril 2023 des projets lauréats de la première vague de l'AAP EDS, pour accueillir des représentants de chacun de ces projets. La liste des membres et invités permanents du groupe de travail, ainsi que des éléments de biographie se trouvent ci-dessous :

#### Membres du groupe de travail

- Geneviève Chêne - Co-pilote du groupe de travail - PU-PH de Santé publique - Université de Bordeaux, CHU de Bordeaux, Inserm

Geneviève Chêne, MD, PhD, est professeur de santé publique à l'université et au CHU de Bordeaux. Elle siège au conseil exécutif de l'Association internationale des instituts nationaux de santé publique (IANPHI) et au comité scientifique indépendant du Lancet Public Health.

Depuis janvier 2023, elle co-anime le GT « socle de données des EDSH » pour le comité stratégique des données de santé. D'octobre 2019 à octobre 2022, elle a été directrice générale de Santé publique France ([www.santepubliquefrance.fr](http://www.santepubliquefrance.fr)), l'agence nationale de santé publique, dont l'ensemble des missions (surveillance épidémiologique, alertes, analyses de risque ; prévention, promotion de la santé et éducation pour la santé ; stocks stratégiques et réserve sanitaire) ont été fortement mobilisés en première ligne de la gestion de crise due au COVID-19, tout en poursuivant le déploiement de son programme national d'action globale pour la protection et l'amélioration de la santé des Français. Entre 2017 et 2019, elle a été directrice de l'école de santé publique de Bordeaux (ISPED, univ Bordeaux) et de 2011 à 2019, cheffe du pôle de santé publique du CHU de Bordeaux. Dans le domaine des politiques publiques, elle a contribué à l'élaboration de la stratégie nationale de santé 2013-2017 en tant que membre du "comité des sages" nommé par le Premier ministre, puis en tant que directrice de l'Institut de santé publique de l'Inserm (2014-2017).

- Cyrille Delpierre - Co-pilote du groupe de travail - Directeur de recherche en épidémiologie - Inserm-Université de Toulouse.

Directeur du CERPOP (Centre d'Épidémiologie de de Recherche en santé des POPulations) les recherches du Dr Cyrille Delpierre portent sur les inégalités sociales dans la santé (ISS), en particulier les déterminants biopsychosociaux de la santé et l'analyse des mécanismes de genèse des ISS, notamment dans le domaine des cancers. Plus précisément, ses recherches portent sur les mécanismes par lesquels l'environnement, au sens large, est biologiquement incorporé, comment il peut affecter le fonctionnement biologique et ainsi favoriser le développement de maladies à long terme. L'objectif de ses recherches est d'expliquer les ISS non seulement par des expositions physico-chimiques, des comportements de santé et une prise en charge par le système des soins socialement distribués, mais aussi par des expositions psychosociales susceptibles de modifier des processus biologiques et favoriser le développement de pathologies à distance.

- Salam Abbara - Membre du groupe de travail - PH de Médecine interne et experte médicale - Health Data Hub

Salam Abbara est interniste et épidémiologiste. Elle exerce comme experte médicale au Health Data Hub, notamment autour de travaux relatifs à l'intelligence artificielle et aux entrepôts de données de santé hospitaliers.

Elle est membre de la Société de Pathologie Infectieuse de Langue Française dont elle co-coordonne le Groupe d'Infectiologie Digitale. Elle mène des travaux de recherche en épidémiologie autour de la résistance des bactéries aux antibiotiques. Au plan pédagogique, elle co-développe un jeu sérieux, Dawaa, pour l'apprentissage du bon usage des antibiotiques aux soignants.

- Elisabeth Fery-Lemonnier - Membre du groupe de travail - Conseillère médicale - DREES

Médecin titulaire d'un DEA de statistiques et d'un master de l'ESSEC, Elisabeth Féry-Lemonnier a travaillé 30 ans à l'assistance publique - hôpitaux de Paris (AP-HP). Praticien hospitalier au sein de l'hôpital Cochin en imagerie médicale, elle a aussi dirigé le CEDIT (comité d'évaluation et de diffusion des innovations technologiques) puis le département de l'organisation des activités médicales à la direction générale de l'Assistance Publique-Hôpitaux de Paris jusqu'en 2007.

