

Paris,
21 mars 2023

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

Objet : financement des entrepôts de données de santé (EDS) au sein des CHU

Les données disponibles dans les établissements de santé sont nombreuses et présentent un important potentiel pour la recherche, l'innovation et l'amélioration des systèmes de soin via le pilotage notamment. Pour autant, elles ne peuvent pas toujours être analysées car elles ne sont pas toujours organisées dans des entrepôts de données de santé (EDS). L'absence de financement dédié aux EDS peut expliquer cette sous exploitation. Dans un contexte où l'utilisation des données de santé revêt un caractère de plus en plus stratégique au niveau international¹ et avec la dynamique enclenchée par la Commission Européenne autour de l'espace européen des données de santé, il est urgent de formuler des modalités de financement soutenable et permettant d'offrir une vision à long terme à toutes les parties prenantes. Il serait essentiel de mettre en place dès 2023 un financement annuel pérenne dédié à la mise en place et au fonctionnement d'environ 30 EDS de taille critique d'établissements de santé, soit une enveloppe globale annuelle estimée de l'ordre de 60 à 90 M€. En conditionnant ce financement à l'ouverture et au partage des données de santé de qualité pour la recherche et l'innovation, dans le respect de la réglementation et des droits des personnes, cette mesure contribuerait efficacement au renforcement du patrimoine national des données de santé et à la mise en œuvre du SNDS élargi.

I. Les enjeux relatifs aux données de santé se diffusent et la mise en place progressive d'entrepôts de données de santé est observée au niveau national depuis quelques années.

Depuis déjà de nombreuses années, la dématérialisation croissante du dossier patient informatisé (DPI) pour la prise en charge individuelle a généré de nombreuses données dans les établissements de santé. Les données des DPI comprennent une grande partie de données médicales spécifiques et originales i.e. présentes uniquement dans ces dossiers médicaux qui complètent les données médico-administratives du SNDS historique. Ces données des DPI sont nécessaires pour l'équipe soignante et le patient pour de la prise en charge individuelle et plus généralement pour le pilotage de l'établissement.

L'un des exemples le plus frappant de pilotage hospitalier par l'analyse de données, ces dernières années, a été la crise sanitaire débutée en 2020. Crise du capacitaire, la gestion a

¹ Une étude commandée par le HDH auprès du cabinet Veltys, spécialisé en économie de la santé, nous apprend que l'exploitation des données de santé représenterait un gain d'environ 7,3 Md€ par an pour l'économie française. A noter que l'étude est centrée sur le potentiel économique de l'utilisation secondaire des données de santé. En revanche, elle dresse le constat selon lequel la France accuse un retard dans la saisie de ce potentiel. C'est le cas en matière de startups fondées sur l'usage de la donnée de santé puisque la France, malgré son patrimoine de données unique au monde, ne compte aucune licorne dans le domaine.

notamment reposé sur l'analyse quotidienne des lits et matériels disponibles, à date, et prévisibles sur les jours et les semaines à venir. Le recours au système SI-VIC, dont la vocation initiale portait sur des événements exceptionnels, a ainsi été un marqueur des difficultés d'analyse des systèmes d'information hospitaliers traditionnels, hors EDS, à répondre à eux-seuls à ce besoin en tableaux de bord à produire en temps réel. Pourtant, comme cela a pu être fait dans certains établissements, dont l'APHP², les EDS ont démontré la capacité de mise en place ultra-rapide de tableaux de bord complets et sécurisés.

Dans un contexte ordinaire, les EDS hospitaliers offrent de nouvelles possibilités pour contribuer au pilotage de l'activité ou de la qualité et sécurité des soins en routine – avec notamment la perspective d'automatisation de la production de certains indicateurs aujourd'hui très « coûteux », tels que les indicateurs IFAQ.

Au-delà de cet usage emblématique, le recueil, la standardisation et le rassemblement des données des patients au sein d'un EDS sont des leviers d'amélioration de la prise en charge, de constitution de bases de données uniques pour les équipes de recherche en interne aux établissements ou en partenariat avec l'extérieur (public comme privé) dans des délais attractifs. Cumulées entre plusieurs établissements et/ou appariées à d'autres sources, notamment les données médico-administratives déjà rassemblées du SNDS historique, ou des données de cohortes ou registres préexistants, ces données peuvent être déterminantes pour le développement de la recherche et de l'innovation en santé à l'échelle non seulement de l'établissement, mais également de son territoire, voire du système de santé dans son ensemble (**annexe 1**).

L'investissement dans des outils permettant la constitution de bases structurées pour la réutilisation de données, à partir de bases de données initialement structurées pour les soins, représente donc un important enjeu qui est progressivement identifié au sein de l'écosystème hospitalier qui y consacre une part de ses fonds propres³ depuis plusieurs années⁴ ou qui postule à des financements à travers des dispositifs ponctuels (par exemple via des AAP). Plus récemment, l'enjeu des EDS a été inscrit au sein de la « Stratégie nationale d'Accélération Santé Numérique » (SASN) et le plan France 2030 qui aspirent à faire de la France un leader international sur cette thématique. Dans le cadre de France 2030, le gouvernement a ainsi lancé en 2022 un premier appel à projets « Accompagnement et soutien à la constitution d'entrepôts de données de santé hospitaliers » avec pour objectif de contribuer aux financements d'amorçage d'EDS à hauteur de 50 millions d'euros.

II. Cette progression reste néanmoins fragile. Une absence de soutien pérenne pourrait mener à une prise de retard du système de santé français sur les volets amélioration des parcours de soin et recherche / innovation, et donc une perte de

² <https://gestions-hospitalieres.fr/entrepot-de-donnees-de-sante-et-pilotage-de-crise/>

³ Les fonds propres mentionnés ici font référence à l'ensemble des produits perçus par les établissements de santé hospitaliers comprenant les produits versés par l'assurance maladie, les autres produits issus de l'activité hospitalière (produits de tarification non pris en charge par l'assurance maladie, produits à la charge de l'Etat, collectivités territoriales et autres organismes publics, etc) et les autres produits (produits financiers, subventions d'exploitation, etc)

⁴ Une estimation réalisée début janvier 2022 au CHU de Rennes permet de chiffrer l'ensemble des coûts (y compris les coûts de support) correspondant à la construction de l'EDS et son exploitation sur 4 années : 4 013 986,48€ soit 1 million d'euros par an en moyenne. L'APHP, dont l'EDS intègre le plus grand volume de données en Europe, fait état d'un coût de développement et d'exploitation à hauteur de 34,8 millions d'euros sur 5 ans soit environ 7 millions d'euros par an pour l'équivalent de 6 CHU. La CRI a d'autre part mené une enquête sur les coûts de fonctionnement d'un EDS (hors frais liés à la mise en place) auprès d'un échantillon représentatif de 6 CHU en France. Le résultat de cette enquête présente des coûts de fonctionnement annuels par EDS allant de 707 000€ à 1 940 000€ pour un CHU de grande taille. Cela correspond à une moyenne annuelle de 1 098 000€ et une médiane à 974 000€.

chance importante pour la compétitivité française dans un contexte international très dynamique.

Alors que le sujet prend une ampleur grandissante à l'international et que la France dispose d'atouts considérables, les moyens mobilisés restent aujourd'hui insuffisants.

Au niveau international, on peut citer par exemple la vingtaine de molécules autorisées à date par la FDA aux Etats-Unis, dont l'élaboration a reposé sur des approches innovantes d'essais cliniques mobilisant des données de vie réelles. On dispose également de l'exemple de l'initiative CPRD⁵ au Royaume-Uni, qui a déjà donné lieu à plus de 3 000 publications portant sur l'étude de la sécurité des médicaments, de leur utilisation, de l'efficacité de la politique de santé, de la prestation des soins de santé et des facteurs de risque des maladies. En parallèle, l'Agence Européenne du Médicament (EMA) lance une plateforme de collecte de données de vie réelle ("Darwin"), la Commission européenne multiplie les initiatives autour de la mise en œuvre d'un espace commun des données de santé à l'échelon européen et a publié un projet de règlement en mai 2022. **Le consortium piloté par la France a d'ailleurs été sélectionné par la Commission Européenne pour mettre en œuvre le pilote de l'espace européen des données de santé pour leur usage secondaire.**

Investir dans l'optimisation de la mise à disposition et de l'utilisation des données de santé en France est par ailleurs tout à fait pertinent du fait des atouts considérables dont elle dispose, avec par exemple sa base centrale de données médico-administratives qui a permis de démontrer l'intérêt et l'utilité des données de santé dans les différents usages : régulation du système, qualité des soins, analyse des parcours de soins, évaluation de la balance bénéfiques/risques des traitements notamment médicamenteux. Toutefois cette base dont l'utilisation a été croissante depuis 10 ans souffre de manques importants, en particulier des informations médicalisées indispensables pour mesurer objectivement l'état de santé (résultats d'examens cliniques ou biologiques) mais aussi l'imagerie ou les compte rendus hospitaliers ou de consultation de ville susceptibles de fournir des éléments de contexte et de guider la prise en charge et des décisions médicales. L'utilisation plus systématique des données cliniques grâce aux entrepôts de données de santé, à leur interconnexion ou à leur croisement avec la grande base médico-administrative ouvre la voie à une variété de nouveaux usages : accélérer la recherche sur données à large échelle, faciliter et accélérer les essais cliniques et développer les essais cliniques du futur (essais adaptatifs, multi-bras et multi-étapes, essais parapluie et panier), offrir des sites d'évaluation de solutions innovantes en appui aux professionnels de santé, améliorer la coordination des soins et des parcours de soin pour les patients, développer des outils en appui des professionnels de santé (prévention, diagnostic...).

Si les bénéfices escomptés pour le système de santé sont ainsi nombreux, l'action de soutien évoquée ci-dessus au sein de la SASN et traduisant un premier investissement en faveur de la nécessité d'accélérer la structuration et les usages de la donnée de santé en milieu hospitalier, ne sera pas suffisante. **De tels moyens requièrent d'être sensiblement accrus et pérennisés pour soutenir la démarche, l'amplifier et la structurer à l'échelle nationale.** L'appel à projets EDS France 2030 constitue un financement d'amorçage considérable mais n'offre cependant pas de perspective pérenne pour les établissements de santé s'engageant dans la construction d'un entrepôt de données de santé, alors même que celui-ci a vocation à perdurer, s'enrichir et à développer une offre de services à destination des utilisateurs des données.

L'objectif principal de la proposition du groupe de travail est de favoriser la mise en place et

⁵ <https://cprd.com/>

la pérennisation d'un EDS et de son écosystème d'exploitation dans chacun des grands établissements de santé du pays ayant des missions de soins, d'enseignement et de recherche, afin de 1) permettre aux établissements porteurs ou partenaires de disposer et de déployer des outils de pilotage et d'amélioration continue de la prise en charge des patients, 2) mettre à disposition, de manière structurée, transparente et encadrée, des données et des services pour des projets de recherche ou d'innovation intra ou inter hospitaliers, et associant des acteurs académiques ou industriels.

Ce financement permettra par ailleurs d'accélérer la mise en œuvre de la politique d'ouverture des données de santé portée par la loi Organisation et Transformation du Système de Santé promulguée en 2019. En effet, cette loi a modifié l'article L1461-1 du Code de la Santé Publique pour élargir le périmètre du SNDS à l'ensemble des données de santé dont le recueil est directement ou indirectement financé par des fonds publics, en particulier les données cliniques recueillies dans le cadre du parcours de soins ce qui inclut les activités de prévention, de diagnostic, des soins, de suivi social et médico-social. La grande majorité des bases de données hospitalières font donc par définition partie du périmètre du SNDS, et se verront appliquer des règles identiques⁶ en termes d'accès, de respect des droits des personnes, ou de respect des exigences de sécurité. Ces bases de données devront également pouvoir être partagées via le catalogue du Health Data Hub pour être enrichies ou venir compléter d'autres sources de données, dont le SNDS historique, et ainsi permettre la réalisation d'études nécessitant des données complémentaires dans un environnement technique à l'état de l'art et sécurisé tout en veillant aux conditions de partage avec les acteurs de l'écosystème.

III. Les entrepôts de données de santé hospitaliers constituent une brique indispensable de l'infrastructure nationale des données de santé et doivent par conséquent bénéficier d'un financement majoritairement public.

Le financement des EDS peut être décomposé en plusieurs niveaux : l'investissement initial pour mettre en place l'organisation et l'outil informatique, l'investissement pour le maintenir (flux de données malgré la modification des sources, disponibilité, qualité des données, renforcement permanent des mesures de sécurisation, standardisation), et les coûts marginaux nécessaires pour l'exploitation des données à des fins de pilotage des établissements ou dans le cadre de projets de recherche ou d'innovation (qu'ils soient techniques ou portant sur la personne humaine).

La réglementation en cours de mise en place au niveau européen via le Data Governance Act (DGA), qui entrera en vigueur dès septembre 2023 ouvre certainement la porte à une possibilité de valoriser économiquement la production et la mise à disposition des jeux de données, tableaux de bord et/ou expertises auprès des acteurs industriels qui en font la demande sous la forme d'une redevance⁷. Toutefois, cette même réglementation semble limiter cette redevance à la couverture des coûts de mise à disposition, en excluant une éventuelle redevance intégrant une contribution à l'investissement, et en restreignant ainsi la capacité des EDS à facturer des redevances au-delà des coûts spécifiquement liés à l'étude sollicitée par l'acteur privé (**annexe 3**). A noter par ailleurs que la facturation des services ne peut être espérée qu'après plusieurs années de construction et fonctionnement des EDS.

⁶ La loi Organisation et Transformation du Système de Santé de 2019 n'a pas seulement voulu créer le Health Data Hub comme service à destination des acteurs dépositaires et utilisateurs des données de santé. Elle a aussi prévu l'élargissement du Système National des Données de Santé initialement résumé aux données médico-administratives de l'Assurance Maladie, de l'ATIH et de l'Inserm, à l'ensemble des bases financées par la solidarité nationale. Cette évolution témoigne d'une volonté forte de considérer les données de santé comme un patrimoine commun devant profiter à l'amélioration de la qualité des soins pour tous.

⁷ Les modalités de calcul de telles redevances bénéficieraient d'une harmonisation à l'échelle nationale, activité qui fera l'objet d'un travail dans un second temps.

Cette évolution réglementaire est en ligne avec le mouvement international de la science ouverte et, en France et en Europe, avec les textes déjà pris s'agissant des données non protégées par un secret (loi Lemaire, Directive open data de 2019, etc.). Le parti pris est donc bien que la donnée brute, qualifiée ou non, lorsqu'elle est détenue par un acteur public et que les étapes de collecte et de traitement ont été réalisées grâce à des financements publics, doit être la plus accessible possible sous réserve de respecter les conditions qui pourraient être requises par son caractère sensible (conditions de traitement dans le respect du RGPD par ex.).

Le contexte évoqué jusqu'ici n'empêche pas l'implication d'acteurs économiques privés dans le financement de projets d'utilisation de l'EDS dont ils seraient commanditaires ou partenaires mais la limitation à la couverture des coûts de mise à disposition et non complets est extrêmement limitante au regard des investissements requis pour développer et maintenir l'infrastructure, sans compter que les revenus issus de tels partenariats ne sauraient donc couvrir les coûts marginaux de projets conduits en interne ou par des acteurs publics. Par conséquent, un engagement de la puissance publique est nécessaire. Un tel raisonnement n'est d'ailleurs pas incohérent avec l'objectif premier des EDS d'une part à contribuer au pilotage des soins et d'autre part avec la vision portée par la puissance publique d'un Etat "plateforme" où les infrastructures sont gérées et financées suivant un modèle régalien tandis que les services qui pourraient s'appuyer sur elles peuvent être proposés par les acteurs économiques. Ainsi, les EDS rendent accessible le patrimoine national des données de santé financé par la solidarité nationale à des acteurs publics comme privés qui peuvent contribuer à la richesse économique du pays en les exploitant (développement de logiciels, dispositifs médicaux connectés, optimisation des essais cliniques etc.).

Pour autant, il ne fait aucun doute que les établissements de santé devront se saisir de l'opportunité laissée par le DGA de refacturer les coûts de mise à disposition auprès des demandeurs industriels et de développer leur stratégie partenariale avec les industriels au regard des enjeux en termes de réutilisation (développement logiciel pour améliorer la pratique, utilisation des données de vie réelle dans le cadre des essais cliniques innovants ou pour l'évaluation post mise sur le marché des produits de santé, etc.). Pour ce faire, la formulation de grilles de tarifs harmonisées permettrait de fluidifier les relations entre établissements et industriels. Cette étape devra faire l'objet d'un travail ultérieur du groupe de travail, avec l'implication de certains acteurs professionnels ayant été mis à contribution sur des problématiques similaires, comme le LEEM par exemple. Les méthodologies utilisées pour harmoniser les prestations des CRB et leurs coûts pourraient aussi être une base de travail.

IV. Les entrepôts de données de santé pourraient bénéficier de financements mixtes visant à assurer à la fois le fonctionnement et produire des incitations au partage .

Ce modèle n'est pas original dans la mesure où plusieurs exemples peuvent être trouvés y compris dans les mécanismes de tarification des activités de recherche clinique, les activités spécifiques telles que celles des centres de ressources biologiques (CRB, cf **annexe 4**), CIC, CRC, etc. Il permet de couvrir un fonctionnement cible de l'infrastructure mais de récompenser l'obtention de résultats dépassant ce fonctionnement, motivant ainsi une utilisation exemplaire et optimale de l'infrastructure.

Le groupe de travail recommande donc d'une part la mise en place d'une part fixe couvrant les coûts d'investissement, de fonctionnement et de maintien de l'EDS permettant l'accompagnement d'une masse critique cible de projets par établissement sur une année. La part fixe doit ainsi permettre de couvrir les coûts de cadrage d'un projet d'EDS, de mise en

place de l'EDS lui-même (« build ») et les coûts de fonctionnement (« run »), lesquels incluent : 1) les coûts incompressibles de maintien en conditions opérationnelles pour ne pas mettre en risque le fonctionnement de l'infrastructure à un niveau standard ainsi que 2) les coûts d'investissement permettant une intégration en continu de nouveaux flux de données au sein de l'EDS et 3) les coûts des activités support nécessaires à l'exploitation. Un certain nombre d'indicateurs pourront être prérequis afin de conditionner l'éligibilité de l'EDS à la part fixe du financement pérenne. On peut citer par exemple l'intégration dans l'EDS d'un périmètre minimal de données prévu dans la doctrine nationale, la conformité de l'EDS au référentiel de la CNIL, ou encore une gouvernance du projet conforme à la doctrine nationale. Un audit régulier de ces prérequis pourra être organisé. Le groupe de travail recommande d'autre part la mise en place d'une part variable qui viendrait soutenir et récompenser les initiatives à fort impact. L'enveloppe allouée à cette part variable pourrait être répartie entre les établissements sur la base de deux typologies principales d'indicateurs quantitatifs et qualitatifs portant sur les résultats et le fonctionnement de l'EDS :

- les indicateurs d'activité, tels que le nombre de partenariats ou de projets réalisés au-delà d'une certaine cible, le nombre de publications de haut rang, le nombre de bases partagées dans le catalogue du HDH, etc.
- les indicateurs de performance, qui permettent d'apprécier le niveau d'engagement des établissements dans l'optimisation des activités liées aux EDS. Ils pourraient par exemple porter sur le délai d'accès aux données, le niveau de qualité des données, de standardisation ou encore de mutualisation de ces dernières.

Ces deux grandes typologies d'indicateurs sont proposées par le groupe pour démarrer la réflexion et ont vocation à être définies avec davantage de précision après analyse du retour d'expérience des lauréats de l'appel à projets.

Pour s'adapter aux besoins de financement des établissements de santé et au niveau de maturité des EDS, le modèle proposé fera intervenir un dialogue de gestion pour évaluer le bon niveau de financement requis pour la part fixe sur la base de documentations et justificatifs adéquats.

Il est proposé de mettre en œuvre ce modèle de manière progressive avec la mise en place de la part variable au bout de quelques années.

Ce financement pourrait viser au démarrage, a minima, l'ensemble des CHU et une réflexion pourrait être menée sur un périmètre élargi ou la connexion d'autres types d'établissements aux CHU dotés d'EDS. En effet, financer en priorité les CHU à travers l'ensemble du territoire permettrait le déploiement rapide d'un premier réseau national d'EDS en capitalisant sur des ressources, infrastructures et capacités de recherche déjà existantes. La montée en compétences des CHU permettrait dans un second temps d'étendre ce réseau aux autres établissements de santé. Les contours précis d'un périmètre élargi du financement pérenne pourront être définis dans un travail ultérieur, nourris par le retour d'expérience des lauréats de l'appel à projets et constituer un des axes de travail de la doctrine nationale pour les EDS qui doit être définie en 2023⁸.

Les CHU devront néanmoins pour être bénéficiaires du financement vérifier un certain nombre de prérequis parmi lesquels : application de la gouvernance nationale d'accès aux données⁹, conformité réglementaire, conformité aux exigences de sécurité, mise en place

⁸ Les travaux relatifs à l'élaboration d'une doctrine nationale relative aux EDS s'inscrivent dans une logique de capitalisation sur les travaux lancés dans le cadre de l'appel à projets, en construisant notamment une meilleure connaissance du niveau d'avancement des établissements de santé en matière d'EDS et de leurs cas d'usage.

⁹ Les lois successives ont fait du partage des données de santé une règle que la Commission Européenne entend élargir à la sphère européenne. Créer et de constituer un patrimoine de données

d'un catalogue de métadonnées interopérable et d'un registre public des projets, mise en place de contrats types, partage des données au catalogue du HDH, etc. Ces derniers correspondent à la définition d'un EDS à laquelle aboutiront les travaux relatifs à la doctrine nationale relative aux EDS en 2023.

Cela représente dans un premier temps une dotation de 60 à 90 millions d'euros par an d'après la méthodologie de chiffrage explicitée en **annexe 2**. En effet, le groupe de travail a estimé en moyenne à 2 M€ par an le coût de développement d'un EDS sur 5 ans, puis à 3 M€ par an le coût de financement de son fonctionnement nominal. La dotation annuelle proposée permettrait alors de financer 30 EDS (un par CHU) sur l'ensemble du territoire (avec possibilité d'organisations mutualisées entre CHU et/ou avec d'autres établissements de santé partenaires en cible).

Pour garantir la réussite du dispositif proposé et d'une manière générale le déploiement des EDS sur le territoire, le groupe de travail préconise par ailleurs :

- que le comité stratégique des données de santé, en collaboration étroite avec la CNIL, s'attache à adresser les freins réglementaires qui rendent actuellement difficile l'accès aux données de santé en France pour les réutilisateurs et porteurs de projet (modalités d'information des patients, délais de délivrance des autorisations de traitement des données, etc.) ;
- que soit engagée une réflexion complémentaire concernant le maillage territorial cible des EDS, afin de favoriser des projets cohérents et en synergie les uns avec les autres ;
- que soit poussée l'idée de l'application d'une contrepartie systématique sous la forme de partage à l'obtention d'un financement public direct ou obtenu par le biais d'un appel à projets ;
- que des appels à projets puissent stimuler et dynamiser les projets de réutilisation des données de santé par le biais de subventions de recherche (voire un volet PHRC) ou en ciblant des innovations permettant d'optimiser le fonctionnement des EDS (outils de création de cohortes, outils de pseudonymisation, extraction ou standardisation de données en opensource et pouvant bénéficier en cible à toute la communauté).

Il apparaît donc intéressant de mettre en place dès 2023 un financement annuel pérenne des EDS de CHU au travers d'une enveloppe globale annuelle estimée de l'ordre de 60 à 90 M€. Il semble pertinent d'investir dans ces infrastructures qui pourront générer un grand nombre de projets d'innovation dont certains avec des acteurs privés susceptibles de générer d'importants bénéfices économiques à l'échelle de la société. Cet investissement devrait être conditionné au respect d'une doctrine nationale en matière de modalités de redevance, délai de mise à disposition des données pour assurer l'attractivité du dispositif dans un contexte international concurrentiel et de s'assurer du partage des bases de données au caractère réutilisable le plus largement possible et notamment dans le catalogue du HDH.

de santé interopérable, partageable et accessible est donc fondamental. Le SNDS au sens élargi renvoie à ce patrimoine, que le HDH a vocation à opérer, et dont les EDS font partie au plan juridique.

Glossaire

AAP	Appel à Projets
APHP	Assistance Publique - Hôpitaux de Paris
CHU	Centre Hospitalier Universitaire
CIC	Centre d'Investigation Clinique
CNIL	Commission Nationale de l'Informatique et des Libertés
CPRD	Clinical Practice Research Datalink
CRB	Centre de Ressources Biologiques
CRC	Centre de Recherche Clinique
DGA	Data Governance Act
DPI	Dossier Patient Informatisé
EDS	Entrepôt de Données de Santé
EMA	European Medicines Agency
FDA	Food and Drug Administration
HDH	Health Data Hub
IFAQ	Incitation Financière à l'Amélioration de la Qualité
PHRC	Programme Hospitalier de Recherche Clinique
RGPD	Règlement Général sur la Protection des Données
SASN	Stratégie d'Accélération Santé Numérique
SI-VIC	Système d'information pour le suivi des victimes d'attentats et de situations sanitaires exceptionnelles
SNDS	Système National des Données de Santé

Liste des Annexes

- Annexe 1A - Etat des lieux
- Annexe 1B - Recensement des cas d'usage d'EDS par la Commission Recherche et Innovation
- Annexe 2 - Méthodologie de chiffrage du coût d'un entrepôt de données de santé
- Annexes 3A et 3B - Contexte juridique - notes de Laurent Cytermann
- Annexe 4 - Modèle de financement pour les centres de ressources biologiques
- Annexe 5 - Lettre de mission du groupe de travail

Annexe 1A - Etat des lieux

1) Définition d'un entrepôt de données de santé

Les établissements de santé en France disposent d'une multitude de sources de données souvent très hétérogènes, non standardisées et dispersées dans les différents logiciels métiers utilisés. La tendance est à un manque d'information en interne et en externe au sein de l'hôpital sur les données disponibles et potentiellement exploitables issues de ce dernier. Face à ce constat, certains établissements ont pris l'initiative de déployer un EDS pour industrialiser le recueil et le rassemblement de données. Cette infrastructure permet à moyen terme l'interconnexion des multiples sources de données de l'hôpital et offre davantage de visibilité aux potentiels utilisateurs sur l'existant en matière de données via un catalogue des données de l'hôpital utilisables pour la recherche, l'innovation ou le pilotage. L'EDS offre également des opportunités de collaboration avec l'extérieur : académiques, industriels, autres établissements de santé en France ou à l'étranger.

Les registres ou cohortes correspondent - quant à eux - au rassemblement de données permettant d'adresser une question scientifique relativement ciblée (ou parfois plus générique mais toujours avec la logique de construire des données spécifiquement -par exemple la cohorte Constances), fondé sur des critères d'inclusion et un protocole strictement nécessaires et suffisants pour répondre à cette question, et à fortiori une durée d'inclusion finie dans le temps. Ils sont généralement considérés comme de qualité supérieure aux entrepôts de données de santé et constituent l'outil de référence pour la conduite des essais randomisés en double-aveugle, qui restent aujourd'hui le "gold standard" dans les essais thérapeutiques. De plus, l'inclusion de patients dans des projets de recherche offre des opportunités si différentes du soin (par exemple la possibilité de faire revenir un patient à rythme régulier, en le soumettant à une batterie d'exams dédiés à chaque passage).