Parallèlement, elle a assuré entre 2002 et 2007 des fonctions de conseillère médicale à la FHF, de membre puis vice-présidente de la commission de l'HAS en charge de l'évaluation des dispositifs médicaux (CEPP : commission d'évaluation des produits et prestations), de rapporteur pour le 5e PCRDT européen.

En 2008 elle est nommée conseillère spéciale pour les affaires médicales auprès de Roselyne Bachelot, alors ministre en charge de la santé, puis conseillère générale des établissements de santé. Dans cette dernière fonction elle a assuré une vingtaine de missions, dont la coordination d'un rapport sur la prise en charge des accidents vasculaires cérébraux (AVC) en France puis la présidence du plan national AVC. Par ailleurs membre du conseil scientifique de la CNAM pendant 10 ans, professeur affilié à l'EHESP pendant 5 ans.

En 2012 elle est nommée chargée de mission auprès du secrétaire général des ministères sociaux, en charge des politiques de parcours de santé puis conseillère médicale du directeur général de l'ARS Auvergne-Rhône Alpes.

En 2018 elle est nommée représentante du ministère de la santé dans le CSIS (conseil stratégique des industries de santé) 2018-2022, membre de la task-force réforme du financement du système de santé 2018-2020 et de l'équipe pilotant la mise en place des expérimentation d'innovations organisationnelles « article 51 ».

Elle est actuellement conseillère médicale du directeur de la DREES (direction de la recherche, des études, de l'évaluation et des statistiques du ministère de la santé).

- Nicolas Garcelon - Membre du groupe de travail - Responsable plateforme Data Science - Institut Imagine - Hôpital Necker Enfants Malades

Nicolas Garcelon a un diplôme d'ingénieur agronome (2000) ainsi qu'un doctorat de santé publique en informatique biomédicale (2017). Il travaille depuis 20 ans dans le domaine médical. De 2001 à 2012, il a travaillé au département d'information médical du Centre hospitalier universitaire de Rennes. Depuis 2012, il dirige la plateforme data science de l'institut Imagine à Paris. Il conçoit et développe des logiciels hospitaliers pour les cliniciens et chercheurs afin de faciliter la réutilisation des données médicales pour la recherche et le soin. Il a notamment créé Dr Warehouse® un entrepôt de données hospitalier en 2014, déposé sous licence open source en septembre 2017. Il est aussi chercheur dans l'équipe INSERM «Recherche en santé numérique pour un système de santé apprenant » à Paris Santé Campus.

- Sophie Georgin-Lavialle - Membre du groupe de travail - PU-PH de Médecine interne - Hôpital Tenon (AP-HP)

Sophie Georgin-Lavialle est interniste et immunologiste, Professeur des Universités-Praticien hospitalier, membre de la Société Nationale Française de Médecine interne (SNFMI). Elle travaille dans le service de médecine interne à l'hôpital Tenon à Paris. Elle coordonne le centre national de référence des maladies autoinflammatoires et des amyloses inflammatoires (AA) (site constitutif adulte de Tenon). Elle fait partie du comité de pilotage de la filière de santé maladies rares autoimunes/autoinflammatoires (FAI2R).

Elle mène des travaux de recherche translationnelle sur les aspects cliniques de ces maladies de l'immunité innée et l'hyperinflammation. Elle travaille en particulier sur le microbiote dans les maladies autoinflammatoires au sein de l'unité INSERM U938 du centre de recherche de l'hôpital St Antoine. Elle est responsable du groupe de recherche clinique de Sorbonne Université sur l'amylose AA. Elle coordonne la cohorte nationale et européenne RaDiCo-ACOSTILL sur la maladie de Still à début pédiatrique et adulte.

Au plan pédagogique, elle coordonne le DIU des maladies autoinflammatoires et une unité d'enseignement sur l'immunité innée.

- Yann Le Strat - Membre du groupe de travail - Directeur appui, traitements et analyses des données - Santé publique France

Yann Le Strat a une thèse de biomathématiques (2000) ainsi qu'une Habilitation à Diriger les Recherches (2011). Il travaille depuis plus de 20 ans dans le domaine de la santé publique. De 2001 à 2016, il a travaillé au département des maladies infectieuses à l'Institut de Veille Sanitaire. Depuis 2016, il dirige la direction data de Santé publique France dont la mission principale est de collecter et de traiter des données de santé issues de nombreux systèmes de surveillance ou d'enquêtes, de produire des indicateurs épidémiologiques et de les restituer en open data.

- Thomas Lefèvre - Membre du groupe de travail - MCU-PH de Médecine légale et sociale - Hôpital Jean Verdier (AP-HP)

Thomas Lefèvre est ingénieur Mines-Télécom (IMT Atlantique 2005), médecin de santé publique et médecin légiste. Il est docteur en épidémiologie sociale, docteur en mathématiques appliquées, sur des sujets touchant notamment aux analyses de données en haute dimensionnalité et l'utilisation secondaire des données hospitalières. Il est actuellement maître de conférences et praticien hospitalier à l'université Sorbonne Paris Nord et à l'AP-HP (hôpital J. Verdier), chercheur à l'Institut de recherche interdisciplinaire sur les enjeux sociaux (Iris – CNRS Inserm EHESS USPN), habilité à diriger les recherches. Il a débuté des développements théoriques et pratiques au début des années 2010 (co-conception de plateforme d'analyse de données) au sein du LaTIM (Inserm) et du CHU de Brest, avec la PME Tekliko (plateforme Spot-Risc, devenue Spe3dlab). Il a créé et pilote le réseau ORFeAD (Outils et réseaux pour la fédération, l'analyse et l'utilisation de données en médecine légale), basé sur 11 centres de médecine légale depuis 2016. Il mène des recherches portant sur l'impact et les transformations induites par le numérique et l'IA à l'échelle micro-sociale. Il a été pilote ou co-pilote de plusieurs groupes de travail portant sur les données en santé : groupe ministériel big data et santé en 2015-16 ; groupe du Haut conseil de la santé publique sur l'avenir des registres à l'ère du numérique ; schéma directeur du SI et mégadonnées à l'APHP 2016-22 entre autres.

Il est membre élu du comité d'éthique de la recherche de l'université de Paris, membre élu du collège du HCSP.

## Représentants des projets lauréats de la vague 1 de l'AAP EDS

### **ACCES AP-HP**

Le projet ACCES AP-HP a pour objectif de développer l'EDS de l'AP-HP grâce à l'intégration, la standardisation et la mise en qualité de données de santé, dans un environnement technique sécurisé, pour la réalisation de projets de recherche et d'innovation, dans les domaines du cancer, du diabète, de l'anesthésie-réanimation, des maladies rares et de la pathologie numérique notamment. Il permettra de développer des composants technologiques mis en open source qui pourront être utilisés par l'ensemble des autres CHU et établissements hospitaliers.

Ces projets bénéficieront d'un accompagnement médico-scientifique et associeront des partenaires publics et privés au bénéfice de la recherche en santé, d'un meilleur dépistage, d'un meilleur diagnostic et d'une amélioration de la qualité de la prise en charge des patients.

Le projet ACCES AP-HP est un projet partenarial associant au travers de cinq projets, huit établissements de santé, des acteurs académiques (tels qu'Inria, l'Inserm, l'institut Imagine et plusieurs universités), ainsi que des industriels partenaires et institutionnels.