Toutefois les EDS ont des atouts complémentaires, faisant qu'ils ont toute leur place dans le dispositif. En effet, contrairement à la plupart des registres, ils reposent sur des procédés d'intégration davantage automatisés - ne requérant ni saisie ni retour aux dossiers. Des données variées (données structurées ou non structurées, données d'imagerie, de pathologie, de signal physiologique, etc.) couvrant des effectifs de grande taille rendent possibles de nouveaux types de recherche et d'innovations parmi lesquels :

- Mieux détecter les symptômes d'une pathologie
- Faire un suivi prédictif de l'évolution d'une maladie
- Formuler des propositions thérapeutiques plus personnalisées
- Optimiser une prise en charge à l'aide un jumeau numérique
- Mettre en évidence des signaux d'alerte d'émergence de pathologies
- Suivre en population générale, ou encore dans des sous-populations spécifiques non toujours incluses dans les essais ou les cohortes, les impacts et les effets des prise en charge

En cible, les EDS peuvent également devenir un socle incontournable à la création de grandes cohortes, puisqu'ils permettent le pre-screening dans des proportions sans comparaison (l'intégralité de la cohorte hospitalière pouvant être passée au crible). Ils devraient par ailleurs devenir rapidement une composante à part entière des cohortes, dans le cadre des essais cliniques dits "du futur" (ex. essais adaptatifs, multi-bras et multi-étapes, etc.).

Les EDS sont construits sur l'idée de réutiliser des données d'ores et déjà présentes. Il n'y a donc pas de coût d'acquisition initial à prévoir. En revanche, les étapes de retraitement de la donnée permettant sa réutilisation représentent une charge qui incombe aux établissements de santé, et donc un coût.

Les EDS sont donc complémentaires des registres ou cohortes et permettent de poursuivre le double objectif suivant :

- la réalisation d'études et recherches pour lesquelles les données réunies dans l'EDS sont suffisantes
- l'automatisation pour partie de la collecte pour la constitution de cohortes et registres lorsqu'une plus grande précision est requise. Cette automatisation se traduit par une économie en termes de coûts mais également en termes de temps.

Comme introduit plus haut, l'EDS représente également en marge des usages orientés recherche un grand potentiel de transformation et d'innovation dans le pilotage des activités et ressources à l'hôpital. Au-delà du pilotage de crise où les EDS ont pu illustrer tout leur intérêt, ou encore du pilotage des activités de flux (comme les services d'urgence, qui de longue date s'appuient sur des tableaux de bord permettant un niveau de prédiction satisfaisant des arrivées, et facilitant ainsi l'organisation des équipes), chaque spécialité médicale, chaque type de soins spécifiques, chaque contexte médical ou de soin gagnerait à s'appuyer sur des tableaux de bord dédiés. Enfin les EDS sont susceptibles de favoriser un pilotage des soins par la valeur (cf. ex.: automatisation et publication des indicateurs requis pour le dispositif IFAQ, aide au repérage des infections du site opératoire pour les EOH, aide à l'évaluation de l'usage des anti-infectieux, tableaux de bord issus des démarches VBHC, etc).

2) Fonctionnement d'un entrepôt de données de santé

Pour l'utilisateur interne comme externe à l'hôpital, l'EDS permet en cible d'explorer, déposer une demande d'accès et traiter les données avec simplicité. Ci-dessous un exemple de parcours utilisateur d'un EDS :

- Exploration. À l'aide d'un catalogue de métadonnées, le porteur de projets peut explorer les données de l'EDS et voir si son projet est faisable et quelles données il pourra mobiliser.
- Remplissage d'un formulaire d'accès. Une fois le projet cadré, le porteur de projet pourra caractériser sa demande d'accès en remplissant un formulaire de demande d'accès et en le déposant auprès de l'équipe de l'EDS qui sollicite l'avis d'un comité éthique et scientifique. Ces démarches de demande d'accès peuvent s'accompagner, notamment pour les projets sortant de cadre de la méthodologie de référence MR004 de la CNIL, d'un dépôt de dossier auprès de la CNIL qui, bien souvent, allonge les délais d'accès aux données pour le porteur.
- Contractualisation. Une fois l'avis favorable formulé, le porteur de projet pourra contractualiser avec l'établissement supportant l'EDS, condition préalable pour se voir autoriser à accéder aux données dans le respect des règles de sécurité, d'utilisation et de valorisation établies par l'établissement de santé et dans le respect du cadre légal et de la gouvernance nationale des données de santé. Pour cette étape seront utilisés des contrats type autant que possible.
- Mise à disposition et exploitation (pour analyse, élaboration d'un modèle, mise au point et entraînement d'un algorithme, etc.). Les données peuvent ensuite être mises

à disposition via l'EDS ou selon les moyens négociés dans le contrat. Selon les cas, le porteur du projet peut travailler de concert avec des experts de l'EDS pour conduire et fiabiliser sa recherche .

- Transparence et valorisation. Tout en s'efforçant de favoriser la protection de la propriété intellectuelle de l'établissement et de ses partenaires (dépôt de brevets, de logiciels, notamment), l'établissement rend publique l'existence et la finalité du projet sur le registre de l'EDS, il sera complété des éléments de résultats le moment venu conformément au code de la santé publique. Les résultats devront citer et remercier les équipes selon les modalités du contrat.

3) Déploiement des EDS sur le territoire

La mise en œuvre d'un EDS présente d'importants défis et mobilise de multiples acteurs au sein de l'établissement de santé. Pour mieux comprendre les activités à mener par chacun, le projet peut être découpé en 3 grandes phases :

- Une phase de cadrage pour définir les modalités de pilotage du projet d'EDS, la mise en conformité réglementaire ainsi que les modes de fonctionnement cible au sein de l'établissement de santé (gouvernance, organisation..) ;
- une phase de déploiement pour définir les solutions techniques envisagées pour répondre aux exigences techniques et fonctionnelles de l'EDS, le circuit des données, les standards employés, ainsi que les modalités prévues pour assurer la sécurité de l'EDS ;
- une phase de valorisation pour définir les modalités de partage des données avec l'écosystème, l'articulation avec le HDH ainsi que les offres de service à destination des futurs utilisateurs de l'EDS.

La mise en place des entrepôts de données de santé en France remonte à la fin des années 2000 et s'accélère à la fin des années 2010. Les autorisations réglementaires par la CNIL (à partir de 2017) succèdent à des premiers travaux techniques plus anciens. L'initiative de la construction des EDS vient très souvent de l'intérieur de l'organisation avec un porteur individuel qui convainc progressivement les autres parties prenantes de l'intérêt de la réutilisation des données. La pérennisation de l'EDS est accompagnée par la construction d'un environnement de coopération entre les différents acteurs (DIM, DSI, DRCI, chercheurs et usagers cliniciens) avec le soutien de la direction et de la Commission Médicale d'Établissement. Elle est accompagnée également par la constitution d'une équipe dédiée au maintien et à l'opérationnalisation de l'EDS ainsi qu'à leur exploitation.

Dans le cadre d'un état des lieux réalisé en Juin 2022, la Commission Innovation et Recherche de la conférence des DG de CHU a dénombré 11 établissements de santé CHU/CHR s'étant dotés d'un EDS structuré et autorisé (ou ayant déclaré sa conformité au référentiel CNIL) : AP-HP, CHU Lille, CHU Rouen, CHU Tours, CHU Nantes, CHU Rennes, CHU Poitiers, CHU Angers, CHU Grenoble, CHU de Brest, CHU de Reims. 19 établissements de santé ont d'autre part déclaré être en cours de structuration pour répondre au référentiel CNIL. Les écarts qui ont pu être constatés par la CRI dans les moyens dédiés aux projets d'EDS s'expliquent en grande partie par des niveaux de maturité différents ou des tailles d'établissement différentes.

En marge de ces initiatives locales, des projets coordonnés à l'échelle régionale / interrégionale et entre établissements ont également vu le jour. C'est le cas du GCS HUGO qui, conscient des enjeux autour de la structuration et du partage de la donnée de santé, a lancé depuis 2016 un programme pluriannuel visant à structurer et organiser l'exploitation des données massives en santé dans l'ensemble des 5 CHU et du CLCC du territoire du

Grand Ouest. Cette stratégie repose sur la coopération et les liens de confiance qui animent et rassemblent les établissements partenaires du GCS HUGO et qui se traduisent par la réalisation de projets communs financés par le GIRCI, la création d'infrastructures pour le partage des données sur projets spécifiques et de ressources financés par les 6 établissements du GCS participant au Ouest Data Hub. Avec le GCS HUGO (5 CHU), deux autres réseaux, G4 (4 CHU) pour le territoire du Nord Ouest et PAISaGE (5 CHU et 1 CHR) pour l'interrégion Grand Est - Bourgogne Franche Comté, coopèrent activement et adoptent des stratégies similaires en les adaptant à leurs propres organisations afin de mener des travaux collaboratifs et encourager une animation scientifique synergique.

4) Exemple : EDS de l'AP-HP

Le groupe a décidé de revenir plus en détail sur l'exemple de l'EDS de l'AP-HP car ses porteurs auditionnés par les membres du GT ont pu fournir des indicateurs d'activité et de résultat pertinents pour la communauté, permettant de retranscrire l'état d'avancement du projet d'entrepôt et son développement sur les dernières années. Cet exemple illustre par ailleurs les impacts potentiellement bénéfiques des EDS à travers le rôle renforcé qu'a joué cet outil pendant la crise Covid.

L'AP-HP a mis en place de sa propre initiative et grâce aux décisions stratégiques d'investissement de son directeur général le premier EDS autorisé en janvier 2017 par la CNIL, aboutissement d'un chantier engagé dès 2014, et doté d'une gouvernance ad hoc, incluant largement la communauté médicale.

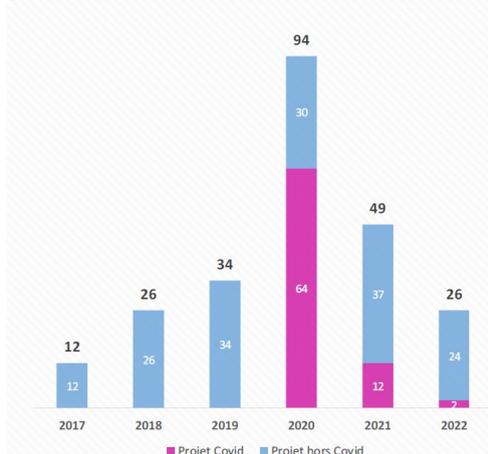
L'[EDS de l'AP-HP](#) intègre aujourd'hui les données de santé de plus 14 millions de patients admis dans les 39 établissements de l'AP-HP (données démographiques, données cliniques : diagnostics, actes, résultats d'examens biologiques, prescriptions médicamenteuses, examens d'imagerie, comptes rendus cliniques, etc.).

En matière de pilotage, l'exploitation de l'EDS participe aux travaux classiques de suivi de l'activité hospitalière mais également, de manière croissante, à l'outillage des équipes de soins pour le suivi de la qualité de la prise en charge de leurs patients (automatisation et publication des indicateurs requis pour le dispositif IFAQ, aide au repérage des infections du site opératoire pour les EOH, aide à l'évaluation de l'usage des anti-infectieux, tableaux de bord issus des démarches VBHC, etc.). **Sur la dernière année, plus de 50000 rapports ont ainsi été obtenus par près de 700 utilisateurs différents, à partir de l'EDS.**

A ce jour, 241 projets ont été déposés au Comité Scientifique et Éthique (CSE) de l'EDS. Le nombre de projets déposés auprès du CSE augmente fortement chaque année. 49 projets ont été déposés en 2021, après un pic à 94 projets déposés en 2020 lié au covid¹⁰.

¹⁰ Sur une seule maladie - le Covid - le nombre de projets s'est élevé à 64 rien qu'en 2020. Cela illustre l'appétence des chercheurs pour la recherche observationnelle sur données rétrospectives lorsque l'accessibilité et la préparation des données est elle-même accrue.

Nombre de projets déposés au CSE (Total = 241) :



34 publications scientifiques (dont 9 de catégorie A) sont issues de l’EDS, et seuls 6% des projets sont terminés.

Par ailleurs, 55 collaborations ont conduit à la signature d’un contrat ou sont sur le point de le faire. Parmi elles, 17 sont relatives à des partenariats avec des institutions privées (industrie pharmaceutique, start-up de la French Health Tech, ...).

A travers ces partenariats engagés ou en cours d’exploration, ou à travers ceux mis en place pour des appels à projets, ce sont de multiples nouveaux usages et le développement de nouvelles pratiques qui s’esquissent, aussi divers que :

- Le pré-screening / études de faisabilité (identification de site susceptible d’inclure des patients dans des essais cliniques, outils d’aide au recrutement) fondés sur l’EDS,
- La constitution facilitée de registres ou de cohortes (et leur rafraîchissement ultérieur), et le peuplement initial d’eCRF pour des essais cliniques ou pour le suivi post autorisation de mise sur le marché, le chaînage de données du soin avec des bases de données de recherche,
- Le partage ou mobilisation de données au sein de réseaux de recherche,
- Le suivi de l’émergence de maladies infectieuses ou de certaines pathologies,
- Etc.

Sur ces différents aspects, l’offre de service de l’EDS AP-HP présente aujourd’hui des degrés de maturité divers, et les chantiers des 5 prochaines années devraient permettre des avancées significatives dans ces domaines.

L’AP-HP travaille en continu à enrichir son EDS de nouveaux flux de données et faciliter leur exploitation grâce à la mise en œuvre d’outils et de partenariats (tels que le lancement de projets pilotes avec le Health Data Hub, la mise en place du laboratoire commun AP-HP – Inria Daniel Bernoulli, etc.). L’objectif est de consolider l’offre de service pour la recherche appliquée en santé et pour l’innovation (notamment évaluation d’algorithmes et autres technologies de santé (HTA)), ainsi que pour le pilotage de l’activité hospitalière.

En pratique, ces chantiers mobilisent aujourd’hui au quotidien au niveau de la seule Direction des Services Numériques (DSN) de l’AP-HP plus de 70 équivalents temps plein (hors appui prestataires) aux expertises spécifiques et diverses (data scientists, médecins experts en informatique médicale, architectes, data managers, experts données, data engineers, experts en cybersécurité, experts en protection des données, etc.). Au sein de la Direction de la Recherche Clinique et de l’Innovation (DRCI) de l’AP-HP et de manière complémentaire avec les équipes de la DSN, une trentaine de personnes interviennent et apportent une expertise

scientifique, méthodologique et réglementaire : aide à la conception des projets de recherche, méthodologie, data management, data science, montage de projets, dépôt réglementaire, juridique, négociation des partenariats dont valorisation, animation des instances et coordination.

Parallèlement à la maintenance des services en place, les travaux prévus pour les trois à cinq prochaines années visent ainsi à assurer la phase d'industrialisation, qui inclut en particulier les chantiers suivants :

- La finalisation de l'intégration de données de nos systèmes d'information hospitaliers, selon le planning d'intégration défini,
- Le développement de flux issus d'applications tierces (objets connectés par exemple) et de données d'exposition (données de pollution de l'air par exemple) pour enrichir l'entrepôt notamment les données générées dans le cadre de la recherche ou du télésuivi des patients de l'AP-HP,
- L'amélioration de la qualité des données pour les rendre exploitables, les SI sources n'étant pas destinés au pilotage ou à la recherche.
- La poursuite de la modélisation, de l'enrichissement et de la transformation des données sur les standards internationaux HL7/ OMOP,
- Le développement d'API FHIR,
- La mise en place d'outils pour faciliter l'usage des données par les utilisateurs pour la recherche sur données, pour l'optimisation de la recherche clinique, l'évaluation de technologies innovantes et le pilotage de la qualité des soins,
- La mise au point d'outils pour assurer la sécurisation des données, avec notamment des techniques de chiffrement, et la totale traçabilité des accès et des actions réalisées,
- La stabilisation de l'infrastructure (interne ou déléguée à des hébergeurs externes agréés pour s'assurer que les performances de calcul soient satisfaisantes),
- Le développement de l'accompagnement méthodologique, médico-scientifique et technico-réglementaire des porteurs de projets de recherche sur données et d'innovation,
- La montée en compétences et en expertise des équipes en charge de l'accompagnement des porteurs de projets (nouvelles méthodologies d'essais cliniques, jumeaux numériques et modélisation, approche multi-échelle),
- La définition de l'offre de services à destination des acteurs académiques et industriels,
- L'ouverture vers un espace européen unique des données de santé en vue de leur partage et leur mise en commun.

Annexe 1B - Recensement des cas d'usage d'EDS par la Commission Recherche et Innovation

Commission Recherche & Innovation

Enquête sur les cas d'usage en lien avec les entrepôts de données de santé

Assistance publique – Hôpitaux de Paris

Cas d'usage interne : CSE-20-63_PROCOVID - Validation des scores pronostiques publiés pour l'infection à SARS-CoV-2 sur la cohorte des patients hospitalisés à l'AP-HP pour ce motif

Résumé du projet et données exploitées

De nombreux scores de risque ont été publiés pour prédire l'évolution défavorable d'une infection à SARS-Cov-2 (responsable de la maladie Covid-19) et adapter ainsi la prise en charge du patient.

Avant de pouvoir utiliser ces scores en pratique clinique, il faut s'assurer que leur valeur prédictive est reproductible dans un échantillon de patients différent de celui ayant servi à leur dérivation. Le but de la recherche sera donc d'analyser la capacité des scores de risque à prédire une évolution défavorable de la Covid-19 sur la cohorte de patients pris en charge à l'AP-HP pour ce motif.

Données exploitées :

- Données démographiques : sexe, âge
- Données de prise en charge (service, etc.) : passages au cours de la visite index
- Données PMSI (codes CIM-10 de la visite index et des visites de l'année précédente) pour la sélection des patients (Covid-19) et pour les antécédents et comorbidités : HTA, cardiopathie ischémique, fibrillation atriale, insuffisance cardiaque, coronaropathie, diabète, hépatopathie, cancer, asthme, bronchopneumopathie obstructive, emphysème, asthme, maladie cérébrovasculaire, infection VIH
- Données physiologiques à l'admission (aux urgences) : pression artérielle, fréquence cardiaque, saturation en oxygène, fréquence respiratoire, température corporelle
- Résultats des examens biologiques et microbiologiques : albumine, lymphocytes, CRP, PCT, ASAT, urée, LDH, IDR (indice de distribution des globules rouges), bilirubine directe, PCR Sars-Cov-2
- Comptes rendus de radiologie (TDM thoracique) : images évocatrices de la Covid-19
- Comptes rendus médicaux du séjour et de l'année précédente (comptes rendus d'hospitalisation, comptes rendus de passages aux urgences) pour les antécédents et comorbidités : HTA, cardiopathie ischémique, fibrillation atriale, insuffisance cardiaque, coronaropathie, diabète, hépatopathie, cancer, asthme, bronchopneumopathie obstructive, emphysème, asthme, maladie cérébrovasculaire,

infection VIH pour les éléments du tableau clinique : troubles de conscience ou de vigilance, dyspnée

Acteurs internes et/ou académiques mobilisés

Acteurs internes uniquement.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Lombardi, Y., Azoyan, L., Szychowiak, P. et al. External validation of prognostic scores for COVID-19: a multicenter cohort study of patients hospitalized in Greater Paris University Hospitals. *Intensive Care Med* 47, 1426–1439 (2021).

<https://doi.org/10.1007/s00134-021-06524-w>

Cas d'usage dans le cadre d'un partenariat avec un industriel :
CSE-19-25_MELANOME - Analyse des facteurs prédictifs de réponse et de toxicité des traitements innovants (thérapies ciblées et inhibiteurs de checkpoint du système immunitaire) dans le mélanome

Résumé du projet et données exploitées

Le traitement du mélanome avancé a été révolutionné depuis 2010 avec l'avènement des inhibiteurs de checkpoint immunitaires (anti-CTLA4 et anti-PD1) et des thérapies ciblées (inhibiteurs de BRAF, MEK, KIT) au stade métastatique et plus récemment en situation adjuvante. Alors que la survie médiane était de 6 mois, actuellement 60% des patients sont en vie à 3 ans. Mais ces traitements s'accompagnent d'un risque d'effets secondaires potentiellement graves (50% sous combinaison anti-PD1 et anti-CTLA4) et parfois même irréversibles. La recherche de facteurs prédictifs de réponse thérapeutique, qui permettraient d'épargner un risque de toxicité à certains patients, et d'effets secondaires graves, est un enjeu majeur de recherche.

Nous étudierons la cohorte de patients atteints d'un mélanome avancé ayant reçu un traitement par anti-PD1 et/ou anti-CTLA4, et/ou inhibiteurs de BRAF, MEK, KIT à l'AP-HP.

L'objectif est d'analyser cette cohorte à travers les caractéristiques cliniques, biologiques, radiologiques et moléculaires des patients, en s'aidant des outils d'intelligence artificielle, dans le but d'identifier les facteurs prédictifs de réponse et de toxicité sévère, et d'identifier la meilleure séquence thérapeutique possible en termes de réponse et de survie.

Données cliniques :

- Caractéristiques démographiques
- Caractéristiques tumorales (histoire du mélanome, stade, site métastatique dont métastases cérébrales, autres traitements reçus pour le mélanome, radiothérapie)
- Données non triées (compte-rendu) pour analyse avec les outils d'IA
- Données biologiques biologie standard (logiciel STARE + compte rendu dans middlecare)
- Données radiologiques : images de TDM, PET-TDM, IRM avant traitement et sous traitement
- Données moléculaires : résultat génotypage BRAF, NRAS, KIT, NGS si disponible, PDL1
- Données histologiques : compte rendu histologique (mélanome, reprise, et tout prélèvement (tumoral ou non) après instauration du traitement), lames virtuelles (à terme)

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

- Collaboration Owkin – APHP
- Méthodes d'apprentissage fédéré

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

- Mise en place d'une solution fédérée
- Soumission d'un article en oncodermatologie
- Entraînement fédéré entre l'AP-HP et le CHU de Nantes à venir.
- Extension à prévoir pour des résultats plus probants.

CHU d'Amiens et de Lille

Cas d'usage dans le cadre d'un partenariat avec un industriel : détection automatique d'amylose cardiaque scintigraphique à l'aide du machine learning

Résumé du projet et données exploitées

Utilisation d'un programme basé sur l'intelligence artificielle (*deep learning*) permettant de détecter automatiquement l'hyperfixation cardiaque en scintigraphie osseuse. Cette hyperfixation cardiaque est pathognomonique de certaines maladies graves, méconnues et potentiellement sous-diagnostiquées.

Données exploitées : images (>30 000 scintigraphies osseuses), données cliniques structurées (PMSI) et non structurées (comptes rendus de scintigraphie), disponibles dans l'EDS à compter de 2008.

Acteurs internes et/ou académiques mobilisés

- Equipe Include (EDS) du CHU Lille
- Equipe EDS du CHU d'Amiens
- Dr Delbarre (CHU Amiens)

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

Sociétés Codoc, Deemea et Pfizer

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

L'objectif est d'utiliser des images scintigraphiques accessibles à partir de l'EDS du CHU de Lille afin de rechercher *a posteriori* des cas de ces maladies.

CHU d'Angers

Cas d'usage interne : Impact des facteurs de risque métabolique et prédictiOn du risque de cOmplication hépatIque dans les maladies chroNIques du foiE - Etude TATOOINE

Résumé du projet et données exploitées

Au même titre que la consommation excessive de boissons alcoolisées, les facteurs de risque métabolique (surpoids/obésité, diabète, hypertension artérielle, dyslipidémie) peuvent favoriser la survenue de la cirrhose et du cancer du foie. Néanmoins, parmi tous ces facteurs, on ne sait pas encore très bien lesquels occasionnent le plus de risque pour les patients. Cette étude a pour objectif d'identifier ces facteurs et ainsi de définir le profil des

patients les plus à risque de développer une cirrhose et/ou un cancer du foie. L'intérêt ensuite sera de réaliser des interventions de prévention chez ces patients afin de leur éviter la cirrhose et le cancer du foie.

Données utilisées : PMSI local, résultats de biologie, Fibroscan (examen mesurant la dureté hépatique), poids et taille, statut vivant/décédé à l'hôpital et après (base nationale ouverte des personnes décédées)

Acteurs internes et/ou académiques mobilisés

Les cinq centres hospitaliers universitaires du Grand-Ouest (Angers, Brest, Nantes, Rennes, Tours). La base de données ainsi obtenue (incluant environ 390 000 patients) permettra de définir les facteurs de risque de cirrhose et de cancer du foie, mais également de développer des équations permettant de prédire le risque pour chaque patient. Ces équations pourront ensuite être utilisées par les médecins chez les patients à risque pour identifier ceux qui nécessitent une prise en charge plus spécifique.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Les bénéfices attendus du projet TATOOINE sont multiples. Une meilleure identification des facteurs de risque de complication hépatique permettra de mieux cibler le sous-groupe de patients dans lequel des procédures de dépistage pourront être envisagées.

Le développement d'équations de prédiction du risque individuel apportera des solutions pour évaluer le pronostic chez ces patients à risque, améliorant ainsi la prise en charge individuelle.

Actuellement beaucoup de patients présentant des anomalies biologiques hépatiques non spécifiques, notamment dans un contexte de facteurs de risque hépatique, sont adressés pour un bilan spécialisé alors que leur risque de complication hépatique est très faible. Ceci occasionne une surcharge des consultations spécialisées et un surplus de bilans coûteux alors que la seule conduite à tenir est l'observance des règles hygiéno-diététiques (réduction de la consommation d'alcool, exercice physique, régime alimentaire, perte de poids), sans nécessité d'une prise en charge spécialisée.

D'un point de vue santé publique, une meilleure identification des patients à risque organisera la prescription des examens complémentaires, améliorera le parcours patient des soins primaires vers les spécialistes et, *in fine*, optimisera l'utilisation des ressources de notre système de santé.

Enfin, les résultats obtenus apporteront des informations très utiles pour élaborer les modèles dans les futures études qui évalueront le coût-efficacité de la prise en charge.

CHU de Besançon

Cas d'usages internes dans le cadre d'un partenariat avec un industriel

Résumé des projets et données exploitées

Dans le cadre de collaborations avec des industriels, le CHU de Besançon est amené à monter des projets en amont des investigations portant sur leurs dispositifs médicaux, ceci afin de renforcer les connaissances sur des sujets spécifiques pour lesquels un consensus dans la définition d'une pathologie n'est pas tout à fait partagé, l'incidence variable d'un établissement de soin à un autre, ou encore lorsqu'une problématique peut être documentée

de manière plus détaillée grâce à l'introduction récent d'un codage CCAM. Des cas d'usage peuvent ainsi être illustrés par :

- l'étude en cours PNOrate
<https://clinicaltrials.gov/ct2/show/NCT05259293?term=PNOrate&draw=2&rank=1>
pour l'étude des pneumothorax consécutifs des biopsies pulmonaires,
- une étude actuellement en montage « Limbo » pour l'étude des reconstructions mammaires par lipomodélage (CCAM « autogreffe de tissus adipeux au niveau du sein »).

Acteurs internes et/ou académiques mobilisés

Le centre d'investigation clinique Inserm CIC 1431 et notamment sa plateforme d'évaluation clinique des dispositifs médicaux « DinaMic » orchestre ces deux sujets, avec une implication forte de l'unité de méthodologie « uMETH », du DIM, et des médecins spécialistes concernés.

Pour le sujet PNOrate les centres investigateurs sont le CHU de Besançon, l'Institut Bergonié (Bordeaux), le CHU de Tours, le Saint James Hospital (Dublin, Ireland).