Le pilotage médico-scientifique du projet ACCES AP-HP sera assuré par le Pr Etienne Audureau, médecin de Santé Publique, épidémiologiste et responsable de l'Unité de Recherche Clinique de l'hôpital Henri-Mondor, AP-HP, UPEC, INSERM, par le Pr Raphaële Renard-Penna, radiologue, responsable du département d'imagerie Uro-Néphrologique du Groupe Hospitalier Pitié Salpêtrière – Charles Foix AP-HP, Sorbonne Université et par Gaël Varoquaux, directeur de recherche et responsable de l'équipe SODA d'Inria Saclay.

Le projet est représenté par :

- Stéphane Bréant - Invité permanent du groupe de travail - Responsable de l'équipe SI recherche du pôle innovation et données - AP-HP

Stéphane Bréant a un diplôme d'ingénieur Ingénierie des systèmes d'information (CNAM – 2015). Il travaille depuis plus de 20 ans au sein de l'Assistance Publique – Hôpitaux de Paris. Après avoir travaillé sur les programmes stratégiques AP-HP de mise en place du système de gestion de laboratoire commun GLIMS, puis sur la mise en place du dossier patient informatisé ORBIS, il rejoint, en 2014, l'équipe à l'origine du projet Entrepôt de Données de Santé AP-HP, en vue notamment d'une réutilisation des données de soins pour la recherche. Depuis septembre 2022, il dirige l'équipe SI Recherche du pôle Innovation & Données de la Direction des Services Numériques, assurant une offre de services au soutien de la recherche et l'innovation, aussi bien dans le cadre de la recherche sur données (EDS) que dans celui de la recherche interventionnelle (gestion e-crf et vigilance, accompagnement grands projets, ...).

- Catherine Duclos - Invitée permanente du groupe de travail - PU-PH de Santé publique - AP-HP

Catherine Duclos est Professeur des Universités-Praticien Hospitalier en informatique médicale à l'Université Sorbonne Paris Nord, membre du Laboratoire d'Informatique Médicale et Ingénierie des Connaissances pour la e-Santé (LIMICS, INSERM 1142) et coordinatrice de l'entrepôt de données de santé dans le groupe hospitalier Hôpitaux Universitaires Paris Seine Saint Denis (Assistance Publique des Hôpitaux de Paris). Ses sujets de recherche portent sur

la représentation des connaissances médicales et pharmaceutiques et les systèmes d'aide à la décision médicale. Elle a mené de nombreuses études publiées dans des revues internationales sur la représentation des connaissances sur les médicaments (indication, pharmacocinétiques, pharmacodynamie des antibiotiques, effets indésirables) et leur formalisation ontologique. Elle a coordonné plusieurs projets scientifiques nationaux (ANR, ANSM) et participé à plusieurs projets nationaux ou européens liés à la sécurité du médicament, l'identification du médicament, la décision thérapeutique. Elle est à l'origine avec A. Venot et JB Lamy du langage iconique VCM qui exprime iconiquement les concepts médicaux et utilisé pour synthétiser les dossiers de santé électroniques, indexer graphiquement les contenus médicaux, faciliter la saisie des données. Elle maîtrise le développement d'ontologies et elle a développé une ontologie des bactéries et des antibiotiques, une ontologie des effets secondaires des médicaments. Elle est experte en bases de données sur les médicaments et en représentation des médicaments pour les autorités sanitaires et participe actuellement à la création du référentiel interopérable unique du médicament de l'Agence du Numérique en Santé.

## **EDEN4HEALTH**

Le projet EDEN4Health ambitionne de créer un hub inter-régional d'intelligence en données de santé agrégé dans 7 entrepôts de données de santé avec 5 CHU (Besançon, Dijon, Nancy, Reims et Strasbourg), le CHR Metz-Thionville et le CH Troyes) et leurs expertises en données de santé afin de mener à bien l'exploitation des données de santé. Ce projet comprend la structuration des données qui seront interopérables avec les standards nationaux grâce à une surcouche d'interopérabilité. Une plateforme mutualisée sera en capacité d'accueillir des données externes. Enfin, ce projet intègre un environnement de traitement sécurisé et une gouvernance simple et réactive garantissant leur accessibilité dans le respect de la réglementation et de l'éthique.

Le projet est représenté par :

- Alban Dupoux - Invité permanent du groupe de travail - Chargé de mission du GCS Grand Est - GCS Grand Est

Alban Dupoux est Docteur en Biologie Cellulaire et Moléculaire (2008) et travaille depuis 2011 dans le domaine de la recherche clinique d'abord à l'accompagnement de projets de recherche au CHU Dijon Bourgogne dans les données de maladies neuro-dégénératives et de la réutilisation des données de santé puis à la constitution du GIRCI EST. D'abord impliqué dans le soutien aux missions de l'investigation clinique industrielle avec l'animation de réseaux interrégionaux, il est poursuivi avec des actions pour le développement de la recherche dans les établissements de santé non universitaires, la mutualisation de logiciels de recherche en vue de l'amélioration de la qualité et de la sécurité, la formation des professionnels. Depuis 2018, il assure la coordination du GCS Groupement du Grand Est (GGEST) pour le développement d'actions de pointe en soins, enseignement et recherche par la collaboration entre les CHU Besançon, Dijon Bourgogne, Nancy, Reims, Strasbourg et le CHR Metz-Thionville. Dans ce cadre, il est impliqué dans la coordination du réseau PAISaGE des plateformes d'accessibilité, d'intelligence des données de santé présentes dans les établissements du GGEST. Il assure la co-animation du Groupe National de Travail « GIRCI et Établissement de Santé non Universitaires » de la Commission Recherche et Innovation de la Conférence Nationale des Directeurs Généraux de CHU.

- Erik Sauleau - Invité permanent du groupe de travail - PU-PH de Santé publique - CHU de Strasbourg

Erik Sauleau est PU-PH de Biostatistique, informatique médicale et technologies de communication (CNU 46-04) à l'Université de Strasbourg (Faculté de Médecine, maïeutique et sciences de la santé) et aux Hôpitaux Universitaires de Strasbourg. Son activité hospitalière est tournée vers (1) la méthodologie de la recherche clinique (aide à la rédaction de protocole, plan d'analyse statistique, réalisation des statistiques, aide à l'interprétation des résultats et à leur publication) et (2) la mise en œuvre et l'exploitation de l'entrepôt de données des hôpitaux universitaires de Strasbourg (responsable de la plateforme SciDonHUS de sciences des données). Pour son activité universitaire, en plus d'une charge d'enseignement de plus de 100h en biostatistique, science des données et statistiques de base (niveaux licence et master) et le portage d'un parcours de master Intelligence en données de santé. Il est vice-doyen Transition digitale et communication à la faculté de médecine. Son activité de recherche, essentiellement tournée vers les modèles spatiaux et géographique et vers l'inférence bayésienne, s'exerce au sein du laboratoire ICube, Sciences de l'ingénieur, de l'informatique et de l'imagerie (CNRS, UMR7357), où il anime le thème de recherche RADoSCauBa, Recueil et analyse de données de santé, causalité et inférence bayésienne. Il est d'autre part président du Comité de Protection des Personnes (CPP) Est4 et membre désigné de la Commission nationale des recherches impliquant la personne humaine (CNRIPIH). Relecteur de plusieurs journaux scientifiques, il est aussi membre de plusieurs sociétés savantes et siège au conseil d'administration de la Société Française de Biométrie et au bureau des groupes Socio-histoire de la statistique et des probabilités et Enseignement de la statistique.

## **eDOL**

L'entrepôt de données de santé du Languedoc 'eDOL' a pour objectif de valoriser les données des patients du CHU de Montpellier et ceux de son Groupement Hospitalier de Territoire, et s'inscrit dans le cadre de la politique de soutien à la recherche de l'établissement. Le but de cet entrepôt est d'améliorer la prise en charge des patients via des projets concrets, construits en partenariat avec les établissements de santé du territoire et le riche environnement scientifique de l'Université de Montpellier. eDOL doit également faciliter la mise à disposition de données à différents partenaires industriels ou académiques, et contribuer à l'enrichissement des données nationales.