Promoteur : CHU Besançon.

Pour le sujet « Limbo » concernant la chirurgie reconstructrice, les centres investigateurs sont le CHU de Besançon et l'hôpital nord Franche-Comté, le promoteur le CHU de Besançon.

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

Un partenaire industriel irlandais porte le projet associé à l'étude PNOrate, le financement du projet étant assuré par l'obtention de fonds européens EIT Health. L'Inserm et le CHU de Besançon sont partenaires du projet en charge de la stratégie d'évaluation clinique.

Le sujet de chirurgie reconstructrice est financé par un industriel bisontin, qui développe donc des dispositifs médicaux sur cette thématique. Le CHU de Besançon est en charge de monter les investigations cliniques.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Pour le premier sujet PNOrate les résultats espérés permettront une meilleure connaissance des pneumothorax liés aux biopsies pulmonaires, des pratiques associées pour leur prise en charge. La future première étude chez l'homme du dispositif médical développé par l'industriel bénéficiera de ces précisions.

Pour le sujet « Limbo », la documentation de la problématique introduira un second volet du projet collaboratif, sous forme d'une investigation clinique sur le dispositif médical (qui aura déjà le marquage CE) ; les résultats espérés permettront un positionnement de l'innovation par rapport aux pratiques actuelles.

CHU de Bordeaux

Cas d'usage interne : suivi de la Covid-19

Résumé du projet et données exploitées

Dans le contexte de la pandémie de Covid-19, deux enjeux majeurs sont rapidement apparus : (i) suivre et anticiper l'évolution de l'épidémie en temps réel afin d'adapter les organisations au sein des établissements de santé en fonction de l'évolution des différentes vagues épidémiques ; (ii) produire de la connaissance sur cette épidémie émergente. Dans ce contexte, le CHU de Bordeaux a exploité son EDS pour permettre la mise en œuvre

d'indicateurs automatiques de suivi de l'épidémie basés sur les données de biologie, des urgences, du SAMU et d'hospitalisation mise à jour quotidiennement.

Ces données ont également été utilisées pour concevoir un modèle de prédiction des hospitalisations à 14 jours, incluant également des données ouvertes régionales (SIDEP et données météorologiques). Par ailleurs, bénéficiant de l'architecture en place, le CHU de Bordeaux contribue au sein du consortium 4CE (<https://covidclinical.net/>) pour produire des connaissances sur la pandémie à partir des données hospitalières.

Acteurs internes et/ou académiques mobilisés

- Service d'information médicale du CHU de Bordeaux, cellule de crise Covid-19 du CHU de Bordeaux
- Centre Inserm U1219 BPH (équipe SISTEM, équipe ERIAS)
- Directions : direction des Systèmes d'information et du numérique, DRCI, direction Qualité et Sécurité

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Les indicateurs conçus sont utilisés en routine par la cellule de crise pour suivre l'évolution de l'épidémie, permettant ainsi une vision globale de la situation. Le modèle de prédiction a été récemment introduit comme une information supplémentaire pour anticiper les évolutions et guider les décisions. Le consortium 4CE a permis de produire des connaissances sur la pandémie en incluant les données agrégées de plus de 300 hôpitaux, en produisant à ce jour 11 publications scientifiques avec des premiers résultats dès mai 2020.

Cas d'usage interne dans le cadre d'un partenariat avec un industriel : PROSIT

Résumé du projet et données exploitées

Dans le contexte du développement de la pertinence des soins au sein du CHU de Bordeaux, le projet PROSIT coordonné par la pharmacologie médicale du CHU de Bordeaux avec l'encadrement méthodologique de l'UMES (service d'information médicale) a été retenu dans le cadre de l'appel à Projet DGOS-PREPS 2019.

L'objectif de ce projet est de détecter et d'analyser, via des méthodes de *machine learning* et de *text mining*, les interactions médicamenteuses potentielles dans les ordonnances de sorties de l'hôpital, puis d'évaluer l'impact de l'utilisation d'une application restituant les résultats des interactions médicamenteuses potentielles détectée pour réduire ce risque d'interactions. Ce projet, conduit avec les CHU de Bordeaux, de Rennes et de l'HEGP, est également réalisé en partenariat avec une start-up *spin off* du CHU de Bordeaux.

Acteurs internes et/ou académiques mobilisés

- Service d'information médicale du CHU de Bordeaux (UMES, IAM), service de pharmacologie médicale du CHU de Bordeaux, centre Inserm U1219 BPH (équipe ERIAS)
- Directions : DRCI

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

Projet conduit avec une *start-up spin-off* du CHU de Bordeaux, de l'Université de Bordeaux et de l'Inserm qui fournit aux professionnels de santé et aux patients une information fiable et utile sur les médicaments, dans l'objectif de prévenir les risques médicamenteux. Le projet a été déposé dans les suites de travaux antérieurs qui ont conduit à la création de la société. Dans ce projet, la start-up met à disposition sa brique logiciel de détection des interactions

au sein de l'architecture d'exploitation des ordonnances de sorties des différents établissements de santé impliqués dans le projet.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Ce projet est encore en cours. Il doit permettre la mise à disposition d'un ensemble d'indicateurs de pertinence concernant l'existence d'ordonnances de sortie à risque d'interaction médicamenteuses. Ces indicateurs seront restitués au sein d'une application dédiée à destination des professionnels de santé. Par ailleurs, une évaluation de l'impact de ces restitutions sur les interactions futures, ainsi que de l'acceptabilité et l'utilisabilité de tels indicateurs par les équipes est prévu dans le cadre de ce projet.

CHU de Brest

Cas d'usage interne : Surveillance des infections du site opératoire - EOH

Résumé du projet et données exploitées

Surveillance des Infections du Site Opératoire (ISO) suite aux interventions chirurgicales. Croisement des données du laboratoire de microbiologie avec les actes chirurgicaux. Ces travaux permettent de contribuer au Réseau d'alerte, d'investigation et de surveillance des infections nosocomiales (Raisin) en fournissant la matière première à l'EOH qui retourne aux dossiers pour qualifier médicalement les cas. L'extension de ces travaux en 2022 a permis de participer au programme SPICMI (Surveillance et prévention du risque infectieux en chirurgie et médecine interventionnelle) coordonné par Santé Publique France. Il s'agit notamment de prendre en compte les reprises chirurgicales et les ré-hospitalisations ultérieures

Acteurs internes et/ou académiques mobilisés

Equipe de l'EOH (Pr Saliou) et du CDC (M. Happe)

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Les résultats préliminaires ont permis de répertorier les prélèvements positifs ainsi que les reprises médicales et ré-hospitalisations sur le premier semestre 2021 et les analyses en cours permettront de mettre en place les actions correctrices appropriées.

Cas d'usage dans le cadre d'un partenariat externe : collecte de données en vie réelle

Résumé du projet et données exploitées

Dans le cadre des travaux d'Irène Frachon autour du Médiateur, la recherche textuelle a été développée au sein du CHU. Ces premiers travaux ont été la base pour l'obtention en 2015 d'un AAP ANSM permettant de développer des outils de fouille au sein de l'entrepôt de données de santé avec pour finalité la collecte des données de pharmacovigilance qui peuvent être croisées avec les notifications réalisées par les soignants.

Acteurs internes et/ou académiques mobilisés

Ansm

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

La collecte des données a permis d'abonder les cohortes ci-dessous :

- Cohorte « SECUKINUMAB-USTEKINUMAB », 539 patients concernés à Brest
- Cohorte « IBRUTINIB », 87 patients concernés à Brest
- Cohorte « BOSUTINIB », 34 patients concernés à Brest

Ces travaux ont permis de colliger des données de vie réelle sur des traitements innovants pour lesquels aucune information n'était disponible permettant ainsi de mettre en place un suivi de pharmacovigilance au fil de l'eau en plus de celui réalisé par les soignants eux-mêmes via les notifications.

CHU de Dijon

Cas d'usage interne : étude de la prise en charge des leucémies aiguës myéloïdes dans le département de la Côte-d'Or

Acteurs internes et/ou académiques mobilisés

- Registre des hémopathies malignes de Côte d'Or
- Plateforme Hospitalo-Universitaire de Biologie

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

La survie à 5 ans des patients atteints de LAM est très variable passant de 20 à 55%. Le sous-type de la LAM et l'âge sont des éléments importants influant sur la survie mais plusieurs autres facteurs, non encore élucidés, jouent un rôle très probable. Il s'agit du délai de prise en charge, du type de structure où se fait la prise en charge, du schéma thérapeutique adopté mais aussi peut être du niveau socio-économique du patient ou de son éloignement d'un centre de prise en charge. Notre étude cherche à identifier quels sont les facteurs qui influent sur la survie des LAM dans la population de Côte-d'Or.

CHU Grenoble Alpes

Cas d'usage interne : ODIASP - L'intelligence artificielle au service de l'Optimisation du DIAGNOSTIC de la SarcoPénie via la détermination automatisée de l'index de surface musculaire en imagerie TDM

Résumé du projet et données exploitées

La sarcopénie constitue une vraie problématique de santé publique avec de graves conséquences. Pourtant, son diagnostic systématique reste complexe en pratique clinique routinière : les méthodes classiques pour mesurer la masse musculaire sont difficilement réalisables, en particulier de manière systématique. Les objectifs du projet consistent à développer de nouvelles approches en réponse à cette problématique clinique. Il s'agit, en particulier, et à partir de données d'imagerie acquises dans le cadre du parcours habituel du patient, de :

- Valider les performances de l'outil d'intelligence artificielle ODIASP pour la détermination automatique de la quantification de la surface musculaire par rapport à une méthodologie de référence (recherche et segmentation manuelle par le médecin) ;
- Identifier les valeurs normales de l'Index de Surface Musculaire (l'ISM) et les seuils pour le diagnostic de la sarcopénie ;
- Déterminer les facteurs associés à l'ISM à partir des données clinico-biologiques (données usuelles morphologiques, données de biologie courantes).

Acteurs internes et/ou académiques mobilisés

- Le service d'endocrinologie, diabétologie, maladies de la nutrition du CHU de Grenoble Alpes ;
- Le service de radiologie diagnostique et interventionnelle Nord du CHU Grenoble Alpes ;
- Le CIC-IT du CHU Grenoble Alpes ;
- Le DRCI du CHU Grenoble Alpes ;
- L'équipe opérationnelle de l'EDS PREDIMED ;
- L'équipe gestes médico-chirurgicaux assistés par ordinateur (GMCAO) du laboratoire TIMC, UMR 5525, CNRS, UGA, Vet'AgroSup
- Le laboratoire de bioénergétique fondamentale et appliquée (LBFA).

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Cette étude en automatisant l'ISM permettra d'optimiser le diagnostic de la sarcopénie, à partir de réutilisation de données produites dans le cadre de la prise en charge habituelle du patient.

L'outil d'IA pourra être mis à disposition des cliniciens après marquage CE pour permettre le calcul de l'ISM automatiquement chez tous patients ayant un scanner abdomino-pelvien.

Il pourra également être mis à disposition en recherche clinique pour créer de larges cohortes de patients avec des données d'ISM et ce afin de préciser les conséquences de la sarcopénie en termes médico-économiques.

Cas d'usage interne : pré-screening en imagerie

Résumé du projet et données exploitées

L'objectif principal de ce cas d'usage était d'identifier un groupe de patient pour pré-screening remplissant les critères suivant : patient majeur, passé aux urgences entre le 15/10/2020 et le 31/01/2021, ayant eu une PCR Covid-19 positive dans les 30 jours précédant le passage aux urgences ou sur le prélèvement réalisé aux urgences, ayant bénéficié d'un angioscanner thoracique pour suspicion d'embolie pulmonaire lors du passage aux urgences, ayant bénéficié d'un dosage des D-Dimères lors du passage aux urgences.

Les objectifs secondaires consistaient à (i) étendre ces résultats sur la période 01/03/2020 – 01/07/2021, (ii) créer un flow chart pour les 2 périodes des patients screenés (passage SAU avec angioscan pour suspicion d'EP dans le cadre d'une infection covid prouvée ou suspectée), des patients éligibles (échantillon précédant avec PCR + dans les 30 jours), inclus (échantillon précédant avec dosage D-Dimères et scanner contributif).

Les données mobilisées sont : les métadonnées des angioscans (RIS) pour la période donnée, le DPI pour identifier les personnes majeures et les dates de venues, les résultats d'examen biologiques sur les D-Dimères, analyse des textes indexés associés aux CR d'imagerie.

Acteurs internes et/ou académiques mobilisés

- Le service de radiologie diagnostique et interventionnelle Nord du CHU Grenoble Alpes ;
- Le CIC-IT du CHU Grenoble Alpes ;
- L'équipe opérationnelle de l'EDS PREDIMED ;

Le pré-screening s'est fait de 2 façons différentes pour démontrer la « puissance » des outils développés au sein de PREDIMED. D'un côté, 2 internes en fin de période d'internat ont établi un *Gold Standard* sur les dossiers de la période initiale par analyse manuelle de ces

dossiers (2 internes pendant 1 mois sur la période donnée), de l'autre côté, un interne en santé publique attaché à PREDIMED et des ingénieurs PREDIMED (CHUGA et UGA) ce qui représente, pour la définition de la méthodologie, l'analyse des résultats et la mise en forme de la pré-étude, l'équivalent de 1ETP pendant 2 semaines.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Le résultat attendu était une comparaison entre l'automatisation de la démarche de recherche via l'EDS et le *gold standard* réalisé manuellement. Le nombre de patients appartenant aux groupes *screening*, éligibilité et inclusion étaient similaires à 10% près entre le recueil manuel (*gold standard*) et le recueil PREDIMED.

226 (/286 79% PREDIMED et thèse) étaient communs aux 2 groupes inclus. 4/22 patients du groupe EP + gold standard ont bénéficié d'une relecture de leur scanner et ne pouvaient donc être identifiés dans ce groupe par PREDIMED. 12 (/22 GS, /18 PREDIMED) patients étaient communs aux 2 groupes par la méthode flow chart et 10 (/22 GS, /12 PREDIMED) l'étaient par la méthode feu vert dans le groupe EP +. Impact : économie en termes de RH (en particulier, médical).

Cas d'usage dans le cadre d'un partenariat avec un industriel : DEMETER - Development of a machine learning model to assess patient safety at the university hospital of Grenoble Alpes

Résumé du projet et données exploitées

Les données européennes montrent que les événements indésirables liés aux soins de santé (EIS) se produisent dans 10 % des hospitalisations et que la moitié de ces dommages pourrait être prévenue. Un EIS est défini comme un événement défavorable pour le patient, consécutif aux stratégies et actes de diagnostics et de traitements, et qui ne relève pas d'une évolution naturelle de la maladie. Les EIS représentent non seulement un problème de sécurité et de qualité des soins pour les patients, mais aussi un problème économique, leurs conséquences financières étant considérables.

L'objectif principal du projet est d'identifier à un stade précoce des sous-groupes de la population de patients du CHUGA présentant un risque accru d'événements indésirables liés aux soins (EIS) prédéfinis. Un objectif secondaire est de développer un système d'alerte qui permettrait une meilleure anticipation d'actions plus efficaces de prévention et de récupération de ces EIG.

Acteurs internes et/ou académiques mobilisés

- L'équipe opérationnelle de l'EDS PREDIMED ;
- Le CIC-IT du CHU Grenoble Alpes
- la DRCI du CHU Grenoble Alpes ;
- Le département d'information médicale du CHU Grenoble Alpes ;
- Le service d'hygiène hospitalière du CHU Grenoble Alpes.

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

Elsevier: analyse de données et élaboration des modèles de *machine learning*. Contact avec l'industriel lors du congrès mondial MEDINFO 2019.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Il s'agit d'identifier, de manière précoce, les patients à risque, en vue d'améliorer la qualité des soins tout en renforçant le rapport coût-efficacité du CHUGA. Il s'agit également de

comprendre les mécanismes conduisant à de mauvaises décisions et à des préjudices pour les patients, ce qui constitue une des premières étapes pour prévenir ces événements. Enfin, il s'agit de concevoir et mettre en œuvre de potentielles modifications des procédures hospitalières ou d'installer un système d'alerte pour aider à prévenir ces événements indésirables graves.

Cas d'usage dans le cadre d'un partenariat avec un industriel : Orthotwin-AI

Résumé du projet et données exploitées

L'enjeu du projet consiste à développer des modèles biomécaniques dynamiques, calibrés et adaptable basés sur des approches IA. Plus précisément, il s'agit d'entraîner de premiers modèles à partir données d'imagerie (CT-Scan des membres inférieurs), produites dans le soin et réutilisées pour ce projet. Ces modèles constitueront les briques élémentaires d'une plateforme, mis à disposition de la communauté, à des fins de recherche ou d'amélioration des soins.

Acteurs internes et/ou académiques mobilisés

- L'équipe opérationnelle de l'EDS PREDIMED ;
- Le CIC-IT du CHU Grenoble Alpes ;
- la DRCI du CHU Grenoble Alpes ;
- Le département de radiologie du CHUGA ;
- Le département d'orthopédie du CHUGA ;
- Le laboratoire TIMC, UMR 5525, CNRS, UGA, Vet'AgroSup

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

Jeune startup TwinSight-medical, avec création d'un laboratoire commun entre UGA / CHUGA / Twinsight-medical.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Plateforme logicielle marquée CE intégrant des briques logicielles en réponse à la demande de la communauté médicale, à des fins de recherche et d'amélioration des soins.

CHU de Lille

Cas d'usage interne : valeur diagnostique du cortisol plasmatique mesuré à 16h00

Résumé du projet et données exploitées

Le diagnostic positif de syndrome de Cushing repose sur la mesure du cortisol libre urinaire des 24h, le test de freinage minute du cortisol et le cortisol salivaire à minuit. Ce dernier n'est pas remboursé et est ainsi réalisé en hospitalisation au niveau plasmatique. Le virage ambulatoire impose de trouver une alternative. Le cortisol plasmatique à minuit est souvent réalisé avec le cortisol à 16h chez le patient hospitalisé (cycle du cortisol). Dans l'hypercortisolisme, il existe en effet une perte du rythme nyctéméral du cortisol avec en particulier une élévation des valeurs en fin de journée et en première partie de nuit. L'objectif de ce travail est d'étudier la valeur diagnostique du cortisol plasmatique à 16h et du différentiel entre le cortisol à 8h et le cortisol à 16h pour le diagnostic d'hypercortisolisme.

Données exploitées = données de biologie (cortisol dynamique à 8h, 16h et 00h)

Acteurs internes et/ou académiques mobilisés

- Equipe Include (EDS) du CHU de Lille

- Service d'endocrinologie et métabolisme du CHU de Lille

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Cette étude permettra pour la première fois d'étudier la valeur diagnostique du cortisol à 16h et du différentiel entre le cortisol à 8h et 16h. La mise en évidence d'une bonne performance diagnostique permettra de positionner le dosage du cortisol à 16h comme une alternative au dosage du cortisol à minuit pour le dépistage du syndrome de Cushing, alternative réalisable facilement en ville et en ambulatoire.

Cas d'usage dans le cadre d'un partenariat académique et industriel : création de cohortes de patients synthétiques (projet SynPatient dans le cadre de la chaire eLodi)

Résumé du projet et données exploitées

Le développement d'outils de transformation digitale des pharmacies hospitalières est complexe et peine à progresser. Plusieurs obstacles l'expliquent, en particulier le manque d'interopérabilité physique et numérique des solutions entre elles, mais également l'absence d'intégration de ces outils dans les cursus de formation. Pour faciliter l'essor de tels outils (par exemple d'aide à la décision ou de gestion des risques) dont la littérature a clairement mis en évidence l'impact sur la qualité et la pertinence des soins, le projet SynPatient vise à générer des données de patients synthétiques, c'est-à-dire présentant des caractéristiques globales identiques aux données réelles, mais sans possibilité de réidentifier des patients, utilisables dans une interface identique à celle du SI hospitalier réel. Ces données fictives seront donc, d'une part, plus largement accessibles pour la recherche et l'expérimentation de nouveaux outils et permettront, d'autre part, une diffusion facilitée dans le cadre des cursus de formation.

Ces données synthétiques seront générées à partir de données réelles de prescriptions et d'administrations de médicaments au CHU de Lille, en cours d'intégration à son entrepôt de données (Include). La méthode de création des données synthétiques émanera également des travaux de recherche de l'école Centrale de Lille dans ce domaine.

Acteurs internes et/ou académiques mobilisés

- Service de pharmacie du CHU de Lille
- Chaire eLodi (E-Logistique et Digitalisation de la prise en charge thérapeutique en établissements de santé), partenariat entre le CHU de Lille, l'Université de Lille et son UFR Sciences de Santé et du Sport, l'École Centrale de Lille et les entreprises ALTAO (Lille) et Computer engineering (Paris)
- Équipe Include (EDS) du CHU de Lille
- École Centrale de Lille

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Cette étude permettra un accès facilité à des données indispensables pour le développement d'outils notamment d'aide à la décision médicale. Ils permettront donc d'accélérer le développement de telles solutions. De plus, la mise à disposition de ces données à des étudiants dès le cursus de formation permettra une meilleure appropriation, y compris des outils informatiques utilisés en temps réel au CHU de Lille. En outre, des réflexions sur une éventuelle généralisation de la méthodologie de génération des données pourront être entamées.

Hospices civils de Lyon

Cas d'usage interne : MEMAURA (Mémoire Auvergne-Rhône-Alpes)

Résumé du projet et données exploitées

Le projet vise à étudier les facteurs prédictifs et pronostiques associés à la dépendance fonctionnelle et son évolution dans le temps auprès de patients non institutionnalisés reçus dans une consultation mémoire et le service de neuro-cognition participant, quelle que soit la présence ou le type de trouble de mémoire ou apparenté. Le recueil de données de soins est complété par des échelles spécifiques de la recherche permettant d'intégrer un nombre important de patients. Il s'inscrit dans le parcours de soins des patients, rendant généralisables les observations issues de ces données.

Les données sont :

- Recueillies
 - (i) Par questionnaire pendant l'examen clinique, et lors d'entretiens entre le personnel soignant, paramédical, les patients et leur aidant principal
 - (ii) À chaque passage des patients en Consultation Mémoire, excepté le profil neuropsychologique.
- Saisies informatiquement par le personnel soignant, paramédical et les secrétaires de consultation, dans le Système d'Information Hospitalier.

Acteurs internes et/ou académiques mobilisés

Le projet est conduit par :

- le Centre Mémoire de Ressources et de Recherche (CMRR de Lyon) des Hospices Civils de Lyon (HCL) en interaction avec
- les centres de Recherche Clinique VCF (Vieillesse, Cerveau, Fragilité),
- les Consultations Mémoires de la région AuRA
- la Direction des Services Numériques des HCL.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

MEMAURA permettra d'améliorer la connaissance sur l'évolution des troubles neurocognitifs et d'élaborer des stratégies d'interventions visant à ralentir le déclin fonctionnel et cognitif des patients, ainsi que les troubles du comportement, ce qui constitue un enjeu majeur de santé publique. L'amélioration de la compréhension des différents stades de la maladie et leurs déterminants permettra de préparer et mobiliser les ressources adéquates pour améliorer la prise en charge de ces patients et justifier de certaines interventions diagnostiques et thérapeutiques nouvelles. Enfin ce projet permettra de collecter des données de pilotage dans les centres de soins ainsi que des informations sur les parcours de soins pouvant contribuer à leur optimisation.

Cas d'usage dans le cadre d'un partenariat avec un industriel : PRévention, Epidémiologie, et COntôle des VIRus Respiratoires : PRECOVER

Résumé du projet et données exploitées

Les infections respiratoires virales, d'acquisition aussi bien hospitalière que communautaire, ont un impact majeur en santé publique à de multiples égards (morbidité, mortalité, coûts). L'objectif de ce projet est d'évaluer le poids de ces infections, leur histoire naturelle, les

déterminants de leur acquisition, leur pronostic et leur impact en termes de morbidité et mortalité ainsi que la consommation des ressources hospitalières qui leur sont liées.

Les données proviendront des bases suivantes :

- Base de données d'hospitalisation (date d'entrée, de sortie, signes cliniques, examens complémentaires, imagerie, séjours en soins intensifs, etc comorbidités, etc.)
- Base de données de microbiologie : virologie (PCR), bactériologie (diagnostic différentiel) Service de médecine du travail : proportions de soignants vaccinés. Données hebdomadaires d'incidence des infections respiratoires en communauté issues de l'ARS (Nombre de cas, période et calcul de taux).

Acteurs internes et/ou académiques mobilisés

- Le Service d'Hygiène, Epidémiologie, Infectio-vigilance et Prévention (SHRIP) des HCL
- Le Centre National de Référence des virus respiratoires (HCL-Institut des Agents Infectieux) et le service de Santé au Travail
- Le service de bio statistique et le Laboratoire de biométrie & biologie Evolutive DE BIOMÉTRIE ET BIOLOGIE ÉVOLUTIVE (HCL-LBBE),
- La Direction des Services Numériques
- Le Centre d'Excellence des Virus Respiratoires (CERP) représenté par les Hospices Civils de Lyon, l'Université et le laboratoire Sanofi (HCL-UCBL-Sanofi).

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

A travers le CERP, tout industriel (comme Sanofi) via un partenariat, pourra demander l'usage de requête de données issues de l'EDS à des fins de recherche, moyennant une compensation financière. Cette demande sera soumise à la validation du Comité Stratégique du CERP.

Les partenariats nationaux et internationaux prévus via le CERP pourront permettre la création de réseau de centres, permettant le partage de méthodologies et de résultats, l'initiation d'études sur des questions de recherches spécifiques et la comparaison de données à une échelle mondiale.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

1. En temps quasi réel :

- 1) Evaluation des flux de patients (urgences ou mutations dans les services);
- 2) Evaluation du fardeau de la maladie (morbidité, mortalité);
- 3) Proposition de mesures de contrôle appropriées selon les diagnostics et le potentiel épidémique des agents infectieux identifiés et 4) Couverture vaccinale (patient vacciné ou non).

2. En temps différé mais à court terme :

- 1) Identification précoce de nouvelles populations à risque;
- 2) Exploration des facteurs d'émergence épidémique;
- 3) Dimensionnement de l'épisode épidémique et

4) Définition de l'efficacité des mesures de prévention mises en œuvre et mesure de l'impact de la vaccination.

3. En temps différé à moyen terme :

- 1) Pronostic de la maladie ;
- 2) Dimension et extension de l'épisode épidémique;
- 3) Impact des interventions vaccinales sur les patients et sur le système hospitalier;
- 4) Modélisation prédictives relatives aux épidémies pour les différents agents infectieux;
- 5) Résultats ayant un apport pédagogique pour les soignants;
- 6) Renforcement des collaborations.

Hôpitaux universitaires de Marseille

Cas d'usage dans le cadre d'un partenariat avec un industriel : FranceCoag

Résumé du projet et données exploitées

En France, plus de 3 millions de nos concitoyens sont atteints par environ 7 000 maladies rares. La prise en charge de celles-ci constitue un enjeu de santé considérable dont témoignent les Plans Nationaux Maladies Rares (PNMR), dont le PNMR 3 lancé en juillet 2018. L'amélioration de la prise en charge de ces patients, en vue d'améliorer leur qualité de vie et leur autonomie passe par la compréhension et l'identification des leviers d'amélioration pour permettre la modernisation des organisations et optimiser les financements nationaux.