Le projet est représenté par :

- Caroline Dunoyer - Invitée permanente du groupe de travail - Responsable de l'unité Sciences des Données de Santé, SSP - CHU de Montpellier

Caroline Dunoyer est ingénieure en informatique (ISIM) et titulaire d'un diplôme de troisième cycle en intelligence artificielle (LIRRM). Elle est actuellement co-responsable de l'Unité de Science des Données de Santé du Service de santé publique (SSP) du CHUM et responsable technique de l'entrepôt eDOL. Spécialiste du traitement des données de santé depuis plus de 30 ans, elle a été à l'initiative de la création et du développement de l'entrepôt de données du CHUM. Son équipe conçoit des modèles et des algorithmes pour la collecte, le stockage, le traitement et l'extraction des données, et développe des outils informatiques spécifiques pour soutenir les missions du SSP. Elle collabore à des projets de recherche dans le domaine de l'IA en santé de l'équipe ADVanced Analytics for data Science (ADVANSE) du Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRRM). Depuis 2021, Caroline Dunoyer est membre de l'Institut Desbrest d'épidémiologie et de santé publique (IDESP, UMR INSERM - Université de Montpellier) qui vise à comprendre le développement et l'évolution des maladies chroniques en s'appuyant

sur l'utilisation de larges bases de données de santé et de données environnementales nationales et européennes.

## **EDS@NOVA**

Les CHU de Bordeaux, CHU de Limoges et CHU de Poitiers sont membres fondateurs du GCS NOVA. Ce dernier est le chef de file du projet EDS@NOVA et s'appuiera sur les compétences de ces trois établissements.

Le projet EDS@NOVA a pour objectif de déployer et mettre à niveau les entrepôts de données des établissements partenaires du projet, de mettre à disposition du Health Data Hub le catalogue de ses données identifiées comme partageables et de mettre en place les interfaces organisationnelles, techniques et réglementaires pour collaborer à l'échelon local, régional et national. La mobilisation des CHU de la Région Nouvelle-Aquitaine permettra ainsi de maximiser l'utilisation secondaire des données de santé pour la recherche, le pilotage et l'innovation dans le but, à terme, d'apporter une meilleure qualité de soins aux patients.

Le projet est représenté par :

- Vianney Jouhet - Invité permanent du groupe de travail - PH de Santé publique et d'Informatique médicale / Responsable de l'EDS du CHU de Bordeaux - CHU de Bordeaux

Vianney Jouhet est praticien hospitalier du CHU de Bordeaux. Médecin de Santé Publique spécialiste en Informatique Médicale il est responsable de la mise en œuvre de l'EDS du CHU de Bordeaux. Vianney Jouhet contribue également au sein de l'équipe AHead (Assessing Health in a Digitalizing Real-World Setting Pharmacoepi & beyond) au sein du centre INSERM U1219 Bordeaux Population Health à l'Université de Bordeaux. Dans ce cadre de ces travaux il développe des méthodes d'intégration et d'exploitation secondaires des données issues du soin. L'objectif est de permettre à terme la mise à disposition de données exploitables à des fins de recherche, d'études et d'évaluation. Dans ce contexte de recherche appliquée, il développe et met en œuvre et exploite des ressources sémantiques et des modèles formels et applique des méthodes issues du traitement automatique du langage naturel et de la recherche d'information.

## **ODH 2.0**

Avec le projet ODH 2.0, le GCS HUGO et ODH souhaitent à consolider la dynamique et poursuivre leur développement :

- Sur le plan de leur aire d'influence géographique en intégrant quatre établissements de l'interrégion pas encore dotés d'entrepôts de données mais engagés dans des travaux de recherche HUGO, et ainsi diversifier les sources et renforcer le maillage territorial en matière de recherche industrielle ;
- Sur le plan des usages en intégrant de nouveaux types de flux de données notamment en oncologie et en imagerie pour offrir un champ de recherche le plus large possible au bénéfice des usagers, et répondre aux défis de santé de demain.