L'hémophilie et les MHC sont des maladies rares mais qui à elles seules touchent près de 40 000 personnes en France. Leur prise en charge est économiquement lourde du fait notamment du coût des traitements par facteurs de coagulation. Ces pathologies peuvent occasionner, au-delà des saignements itératifs, des complications majeures, provoquant possiblement un handicap physique, avec des conséquences psychologiques et sociales importantes. Les conséquences de prises en charge inadaptées peuvent donc être très lourdes et très coûteuses. Elles sont source de complexité supplémentaire dans la prise en charge de situations cliniques variées, chirurgie, pathologies cardiovasculaires, cancers.

La prise en charge de ces maladies hémorragiques a été considérablement améliorée grâce à la création de centres spécialisés multidisciplinaires qui travaillent en collaboration avec les médecins traitants dans le cadre de réseaux de soins. L'hétérogénéité des pratiques de soins a très peu été étudiée dans ces pathologies.

L'objectif est de décrire les parcours de soins des patients atteints de maladie hémorragique constitutionnelle (MHC). Plus précisément, seront comparés les parcours observés aux prises en charge de référence ; et identifiés les facteurs associés à l'adhésion à la prise en charge recommandée (caractéristiques sociodémographiques, cliniques et thérapeutiques des patients et organisation des soins).

Acteurs internes et/ou académiques mobilisés

Le dispositif FranceCoag constitue un outil d'une importance majeure pour évaluer la prise en charge des patients atteints de MHC car il comporte des données très précises sur le profil diagnostique et la description clinique des patients. Cependant, il comporte peu de données sur le parcours des patients et les consommations de soins. Les données du dispositif FranceCoag et les données du système national des données de santé (SNDS) sont très

complémentaires et leur appariement permettrait d'étudier les parcours de soins de ces patients à la fois dans le cadre de la prise en charge de leur pathologie hémorragique, mais aussi de la prise en charge de leurs comorbidités. Elle permettra également d'évaluer la qualité de la prise en charge de ces pathologies en France et d'identifier des profils de patients peu adhérents à la prise en charge recommandée en vue d'analyser les facteurs associés à ces prises en charge non optimales.

11 462 patients inclus dans la cohorte FranceCoag au 31/12/2018

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

Un consortium d'industriels est associé à ce projet : CSL, LFB Biomédicaments, Novonordisk, Pfizer, Roche, Sobi, Takeda contribuant financièrement au recueil de données dans l'entrepôt FCoag.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Le dispositif FranceCoag constitue un outil d'une importance majeure pour évaluer la prise en charge des patients atteints de MHC car il comporte des données très précises sur le profil diagnostique et la description clinique des patients. Cependant, il comporte peu de données sur le parcours des patients et les consommations de soins. Les données du dispositif FranceCoag et les données du système national des données de santé (SNDS) sont très complémentaires et leur appariement permettrait d'étudier les parcours de soins de ces patients à la fois dans le cadre de la prise en charge de leur pathologie hémorragique, mais aussi de la prise en charge de leurs comorbidités.

Elle permettra également d'évaluer la qualité de la prise en charge de ces pathologies en France et d'identifier des profils de patients peu adhérents à la prise en charge recommandée en vue d'analyser les facteurs associés à ces prises en charge non optimales. Ces données sont utiles dans le montage de dossiers auprès de la CESP de la HAS (post inscription).

CHU de Montpellier

Cas d'usage interne : PredIC

Résumé du projet et données exploitées

L'objectif du projet est la conception et la validation d'un algorithme prédictif du risque de réhospitalisation pour les patients insuffisants cardiaques chroniques suivis au CHUM, à partir de sources de données multiples, y compris non structurées en utilisant l'Intelligence Artificielle.

Les données cliniques structurées (âge, sexe, comorbidités, médicaments, actes, biologie, parcours antérieur, couverture CMUc/AME) sont issues de l'entrepôt. Les données non structurées (statut tabagique et FEVG) seront extraites des textes libres de l'entrepôt par des méthodes de fouille de données (*Natural Language Processing*). Le projet intégrera des données de référence de parcours issues du SNDS, et des données contextuelles socio-économiques.

Acteurs internes et/ou académiques mobilisés

Pr François Roubille, Service de Soins Intensifs Cardiologiques - Département de Cardiologie et Maladies Vasculaires, Dr Grégoire Mercier, Unité Sciences des Données de Santé, Ulysse Rodts, président de la société KanopyMed. Le projet est financé par la Région Occitanie dans le cadre du dispositif READYNOV.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

A la fin du projet, nous aurons donc (1) un outil de structuration des données issues des dossiers patients (2) un jeu de données complet interopérable et (3) un algorithme prédictif performant. Le caractère interopérable garantira le fonctionnement de l'outil dans les autres hôpitaux et cliniques.

PredIC permettra aux hôpitaux et cliniques d'identifier les patients à risque élevé pour personnaliser leur parcours de soins en déclenchant des interventions existantes. En France, aucun outil de stratification ou d'évaluation du risque n'existe à ce jour. PredIC s'intègre pleinement dans les évolutions en cours du système de santé associant de nouveaux modèles de financement des soins (Article 51) et une place renforcée du numérique et de l'interopérabilité (MaSanté2022).

Cas d'usage interne : eMatch

Résumé du projet et données exploitées

Ce projet a pour objectif, à terme, de rendre accessible, pour la recherche biomédicale, des données clinico-biologiques issues du soin en lien avec des échantillons des biobanques provenant des trois établissements cliniques : CHU de Montpellier, CHU de Nîmes et Institut du cancer de Montpellier (ICM).

Les données sont extraites du logiciel de gestion des biobanques et complétées par des données cliniques issues du Dossier Patient Informatisé, liées à ces prélèvements selon une temporalité définie. Ces données sont organisées dans un entrepôt au format OMOP pour favoriser l'interopérabilité entre les différents établissements.

Acteurs internes et/ou académiques mobilisés

eMATCH est un projet initié et financé par la Key Initiative « Biomarkers & Therapy » de l'Université d'Excellence de Montpellier (MUSE), en collaboration avec le Département d'Information Médicale (DIM) du CHU de Montpellier. Il est issu d'une réflexion menée au sein des communautés hospitalo-universitaires de Montpellier et Nîmes.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

A la fin du projet, le Centre de Ressource Biologique sera en mesure de fournir à des investigateurs internes ou externes, après approbation du projet par le Comité Scientifique et Ethique, en plus des échantillons, un ensemble de données d'intérêt. Des corpus de données ont été définis pour chaque champs pathologiques correspondant aux FHU-COEN (cancérologie, maladies Infectieuses, médecine expérimentale et régénératrice, neurologie, maladie métabolique).

Case d'usage interne : prédiction des urgences vitales au cours de séjour en MCO

Résumé du projet et données exploitées

Thèse CIFRE Prédiction des urgences vitales au cours de séjour en MCO : La sécurité des patients doit être un objectif ancré au cœur des pratiques des professionnels et du système de santé, pour prévenir la survenue d'évènements indésirables graves au cours de l'hospitalisation comme l'arrêt cardio-circulatoire ou le décès inattendu. Généralement, ces évènements indésirables graves sont précédés d'une accumulation de signes physiologiques circulatoires, respiratoires ou neurologiques, qui surviennent dans les 12 à 36h précédant

l'urgence vitale. Des scores de prédiction (Churpek 2015, Davies 2014), plus ou moins performants, ont été développés pour prendre en compte ces signes.

L'objectif de cette thèse est d'identifier et de valider les éléments d'informations présents dans les dossiers des patients, composés de données hétérogènes (numériques et textuelles), multi-sources et multi-auteurs, pour prédire des urgences vitales au cours de séjours en Médecine-Chirurgie-Obstétrique, à partir de jeux de données faiblement étiquetés.

Acteurs internes et/ou académiques mobilisés

- Pr Boris Jung, Service de Médecine Intensive, Réanimation Médicale et Grands Brûlés
- Caroline Dunoyer, Unité Sciences des Données de Santé,
- Nicolas Molinari, Unité de Recherche Clinique et Epidémiologie
- Pr Pascal Poncelet, Directeur de l'équipe ADVANSE au LIRMM (Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier).

La société INGIMA experte en conception et développement de projets innovants de Data Science finance la bourse CIFRE.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Les travaux doivent permettre d'améliorer la représentation des données, mieux gérer les échelles temporelles, appréhender et modéliser la confiance. Le développement et la validation d'un tel outil de screening des patients à risque permettra de conduire des travaux ultérieurs afin de proposer un outil d'alerte et/ou mettre en place des outils de surveillance clinique spécifiques.

Cas d'usage dans le cadre d'un partenariat avec un industriel : analyse du temps de travail

Résumé du projet et données exploitées

Les pôles pharmacies des centres hospitaliers se focalisent sur la délivrance des traitements et ont une action limitée de l'analyse des prescriptions, ce qui génère un risque de surdosage important et/ou un coût de prise en charge qui pourrait être diminué par un fléchage vers un choix thérapeutique optimisé.

Dans ce contexte, la SATT AxLR, la Société Quinten, le Centre Hospitalier de Valenciennes, l'Université de Montpellier et le CHU de Montpellier ont mis en œuvre un projet commun dans l'objectif de **développer une plateforme numérique permettant l'analyse du temps de travail des pharmaciens sur des ordonnances liées au diabète, thrombose et antibiotique en le croisant avec les prescriptions et de l'interfacer avec les systèmes d'informations du CHU de Montpellier et du CHV.**

Données : données médicales « pseudonymisées » et leurs annotations extraites pendant le Projet des bases de données constituant les Connaissances Princeps des Etablissements, concernant les environnements thrombose et antibiotique s'agissant du CHV et l'environnement diabète et s'agissant du CHU de Montpellier, comprenant notamment les prescriptions, les données biologiques, les données diagnostiques (PMSI), les données de consommation des médicaments, les pancartes, les interventions pharmaceutiques, les données d'administration des médicaments, les données cliniques et constantes physiologiques.

Les Connaissances Princeps du CHU sont :

- ses bases de données « prescriptions » et « IP » (Interventions pharmaceutiques) concernant les environnements diabète, et
- ses règles d'analyse de prescriptions de la pharmacie.

Acteurs internes et/ou académiques mobilisés

Avec l'implication des partenaires (Centre Hospitalier de Valenciennes, l'Université de Montpellier et le CHU de Montpellier), le projet PharmIA a fait l'objet d'un investissement en maturation par la SATT AXLR qui a abouti à la création de la start up PharmIA.

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

SATT AxLR et Société Quinten

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Les porteurs de projet M. Antoine AVIGNON, PU-PH au CHU de Montpellier, M. Etienne COUSEIN, pharmacien au CH de Valenciennes, M. Philippe ARANY, Directeur Général de la société Quinten et M. Frédéric COURIOL, Directeur de la Société PharmIA, ont développé la Plateforme d'analyse d'ordonnance augmentée afin de :

- détecter automatiquement des prescriptions sensibles et/ou à risque à traiter en priorité ;
- accélérer l'investigation via l'identification des critères de risque associés à une prescription ;
- optimiser le processus de validation du pharmacien en organisant le flux de traitement des prescriptions.
- générer des économies sur les dépenses de médicaments inclus dans la T2A (Tarification à l'Activité).

La plateforme PharmIA aujourd'hui commercialisée est un dispositif médical innovant marqué CE.

Un partenariat stratégique a été signé entre PharmIA et le leader européen de l'édition et de l'intégration de solutions numériques pour le monde de la santé permettant d'accélérer le développement commercial et l'intégration de la plateforme PharmIA auprès des hôpitaux européens. La plateforme PharmIA s'intègre directement dans les DPI facilitant ainsi son utilisation par le pharmacien hospitalier.

CHU de Nancy

Cas d'usage interne : French Paradise

Résumé du projet et données exploitées

Le pronostic des dyspnées aiguës est grave, plus grave que celui de l'infarctus. Le diagnostic étiologique de dyspnée aigue est parfois difficile à porter. C'est particulièrement le cas pour l'insuffisance cardiaque aigue. Pour autant, peu de données existent sur cette entité clinique majeure. Il y a notamment peu de données sur le pronostic à moyen terme de ces patients. Par ailleurs les descriptions se bornent souvent à une présentation spécifique à une étiologie donnée de dyspnée (p.e. insuffisance cardiaque, exacerbation de bronchite chronique...) plutôt qu'à l'entité clinique « dyspnée aigue » dans son ensemble.

Nous nous proposons d'étudier les patients pris en charge pour dyspnée aiguë par une équipe médicalisée des urgences des centres investigateurs. L'étude devrait nous permettre de mieux identifier les caractéristiques des patients dyspnéiques pris en charge par une équipe médicalisée des urgences. Les résultats permettront de dégager les éléments clés de la filière de soins en urgences des patients dyspnéiques. Nous pourrions identifier les déterminants des difficultés diagnostiques aux urgences, ainsi que les facteurs pronostiques, globaux et par pathologie sous-jacente, des patients admis pour dyspnée aiguë. L'objectif final est de proposer des pistes d'amélioration de la filière de soins des dyspnées aiguës.

Pour ce faire, le projet évaluera le risque de morbi-mortalité (réhospitalisation/décès) à court et moyen terme de ces patients admis pour dyspnée aiguë et ses déterminants ainsi que l'association d'un diagnostic erroné avec le risque de réhospitalisation et/ou décès. Enfin, un score de risque de morbi-mortalité dans le cadre d'une dyspnée aiguë admise aux urgences sera construit.

Données mobilisées :

Les bases de données qui seront mobilisées :

- Données intra-hospitalières extraites des dossiers patients informatisés des centres hospitaliers participants
 - Données du logiciel des urgences (RESURGENCES / 30% des centres d'urgence français environ ou autres logiciels)
 - Données du PMSI (Circuit hospitalier/services d'hospitalisation, diagnostic finaux, mortalité intra-hospitalière)
 - Données de biologie aux urgences et en hospitalisation
- Données extraites du SNDS (données post-hospitalisation)
 - Mortalité toute cause et spécifique
 - Réhospitalisation toute cause avec codage des causes d'hospitalisation
 - Consommation de soins
- Consultations ambulatoire
- Echocardiographie
- Réalisation de biologie (ionogramme, peptides natriurétiques)

Acteurs internes et/ou académiques MOBILIS2S

- Etude conduite en partenariat avec le Health Data Hub (convention de novembre 2021)
- CHRU de Nancy : Investigateurs coordonnateurs, Pr Nicolas Girerd (cardiologie), Pr Tahar Chouied (Urgences). Responsable de l'entrepôt : Pr Nathalie Thilly
- Hôpitaux Universitaires de Strasbourg
- Hospices Civils de Lyon
- CHU de Tours (sous réserve, en attente de confirmation)

Dans chacun des 3 centres associés, 1 investigateur principal cardiologue + 1 urgentiste

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

L'objectif final est de proposer des pistes d'amélioration de la filière de soins de ces patients admis pour dyspnée aiguë. Le bénéfice ultime de ce projet est une **amélioration des prises en charge** des patients admis pour dyspnée aiguë aux urgences.

Plusieurs **publications** internationales pourront émerger du projet. L'expérience du CIC-P dans ce segment est déjà importante (p.e. Kobayashi, ... Girerd, Plos Med, 2020 et Buessler ... Girerd, CHEST, 2020).

Ce projet permettra de **finaliser et valider des outils de text-mining issus du NLP** indispensable à l'utilisation optimale des données du HDH.

CHU de Nantes

Cas d'usage étude de la physiologie des anévrysmes intracrâniens

Résumé du projet et données exploitées

Sur la base de la cohorte nationale multicentrique ICAN (*Understanding the pathophysiology of intracranial aneurysm*). Elle a pour objectif initial d'identifier les facteurs cliniques, morphologiques, biologiques et génétiques impliqués dans le développement des AIC et rassemble un réseau national comptant la quasi-totalité des centres où des AIC sont pris en charge en France (29 centres). La base est enrichies de données de la cohorte CONSTANCES et de l'EDS du CHU de Nantes

Acteurs internes et/ou académiques mobilisés

Equipes CHU de Nantes (Clinique des données, DRI)

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

Octopize pour la création d'une base de données synthétiques mise en licence libre.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Location of intracranial aneurysms is the main factor associated with rupture in the ICAN population - J Neurol Neurosurg Psychiatry. 2021 Feb;92(2):122-128 PMID 33097563

Amélioration des facteurs prédictifs de rupture des anévrysmes intracrâniens sur la base de données d'Imagerie, de biologie, de génétique et cliniques permettant d'anticiper les traitements chez des patients à risques de ruptures prioritairement.

Cas d'usage application au suivi de la politique de dé-confinement COVID-19 – Projet ARGOS

Résumé du projet et données exploitées

Suite à la demande du directeur général du CHU de Nantes et du président de Commission médicale d'établissement, les équipes du CHU de Nantes et WeData ont mis sur pied le projet ARGOS de pilotage par la multi-data du dé-confinement. Durant la crise, le CHU de Nantes a accéléré son passage au numérique, une médecine 4P qui fait toute sa place aux données.

Avec Argos, le CHU de Nantes surveille la reprise de la pandémie COVID grâce aux données de santé. Dans l'esprit de jeu collectif qui est le sien, il intègre des sources de *big-data* de plusieurs établissements juridiquement indépendants : CHU, SAMU, SOS Médecins, Bioliance (fermées, individuelles pseudonymisées), Santé Publique France (ouvertes, stratifiées ou agrégées)) pour présenter sous formes de séries temporelles, les indicateurs de reprise de l'épidémie dans la région et appuyer la prise de décision sur des faits quantifiés.

Comme dans tout hôpital numérique, le respect de la confidentialité des données est au cœur de la démarche, les données sont toutes pseudonymisées. En responsabilité, le CHU de Nantes a choisi de garder les données sur ses serveurs internes « hébergeurs de données de santé » en y déployant des technologies *open-source* de pointe de type « Elastic Search » et « Kibana ». Soutenu par Nantes Métropole, le projet unit le CHU de Nantes, SOS-Médecins, le SAMU 44, Santé publique France, WeData, les laboratoires Bioliance.

Acteurs internes et/ou académiques mobilisés

Equipes CHU de Nantes (Clinique des données, DSN, Santé publique),

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

Laboratoires de ville (Réseau BIOLIANCE), Réseau Sentinelles, SOS Médecins

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Preuve de concept de la capacité de croiser des données de sources internes et externes dans un but de pilotage. Cet essai va être étendu à des cas d'usages plus variés.

<https://www.data.gouv.fr/fr/reuses/projet-argos/>

CHU de Rennes

Cas d'usage interne : analyser les interactions et les événements indésirables liés aux statines en ville et à l'hôpital

Résumé du projet et données exploitées

Les statines sont des molécules extrêmement prescrites et pour lesquelles des interactions médicamenteuses peuvent entraîner des événements indésirables graves. La mesure de la prévalence de telles interactions à partir des données de vie réelles est d'usage courant. Dans cette étude pionnière en France, l'approche a permis de mesurer cette prévalence au cours du parcours de soin du patient « en ville » et à l'hôpital. Pour cela, les données de l'EDS du CHU de Rennes ont été chaînées avec une extraction du SNDS. Ce cas d'usage conduit dans le cadre du projet ANR INSHARE a été également l'occasion de mettre au point un algorithme de chaînage des données.

Acteurs internes et/ou académiques mobilisés

CHU, Centre de données cliniques, LTSI (UMR 1099 INSERM UR1) etc.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Publication scientifique : Bannay, Aurélie, Mathilde Bories, Pascal Le Corre, Christine Riou, Pierre emordant, Pascal Van Hille, Emmanuel Chazard, Xavier Dode, Marc Cuggia, et Guillaume Bouzillé. « Leveraging National Claims and Hospital Big Data: Cohort Study on a Statin-Drug Interaction Use Case ». JMIR Medical informatics 9, no 12 (13 décembre 2021): e29286. <https://doi.org/10.2196/29286>.

Cette étude a démontré la valeur ajoutée d'une combinaison de données clinique et de remboursement pour produire une évaluation à grande échelle de la prévalence des interactions médicamenteuses et des parcours de soins en dehors des hôpitaux. Elle a ouvert la voie à l'évolution du système de santé vers un système de santé apprenant utilisant les connaissances issues des données de la recherche et des données de vie réelle.

Cas d'usage interne : aider à la production du PMSI – codage automatique des diagnostics en CIM à partir des compte-rendu d'hospitalisation

Résumé du projet et données exploitées

L'EDS peut être utilisé en routine pour aider à la production du PMSI. En étant une source de données mobilisable pour les outils de codage, il permet, grâce à la mise en œuvre de règles expertes, l'accès rapide et ciblé à des éléments du dossier médical pour coder des comorbidités et mieux valoriser les séjours. Dans ce travail cité ci-dessous on combine l'usage des données de l'EDS avec l'usage d'une base de connaissances sur les médicaments

(Base Thériaque) pour détecter des comorbidités pouvant permettre une meilleure valorisation des séjours.

Acteurs internes et/ou académiques mobilisés

CHU, Centre de données cliniques, DIM

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Publication scientifique : Sylvestre E, Bouzillé G, Chazard E, His-Mahier C, Riou C, Cuggia M. Combining information from a clinical data warehouse and a pharmaceutical database to generate a framework to detect comorbidities in electronic health records. BMC Med Inform Decis Mak. 2018 Jan 24;18(1):9. doi: 0.1186/s12911-018-0586-x. PMID: 29368609; PMCID: PMC5784648.

Un algorithme simple basé sur la combinaison de données accessible et directement réutilisable issues de bases de connaissance, des prescriptions médicamenteuses et des résultats de laboratoire peut détecter les comorbidités.

CHU de Reims

Cas d'usage interne : CovCost

Résumé du projet et données exploitées

Objectif du projet : déterminer pour les patients admis pour une infection au Covid-19, des parcours types au sein de l'établissement en fonction de leur profil médical.

Données exploitées : données médicales contenues dans le dossier patient informatisé ; données administratives du patient

Acteurs internes et/ou académiques mobilisés

Institut de l'intelligence artificielle ; Direction des finances ; Direction des services informatiques

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

- Identification de parcours types par profil patient
- Anticipation de la gestion des flux dans le Nouvel Hôpital
- Identification des indicateurs économiques par type de parcours permettant identification des surcoûts Covid auprès de l'ARS

Cas d'usage dans le cadre d'un partenariat avec un industriel : identification du risque sanitaire et minimisation des risques de contamination Covid-19 des patient en consultation externe grâce la modélisation de flux patients par intelligence artificielle

Résumé du projet et données exploitées

Identification du risque sanitaire et minimisation des risques de contamination COVID19 des patient en consultation externe grâce la modélisation de flux patients par intelligence artificielle.

Utilisation des données cartographiques de l'établissement ainsi que des consultations et passages au bureau des entrées.

Acteurs internes et/ou académiques mobilisés

Institut de l'intelligence artificielle et Direction du projet Nouvel Hôpital

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

- Société Onhys
- Mise à disposition d'un outil algorithmique par la société

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

- Identification des zones de saturation et des points de passage à haute fréquentation avec risque de contamination.
- Proposition de réorganisation des circuits patients afin de limiter le risque sanitaire.

CHU de Rouen

Cas d'usage interne : identification automatique de patients à prioriser pour la vaccination Covid-19

Résumé du projet et données exploitées

Nous avons cherché, dans EDSaN (EDS du CHU de Rouen), des patients à haut risque de développer une forme grave de COVID-19, à partir d'une liste de pathologies particulières (publiées par la HAS début 2021). Chaque pathologie a fait l'objet d'une requête spécifique. La particularité de ces requêtes est que nous avons pu croiser des sources de données différentes pour identifier les dossiers (comptes rendus textuels et codages PMSI). De plus, il a fallu réidentifier ces patients à partir d'EDSaN pour savoir s'ils étaient décédés ou non avant de les contacter. Données structurées (PMSI) et non structurées (comptes-rendus médicaux).

Acteurs internes et/ou académiques mobilisés

- Département d'Information et d'Informatique Médicales du CHU de Rouen (Pr Darmoni)
- Service d'infectiologie du CHU de Rouen (Pr Etienne)

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Ce travail a permis l'identification de plus de 6000 patients à haut risque de développer une forme grave de Covid-19. Même si la construction des requêtes nécessaires à l'identification de ces patients peut prendre du temps, cette méthode de détection automatique reste beaucoup plus rapide qu'une recherche manuelle dans les dossiers des patients. Ce projet a donc permis un gain de temps considérable en ce contexte épidémique.

Ce travail a été publié à la conférence internationale MedInfo 2021 :

<https://www.cismef.org/cismef/wp/wp-content/uploads/pdf/Grosjean2021.pdf>.

Hôpitaux universitaires de Strasbourg

Cas d'usage interne : recherche de patient « *similaire au mien* » par deep phenotyping - DeepSimilarPatient

Résumé du projet et données exploitées

Le dossier patient informatisé (DPI) est une source riche en données cliniques qui est renseignée et exploitée de façon routinière dans l'activité de soin hospitalière. Cependant la nature hétérogène et complexe des données du DPI (structurée et non-structurée, numérique et symbolique, temporelle et incomplète) soulèvent de nombreux défis pour leur utilisation à des fins de recherche clinique, ou de développement d'outils d'intelligence artificielle (i.e., modèles prédictifs, système d'aide à la décision).

Une des problématiques soulevées par la nature et structuration hétérogènes des données est la caractérisation des patients. Le phénotypage profond (*deep phenotypings*) des patients consiste à extraire automatiquement à partir du DPI des caractéristiques fines qui décrivent le patient et son parcours de soins, et qui ne sont pas accessibles directement par une requête simple sur les données structurées du DPI. Ces caractéristiques fines peuvent inclure la gravité d'une affection, la longueur d'un traitement, un antécédent mentionné uniquement dans un courrier et doivent permettre de répondre à des questions comme : De quelles maladies souffre le patient ? Depuis quand évoluent ces pathologies ? Quels en sont les symptômes ? Quelles en sont les thérapeutiques ? Quelle est l'évolution ?

En plus de l'extraction d'information en tant que telle, le phénotypage profond assure que les caractéristiques des patients soient décrites de façon homogène sur l'ensemble des individus considérés. Cette homogénéisation des descriptions permet de comparer les patients entre eux, notamment en permettant de mesurer une similarité entre patients. Cette notion de patients similaires est particulièrement intéressante puisqu'elle permettrait, à partir d'un DPI (celui de « mon » patient) de constituer une cohorte des patients similaires (les patients « similaires au mien ») qui peuvent être retrouvés au sein d'un entrepôt de données cliniques local ou d'un autre établissement. La constitution d'une telle cohorte peut s'avérer extrêmement utile à la fois en pratique et en recherche clinique. En effet, à partir d'une telle cohorte il est possible :

- de produire des statistiques descriptives (combien de patients avec les mêmes antécédents que mon patient et le même diagnostic ont été traités avec le médicament x ?) ;
- d'évaluer un risque (quel est le risque pour un patient avec les mêmes antécédents que le patient que je prends en charge d'avoir une toxicité aiguë au médicament x ?) ;
- d'évaluer un effet causal (quel est l'effet du médicament x chez les patients avec les mêmes antécédents que le patient que je prends en charge ?)
- de décrire les profils évolutifs des patients qui sont similaires au mien.

Dans un premier temps, le projet portera sur la polyarthrite rhumatoïde. Il s'agit d'une pathologie relativement fréquente qui présente des difficultés de choix thérapeutiques lorsqu'elle est réfractaire au traitement de première intention. Cette pathologie est particulièrement importante dans un cadre hospitalier, car les patients complexes sont souvent suivis dans des services spécialisés et le CHU de Strasbourg est un centre de référence de cette pathologie.

Les données exploitées sont l'ensemble des données disponibles au sein du dossier patient informatisé. La période d'étude s'étale de 2015 à 2020. Les premières étapes du projet se concentrent sur les données structurées (biologie, médicament) et le texte libre présent dans

le DPI. Le caractère fin du phénotypage réalisé rend indispensable l'accès à des données non structurées plus riches en informations.

Les techniques d'apprentissage fédératif seront utilisées. Plusieurs étapes ont été identifiées :

- phénotypage des patients et constitution de cohorte « gros grain » de patients par pathologie (classification supervisée et semi-supervisée) ;
- représentations des patients présélectionnés dans le contexte de leur maladie (extraction d'entités nommées à partir du DPI, temporalisation des données et représentation des données) ;
- recherche de similarité entre les patients (clustering, calcul de distance entre les patients représentés sur un espace vectoriel à n-dimensions).

Acteurs internes et/ou académiques mobilisés

- Service de santé publique, Hôpitaux universitaires de Strasbourg et Laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie (ICube, UMR7357).
- Direction : DRCI
- Établissements du réseau PAISaGE
- Établissements du réseau HUGO

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

L'intérêt principal de ce projet est le développement de la médecine personnalisée. La constitution de cohorte de patients similaire permettant d'adapter les choix thérapeutiques de façon plus personnalisée que les études cliniques sur données rétrospectives. De telles cohortes donnent des arguments supplémentaires aux cliniciens pour le choix des thérapeutiques le plus adapté aux patients. La réalisation d'étude clinique étant onéreuse, elles sont en général réalisées sur un échantillon spécifique de la population et pendant des durées courtes.

Cas d'usage interne : Intervention pharmaceutique, priorisation et intelligence artificielle - InterPharmIA

Résumé du projet et données exploitées

La pharmacie clinique est une discipline de santé centrée sur le patient qui contribue à la sécurisation, la pertinence et à l'efficacité du recours aux médicaments. En collaboration avec les autres professionnels, le pharmacien, en tant que spécialiste, joue un rôle clé dans l'analyse et la sécurisation des thérapeutiques médicamenteuses. L'analyse pharmaceutique (ou « expertise pharmaceutique clinique des prescriptions ») fait partie intégrante des activités de pharmacie clinique. Elle se définit comme une « expertise structurée et continue des thérapeutiques du patient, de leurs modalités d'utilisation et des connaissances et pratiques du patient ». Elle peut donner lieu à la formulation d'interventions pharmaceutiques, « propositions de modification de la thérapeutique médicamenteuse » à l'égard du prescripteur. Cette activité a été précisée récemment dans un décret [n°2019-489 du 21 mai 2019] relatif aux pharmacies des structures hospitalières dites à usage intérieur qui en fait une obligation réglementaire.

Depuis une décennie, l'implémentation de logiciels d'aide à la prescription dans les hôpitaux français a permis de s'affranchir des prescriptions manuscrites disponibles exclusivement dans les unités de soins et de faciliter ainsi le déploiement de l'analyse pharmaceutique à partir des pharmacies à usage intérieur. Toutefois, comme le rapporte une étude nationale réalisée par la Direction Générale de l'Offre de Soins, il est actuellement impossible pour les pharmaciens cliniciens d'effectuer en temps réel une analyse approfondie et exhaustive des

prescriptions de tous les patients hospitalisés. Les freins majeurs au déploiement systématique de cette expertise sont d'une part le manque de ressources humaines pharmaceutiques et d'autre part l'absence d'outils pertinents pour identifier les prescriptions les plus à risques en termes d'iatrogénie médicamenteuse. En conséquence, l'organisation de l'analyse pharmaceutique et en corolaire l'évaluation de la pertinence des prescriptions, voire la détection d'erreurs médicamenteuses reposent principalement sur des décisions humaines.

Les techniques d'apprentissage automatique offrent une opportunité unique de développer des outils permettant d'assister le travail des pharmaciens hospitaliers. Dans ce contexte, nous avons initié le développement d'algorithmes pharmaceutiques prédictifs capables de détecter les prescriptions nécessitant une intervention pharmaceutique à forte valeur ajoutée. L'objectif de ce projet de recherche est de proposer pour les hôpitaux un outil capable d'identifier par des méthodes d'apprentissage automatique les prescriptions les plus à risque pour leur donner la priorité. L'identification en temps réel se fera sous forme d'alerte indiquant ces prescriptions à risque qui seront alors intégrées dans un plan de travail et redistribuées au sein des équipes pharmaceutiques.

Aujourd'hui de tels systèmes existent, mais n'utilisent pas ces techniques d'apprentissages automatiques. Les systèmes d'aide à la décision s'appuient sur des règles co-écrites par des pharmaciens cliniciens à partir des différentes recommandations à disposition, émettant une alerte lorsqu'une prescription rentre dans les critères prédéfinis. Cet algorithme se différencie par sa construction qui s'appuie sur des données antérieures analysées et validées par les pharmaciens et annotées des interventions réalisées. A partir de ces données annotées, l'algorithme pourra apprendre un modèle prédictif pouvant être transposé à de nouvelles données pour mettre en avant les prescriptions nécessitant l'intervention d'un pharmacien. En complément, nous souhaitons créer une ontologie globale (contre-indications, les précautions d'emploi, les propriétés pharmacodynamiques des médicaments, en plus des interactions médicamenteuses) reprenant chaque principe actif, tout en y associant des données issues des RCP, afin que l'algorithme apprenne le contexte du médicament qu'il pourra ensuite transposer dans le jeu de données étudié.

Les données sont issues du dossier patient informatisé et du serveur de biologie ainsi que des Résumés des Caractéristiques des Produits (RCP). Le corpus des données est constitué de :

- Données administratives : numéro interne d'identification du patient
- Données cliniques : date de naissance, sexe, poids, PAS/PAD, SpO2, EVA, allergies
- Données de mouvement : code UF, code pôle
- Données de prescription : libelle du principe actif, classe ATC 4, dosage, voie d'administration, statut prescription, commentaire médecin, commentaire IDE, titre du prescripteur, dose jour, date de saisie de la prescription
- Données pharmaceutiques : statut pharmaceutique de la prescription, date de validation de la prescription par les pharmaciens
- Données biologiques : glycémie, albumine, calcium, créatinine, CRP, DFGe (CKD-EPI), ferritine, hémoglobine, INR, leucocytes, P. neutrophiles, plaquettes, potassium, sodium, TGO, TGP
- RCP des médicaments : contre-indication, précautions d'emploi, interactions médicamenteuses, propriétés pharmacodynamique (code ATC)

Acteurs internes et/ou académiques mobilisés

- Service de santé publique et Pôle de Pharmacie, Hôpitaux universitaires de Strasbourg et Laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie (ICube, UMR7357).
- Direction : DRCI
- Établissements du réseau PAISaGE

- Établissements du réseau HUGO
- Agence Régionale de Santé Grand Est

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Cet algorithme pharmaceutique prédictif est pensé et mis en œuvre pour lutter contre l'iatrogénie médicamenteuse, diminuer le mésusage et viser la pertinence des soins. Son utilisation doit également contribuer à l'amélioration continue des pratiques et à l'efficacité des organisations.

CHU de Toulouse

Cas d'usage interne : formalisation d'un parcours patient

Résumé du projet et données exploitées

Dans le cadre de la mise en place de la tarification liée à un parcours complet de soins, l'utilisation d'un EDS permettra de croiser efficacement l'ensemble des données issues des différentes sources nécessaires à l'établissement pour reconstituer le parcours complet du patient.

Acteurs internes et/ou académiques mobilisés

Département d'Information Médicale, Service de néphrologie, DSIO (équipe Data). Potentiellement acteur industriel ou académique pour l'extraction automatique des données du dossier médical (extraction du DFG des comptes-rendus).

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Simplification du travail de recueil des données et de leur collation, aujourd'hui effectués « à la main », et source d'une forte charge de travail.

Cas d'usage dans le cadre d'un partenariat avec un industriel : anesthésie-réanimation

Résumé du projet et données exploitées

Récupération automatique des données d'anesthésie et de réanimation du logiciel ICCA de Philips à l'intérieur de l'EDS. Potentiellement récupération de données supplémentaires à plus haute fréquence que les données ICCA. Utilisation de ces données dans le cadre d'une collaboration avec une startup spin-off d'Airbus pour le développement d'un respirateur français (projet CONFIANCE dont la phase 1 est financé par BPI France à hauteur de 6M€).

Acteurs internes et/ou académiques mobilisés

Service d'anesthésie réanimation, DSIO (service Data), Institut de Recherche en Informatique de Toulouse

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

Philips (logiciel ICCA, partenaire), Hilum healthcare (spinoff Airbus, porteur du projet), Initiative de Recherche Technologique Saint-Exupéry

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

L'intérêt du recueil des données pour l'établissement est multiple : intérêt pour le soin et le suivi patient, et participation à un projet de R&D industriel soutenu nationalement.

GCS G4

Cas d'usage dans le cadre d'un partenariat académique : PARADE (Prédiction Automatisée de la Ré-hospitalisation par Apprentissage Décentralisé)

Résumé du projet et données exploitées

Le projet vise à mettre en place des modèles prédictifs de la ré-hospitalisation de patients, à partir des seuls codes diagnostics issus du PMSI. L'objectif est par ailleurs d'apprendre ces modèles prédictifs de manière fédérée, en mobilisant les données de plusieurs centres hospitaliers sans recourir à leur centralisation, par la réalisation synchrone de calculs au sein de chacun des centres détenant de la donnée et le partage de résultats agrégés ne comportant pas d'information individuelle ni directement identifiante.

Données exploitées = données structurées

- PMSI : dates entrée et sortie des séjours MCO (uniquement dans la phase de préparation des données, pas dans l'analyse), codes CIM-10,
- Âge, sexe.

Acteurs internes et/ou académiques mobilisés

CHU de Lille :

- Equipe EDS (data scientist)
- Direction des Ressources Numériques
- Juriste et DPO
- Direction Recherche et Innovation

INRIA, équipe Magnet : développement informatique de la méthode d'apprentissage fédérée

CHU de Rouen :

- Equipe EDS (ingénieur)
- Direction du Système d'Information
- DPO
- Direction Recherche et Innovation

GCS G4 : chargée de mission EDS

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Ce projet permet de réaliser une preuve de concept sur une méthode d'apprentissage fédéré sur des données hospitalières. L'apprentissage fédéré pourrait permettre de mettre en place des études entre CHU, notamment au sein du G4, réseau de CHU du Nord-Ouest, réunissant Amiens, Caen, Lille et Rouen dans un GCS (Groupement de Coopération Scientifique), sans avoir à échanger des données individuelles entre les établissements participants.

GCS HUGO

Cas d'usage interne : GAVROCHE

Résumé du projet et données exploitées

Le projet décrit met en oeuvre des outils d'exploration des données innovants à base d'apprentissage, notamment pour analyser les comptes rendus d'examens avec des techniques de traitement automatique du langage naturel (TALN). Il exploitera également les données structurées (sociodémographiques, biologiques). Il démontre la pertinence des hub

interrégionaux qui, par le volume et la diversité de leurs données, permettent de mener des études ambitieuses en épidémiologie ou sur maladies rares.

L'étude portera sur plus de 20000 patients hospitalisés dans les 5 CHU d'HUGO entre 2011 et 2019.

Le projet GAVROCHE a pour objectif d'étudier le lien entre la variation du taux de sucre dans le sang (glycémie) et le pronostic des patients hospitalisés pour insuffisance cardiaque. L'amélioration du pronostic doit s'appuyer sur l'identification précoce des sujets les plus à risque. Parmi les facteurs qui peuvent déterminer son évolution, l'hyperglycémie à l'admission semble associée au risque de décès. Mais, au-delà de la valeur brute, ce sont les variations de cette glycémie qui seront spécifiquement explorées dans cette étude comme un facteur indépendant qui conditionne des événements cardio-vasculaires ultérieurs.

Acteurs internes et/ou académiques mobilisés

Investigateur coordonnateur 1 : Samy Hadjadj, PU-PH, Endocrinologie, CHU de Nantes

Investigateur coordonnateur 2 : Matthieu Wargny, Assistant Hospitalo-Universitaire, Santé Publique, CHU de Nantes.

Structures impliquées :

- GCS HUGO, Animation du GIRCI et organisation des responsabilités juridiques des flux de données.
- HUGO-ED (Endocrino-Diabétologie), réseau thématique HUGO, expertise en recherche clinique en diabétologie
- L'institut du thorax, UMR INSERM 1087/CNRS 6291, expertise clinique et scientifique dans la prise en charge de l'ICA
- Equipe Traitement automatisée du langage Naturel, Laboratoire des sciences du numérique de Nantes, UMR CNRS 6004
- Equipes des Centres de données cliniques et EDS des CHU d'Angers, Brest, Nantes, Rennes et Tours.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Ce projet, comme le projet HUGO-RD décrit ci-après, a permis de valider le parcours réglementaire des projets devant être menés sur le ODH : avis de notre CSE local dédié au ODH, transmission de cet avis avec le dossier de demande d'autorisation auprès du CESREES et de la CNIL, délivrance de l'autorisation par la CNIL.

Cette recherche nous semble importante du point de vue du patient. Négative, elle limitera des dosages itératifs de la glycémie, inutiles et coûteux. Positive, en confirmant le caractère péjoratif de la variation glycémique, elle permettra d'établir des scores pronostiques favorisant les prises en charge ciblées et dessinant les futures études d'intervention.

La conduite de ce projet est aujourd'hui possible grâce à la convergence de 3 facteurs (i) de la standardisation des pratiques au sein des 5 CHU concernés dans le cadre du GCS HUGO (dosages biologiques, codage PMSI) ; (ii) de la disponibilité presque instantanée des informations (structurées et non structurées) via l'outil Ehop déployé dans les 5 CHU dans le cadre du GIRCI Grand Ouest ; (iii) de la gouvernance locale (recherche, soin et numérique) et inter-régionale (réseau RiCDC/HUGO), sans laquelle un tel projet n'est pas réalisable à l'échelle inter-régionale.

Résumé du projet et données exploitées

Le projet HUGO-RD propose d'ouvrir une voie vers l'intégration et l'exploitation systématique des données issues des consultations de génétique clinique afin d'améliorer et faciliter le diagnostic en génétique moléculaire par séquençage d'exomes. L'étude portera sur 20 000 patients répartis sur 3 CHU (Angers, Nantes et Rennes) et ayant bénéficié d'une consultation de génétique de 2014 à 2020. Cette étude nécessitera de développer un algorithme de Traitement Automatique du Langage Naturel capable d'extraire un phénotype patient standardisé à partir de comptes rendus de génétique clinique. Ceci permettra de garantir une amélioration de la pertinence des interprétations clinico-moléculaires et ainsi d'aider les médecins généticiens dans leur prise en charge des personnes atteintes de maladies rares et de leur famille afin de diminuer l'errance diagnostique.

Acteurs internes et/ou académiques mobilisés

Investigateur coordonnateur 1 : Marie de Tayrac, PU-PH, Génétique, bio-informatique, CHU de Rennes

Responsable CDC : Marc Cuggia, PU-PH, Santé Publique, CHU de Rennes.

Structures impliquées :

- ANDDI RARES, DEFISCIENCE, BRAINTEAM (coordonnée par Pr Christophe Verny, CHU d'Angers)
- Centres de référence anomalies du développement et syndromes malformatifs de l'Ouest (CLAD Ouest), coordonné par le Pr Sylvie Odent au CHU de Rennes, qui associe les CRM d'Angers, Nantes et Tours, ainsi que les centres de compétence de Brest, Vannes et Le Mans
- UMR 6290 INSERM, Institut de Génétique et Développement de Rennes (IGDR) et
- UMR 1099 INSERM LTSI - Equipe Données Massives en Santé (Health Big Data), Pr Marc Cuggia.
- GCS HUGO, Animation du GIRCI et organisation des responsabilités juridiques des flux de données.
- Equipes des Centres de données cliniques et EDS des CHU d'Angers, Nantes et Rennes.

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Le projet HUGO-RD propose d'ouvrir une voie vers l'intégration et l'exploitation systématique des données issues des consultations de génétique clinique afin d'améliorer et faciliter le diagnostic en génétique moléculaire par séquençage d'exome. Dans ce projet, nous proposons de dépasser les standards actuels du diagnostic clinique et moléculaire des maladies rares en permettant, dans le cadre du soin, le chaînage et la structuration des données cliniques (phénotype) et des données de NGS dans le Grand-Ouest. Le projet HUGO-RD permettra de structurer au sein des CDC les données nécessaires à un diagnostic génétique innovant pour les maladies rares, et ainsi la création d'un corpus unique et multicentrique regroupant données génétiques (exome), comptes rendus médicaux annotés et traductions en terme.

Pour la région Grand-Ouest, le projet HUGO-RD sera une opportunité pour se positionner au rang de leader sur ces thématiques, au niveau national mais également au niveau européen (ERN ITHACA, H2020 Solve-RD, RD-connect, IRDiRC).

Cas d'usage dans le cadre d'un partenariat avec un industriel : études de faisabilité

Il n'y pas eu à ce jour d'études menées avec un partenaire industriel sur la plateforme Ouest DataHub, mais des études de faisabilité actuellement menées permettent de connaître certains besoins de grands groupes pharmaceutiques en matière d'études sur données de santé de vie réelle et d'évaluer le potentiel de notre plateforme. Une étude de faisabilité type est décrite ci-dessous.

Résumé du projet et données exploitées

L'objectif de l'étude est d'évaluer l'espérance de vie de patientes sous Molécule_anticancéreuse_1 en 1ère ligne en association à un inhibiteur de l'aromatase. La molécule a été mise sur le marché récemment (4 ans). Dans le cadre du renouvellement de son AMM, l'industriel souhaite exploiter des données de vie réelle issues de CHU et de CRLCC. Les données nécessaires à l'étude pour réaliser une analyse fine du parcours patient sont les suivantes :

- Age / date de naissance
- Ligne de traitement au stade métastatique
- Traitement adjuvant précédent (lequel) / date de fin du traitement adjuvant
- Date de début du traitement avec IBRANCE
- Traitements associés à IBRANCE
- Date de fin du traitement par IBRANCE ou date de début du traitement ultérieur
- Traitements ultérieurs
- Décès (oui / non) et date de décès

Les données issues d'au minimum 1000 patients est attendu pour réaliser cette étude, nombre de patients qu'aucun établissement ne peut à lui seul inclure dans l'étude.

Acteurs internes et/ou académiques mobilisés

Les EDS-CDC mobilisés (médecin de santé publique, datascientist, datamanager,...) pour cette étude sont issus des établissements du Grand Ouest suivants : Institut et de cancérologie (Nantes et Angers), Centre Eugène Marquis (Rennes), CHU de Brest, CHU de Tours. Par ailleurs, un clinicien référent/cancérologue des établissements sollicités est associé à chaque EDS-CDC pour la définition et la conduite de l'étude, l'analyse des résultats.

Industriels (partenaires et rôles) et modalités de sollicitation par/de l'industriel

L'industriel fait partie des 5 premiers groupes pharmaceutiques mondiaux et a pris contact directement auprès du guichet unique du ODH. Il apporte la problématique (besoin de données de vie réelle pour compléter un dossier réglementaire d'une molécule déjà sur le marché) avec un certain nombre d'exigences : nombre minimum de patients « inclus », périmètre de l'étude (2 CRLCC, 2 CHU), niveau d'exigence sur la qualité des données, délai de restitution des résultats. La plateforme ODH prend en charge la rédaction du protocole, le parcours réglementaire, la conduite de l'étude (méthodes de collecte des données, coordination des CDC-EDS, gestion de la plateforme d'hébergement des données issues des EDS-CDC, analyse des données, synthèse des résultats).

Résultats principaux attendus et/ou obtenus et impacts potentiels pour le système de santé et/ou l'établissement

Ces premières études avec des partenaires industriels ont pour vocation, d'une part, à servir de preuve de concept pour des études menées sur le ODH avec des objectifs différents de celles menées dans le cadre de la recherche académique : elles permettent d'éprouver notre organisation et notre capacité à mener des études dans des temps contraints avec un niveau de qualité élevé. Par ailleurs, ces études s'inscrivent directement dans le modèle économique

de notre plateforme et ont vocation à compléter les financements publics dont elle bénéficie et à permettre de renforcer sa structuration et son développement.

Annexe 2 - Méthodologie de chiffrage du coût d'un entrepôt de données de santé

Introduction & contexte

Cette note a pour objectif de détailler les différents postes de coûts qui accompagnent la construction d'un Entrepôt de Données de Santé (EDS). Les estimations documentées dans cette note ont permis la construction de la part fixe du modèle de financement proposé par le groupe de travail.

Afin d'objectiver le besoin de financement pérenne des EDS, les membres du groupe ont entrepris sur plusieurs mois un travail d'alignement autour du chiffrage d'un EDS dit "type", c'est à dire intégrant les composants, flux et ressources requis a minima pour constituer un EDS de taille moyenne. Afin de s'inscrire au mieux dans la logique d'un financement pérenne public qui interviendrait en relai de la première initiative d'appel à projets actuellement portée par le gouvernement, les membres du groupe se sont placés dans une logique prospective pour estimer les coûts de l'EDS type. Cela signifie que les membres du GT ont cherché à estimer les coûts de ce à quoi devrait ressembler un EDS moyen en cible, plutôt que de raisonner dans les limites des entrepôts de données de santé existant actuellement dans l'hexagone, lesquels sont souvent encore en cours de développement ou pas encore passés à l'échelle.

La première étape a été de s'aligner sur les trois grandes phases du développement d'un EDS et leur périmètre d'activités respectif. Ci-dessous les trois phases qui ont été retenues par leur groupe et un résumé des activités principales concernées :

- **Cadrage** : investiguer le cadre réglementaire, définir une solution technologique cible, identifier les besoins en développement, déterminer les partenariats technologiques à conclure et licences à acquérir le cas échéant, qualifier et prioriser les sources de données à ingérer, etc.
- **Build** : déployer et interconnecter les différentes briques techniques (infrastructure, licences), mettre en place un catalogue de métadonnées, mettre en place une stratégie de test et d'audit de sécurité, recruter les ressources humaines nécessaires à l'administration de l'EDS, définir les processus de gouvernance des données, etc.
- **Run** : maintenir les processus de mise en qualité des données, intégrer de nouveaux flux de données afin d'élargir le périmètre d'activité de l'EDS, financer les coûts de consommation liés aux projets et aux prestations de sécurité, assurer le respect des droits des citoyens dans la durée, promouvoir l'EDS à l'écosystème, accueillir un nombre croissant de projets sur les données de l'EDS, etc.

La seconde étape a consisté en l'identification fine des postes de coûts et activités associés à chaque phase. Pour lancer ce chantier, le Health Data Hub a proposé au groupe un premier chiffrage basé sur les coûts de sa plateforme technologique et de son offre d'accompagnement aux porteurs de projets de réutilisation secondaire des données de santé. A partir de ce premier modèle, les CHU membres du GT ont mobilisé en sous-groupes les profils pertinents au sein de leurs équipes pour permettre au groupe de converger vers un modèle de chiffrage au plus proche de la réalité des établissements de santé. Une grille de coûts commune a ainsi pu être validée en séance plénière par les membres. Les membres du groupe se sont également mis d'accord sur une grille salariale unique décrivant les rémunérations annuelles observées en moyenne des différents profils mobilisés dans le cadre du développement d'un EDS hospitalier.

Dans un troisième temps, l'objectif du groupe a été de déterminer la durée de chaque phase, le nombre de ressources humaines à mettre en face des activités identifiées pour chaque phase, et le dimensionnement financier associé. Plusieurs établissements de santé se sont prêtés à l'exercice de remplissage de la grille de coûts commune : AP-HP, CHU de Brest, CHU de Grenoble, CHU de Lille, CHU de Reims, CHU de Rouen, GCS HUGO (établissements membres : CHU d'Angers, le CHU de Brest, le CH du Mans, le CHU de Nantes, le CHR d'Orléans, le CHU de Rennes, le CHU de Tours, le CH de Vendée et l'Institut de Cancérologie de l'Ouest). Ces multiples contributions ont été consolidées pour permettre la construction d'une fourchette de coûts représentative de la diversité des établissements de santé mobilisés au sein du groupe. Les membres du groupe se sont également accordés pour fixer la durée de développement moyenne d'un EDS à cinq années.

1) Construction d'un EDS - Cadrage

Durée moyenne de la phase de cadrage : 1 année

Le projet d'EDS débute par une phase de cadrage, dont la durée varie en fonction du niveau de maturité de l'établissement en matière d'initiatives autour de la donnée. Cette phase mobilise fortement les acteurs en interne à l'établissement de santé autour des activités suivantes :

- Stratégie de l'EDS : définition des objectifs et finalités de l'EDS, élaboration de la feuille de route, identification de partenariats, etc.
- Organisation du projet: cadrage du modèle opérationnel, des instances, etc.
- Cadrage médical : mise en place de l'écosystème de l'EDS, mise en place d'une stratégie d'adhésion au projet par la communauté médicale, etc.
- Cadrage de la gouvernance et des modalités de l'accès aux données : modalités de demande d'accès aux données de l'EDS, modalités de valorisation scientifique de l'EDS, modalités pratiques de mise à disposition des données en interne et en externe à l'établissement, etc.
- Cadrage des flux de données : participation à l'identification des sources de données prioritaires, élaboration des spécifications fonctionnelles et d'une cartographie des flux, etc.
- Cadrage réglementaire et juridique : identification du cadre réglementaire adapté, lancement des démarches, définition de la stratégie d'information aux patients, définition des processus de conventionnement avec les porteurs de projets internes et externes, etc.
- Définition de l'infrastructure technique cible : définition de la solution technologique cible, identification des besoins en développement, des partenariats technologiques à conclure et licences à acquérir le cas échéant, qualification et priorisation des sources de données à ingérer, définition de la stratégie de sécurité informatique, etc.
- Cadrage de l'offre de service à destination des utilisateurs : identification des besoins des utilisateurs et définition d'une offre de service, cadrage du catalogue de métadonnées de l'EDS afin de mettre en visibilité les données disponibles, etc.

a) Stratégie de l'EDS

Le cadrage d'un EDS mobilise fortement la direction générale de l'établissement, la direction de la recherche et innovation, et la direction des services numériques. L'alignement en interne et la mobilisation des différents décideurs représente une charge à ne pas sous-estimer afin que cette phase de cadrage ne se prolonge pas trop dans le temps.

Les profils mobilisés sont :

- Chef de projet maîtrise d'ouvrage

Les principales activités couvertes sont :

- Conduite d'ateliers pour aligner les directions de l'établissement autour d'une vision cible, des usages visés, et création d'une feuille de route globale du projet
- Cadrage et animation de la comitologie : comité de pilotage, comité scientifique et éthique, interactions entre les différentes instances, etc.
- Cadrage du modèle opérationnel : Dimensionnement de l'équipe cible, son rattachement hiérarchique et fonctionnel et les interactions avec les autres directions de l'établissement
- Coordination des ressources internes et éventuellement externes
- Modèle de financement de l'entrepôt de données de santé, description des pistes de pérennisation sur tout ou partie des coûts de l'EDS et des projets de recherche
- Conduite du changement et lancement de la communication en interne aux professionnels de santé autour du projet d'EDS
- Définition d'une stratégie d'interopérabilité de l'EDS avec d'autres établissements de santé et le Health Data Hub

b) Organisation du projet

La phase de cadrage permet également d'établir une gouvernance opérationnelle du projet afin de coordonner les équipes chargées du développement et du déploiement technique de l'EDS.