Le projet est représenté par :

- Denis Delamarre - Invité permanent du groupe de travail - Ingénieur bases de données - CHU de Rennes

Denis Delamarre travaille comme ingénieur des données au centre de données clinique du CHU de Rennes. Il participe à la recherche et développement de l'entrepôt de données EHOP.

- Pierre-Antoine Gourraud - Invité permanent du groupe de travail - Responsable du pôle de santé publique et de la Clinique des Données du CHU de Nantes - CHU de Nantes

Pierre-Antoine Gourraud est professeur des universités, praticien-hospitalier de la faculté de médecine de la Nantes Université (France). C'est un ancien élève de l'école Normale Supérieure de Lyon (France) du département de biologie. Arrivé comme chercheur post-doctorant du département de neurologie à l'université de Californie (UCSF, Etats-Unis), il le quitte en tant que professeur associé en 2018. Auteur de plus de 180 publications cité plus de 11000 fois (h-index 55), ses activités de recherche se positionnent au carrefour de l'immunologie, de la génétique et du traitement informatique des données de santé.

- Pascal Van Hille - Invité permanent du groupe de travail - Coordinateur Technique du Réseau interrégional des Centres de Données Cliniques du GIRCI Grand Ouest - CHU de Rennes

Pascal van Hille est ingénieur en système d'information du Conservatoire National des Arts et Métiers (2011). Il a travaillé 20 ans dans l'industrie en tant que responsable technique. En 2010, il a rejoint l'unité Inserm U936 de Rennes (Modélisation conceptuelle des connaissances biomédicales), laboratoire d'informatique médicale. Depuis 2015 au CHU de Rennes, il est responsable de l'équipe technique de l'équipe de recherche DONNÉES MASSIVES et Système d'Information Apprenant en santé (DOMASIA), Laboratoire Traitement du Signal et de l'Image (LTSI), INSERM UMR-1099. En tant qu'ingénieur d'étude, il a participé à différents projets de recherche nationaux et européens. Il a eu en charge l'industrialisation de la solution d'entrepôt eHOP issue d'un projet de recherche du laboratoire et a participé en 2018 à la création du LabCom LITIS, associant le CHU de Rennes, le LTSI et la société ENOVACOM, qui assure depuis la R&D de la solution. Il a coordonné de 2016 à 2022 le déploiement des 6 Entrepôts de Données de Santé du GCS HUGO et est depuis 2023 chef de projet technique pour le projet ODH 2.0.

## **OncoDS**

OncoDS est un projet d'entrepôt fédéré piloté par Unicancer s'appuyant sur la mise en réseau d'entrepôts locaux communicants et interopérables de 12 centres de lutte contre le cancer. L'objectif est de répondre aux enjeux actuels de la recherche et de l'innovation en oncologie en garantissant aux acteurs de la recherche un accès simple et sécurisé à ces données pour progresser dans la lutte contre le cancer.

Le projet est représenté par :

- Nabil Benhajkassen - Invité permanent du groupe de travail - Chef de projet scientifique - Unicancer

Nabil Benhajkassen est pharmacien et épidémiologiste. Il travaille actuellement au sein de la direction des data et partenariats d'Unicancer en tant que chef de projet scientifique sur différents projets de constitution d'entrepôts de données de santé de vie réelle.

### Autres invités permanents

- Stéphanie Combes - Invitée permanente du groupe de travail / Représentante du comité stratégique des données de santé - Directrice - Health Data Hub
- Carole Dufouil - Invitée permanente du groupe de travail / Représentante du comité stratégique des données de santé - Chargée de mission - DGRI
- Claude Gissot - Invité permanent du groupe de travail / Représentant du comité stratégique des données de santé - Directeur de projet AMDAC - DREES
- Javier Nicolau - Invité permanent du groupe de travail - Adjoint au Bureau "Lab innovation et évaluation en santé" - DREES