Les profils mobilisés sont :

- Chef de projet maîtrise d'oeuvre SI

Les principales activités couvertes sont :

- Cadrage et animation de la comitologie opérationnelle : comité de suivi opérationnel
- Déploiement du modèle opérationnel : recrutement des profils cibles

c) Cadrage de la gouvernance et des modalités de l'accès aux données

La phase de cadrage d'un EDS implique une réflexion autour de comment sera géré l'accès aux données de ce dernier par les futurs utilisateurs internes et externes à l'hôpital. La gouvernance des données préfigurée en phase de cadrage jouera un rôle clé dans l'attractivité future de l'EDS et l'accélération de l'utilisation de ses données, en prévoyant des accès aux données fluides, clairs et sécurisés pour les utilisateurs.

Les profils mobilisés sont :

- Chef de projet maîtrise d'ouvrage

Les principales activités couvertes sont :

- Cadrage des modalités de demande d'accès aux données de l'EDS, modalités et procédure d'instruction des demandes d'accès aux données par des porteurs de projet internes à l'hôpital ou externes, outillage et publicité (site internet, etc.)
- Cadrage du comité scientifique et éthique : composition, règles de fonctionnement, et outillage
- Cadrage des modalités de partage de données (tarification, valorisation scientifique, délais etc.)

- Cadrage des modalités pratiques de mise à disposition des données en interne et en externe à l'établissement
- Mesures prises pour assurer la conformité des modalités décrites ci-dessus avec la gouvernance nationale des données de santé

d) Cadrage des flux de données

Avant de lancer la phase de déploiement technique de l'entrepôt, il est essentiel de cadrer précisément le socle de données prioritaires à inclure au sein de l'EDS et d'établir l'ensemble des spécifications techniques permettant leur ingestion et leur mise en qualité.

Les profils mobilisés sont :

- Expert fonctionnel

Les principales activités couvertes sont :

- Participation à l'identification des sources de données prioritaires à inclure dans l'EDS et identification des logiciels source
- Rédaction des spécifications encadrant la création et le déploiement des canaux de transfert entre les centres de production des données et l'EDS
- Elaboration d'une cartographie des flux de données
- Cadrage des modalités de mise en qualité et de standardisation des données afin de faciliter leur réutilisation future

e) Cadrage réglementaire et juridique

Le lancement d'un EDS dans un établissement de santé nécessite le montage d'un dossier réglementaire permettant son autorisation. La CNIL a récemment publié un référentiel simplifié permettant de s'abstenir de l'autorisation formelle mais l'évaluation de la conformité au référentiel implique un travail conséquent du service juridique similaire (environ 120 items à vérifier). Par ailleurs, l'utilisation de l'EDS par des tiers ayant vocation à utiliser les données de santé pour leur projet requiert également un cadre juridique (contrat type d'utilisation par exemple). Les activités présentées ci-dessous ont vocation à se prolonger en phase de build.

Les profils mobilisés sont :

- Délégué à la protection des données (DPO)

Les principales activités couvertes sont :

- Evaluation du cadre réglementaire
- Définition d'une stratégie d'information des personnes et notamment cadrage des métadonnées du répertoire public
- Définition d'une stratégie permettant aux personnes d'exercer leurs droits
- Réalisation d'une AIPD
- Préparation du dossier et du passage aux différentes instances réglementaires le cas échéant

Nota Bene : Selon les retours d'expérience d'aide à la constitution d'EDS ou d'appui à la constitution des dossiers de demande d'autorisation CNIL, le cadrage réglementaire dure 4 à 6 mois.

f) Définition de l'infrastructure technique cible

i) Qualification et priorisation des sources de données

Dès la phase de cadrage, les premières sources de données doivent être identifiées et priorisées pour identifier l'infrastructure technique cible qui permettra de les ingérer et les valoriser de manière optimale. Un EDS sans données ne présente pas d'intérêt. Ces travaux de cadrage sont menés à bien par un ou plusieurs data engineers (appuyés par le ou les chefs de projet) en collaboration étroite avec la gouvernance du projet d'EDS étant donné le caractère stratégique de la priorisation des sources de données pour l'établissement. Ces travaux devront se poursuivre par la suite, puisque l'EDS sera enrichi de nouvelles sources de données dans le temps.

Les profils mobilisés sont :

- Data engineer

Les principales activités couvertes sont :

- Identification détaillée des usages de l'EDS (pilotage, recherche, les deux) et de l'offre logicielle appropriée à ces usages
- Recensement des sources de données existantes et à venir au sein de l'établissement (patients, séjours, diagnostics, etc.)
- Evaluation de la qualité des données
- Définition et application de critères de priorisation pour l'ingestion des sources de données, en cohérence avec la stratégie de l'établissement sur les volets recherche et pilotage
- Planification de vagues d'ingestion sur 2 ans
- Réalisation d'une cartographie des sources de données, des systèmes et des flux prioritaires
- Identification des flux et connecteurs à mettre en place entre les systèmes
- Identification et sélection d'un format de donnée unique partagé par l'ensemble des acteurs

ii) Design de l'architecture technique et stratégie de déploiement

La constitution d'un EDS nécessite la définition d'une architecture cible dans un environnement technologique préexistant (le système d'information de l'établissement). L'adéquation de l'EDS avec ce dernier sera un véritable défi nécessitant une expertise technique forte pour garantir sa pérennité.

Les profils mobilisés sont :

- Expert architecture

Les principales activités couvertes sont :

- Définition de la stratégie de développement et d'intégration des fonctionnalités
- Identification des solutions techniques et de la stratégie d'intégration de ces dernières (identification des intégrateurs si l'intégration est externalisée)
- Formalisation d'un dossier d'architecture technique et fonctionnelle
- Identification d'une organisation pour l'opération de la plateforme et sa maintenance. Plusieurs rôles peuvent devoir être définis pour assurer le cloisonnement des droits et donc la protection des données
- Définition et rédaction des procédures d'opération de la plateforme (sauvegarde des ressources, procédure d'accès, procédure d'ingestion des données, procédure de copie, préparation et mise à disposition des données, procédure d'incidents, etc.)

- Cadrage du dossier d'homologation de la plateforme pour les démarches réglementaires
- Participation au cadrage des modalités d'interopérabilité de l'EDS avec d'autres établissements de santé et le Health Data Hub
- Participation au cadrage du catalogue de métadonnées de l'EDS : identification des solutions
- Cadrage du répertoire public des projets et de l'outil de gestion des demandes d'accès le cas échéant

iii) Cadrage SSI

La constitution d'un EDS nécessite un haut niveau de sécurité au regard des données traitées. L'implication d'un expert en Sécurité des Systèmes d'Information (SSI) est donc indispensable dès la phase de cadrage pour garantir la bonne mise en place de l'EDS et sa pérennité.

Les profils mobilisés sont :

- Responsable de la Sécurité des Systèmes d'Information (RSSI)

Les principales activités couvertes sont :

- Participation aux travaux d'architecture et aux choix des solutions
- MOA de l'évaluation des risques et la mise en place de plans de remédiation
- Définition d'une stratégie de sécurité allant de l'encadrement des transferts de données entre systèmes, des tests de la solution et audits de sécurité avant la mise en production, de la procédure de gestion des logs, plan de surveillance et de contrôle de la solution, à la gouvernance d'accès au système (identification des rôles et de leurs droits associés, habilitation, création de CGU, sensibilisation, etc) etc.
- Aide au cadrage du dossier réglementaire sur les volets techniques et sécurité de l'information et notamment la stratégie d'homologation, l'Analyse d'Impact sur la Protection des Données (AIPD), les procédures de sécurité en vue notamment de la réalisation d'une homologation de l'EDS en tant que système d'information
- Définition d'un plan de Continuité d'Activité (PCA) et Plan de Reprise d'Activité (PRA)
- Pilotage de l'analyse de risque en amont du déploiement de la solution et mise à jour régulière pour apprécier les risques nouveaux liés à l'évolution de la menace, l'intégration de nouvelles sources ou outils
- Analyse de risque (25 000€ en moyenne, prestation externe)

g) Cadrage de l'offre de service à destination des utilisateurs

Pour préparer une utilisation de l'EDS la plus large possible, l'établissement doit sécuriser la mobilisation d'une ressource afin de piloter le déploiement d'une offre de service à destination des futurs utilisateurs internes et externes de l'EDS.

Les profils mobilisés sont :

- Chef de projet maîtrise d'ouvrage

Les principales activités couvertes sont :

- Stratégie d'identification des besoins des utilisateurs et définition d'une offre de service (formation, animation, documentation, appui en expertise)
- Cadrage du catalogue de métadonnées de l'EDS afin de mettre en visibilité les données disponibles

2) Construction d'un EDS - Build

Durée de la phase de Build : 3 années

La structure de coûts de la phase de build d'un EDS peut être décomposée en 7 centres de dépenses principaux :

- Pilotage global : pilotage de l'avancement global du projet, respect des délais, etc.
- Conception des flux de données
- Mise en place de la gouvernance et des modalités de l'accès aux données : Mise en place des outils permettant de réaliser les procédures de demande d'accès, mise en place du comité scientifique et éthique, contrats type
- Mise en oeuvre du chantier juridique et réglementaire : développement d'un kit de contractualisation
- Construction de l'infrastructure technique de l'EDS : identification des premiers développements, construction de l'architecture technique cible, sélection et intégration des outils, prestation de sécurité suite au cadrage SSI
- Construction des premiers flux de données : développement des canaux d'ingestion, mise en qualité des données, transformation dans un format commun et interopérable
- Construction et déploiement de l'offre de service à destination des utilisateurs : conception des différents éléments de l'offre de service à destination des utilisateurs identifiés (formations, documentation ouverte, bibliothèque de programme, appui en expertise, animation de sessions collectives, etc.)

a) Pilotage global

La construction d'un EDS requiert la mobilisation d'un grand nombre d'acteurs pouvant provenir de départements différents avec des méthodes de travail hétérogènes. Le pilotage et l'orchestration d'un tel projet nécessite l'intervention d'un chef de projet afin de garantir l'ambition du projet tout en respectant le calendrier de mise en oeuvre prévu.

Les profils mobilisés sont :

- Chef(fe) de projet maîtrise d'ouvrage
- Chef(fe) de projet maîtrise d'oeuvre SI

Les principales activités couvertes sont :

- Pilotage de la feuille de route du projet
- Organisation et préparation de la comitologie : mise en place d'un comité de pilotage, du comité scientifique et éthique et de leur outillage
- Appui à l'identification des ressources à mobiliser sur les différents chantiers
- Rédaction des différents dossiers internes et externes avec l'appui des experts associés (e.g. demande d'autorisation, etc.)
- Coordination des ressources internes et externes le cas échéant
- Pilotage financier
- Conception d'un plan de passage à l'échelle de l'EDS
- Pilotage de la stratégie d'interopérabilité de l'EDS avec d'autres établissements de santé et le Health Data Hub

b) Conception des flux de données

Les profils mobilisés sont :

- Expert fonctionnel

Les principales activités couvertes sont :

- En lien avec les experts médicaux, écriture des spécifications fonctionnelles pertinentes pour déployer les flux de données priorités en phase de cadrage
- Articulation des flux de données envisagés avec les différentes briques applicatives préexistantes au sein du SI de l'hôpital

c) Mise en place de la gouvernance et des modalités de l'accès aux données

Durant la phase de build, un profil chef de projet MOA est chargé de construire et déployer les outils et instances de gouvernance d'accès aux données de l'EDS préfigurés lors de la phase de cadrage.

Les profils mobilisés sont :

- Chef de projet maîtrise d'ouvrage

Les principales activités couvertes sont :

- Mise en place des outils permettant de réaliser les procédures de demande d'accès
- Mise en place du comité scientifique et éthique
- Conception et rédaction des clauses des contrats type relatives aux conditions d'accès aux données et mise en place des processus de contractualisation

d) Mise en oeuvre du chantier juridique et réglementaire EDS

Une fois le montage du dossier réglementaire réalisé lors de la phase de cadrage, ce dernier doit être déposé auprès de la CNIL pour autorisation le cas échéant. Une fois l'autorisation obtenue, les premiers projets de réutilisation des données de l'EDS pourront être lancés. Pour anticiper ces demandes, un kit de contractualisation doit être développé.

Les profils mobilisés sont :

- DPO

Les principales activités couvertes sont :

- Rédaction des clauses des contrats type relatives à l'utilisation de la plateforme de l'EDS dans le respect des droits des personnes et des exigences de sécurité
- Rédaction des CGU
- Conception des outils de sensibilisation et formation éventuelle des utilisateurs

e) Construction de l'infrastructure technique de l'EDS

Le développement de l'infrastructure d'entrepôt nécessite l'implémentation de fonctionnalités et du socle technique, une mobilisation conséquente de ressources est donc indispensable. Ces profils sont rares et nécessitent régulièrement l'appui d'une prestation externe. Des licences permettant le fonctionnement sont également à prévoir.

Il convient de préciser que l'estimation budgétaire proposée est basée sur le recours à une solution cloud pour l'hébergement de l'EDS, impliquant un investissement moindre lors de la construction.

i) Développement et intégration de l'architecture technique cible

Le développement technique d'un EDS requiert la construction des solutions techniques nécessaires pour répondre aux différents usages visés (en interne avec l'administration de l'EDS et la préparation des données, la préparation des données et en externe si l'EDS est ouvert aux tiers porteurs de projet de recherche par exemple), dans le respect des exigences de sécurité (outils de supervision).

Les profils mobilisés sont :

- DevOps

Les principales activités couvertes sont :

- Cadrage des opérations de gestion de la plateforme
- Construction des environnements de développement, préproduction, et production
- Intégration et développement des briques identifiées en phase de cadrage dans l'architecture fonctionnelle et le dossier d'architecture
- Développement du catalogue de métadonnées de l'EDS, du répertoire public des projets et de l'outil de gestion des demandes d'accès le cas échéant
- Intégration des différentes licences d'outils au SI existant de l'établissement
- Mise en oeuvre de mesures pour assurer l'interopérabilité de l'EDS avec d'autres établissements de santé et le Health Data Hub

ii) Licence/infrastructure

Le socle technologique d'un EDS repose sur plusieurs outils incontournables pour une plateforme data et accessibles via l'acquisition d'une licence auprès d'un tiers.

Les principales licences à acquérir pourraient être :

- Licence d'infrastructure cloud et stockage, comme par exemple :
 - Module de gestion des clés de chiffrement pour une plateforme Data sécurisée (~180 000€ annuel)
 - Licence pour des serveurs GPU (plusieurs centaines de milliers d'euros)
 - Licence Direct connect pour garantir les performances de son réseau (plusieurs centaines de milliers d'euros)
- Licences d'intégration et transformation : nécessaires à la mise en place d'un Datalake, Datawarehouse et Datamarts
- Licences d'outils de traitement et exposition : outils d'analyse de données souvent gratuits mais dont les usages professionnels (avec traçage des log par exemple) peuvent être payants, annotation d'image, outils pour la réalisation de croisement, data visualisation, etc.
- Licences pour les outils de sécurité tels que bastion, outils de gestion des logs pour tracer les activités réalisées

La sélection des licences est une étape importante dans la construction d'un EDS et il apparaît nécessaire de prendre des licences professionnelles pour fournir le niveau de protection maximal aux citoyens et notamment vis-à-vis du RGPD.

Les coûts relatifs à l'acquisition initiale des premières licences et briques techniques nécessaires à la mise en place de l'EDS sont estimés entre 100 000 € (cas d'un EDS utilisant en grande majorité des logiciels et solutions ouverts) et 450 000 €. Cette fourchette reflète les estimations des 7 établissements de santé ayant participé à l'exercice de chiffrage d'un EDS type. Ces coûts sont par ailleurs cohérents avec les pratiques du marché constatées par le HDH sur des outils et licences similaires.

iii) Prestation de sécurité

Sur la base de la stratégie de sécurité informatique définie en phase de cadrage, des analyses indépendantes et externes devront être menées pour garantir un haut niveau de sécurité de l'EDS.

Les principales prestations seront :

- Audits de sécurité réguliers inscrit dans la stratégie de déploiement des fonctionnalités (un audit de sécurité peut être plus ou moins large, le premier peut l'être et contenir un audit technique mais également organisationnel, de configuration et de code)

La prestation de sécurité informatique comprenant les activités citées ci-dessus sera menée par une entreprise spécialisée. Le coût annuel moyen pour cette prestation est évalué à 100 000€ et peut être inclus dans la prestation d'audit général des systèmes d'informations de l'établissement. Suite au cadrage, la mobilisation d'un expert SSI devra être reconduite.

f) Construction des premiers flux de données

Dès la phase de Build, les premières sources de données doivent être restructurées pour faciliter leur intégration au sein de l'EDS. Des canaux de transfert/ingestion doivent ensuite être développés pour permettre les flux de données préfigurés en phase de cadrage. Ces travaux sont principalement menés à bien par des data engineers et devront se poursuivre par la suite, puisque l'EDS sera enrichi de nouvelles sources de données dans le temps.

Les profils mobilisés sont :

- Data engineer

Les principales activités couvertes sont :

- Construction des flux et connecteurs à mettre en place entre les systèmes
- Mise en qualité des données
- Transformation des données au format unique et interopérable

Ces activités seront appuyées par le(s) chef(s) de projet et par la direction des services numériques de l'établissement de santé.

g) Construction et déploiement de l'offre de service à destination des utilisateurs

Les profils mobilisés sont :

- Chef de projet maîtrise d'ouvrage

Les principales activités couvertes sont :

- Conception des différents éléments de l'offre de service à destination des utilisateurs identifiés (formations, documentation ouverte, bibliothèque de programme, appui en expertise, animation de sessions collectives, etc.)
- Pilotage de la création d'une base documentaire pour les utilisateurs, décrivant les scripts de préparation de données, les bonnes pratiques d'utilisation des outils
- Pilotage de la mise en place des informations sur le site internet pour que les porteurs de projet potentiels puissent trouver facilement l'EDS et la manière d'y accéder

3) Construction d'un EDS - Run

La phase de Run débute en parallèle de la seconde année de build, et suit une trajectoire de montée en charge progressive jusqu'à la cinquième année et dernière année de développement de l'EDS. Le groupe a estimé comme décrit ci-dessous les coûts relatifs à une année complète de Run en cible et en a imputé une partie en année 3 (20%) et année 4 (60%) pour refléter cette montée en charge.

La structure de coûts pour la phase de Run d'un EDS peut être décomposée en deux grandes catégories :

- Le Run technique, qui correspond aux activités de maintien en condition opérationnelle, appui aux nécessaires évolutions technologiques de l'infrastructure et évolutions du périmètre de données (intégration de nouveaux flux simples, flux complexes) ;
- Le Run exploitation, qui correspond aux activités de screening, étude de faisabilité, préparation et extraction des données et gestion des environnements de traitement.

Run Technique

a) Coordination technique

Pour accompagner le passage à l'échelle progressif de l'EDS, un pilotage du maintien et de la coordination technique de cette infrastructure, en interne et en externe à l'établissement, devient indispensable.

Les profils mobilisés sont :

- Chef de projet maîtrise d'oeuvre SI

Les principales activités couvertes sont :

- Pilotage de la mise à disposition d'expertise technique pour lever d'éventuels blocages sur les projets de réutilisation des données de l'EDS
- Pilotage de l'alimentation en continu du catalogue de métadonnées de l'EDS en lien avec les besoins des utilisateurs
- Mesures prises pour assurer la mise en open source des éléments de documentation ou outils

b) Gestion de l'infrastructure technique en Run

Une fois le développement du socle technologique et des différentes fonctionnalités de l'EDS terminé, les activités de maintenance en conditions opérationnelle (MCO) et maintenance en conditions de sécurité (MCS) devront prendre le relais avec le développement d'éventuelles corrections ou évolutions à apporter à l'architecture et aux outils de l'EDS. Le chantier SSI a également vocation à se prolonger dans la phase de Run avec notamment le pilotage et le suivi de la stratégie de sécurité informatique de l'EDS. Ce chantier aura même vocation à s'élargir. Il sécurisera dans un premier temps les problématiques SSI autour d'un cas d'usage pilote puis prendra en charge les tâches plus classiques prévues dans le cadre du chantier, notamment la bonne gestion des incidents informatiques et de sécurité dans un contexte d'accroissement fort du nombre de cyberattaques (e.g. fuite de données dans des établissements de santé d'importances ou ransomware). Des compétences data engineering seront d'autre part mobilisées pour accompagner le maintien et l'amélioration des standards d'interopérabilité, du guichet d'accès aux données, de l'espace de mise à disposition de ces dernières et de l'alimentation du catalogue de métadonnées.

Les profils mobilisés sont :

- DevOps
- Ingénieur DSN
- RSSI
- Data engineer

Les principales activités couvertes sont :

- Administration de la consommation des licences. *Nota Bene* : La sélection de licences dans la phase de construction d'un EDS impliquera obligatoirement un coût annuel d'utilisation pour garantir son fonctionnement à niveau de service égal. La ligne de coût prévue dans le chiffrage décrit les coûts de consommation de licence généraux : saut de version EDS, saut de version majeur EDS, outils d'analyse (Rserver, ...), outils techniques (Vmware, ...), outils sécurité (Octopize, ...), Outils métier. Le détail peut être retrouvé dans l'onglet paramètres.
- Maintenance en conditions opérationnelles, maintenance en conditions de sécurité : documentation, pilotage et audit régulier des opérations de gestion de la plateforme, corrections des bugs, mise en place des espaces projets, ouverture des droits et gestion des accès, maintenance des mesures de sécurité mise en place, poursuite des développements continus pour répondre aux enjeux des porteurs de projets.
- Surveillance de la plateforme : Mise à jour de l'analyse de risque et appréciation du caractère personnel/sensible des nouvelles données à ingérer et traiter ou des nouveaux outils (mise à jour régulière), gestion des incidents et mise en place de plans de remédiation, gestion du SOC, pilotage du plan de surveillance, suivi de la stratégie de sécurité informatique (encadrement des transferts de données, tests, audits, revue des habilitations, etc.), maintien à jour des procédures et suivi d'un plan de contrôle, organisation et animation de sessions de sensibilisation auprès des utilisateurs.
- Prestation d'audit de sécurité et pilotage de cette dernière
- Gestion interopérabilité
- Gestion catalogue de métadonnées et guichet d'accès aux données
- Gestion environnement mise à disposition données

c) Evolution technologique de l'EDS

L'EDS est une infrastructure technologique encore relativement récente. De nombreuses évolutions technologiques sont à anticiper et leur intégration requiert souvent la mobilisation de compétences rares non préexistantes au sein de l'établissement de santé. Il paraît alors raisonnable de provisionner ces sauts technologiques importants et de prévoir une compétence DevOps pour y faire face.

Les profils mobilisés sont :

- DevOps

Les principales activités couvertes sont :

- Optimisation et évolution architecture
- Intégration des différentes API,
- Diverses intégrations technologique innovantes

d) Préparation et optimisation en continu des données issues de l'alimentation de nouveaux flux

Les activités de préparation menées lors du cadrage sur les premières sources de données auront vocation à se prolonger en phase de run avec l'identification de nouvelles sources de données à prioriser et restructurer avant ingestion dans l'EDS. L'équipe de l'EDS sera garante de la bonne conduite des travaux d'interopérabilité, mise en qualité et ingestion sur les nouvelles sources de données à venir. A noter que l'ambition du projet dépendra de manière importante du dimensionnement de cette équipe, qui pourra également s'appuyer sur les ressources de la direction des services numériques de l'établissement. Une majoration incrémentale liée aux nouveaux types de sources de données ingérées (distinction flux simples versus flux complexes décrite plus bas) a également été prévue par le groupe dans le chiffrage, à hauteur de 25% des coûts totaux liés à l'ingestion de ces flux supplémentaires (lignes 43 à 57).

Les profils mobilisés sont :

- Data Engineer
- Data Scientist

Les principales activités couvertes sont :

- Enrichissement, évolution interopérabilité, harmonisation : identification en continu des types de données nécessaires (patients, séjours, diagnostics, etc.), évaluation de la qualité des données, application de critères de priorisation, mise en qualité, transformation au format unique et interopérable
- Préparation des données en cohérence avec l'évolution du périmètre de données : maintien et mise à jour de la stratégie d'ingestion, etc.

e) Mise en place de trois flux simples par an

Le nombre de flux construits et ingérés par an au sein de l'EDS impacte considérablement le dimensionnement financier de cette infrastructure. A titre d'illustration, les membres du GT ont convenu de chiffrer les coûts associés à la mise en place de trois flux simples par an. Un flux simple est un flux normalisé, qu'on va pouvoir retrouver dans les référentiels d'interopérabilité par exemple, dont l'alimentation implique un coût et un effort moyen.

Les profils mobilisés sont :

- Médecin référent métier
- Expert fonctionnel
- Data engineer
- DevOps

Les principales activités couvertes sont :

- Alimentation des flux simples supplémentaires : Expression du besoin sur les nouveaux flux de données, conception et implémentation des nouveaux flux
- Administration des licences spécifiques aux flux simples le cas échéant
- Evolution de l'architecture technique liée à la mise en place de flux simples

f) Mise en place d'un flux de données complexes par an

Un flux complexe est un flux non normalisé, dont l'intégration et le maintien impliquent un investissement humain et financier important. L'intégration de cette notion dans le chiffrage permet également d'anticiper des sources de données nouvelles dont l'émergence reste à venir. Exemples : données d'imagerie, données de monitoring périphériques internes ou externes implantés aux patients (RWD), données de biologie moléculaire, etc. A titre d'illustration, les membres du GT ont convenu de chiffrer les coûts associés à la mise en

place d'un flux complexe par an. A noter que l'ingestion de ce type de flux entraîne souvent un redimensionnement conséquent de l'infrastructure technique. Une ligne de coût dédiée a donc été incluse dans le chiffrage (ligne 56).

Les profils mobilisés sont :

- RSSI
- Médecin référent métier
- Expert fonctionnel
- Data engineer
- DevOps

Les principales activités couvertes sont :

- Cadrage métier flux complexe
- Alimentation flux supplémentaire complexe : Expression du besoin sur les nouveaux flux de données, conception et implémentation des nouveaux flux
- Gestion des coûts de licence spécifiques aux flux complexes
- Gestion des coûts matériels et humains liés à l'évolution de l'architecture avec l'ingestion des flux complexes

Run Exploitation

a) Pilotage

Durant la phase de Run, l'exploitation de l'EDS par le personnel médical, le corps de recherche, et les éventuels réutilisateurs externes doit être pilotée selon deux axes : pilotage global médical et pilotage global de maîtrise d'œuvre. En plus de fournir un appui précieux de coordination aux équipes de l'EDS, les profils mobilisés pour le pilotage veilleront à l'amélioration régulière de l'offre de service à destination des utilisateurs et à la conduite du changement au sein de l'établissement de santé.

Les ressources mobilisées sont :

- Médecin référent métier
- Chef de projet maîtrise d'œuvre SI

Les principales activités couvertes sont :

- Mise à jour des éléments relatifs à la procédure d'accès et maintenance des informations disponibles sur le site internet
- Organisation et animation de sessions de sensibilisation à la sécurité informatique auprès des utilisateurs
- Maintien à jour du registre public des projets et du site internet
- Modalités de communication à destination des porteurs de projet externes pour encourager une utilisation des données de l'EDS la plus large possible
- Modalités d'association de la communauté médicale et actions spécifiques à destination des professionnels de santé
- Actions spécifiques à destination des usagers du système de santé

b) Coûts de support

En phase de Run, des coûts "support" liés à la mobilisation de profils clés non dédiés à l'EDS en interne continuent à être supportés par l'établissement de santé. L'activité de ces

ressources permet de maintenir l'entrepôt en conditions opérationnelles et de gérer l'entrepôt dans son mode de fonctionnement nominal. Ces ressources peuvent également être mobilisées pour accompagner les porteurs de projets souhaitant réutiliser les données mises à disposition dans l'EDS.

Les profils mobilisés sont :

- DPO
- Chef de projet maîtrise d'ouvrage
- Chargé de valorisation
- Médecin référent métier
- Biostatisticien
- Chef de projet recherche clinique

Les médecins référents métier assument la fonction de chargé d'études dans le cadre des projets de réutilisation des données de l'EDS.

- Ils accompagnent la conception des études et l'élaboration des méthodologies associées : Support à l'expression du besoin scientifique des porteurs de projet, transcription en données cibles, etc.
- Ils fournissent un suivi médical et une expertise métier sur les données intégrées à l'EDS
- Ils accompagnent le montage de projets et analysent leur faisabilité
- Ils assurent une sensibilisation et une formation des professionnels de santé à l'utilisation des données de santé

Des profils issus de la direction de la recherche et de l'innovation sont également mobilisés afin d'accompagner les porteurs de projet souhaitant réutiliser les données de l'EDS. Leur montée en charge doit prendre en compte la prolifération des projets de recherche sur les données de l'EDS. Parmi les travaux à réaliser :

- L'aide au montage de projet
- La gestion administrative et financière des projets
- La valorisation des résultats
- L'organisation d'appels à projets pour stimuler l'écosystème de recherche et encourager l'utilisation des données de l'EDS
- La rédaction de contrats de partenariat et la stratégie de partenariats industriels

Les autres ressources mobilisées sur cette partie soutiennent les activités de run exploitation en menant à bien les travaux suivants :

- Appui juridique et réglementaire aux équipes de l'EDS (DPO) :
 - Aide au montage des dossiers de traitements de données au sein de l'EDS (e.g. vérification de la conformité des projets aux méthodologies de référence, etc.)
 - Mise à jour régulière du dossier réglementaire, soumission aux instances pertinentes et suivi si nécessaire
 - mise en oeuvre de la stratégie d'information des personnes
 - Mise en oeuvre de la stratégie permettant aux personnes d'exercer leurs droits, gestion des droits des personnes et transparence
 - Réponses aux incidents sur le traitement des données
 - Déclaration auprès des autorités compétentes le cas échéant
- Gestion du PIA par projet accompagné (DPO)
- Analyse juridique de l'impact de l'intégration d'un nouveau flux de données (DPO)
- Gestion du CSE (chef de projet maîtrise d'ouvrage)
- Support à la conception des études et des méthodologies associées (biostatisticien)

- Gestion des données de santé (chef de projet recherche clinique)

L'accompagnement juridique de l'EDS est maintenu afin d'apporter une réponse au nombre grandissant de requêtes juridiques récurrentes et potentiellement bloquantes. Une aide particulière sera également apportée sur les démarches juridiques entourant le déploiement des différents cas d'usage.

c) Mise à disposition des données

La mise à disposition des données aux porteurs de projets s'organise autour de 4 grandes phases permettant d'assurer la disponibilité technique des données et leur maintien en qualité en continu.

i) Préparation des données en continu

Afin d'être en capacité de fournir en continu des données de qualité à un porteur de projet, les équipes de l'EDS doivent mobiliser un ou plusieurs profils de data engineer couvrant les activités principales suivantes :

- Monitoring de l'entrepôt (débugage, optimisation de la mise à disposition des jeux de données, etc.)
- Réalisation d'audits de la qualité des données entreposées
- Qualification en continue de la documentation et mise à jour de la documentation

ii) Optimisation des données en continu

Afin de multiplier les opportunités présentées par l'EDS auprès de l'écosystème de recherche et innovation, celui-ci doit s'inscrire dans une démarche de création de valeur continue en mobilisant des profils de data scientists qui mèneront des travaux visant à :

- Enrichir les bases de données existantes et les harmoniser
- Créer des outils "sur étagère" d'exploitation des données collectées
- Rendre les bases interopérables au sein de l'EDS puis à l'extérieur

iii) Pré-traitement des données

Dans le cadre du déploiement de son offre de service à destination des porteurs de projets, les équipes de l'EDS pourront s'enrichir de profils data scientists afin d'assurer le pré-traitement des données pour leur mise à disposition des porteurs de projet. Leurs objectifs seront de :

- Formater et mettre en qualité les données mobilisées dans le cadre du projet
- Accompagner le porteur du projet dans son traitement des données afin de mener son projet à bien

iv) Gestion de l'environnement de traitement

Une mobilisation de compétences data science est à anticiper autour des activités de gestion de l'environnement de traitement afin de :

- Maintenir un environnement de développement sécurisé
- Mettre à jour et améliorer la bibliothèque d'outils de traitement
- Gérer la création de formations à l'utilisation de l'environnement de traitement

Annexes 3A et 3B - Contexte juridique - notes de Laurent Cytermann

Mission d'appui au Health Data Hub sur le cadre juridique de la tarification de l'accès aux données du système national des données de santé (SNDS)

Résumé des principales conclusions

- Le code de la santé publique (CSP) définit de manière complète le régime du traitement, de l'accès et de la réutilisation des données du SNDS. Ce régime spécial prévaut sur les dispositions générales du code des relations entre le public et l'administration (CRPA) en matière de droit d'accès et de réutilisation des données publiques.
- Le principe de gratuité de la réutilisation des données publiques énoncé par l'article L. 324-1 du CRPA n'est donc pas applicable. La gratuité ne s'impose que dans les hypothèses prévues par les articles L. 1461-2 et L. 1461-5 du CSP.
- Dès lors qu'une personne souhaitant accéder au SNDS obtient une autorisation de la CNIL ou se conforme à ses méthodologies de référence, les personnes responsables de la mise à disposition du SNDS sont tenues de lui donner cet accès.
- Le régime du SNDS s'applique à l'ensemble des bases de données entrant dans le périmètre défini par l'article L. 1461-1 du CSP (« SNDS élargi »), même celles qui ne sont pas inscrites au catalogue. Il s'impose aux personnes publiques ou privées qui contrôlent ces bases.
- Les redevances mises en place doivent se conformer à la jurisprudence du Conseil d'Etat sur les redevances pour service rendu, dont les grands principes sont présentés dans cette note. Le projet de règlement européen sur la gouvernance des données pourrait toutefois imposer un encadrement supplémentaire, notamment une tarification au coût marginal, s'il était adopté en l'état.
- Il apparaît souhaitable de définir un cadre tarifaire cohérent, qui serait mis en œuvre tant par le HDH que par les producteurs de bases de données.

Introduction

La saisine du HDH porte sur le cadre juridique de la tarification de la mise à disposition des données du SNDS. Cette question est étroitement liée à celle des règles d'accès et de réutilisation de ces données et les auditions effectuées au cours de cette mission ont montré que des questions majeures subsistaient à cet égard.

La présente note s'efforce donc d'apporter des réponses aux questions de principe relatives au droit d'accès, à la liberté de réutilisation et à la possibilité de percevoir des redevances pour la mise à disposition des données du SNDS. Elle examine notamment l'articulation entre les règles de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, du code des relations entre le public et l'administration (CRPA) sur le droit d'accès

aux documents administratifs et la liberté de réutilisation, et du code de la santé publique sur le SNDS. La réponse à ces questions préalables permet d'aborder plus en détail la construction du régime de redevances, qui devra faire l'objet d'analyses complémentaires.

1) Les données de santé hors SNDS

Compte tenu de l'élargissement du SNDS opéré par la loi du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé, le périmètre des données de santé hors de ce système est sans doute restreint. On peut mentionner par exemple les données destinées aux professionnels et organismes de santé pour des soins non remboursés par la sécurité sociale ou les données enregistrées par certains appareils connectés. Il est sans doute rare que de telles données soient détenues par des administrations au sens de l'article L. 300-2 du CRPA.

L'intérêt pratique de cette partie est donc sans doute limité, mais elle permet de faire ressortir le contraste entre le régime juridique qui serait applicable en l'absence des dispositions spécifiques du code de la santé publique (CSP) sur le SNDS et celui qui résulte de ces dispositions.

1.1. Des règles spécifiques de protection des données à caractère personnel, caractérisées par une condition d'intérêt public et un régime d'autorisation par la CNIL

Les données concernant la santé d'une personne physique (« données de santé ») sont au nombre des données dites « sensibles » énumérées par l'article 6 de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés et par l'article 9 du RGPD, dont le traitement est en principe prohibé. En dehors des traitements nécessaires à la prise en charge médicale, les dérogations ouvertes par l'article 9 du RGPD concernent principalement les traitements nécessaires à des fins de recherche scientifique (article 9.2.j), de santé publique (article 9.2.i) ou pour d'autres motifs d'intérêt public important (article 9.2.g). L'article 9 du règlement renvoie largement au droit national pour définir les garanties entourant ces traitements, et le cas échéant introduire des conditions supplémentaires, de sorte que le cadre juridique du traitement des données de santé en France résulte essentiellement de la subdivision de la loi du 6 janvier 1978 qui leur est consacrée (section 3 du chapitre III du titre II).

Mis à part certaines exceptions énumérées par l'article 65 de la loi (traitement mis en œuvre par les organismes d'assurance-maladie, conduite des politiques publiques de santé par l'Etat, etc) et par l'article 67 (traitements mis en œuvre par des organismes publics pour répondre à une urgence sanitaire), l'article 66 énonce comme condition générale que les traitements de données de santé « *ne peuvent être mis en œuvre qu'en considération de la finalité d'intérêt public qu'ils présentent* ». Ces traitements sont subordonnés à une autorisation de la CNIL (article 66-III), sauf lorsqu'ils sont conformes à des référentiels ou règlements-types édictés par celle-ci en concertation avec le HDH et des organismes publics et privés représentatifs des acteurs concernés (article 66-II).

Lorsque les traitements ont pour finalité la recherche ou les études dans le domaine de la santé, ou l'évaluation ou l'analyse des pratiques ou des activités de soins ou de prévention, l'autorisation de la CNIL est donnée après avis du comité de protection des personnes (CPP)

pour les recherches impliquant la personne humaine (RIPH) et du comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES) pour les autres recherches ou études (article 76). Tous les dossiers doivent être déposés auprès du secrétariat unique assuré par le HDH, qui assure leur orientation vers les instances compétentes. Dans ce domaine de la recherche et de l'évaluation, les référentiels mentionnés à l'article 66-II sont nommés « *méthodologies de référence* » (MR) et sont homologués et publiés par la CNIL. Un traitement conforme à une MR peut être mis en œuvre sur simple déclaration du responsable de traitement à la CNIL, attestant de cette conformité (article 73). La CNIL a à ce jour homologué 6 MR, concernant notamment les RIPH et les recherches impliquant l'examen des caractéristiques génétiques (MR-001), les recherches non RIPH et ne nécessitant pas le traitement du NIR (MR-004) ou encore l'accès des établissements de santé aux données du PMSI (MR-005).

Les traitements de données nécessités par des RIPH, à l'exception de celles qui ne comportent aucun risque ni contrainte et dans lesquelles tous les actes sont pratiqués et les produits utilisés de manière habituelle (article L. 1122-1-1 du CSP), ainsi que ceux nécessités par des recherches impliquant l'examen de caractéristiques génétiques (article 75 de la loi du 6 janvier 1978), sont subordonnés au consentement éclairé et exprès des personnes concernées. Les autres traitements de données de santé sont soumis à un droit d'opposition (article 74).

1.2. Des règles de droit commun applicables en matière de droit d'accès aux documents administratifs et de réutilisation des données à caractère personnel, mais dont la portée est restreinte par les exigences de protection de la vie privée

Lorsque les données de santé sont contenues dans des documents produits et reçus par des administrations au sens de l'article L. 300-2 du CRPA, elles entrent dans le champ d'application du droit d'accès aux documents administratifs. La portée de cette inclusion est toutefois limitée par l'article L. 311-6, en vertu duquel ne sont communicables qu'à l'intéressé les documents administratifs dont la communication « *porterait atteinte à la protection de la vie privée* » ou « *au secret médical* ». De manière générale, il ne suffit pas qu'un document comporte des DCP pour que la communication porte atteinte à la protection de la vie privée (cf. par exemple Conseil d'Etat, 4 novembre 2020, *M. Appert*, n° 427401, Tab., sur le bulletin de salaire d'un agent public), mais les données de santé appellent une protection renforcée, en particulier en vertu du secret médical.

La question la plus complexe est de savoir si et à quelles conditions des données de santé pseudonymisées, c'est-à-dire qui ne peuvent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires¹¹, peuvent être communiquées à des tiers sans porter atteinte à la protection de la vie privée au sens de l'article L. 311-6 du CRPA. A partir de 1998, la Commission d'accès aux documents administratifs (CADA) a donné des avis favorables à la communication d'extraits du PMSI, au motif que les enregistrements individuels constituant les fichiers de résumés de sortie du

¹¹ Cf. la définition donnée par l'article 4.5 du RGPD : « *le traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable;* ».

PMSI y figuraient sous une forme anonyme ; c'est ainsi notamment qu'a pu être élaboré le palmarès des hôpitaux publié chaque année par le magazine l'Express¹². Toutefois, dès 1999, des travaux statistiques réalisés au sein du ministère de la santé ont mis en évidence le risque de réidentification du séjour d'une personne, l'identification pouvant résulter du rapprochement d'informations connues par ailleurs sur la personne avec les données "administratives" contenues dans le fichier¹³. Dans un avis plus récent (n° 20144116, séance du 27 novembre 2014), la CADA a considéré que seuls les extraits du PMSI « *dont auront été supprimées toutes les données susceptibles de permettre l'identification des personnes concernées* », tout en estimant que les opérations d'anonymisation pouvaient être conduites par l'ATIH au moyen d'un traitement automatisé d'usage courant.

Il paraît en effet difficile d'admettre que la communication de documents comportant des données de santé et pour lesquelles subsisterait un risque de réidentification ne porte pas atteinte au droit à la protection de la vie privée. La communication de tels documents est donc subordonnée à une anonymisation complète, allant au-delà d'une simple pseudonymisation. La pratique décisionnelle de la CADA est de donner un avis favorable lorsque les opérations peuvent être réalisées au moyen de traitements automatisés d'usage courant, dont l'ampleur n'est pas telle qu'ils peuvent être considérés comme impliquant la constitution d'un nouveau document (cf. notamment conseil n° 20133264, séance du 10 octobre 2013). L'ampleur et la complexité des opérations d'occultation, et en conséquence la communicabilité du document, dépendent du fichier en cause (Conseil d'Etat, 27 mars 2020, *Association contre l'extension et les nuisances de l'aéroport de Lyon-Saint-Exupéry*, n° 426623, Tab.).

Le régime de la réutilisation des données de santé dépend étroitement de la communicabilité du document dans lequel elles sont contenues. En effet, si l'article L. 321-1 du CRPA énonce un principe de liberté de réutilisation des informations publiques, qui sont définies comme les informations figurant dans les documents administratifs, l'article L. 321-2 dispose que ne sont pas considérées comme telles les informations contenues dans des documents « *dont la communication ne constitue pas un droit pour toute personne en application du titre Ier ou d'autres dispositions législatives* ». Il en va notamment ainsi des documents dont la communication porterait atteinte à la protection de la vie privée ou au secret médical¹⁴.

De même, dans le cadre du droit de l'Union européenne, la directive « open data » du 20 juin 2019¹⁵ exclut de son champ d'application les documents « *dont l'accès est exclu ou limité en application de règles d'accès pour des motifs de protection des données à caractère personnel* » (article 1.2.h).

Au total, le régime de droit d'accès aux documents administratifs et de libre réutilisation des informations publiques qu'ils contiennent ne s'applique aux données de santé que dans des cas circonscrits, dans lesquels l'anonymisation peut être pratiquée sans impliquer

¹² F. Malye et J. Vincent, « Le palmarès des hôpitaux », Les Tribunes de la santé, 2015/2 n° 47.

¹³ Circulaire DH/PMSI/99 n° 133 du 2 mars 1999 relative aux conditions d'utilisation et de communication des fichiers des résumés du PMSI transmis par les départements d'information médicale des établissements de santé.

¹⁴ Cf. en ce sens A. Lallet et P. Nguyen Duy, « Diffusion et réutilisation des informations publiques : « Open data » – données ouvertes » », Répertoire de contentieux administratif, Dalloz, septembre 2020, §223.

¹⁵ Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public.

d'opérations trop complexes.

1.3. Des règles de tarification qui laissent une latitude importante aux administrations fournissant les données

Depuis une loi du 28 décembre 2015¹⁶, il existe un principe de gratuité de la réutilisation des informations publiques (article L. 324-1 du CRPA), assorti de quelques exceptions notamment pour les administrations « *tenues de couvrir par des recettes propres une part substantielle des coûts liés à l'accomplissement de leurs missions de service public* », qui peuvent instituer une redevance de réutilisation de ces informations. Toutefois, ce principe ne s'applique qu'aux informations publiques et pour les motifs exposés ci-dessus, les données de santé ne peuvent être ainsi qualifiées qu'à des conditions d'anonymisation restrictives.

Lorsque les données de santé ne sont pas des informations publiques, il est loisible aux administrations qui les contrôlent d'instaurer une redevance de réutilisation. Il convient cependant de souligner que la communication de ces données par l'administration en vue de leur réutilisation est un traitement de données à caractère personnel et que lorsque cette communication ne résulte pas du droit d'accès aux documents d'administratifs¹⁷, elle doit être conforme au RGPD et à la loi du 6 janvier 1978. Le réutilisateur est lui-même soumis au respect de ces obligations en tant que responsable du traitement constitué par la réutilisation.

Le droit des redevances pour services rendus, dont font parties les redevances perçues pour la réutilisation de données détenues par l'administration, résulte essentiellement de la jurisprudence du Conseil d'Etat. On peut notamment mentionner les principes suivants :

- Chaque personne publique est compétente pour instituer une redevance rémunérant les services qu'elle rend. Pour l'Etat, les redevances doivent être créées par décret en Conseil d'Etat pris sur le rapport du ministre chargé des finances et du ministre intéressé (article 4 de la loi organique du 1^{er} août 2001 relative aux lois de finances). Pour les établissements publics et les GIP, l'organe compétent est défini par l'acte énonçant l'organisation de cette personne publique, à savoir des dispositions législatives ou réglementaires pour les établissements publics et la convention constitutive pour les GIP¹⁸.

- Une redevance pour service rendu peut être légalement établie à la condition, d'une part, que les opérations qu'elle est appelée à financer ne relèvent pas de missions qui incombent par nature à l'Etat et, d'autre part, qu'elle trouve sa contrepartie directe dans une prestation rendue au bénéfice propre d'usagers déterminés (CE, 28 novembre 2018, *SNCF Réseau*, n° 413839, Rec.). Les missions qui incombent par nature à l'Etat sont notamment les activités de réglementation, de police et de contrôle qui impliquent la mise en œuvre de prérogatives de puissance publique¹⁹.

¹⁶ Loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public.

¹⁷ Le RGPD autorise en effet de manière générale la communication lorsqu'elle résulte du droit d'accès aux documents administratifs (article 86).

¹⁸ Cf. sur ce point Conseil d'Etat, *Redevances pour service rendu et redevances pour occupation du domaine public*, Documentation française, octobre 2002, p. 23.

¹⁹ Comme l'exposent les conclusions de Guillaume Odinet sur cette décision, faisant le lien avec la jurisprudence Blanckeman (Tribunal des conflits, 2 février 2004, n° 247369, Rec.).

- Les services consistant dans la fourniture de données détenues par une administration peuvent faire l'objet d'une rémunération (CE, Ass., 10 juillet 1996, *Société Direct Mail Promotion et autres*, n° 168702, Rec., affaire qui portait sur les redevances perçues par l'INSEE pour la réutilisation du fichier SIRENE des entreprises).
- Le montant de la redevance doit être proportionné aux services rendus (cf. par exemple CE, 10 avril 2002, *GISTI*, n° 227746, Tab.). Le juge administratif n'exerce qu'un contrôle restreint sur ce rapport de proportionnalité (CE, 2 novembre 1987, *Mansier*, n° 57051, Rec.).
- Les bases de calcul de la redevance ne peuvent intégrer des dépenses ne correspondant à des prestations directement rendues aux usagers (CE, Sect., 10 février 1995, *Chambre syndicale du transport aérien*, n° 148035, Rec.).
- La rémunération pour services rendus peut être assortie de produits associés à la cession de droits de propriété intellectuelle, notamment pour l'utilisation de bases de données protégées par un tel droit (arrêt *Société Direct Mail Promotion* précité).
- Traditionnellement, le Conseil d'Etat considérait que le prix ne pouvait excéder le coût du service rendu par l'administration (règle du plafonnement). Toutefois, depuis une décision *Syndicat national de défense de l'exercice libéral de la médecine à l'hôpital* (CE, Ass., 16 juillet 2007, n° 293229, Rec.), le Conseil d'Etat considère que « *le respect de la règle d'équivalence entre le tarif d'une redevance et la valeur de la prestation ou du service peut être assuré non seulement en retenant le prix de revient de ce dernier, mais aussi, en fonction des caractéristiques du service, en tenant compte de la valeur économique de la prestation pour son bénéficiaire* ».
- Des différenciations du tarif entre catégories d'usagers sont possibles dans le respect du principe d'égalité devant le service public. En vertu de ce principe, « *la fixation de tarifs différents applicables, pour un même service rendu, à diverses catégories d'usagers d'un service public implique, à moins qu'elle ne soit la conséquence nécessaire d'une loi, qu'il existe entre les usagers des différences de situation appréciables ou que cette mesure soit justifiée par une nécessité d'intérêt général en rapport avec les conditions d'exploitation du service* » (CE, Sect., 10 mai 1974, *Sieur Denoyez et Sieur Chorques*, n° 88032, Rec. ; 28 février 1996, *Etablissement public du musée du Louvre*, n° 163528, Tab.).

2) Les données de santé du SNDS

2.1. L'application des règles de protection des données à caractère personnel relatives aux données de santé

L'article L. 1461-3 du code de la santé publique définit deux régimes d'accès au SNDS :

- Un accès dit « permanent », réservé aux institutions sanitaires pour l'accomplissement de leurs missions de service public (cf. les articles R. 1461-11 à R. 1461-17 du CSP pour la liste et le périmètre de ces accès) ;
- Un accès sur autorisation pour les autres traitements, contribuant à l'une des finalités du SNDS (évaluation des politiques de santé, veille sanitaire, recherche, etc) et répondant à un motif d'intérêt public.

Dans le cadre de ce second régime, les traitements sont autorisés selon les règles définies

par la loi du 6 janvier 1978 pour les données de santé : ils sont donc subordonnés à une autorisation de la CNIL, sauf lorsqu'ils correspondent à l'une des méthodologies de référence homologuées par celle-ci. Par rapport au droit commun des données de santé, les spécificités du SNDS sont donc limitées du point de vue de la protection des données à caractère personnel : il en résulte seulement quelques exigences supplémentaires, à savoir l'inscription dans l'une des finalités du SNDS énumérées par l'article L. 1461-1-III et la définition de deux « finalités interdites », qui sont la promotion des produits de santé en direction des professionnels ou établissements de santé et la modification des conditions d'assurance d'un individu ou d'un groupe d'individus en fonction de leur risque (article L. 1461-1-V).

2.2. L'autorisation de traiter les données du SNDS emporte obligation de les rendre accessibles au bénéficiaire de celle-ci

La singularité du régime du SNDS se situe sur un autre plan. En règle générale, lorsque des données sont contrôlées par une administration, l'autorisation de les traiter délivrée par la CNIL n'emporte pas d'obligation pour l'administration de les ouvrir : cette question relève d'un autre corpus législatif, celui du titre Ier du livre III du CRPA relatif au droit d'accès aux documents administratifs. Il s'agit d'une illustration du principe d'indépendance des législations. Comme exposé en première partie, la communicabilité des données de santé doit souvent être écartée en raison des risques de réidentification.

Il en va différemment pour le SNDS : dans le cadre de cette législation spécifique, l'autorisation de traiter les données à caractère personnel emporte obligation pour les organismes contrôlant ces données de les rendre accessibles aux bénéficiaires de l'autorisation. Le chapitre Ier du titre VI du livre IV de la première partie du CSP définit de manière complète le régime du traitement, de l'accès et de la réutilisation des données du SNDS, sans qu'il y ait lieu de se référer pour ces deux derniers aspects aux règles de droit commun du CRPA.

Ceci résulte d'abord des raisons même de la création du SNDS par le législateur. Selon l'exposé des motifs du projet de loi de modernisation de notre système de santé, l'objet de ces dispositions est de « *créer les conditions d'un accès ouvert aux données de santé* » ; le SNDS « *centralisera les données des bases existantes en matière sanitaire et médico-sociale et assurera leur mise à disposition* ». Le SNDS a été créé pour rendre les données de santé accessibles : comme le résume l'étude d'impact du projet de loi, le but est d'organiser « *l'accès pour des recherches, études et évaluation d'intérêt public le plus large compatible avec le secret des données personnelles* ». De même, lors de la discussion de cet article en première lecture à l'Assemblée nationale, la ministre de la santé a indiqué que les données « *qui présentent un risque de ré-identification doivent être mises à disposition de certaines personnes, associations, entreprises ou journalistes, mais dans un cadre garantissant la protection de la vie privée* ». Dès lors que la procédure d'autorisation par la CNIL après avis du CESREES ou du CPP permet de s'assurer de la protection des données à caractère personnel, de la rigueur méthodologique du projet et de sa contribution à l'intérêt public, il n'existe aucune raison valable de ne pas ouvrir le SNDS à une personne ayant franchi cette procédure avec succès.

Ceci résulte en second lieu des termes de l'article L. 1461-3, qui définit les conditions de « *l'accès* » aux données du SNDS. Si ces dispositions renvoient aux conditions de procédure et

de fond de la loi du 6 janvier 1978, l'objet de l'autorisation délivrée par la CNIL est propre au régime du SNDS : il ne s'agit pas seulement d'autoriser le traitement des données mais aussi l'accès à celles-ci. La loi du 6 janvier 1978 n'emploie jamais en ce sens le terme d'accès, celui-ci ne concernant que le droit de la personne physique d'accéder à ses propres données (article 49).

Il s'agit ainsi d'un régime d'accès dérogatoire à celui du CRPA : l'accès est donné non au vu de la communicabilité intrinsèque du document dans lequel sont contenues les données, mais du projet du demandeur qui a bénéficié de l'autorisation de la CNIL ou qui atteste de sa conformité à une méthodologie de référence. La création d'un régime spécial d'accès à des documents administratifs n'est pas sans précédent : le professeur Jean-Bernard Auby en dénombre plus d'une dizaine et indique qu'il existe un « *principe de prévalence des régimes spéciaux d'accès* », selon lequel « *les dispositions de la loi spéciale ayant le même objet que celles de la loi générale l'emportent sur celles-ci, du moins si elles sont inconciliables* »²⁰.

2.3. Des exigences de gratuité circonscrites

Le législateur n'a pas prévu la gratuité de la mise à disposition des données du SNDS que dans deux hypothèses : pour les données mises en ligne en raison de l'absence de tout risque de réidentification (article L. 1461-2) et pour « *les traitements de données concernant la santé demandés par l'autorité publique* » et « *les recherches réalisées exclusivement pour les besoins de services publics administratifs* » (article L. 1461-5). En dehors de ces hypothèses, les personnes publiques responsables de la mise à disposition des données peuvent instaurer une redevance dans les conditions définies par la jurisprudence administrative (cf. 1.3 de cette note).

Il convient d'ajouter que si la circulaire du Premier ministre du 24 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources énonce un principe « d'ouverture libre, gratuite et par défaut des données », il ne s'agit que d'un rappel des dispositions législatives en ce sens du CRPA, dont la circulaire ne peut élargir le champ d'application.

2.4. Le régime d'accès au SNDS s'applique à l'ensemble des données entrant dans son champ, indépendamment de leur inscription au catalogue

Depuis l'élargissement du SNDS opéré par la loi du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé, deux acceptions du périmètre de celui-ci coexistent dans les dispositions législatives aux réglementaires :

- Au niveau législatif, l'article L. 1461-1 définit le SNDS en prévoyant qu'il « *rassemble et met à disposition* » les 11 catégories de données qu'il énumère, y compris l'ensemble des données recueillies à l'occasion des activités de soins remboursées par la sécurité sociale (6°) et les données à caractère personnel des enquêtes dans le domaine de la santé (8°), lorsqu'elles sont appariées à certaines autres catégories de données ;
- Au niveau réglementaire, l'article R. 1461-2 prévoit que le SNDS comprend une base principale couvrant l'ensemble de la population et un ensemble de bases de données ne couvrant pas l'ensemble de la population dénommé " catalogue ". La définition de la base

²⁰ J.-B. Auby, « Données publiques. Droit d'accès – Etendue », Jurisclasseur administratif, juin 2021.

principale comprend une partie stable, qui correspond au périmètre du SNDS dit « historique » issu de la loi du 26 janvier 2016 (SNIIRAM, PMSI, causes de décès et SI des MDPH), et une partie évolutive, puisque selon l'article R 1461-2, elle est « *complétée progressivement de données mentionnées aux 5° à 11°* » du I de l'article L. 1461-1. La définition du catalogue est entièrement évolutive, un arrêté du ministre chargé de la santé pris après avis de la CNIL désignant les bases de données inscrites dans le catalogue.

Dans son avis sur le projet de décret d'application de la loi du 24 juillet 2019²¹, la CNIL qualifie le premier périmètre de « SNDS élargi » et le second de « SNDS centralisé ».

La question qui se pose est de savoir si le régime d'accès au SNDS, exposé ci-dessus (cf. 2.1 et 2.2), s'applique aux bases de données qui sont dans le périmètre du SNDS élargi mais qui ne sont pas encore ajoutées à la base principale ou inscrites dans le catalogue.

Trois arguments principaux conduisent à une réponse positive :

- La lettre de la loi : au niveau législatif, seul existe le périmètre du SNDS élargi. Le régime d'accès au SNDS, qui résulte principalement de l'article L. 1461-3, s'applique donc à l'ensemble de ce périmètre législatif. La CNIL retient cette interprétation dans son avis sur le projet de décret d'application : « *L'ensemble de ces données entre dans le périmètre du «SNDS élargi». Lorsqu'elles sont utilisées pour l'une des finalités visées à l'article L. 1461-1 III du CSP, elles sont soumises au respect de l'ensemble des dispositions de ce code, tout particulièrement au respect du référentiel de sécurité, à l'interdiction de poursuite de finalités interdites, aux conditions d'accès aux données via un accès permanent ou à la réalisation d'une formalité ainsi qu'aux modalités de transparence. Ce rattachement juridique au «SNDS élargi» n'emporte pas pour autant à lui seul de migration des données au sein d'une base centralisée.* »

Certes, l'article L. 1461-7 habilite pouvoir réglementaire à dresser, par décret en Conseil d'Etat, « *la liste des catégories de données réunies au sein du système national des données de santé* ». Il s'agit cependant de définir, au sein des ensembles de données énumérés par l'article L. 1461-1, le type de variables pouvant être traités dans le cadre du SNDS (cf. l'article R. 1461-4 : informations relatives à la personne, à sa santé, aux professionnels de santé, etc), et non de permettre au pouvoir réglementaire de moduler de manière discrétionnaire le champ d'application de la loi. L'inscription dans le « SNDS centralisé » emporte des conséquences techniques, notamment en termes de responsabilité du HDH sur le traitement de données et de modalités de mise à disposition des données, mais elle ne modifie pas le régime juridique des données du SNDS qui résulte de la loi.

- L'intention du législateur : selon l'exposé des motifs du projet de loi relatif à l'organisation et à la transformation du système de santé, « *le système national des données de santé sera enrichi de l'ensemble des données collectées lors des actes pris en charge par l'assurance maladie* ». De même, dans son avis public sur le projet de loi, le Conseil d'Etat avait estimé que celui-ci « *modifie d'abord le périmètre des données incluses dans le SNDS, en y ajoutant des données cliniques qui seront recueillies par les praticiens, en secteur hospitalier comme en médecine de ville, lors de la réalisation d'actes pris en charge par*

²¹ Délibération no 2020-106 du 29 octobre 2020 portant avis sur un projet de décret relatif au système national des données de santé (demande d'avis no 20011090).

l'assurance maladie au titre de la maladie, la maternité, les accidents du travail et les maladies professionnelles ». Ceci correspond aux propositions du rapport de préfiguration du Health Data Hub²², qui invitait le législateur à constituer un « *patrimoine national des données de santé* ».

- Les possibilités d'accès auprès du producteur reconnues par les dispositions réglementaires : si l'article R. 1461-2-IV permet aux organismes responsables des données alimentant le système national des données de santé peuvent les mettre à la disposition d'autres responsables de traitement, sans passer par le HDH, c'est en imposant que cette mise à disposition se fasse « *dans les conditions prévues aux articles L. 1461-1 à L. 1461-6* ». Ceci confirme que le régime législatif d'accès au SNDS s'applique à l'ensemble des bases entrant dans le périmètre élargi, indépendamment de leurs modalités techniques d'accès.

Les conséquences de cette application du régime d'accès du SNDS à l'ensemble de son périmètre doivent être bien appréhendées. Un certain nombre d'institutions responsables de grandes bases de données, telles que l'AP-HP (par exemple pour l'entrepôt de données de santé ou la banque nationale de données maladies rares (BNDMR)), INSERM-Transfert ou la CNAM et ses partenaires pour la cohorte Constances, ont défini leur propre gouvernance (comités éthiques et / ou scientifiques) pour décider des projets de recherche autorisés à accéder à ces données. Or, la gouvernance nationale prime sur cette gouvernance locale : le gestionnaire d'une base de données relevant du SNDS ne peut interdire son accès à un porteur de projet bénéficiant d'une autorisation de la CNIL après avis du CESREES ou du CPP ou s'étant déclaré conforme à une méthodologie de référence de la CNIL, même si son instance locale avait donné un avis défavorable. Toutefois, un rôle délégué peut être reconnu à cette gouvernance locale : ainsi, le projet de méthodologie de référence de la CNIL relatif aux entrepôts de données de santé²³ prévoit la mise en place d'une gouvernance spécifique à chaque entrepôt par le responsable de traitement, comportant un comité scientifique et éthique qui doit donner un avis préalable et motivé sur chaque projet de réutilisation. Un projet ayant reçu un avis défavorable de ce comité local pourrait toujours néanmoins prétendre à l'autorisation de la CNIL.

2.5. Les dispositions du code de la recherche n'interfèrent pas avec le régime d'accès au SNDS

Issu de la loi pour une République numérique du 7 octobre 2016, l'article L. 533-4-II du code de la recherche définit un régime de liberté de réutilisation pour les données issues d'une activité de recherche financée au moins pour moitié par des fonds publics. Ces dispositions assurent la transposition en droit français de la directive « open data » du 20 juin 2019 en ce qui concerne les données de la recherche. Toutefois, il s'agit de dispositions supplétives, puisqu'elles s'appliquent aux données qui « *ne sont pas protégées par un droit spécifique ou une réglementation particulière* » : le SNDS étant protégé à l'évidence par un droit spécifique, l'article L. 533-4-II du code de la recherche ne lui est pas applicable.

²² M. Cuggia, D. Polton et G. Wainrib, rapporté par S. Combes, *Health Data Hub. Mission de préfiguration*, octobre 2018.

²³ Mis en consultation publique en mars 2021 : <https://www.cnil.fr/fr/entrepots-de-donnees-de-sante-la-cnil-lance-une-consultation-publique-sur-un-projet-de-referentie>

2.6. Si quelques incertitudes existent sur les droits de propriété intellectuelle afférents à certaines composantes du SNDS, elles ne paraissent pas faire obstacle à l'accès à celui-ci

Le code de la propriété intellectuelle reconnaît l'existence d'un droit de propriété dit « *sui generis* » du producteur de bases de données (titre IV du livre III de ce code). Dans le cadre du CRPA, l'existence de droits de propriété intellectuelle d'un tiers fait obstacle à l'accès aux documents administratifs et à la réutilisation des données publiques (cf. les articles L. 311-4 et L. 321-2).

Si les dispositions du code de la santé publique ne prévoient quant à elle aucune prise en compte de l'existence de tels droits, le titre IV du livre III du CPI assure la transposition en droit français de la directive du 11 mars 1996 concernant la protection juridique des bases de données²⁴. La directive s'imposant au législateur comme l'ensemble du droit dérivé de l'Union européenne, le code de la santé publique ne peut méconnaître la directive. La protection d'une base de données est subordonnée à la démonstration d'un « *investissement financier, matériel ou humain substantiel* » dans « *la constitution, la vérification ou la présentation* » du contenu de la base (article L. 341-1 du CPI). Le producteur d'une base de données protégée a le droit d'interdire l'extraction ou la réutilisation de la totalité ou d'une partie qualitativement ou quantitativement substantielle de celle-ci (article L. 342-1).

Néanmoins, deux éléments limitent la portée de ces dispositions :

- Selon la jurisprudence de la CJUE, les moyens mis en œuvre pour la création des éléments constitutifs du contenu d'une base de données ne sont en revanche pas pris en compte (CJUE, Grande chambre, 9 novembre 2004, *The British Horseracing Board Ltd e.a.*, C-203/02), ce qui exclut de la protection les bases de données qui ne sont que le sous-produit d'une activité principale conduisant à la création de données. Ainsi, la Cour de cassation a écarté l'existence d'une protection pour une base de données d'annonces immobilières sur le site internet du journal Ouest France (Cass. Civ. 1re, 5 mars 2009, *Ouest France*, n° 07-19.734 et 07-19.735). Selon une mission du Conseil supérieur de la propriété littéraire et artistique (CSPLA), « *les bases de données issues d'une source unique, qui est aussi à l'origine des données, seront rarement protégées par le droit sui generis, en tant qu'il sera difficile pour cette source d'isoler l'investissement spécifiquement dédié à l'assemblage des données de l'investissement consacré à leur production de sorte que ceux-ci ne paraissent pas fort obstacle à l'accès aux composantes du SNDS qui bénéficieraient de cette protection* »²⁵. La question de savoir si des bases de données issues d'enquêtes ou d'entrepôts liés à l'activité de soins, tels que celui de l'AP-HP, pourraient bénéficier de la protection est délicate.

- A supposer que la protection existe, le CPI prévoit plusieurs exceptions en faveur de la recherche scientifique (articles L. 342-3-4° et 5°).

En tout état de cause, dès lors qu'une convention serait conclue avec les producteurs pour l'inscription de leur base au catalogue du SNDS, cette convention emporterait accord de celui-ci pour l'extraction et la réutilisation de sa base. Ceci souligne toutefois l'importance

²⁴ Directive 96/9/CE du Parlement européen et du Conseil du 11 mars 1996 concernant la protection juridique des bases de données.

²⁵ CSPLA, *Mission sur la réforme européenne du droit des bases de données. Note d'étape*, avril 2021.

d'un certain consensus sur les modalités financières de la mise à disposition (3^e partie de cette note) et pose la question de l'opportunité d'inclure dans la convention des clauses explicites relatives aux droits de propriété intellectuelle (par exemple l'octroi d'une licence allant au-delà des exceptions légales, dont la possibilité est prévue par l'article L. 342-1).

Le cas échéant, un entretien avec la mission en cours du CSPLA sur la réforme du droit des bases de données pourrait être organisé pour approfondir ce point.

3. Les questions qui se posent dans la construction d'un régime de redevances d'accès et de réutilisation du SNDS

Les deux parties précédentes de la note attestent de la possibilité, pour le HDH comme pour les producteurs, de percevoir des redevances en contrepartie de la mise à disposition de données du SNDS. Toutefois, plusieurs questions se posent dans la construction de ce régime, qui appellent pour certaines des analyses juridiques ou économiques complémentaires, et pour d'autres des choix du HDH ou de ses tutelles. La présente note se bornera à en énoncer le cadre.

Pour conduire ces investigations, un exercice de benchmark avec les administrations autorisées par le CRPA à percevoir de telles redevances (IGN, Météo France et le SHON) pourrait être d'une grande utilité. Il pourrait être accompagné d'un benchmark international sur les pratiques de tarification de gestionnaires de grandes bases de données de santé (NHS au Royaume-Uni, Clalit en Israël ?). Ce benchmark international pourrait s'appuyer sur les travaux conduits par la Commission dans le cadre de la mise en place de l'espace européen des données de santé (cf. infra).

3.1. La prise en compte du cadre en construction du DGA et de l'espace européen des données de santé

En l'état actuel, le droit de l'Union européenne n'impose aucune obligation concernant la mise à disposition de données telles que celles du SNDS, la directive « open data » du 20 juin 2019 n'incluant pas les données à caractère personnel dans son champ d'application. En revanche, le projet de règlement sur la gouvernance des données (Data Governance Act) comporte des dispositions sur la réutilisation de certaines catégories de données protégées détenues par des organismes du secteur public, notamment celles qui relèvent de la protection des données à caractère personnel. La proposition de règlement de la Commission européenne²⁶, si elle était adoptée par le Parlement européen et le Conseil, emporterait plusieurs conséquences :

- Selon l'article 5.7, les organismes du secteur public ne pourraient plus exercer le droit *sui generis* sur leurs bases de données relevant du règlement, ce qui lèverait les interrogations exposées ci dessus (cf. 2.6 de cette note). Les seules conditions applicables à la réutilisation devraient être non discriminatoires, proportionnées et objectivement justifiées en ce qui concerne les catégories de données et les finalités de réutilisation (article 5.2), et rendues publiques (article 5.1), conditions qui paraissent déjà remplies aujourd'hui par le cadre législatif et réglementaire du SNDS.
- L'article 6, relatif aux redevances, prévoit qu'elles peuvent être prévues par les organismes

²⁶ 2020/0340 (COD).

du secteur public pour autoriser la réutilisation des données, qu'elles doivent être non discriminatoires, proportionnées et objectivement justifiées, qu'elles doivent être « *calculées sur la base des coûts liés au traitement des demandes de réutilisation* », c'est-à-dire au coût marginal, et que la méthode de calcul doit être publiée à l'avance.

La proposition de règlement prévoyant une application 12 mois après la publication, et l'adoption du règlement ne pouvant sans doute être envisagée avant l'année 2022, l'entrée en vigueur de ces nouvelles ne pourrait intervenir avant l'année 2023. Il importe de déterminer si la construction du régime de redevances du SNDS doit se faire en anticiper l'application de ce nouveau cadre, ce qui paraît opportun mais ne sera pas aisé, les dispositions sur le calcul des redevances étant sans doute appelées à évoluer au cours de la discussion.

En outre, la Commission européenne a engagé la constitution d'un Espace européen des données de santé (European Health Data Space), afin de promouvoir la circulation de ces données à l'échelle européenne. Une évaluation des règles des Etats membres en matière de données de santé a été rendue publique par la Commission²⁷. Cette étude explore différentes pistes d'action à l'échelle européenne, dont une intervention législative sectorielle dans le cadre de l'application du règlement sur la gouvernance des données.

Afin d'être davantage éclairés sur le calendrier d'adoption de ces mesures et sur les principaux points de discussion, des entretiens pourraient être organisés avec le SGAE et avec la Commission européenne.

3.2. La transparence des règles applicables aux redevances

Dans l'attente de l'adoption du DGA, aucune norme européenne n'impose la publication des règles de calcul des redevances. Toutefois, les actes administratifs instituant des redevances sont des actes réglementaires, or un acte réglementaire n'est opposable que s'il a été publié (CE, Ass., 13 décembre 1957, *Barrot et autres*, Rec. p. 675). Il semble donc que la publication des règles relatives aux redevances s'impose déjà dans l'état du droit pour assurer leur effectivité.

3.3. La prise en compte des droits de propriété intellectuelle

La jurisprudence du Conseil d'Etat autorise la rémunération des droits de propriété intellectuelle des administrations dans le cadre des redevances de réutilisation des données (arrêt *Direct Mail Promotion* précité). Cette prise en compte se heurte toutefois à deux difficultés : d'une part, l'identification des bases de données bénéficiant d'un tel droit (cf. §2.6) ; d'autre part, la possible interdiction d'exercice de ces droits qui découlerait du DGA (§ 3.1).

3.4. La différenciation des tarifs : chercheurs, industriels, UE / hors UE, etc

Selon la jurisprudence précitée du Conseil d'Etat, « *la fixation de tarifs différents applicables, pour un même service rendu, à diverses catégories d'utilisateurs d'un service public implique, à moins qu'elle ne soit la conséquence nécessaire d'une loi, qu'il existe entre les usagers des*

²⁷ Assessment of the EU Member States' rules on health data in the light of GDPR, 2021.

différences de situation appréciables ou que cette mesure soit justifiée par une nécessité d'intérêt général en rapport avec les conditions d'exploitation du service ».

Plusieurs différenciations peuvent être envisagées :

- En faveur des chercheurs : un motif d'intérêt général lié à la promotion de la recherche pourrait justifier des tarifs plus favorables. On peut même se demander si l'article L. 1461-5 du CSP n'impose pas une exonération pour les chercheurs appartenant à des établissements publics administratifs : en effet, il prévoit un accès gratuit pour « *les recherches réalisées exclusivement pour les besoins de services publics administratifs* ». Or, tant le CNRS que l'INSERM sont des établissements publics à caractère scientifique et technologique (EPCST), qui sont des établissements publics administratifs (article L. 311-1 du code de la recherche). Il en va de même des universités et des établissements publics de santé.

- En faveur des PME : qu'il s'agisse d'un accès direct ou par l'intermédiaire d'un laboratoire de recherche ou d'un bureau d'études (article L. 1461-3-II du CSP), il pourrait être envisagé de traiter plus favorablement les PME. Une telle différenciation serait cependant plus solide si le DGA était adopté (l'article 6.4 de la proposition de la Commission prévoyant que lorsqu'ils appliquent des redevances, les Etats prennent des mesures pour encourager la réutilisation par des PME). En l'état du droit, il n'est pas évident que des différences de situation appréciables ou une nécessité d'intérêt général en rapport avec les conditions d'exploitation du service puisse être invoquée en faveur des PME.

- Selon la nationalité ou le pays de résidence du demandeur : le fait que des tarifs plus élevés soient appliqués aux utilisateurs étrangers du SNDS peut être justifié par plusieurs considérations. Le SNDS constitue un patrimoine national des données de santé, financé en grande partie par l'impôt dont s'acquittent les personnes résidant en France. Si la contribution à la connaissance revêt par nature une dimension internationale, nombre des finalités énumérées par l'article L. 1461-1 du CSP (information sur l'offre de soins, évaluation des politiques de santé, veille sanitaire, etc) n'ont de sens que dans un périmètre national. La jurisprudence du Conseil d'Etat a admis que des tarifs plus élevés soient appliqués aux étrangers pour l'inscription dans l'enseignement supérieur (CE, 1^{er} juillet 2020, *Association UNEDESEP et autres*, n° 430121, Rec.) et la visite des musées et monuments nationaux (CE, Sect., 18 janvier 2013, *Association SOS Racisme*, n° 328230, Rec.). La jurisprudence de la CEDH l'admet également dans le domaine universitaire (CEDH, 21 juin 2011, *Ponomaryovi c/ Bulgarie*, n° 5335/05). Le droit de l'Union imposant une égalité de traitement entre citoyens de l'Union, les tarifs plus élevés ne pourraient cependant être appliqués aux ressortissants des pays de l'Union européenne et de l'Espace économique européen.

- Selon la qualité de membre du GIP : la possibilité de pratiquer des tarifs plus faibles pour les adhérents du HDH devra être examinée.

- Selon la valeur économique de la prestation pour son bénéficiaire : contrairement à ce qui est indiqué par le cabinet DS Avocats dans son analyse, la portée de la jurisprudence *Syndicat national de défense de l'exercice libéral de la médecine à l'hôpital* est générale et ne se limite pas au cas de l'exercice privé de la médecine à l'hôpital public. Les cas d'application de cette jurisprudence sont cependant peu nombreux, de sorte que des questions pratiques peuvent se présenter. Une telle modulation impliquerait de pouvoir identifier le chiffre d'affaires généré par l'utilisation du SNDS. En outre, si le DGA était

adopté en l'état, cette modulation serait exclue.

3.5. Le respect du droit de la concurrence

Le juge administratif censure des tarifs qui, par leur caractère excessif, font obstacle à l'activité concurrentielle d'autres opérateurs économiques lorsque les données qui en sont la contrepartie constituent pour ces derniers une ressource essentielle à l'élaboration d'un produit ou d'une prestation qui diffèrent de ceux fournis par la personne publique ; il regarde de tels tarifs comme un abus de position dominante (CE, 29 juillet 2002, *Société CEGEDIM*, n° 200886, Rec. ; 24 juillet 2006, *Société CEGEDIM*, n° 247769, Tab.). Dans ces affaires, les tarifs étaient ceux appliqués par l'INSEE pour l'octroi d'une licence de rediffusion du fichier SIRENE des entreprises, dont le niveau ne devait pas empêcher ces rediffuser de concurrencer l'INSEE sur le marché aval de la fourniture de ce fichier aux consommateurs finals.

L'applicabilité du droit de la concurrence dépend du caractère économique de l'activité au sens du droit de l'Union européenne. En tant que telle, la mise à disposition du SNDS ne paraît pas relever d'une telle qualification : d'une part, il s'agit de l'ouverture de données constituées pour leur plus grande part dans le cadre de services publics administratifs, le fait pour une personne publique de donner accès à ses données pouvant être considéré comme le prolongement de sa mission principale ; d'autre part, les finalités assignées au SNDS par le législateur relèvent pour l'essentiel de la conduite des politiques publiques de santé et ne présentent pas un caractère économique.

Cependant, si le HDH venait à développer des services allant au-delà de la mise à disposition, par exemple des services d'analyse des données, pour lesquels il serait en concurrence avec des réutilisateurs du SNDS, alors ces services constitueraient une activité économique. Il conviendrait alors de veiller à ce que les tarifs pratiqués en amont sur l'accès au SNDS ne constituent pas un abus d'éviction. Compte tenu de son périmètre très large, le SNDS constitue sans doute une ressource essentielle ; néanmoins, il pourrait être tenu compte du fait qu'il est possible d'accéder à ses composantes directement auprès de leurs producteurs sans passer par le HDH. Il conviendra également d'étudier si la circonstance que ces services seraient fournis par des prestataires choisis par le HDH dans le cadre d'une procédure d'appel à la concurrence contribue au respect du droit de la concurrence.

Il conviendra aussi de veiller à ce que les différenciations tarifaires pratiquées le cas échéant en faveur de certaines catégories de réutilisateurs ne soient pas qualifiées d'aides d'Etat.

3.5. Le partage des redevances entre HDH et producteurs

La redevance est instituée par la personne publique qui met à disposition les données. Les données du SNDS pouvant être rendues accessibles par les producteurs ou le HDH, chacun peut instituer ses propres redevances. Il en résulte deux difficultés : d'une part, une incohérence, des tarifs différents pouvant être appliqués aux mêmes données selon le guichet d'accès ; d'autre part, sans compensation adéquate, l'inscription d'une base de données au catalogue se traduirait potentiellement par une perte de ressources pour le producteur, à la mesure de la part d'accès s'effectuant auprès du HDH.

Il conviendrait de prévoir des modalités de rétrocession de tout ou partie de la redevance au producteur. Cette rétrocession pourrait prendre deux formes : la perception de la redevance par le HDH pour le compte du producteur ; une rétrocession conventionnelle, la redevance étant perçue par le HDH pour son compte, mais la convention entre le HDH et le producteur prévoyant que le HDH en reverse tout ou partie au producteur.

3.6. La gouvernance : orientations du ministre, COSDS, intervention d'un décret

La fixation de la redevance relève de la compétence du conseil d'administration du HDH (cf. l'article 9.4 de la convention constitutive). Toutefois, les problèmes de cohérence exposés ci-dessus ne peuvent être réglés dans ce seul cadre. Différentes voies sont envisageables pour définir une politique cohérente de tarification du SNDS :

- Des orientations générales fixées par le ministre de la santé, conformément à l'article R. 1461-10 du CSP. Ces orientations seraient délibérées en comité stratégique des données de santé, où sont représentés un certain nombre de producteurs. Elles n'auraient toutefois pas de portée contraignante à leur égard.
- Un décret pourrait en revanche fixer des règles contraignantes. Il conviendrait d'approfondir l'analyse sur son fondement juridique (dispositions législatives relatives au SNDS ; pouvoir réglementaire autonome) et son contenu (cadre tarifaire unique ou seulement des règles communes sur les critères pouvant être pris en compte par chaque organisme public dans la fixation de la redevance).

Mission d'appui au Health Data Hub sur le cadre juridique de la tarification de l'accès aux données du système national des données de santé (SNDS) – phase 2

Principales conclusions

- Le cadre commun de tarification devrait être défini en conformité avec le règlement sur la gouvernance européenne des données (Data Governance Act), dont le texte est stabilisé et qui est susceptible d'entrer en vigueur au milieu de l'année 2023.
 - L'article 6 du règlement limite les coûts pouvant être pris en compte par les redevances aux coûts de traitement des demandes de réutilisation. S'agissant du SNDS, pourraient être couverts en pratique les coûts de l'interface technique de mise à disposition, les coûts liés à l'aide aux utilisateurs concernant les procédures d'accès au SNDS et les explications données sur les bases de données, les coûts d'appariement des bases de données et les coûts supportés par les producteurs dans le cadre de l'inscription de la base au catalogue.
 - Il conviendra de publier non seulement la grille tarifaire mais aussi la méthodologie du calcul, la description des principales catégories de coûts et les règles utilisées pour la répartition des coûts.
 - La prise en compte des droits de propriété intellectuelle *sui generis* sur les bases de données est exclue par l'article 5 du règlement lorsque le producteur est un organisme du secteur public.
 - Des services complémentaires comme la réalisation d'analyses de données à la demande du réutilisateur pourront être facturés sans être soumis aux limites du règlement DGA. Il conviendra toutefois de veiller à ce que les tarifs pratiqués ne fassent pas obstacle à l'activité concurrentielle d'autres acteurs (bureaux d'études, organismes de recherche, etc), susceptibles de conduire également de telles analyses.
 - Outre les cas de gratuité imposés par le code de la santé publique, le règlement DGA impose de pratiquer des tarifs réduits ou nuls pour certains acteurs : PME, organismes de la société civile.
 - Compte tenu des limites imposées par le règlement DGA, qui interdit de couvrir les coûts de production, il apparaît logique de définir une grille tarifaire uniforme qui s'appliquerait à l'ensemble des bases de données du SNDS. Un décret pourrait prévoir que la grille définie par le HDH s'impose à l'ensemble des producteurs de bases de données appartenant au secteur public. Une loi apparaît nécessaire pour que cette grille s'impose également aux producteurs appartenant au secteur privé.
- Une autre possibilité est d'inscrire cette clause d'application de la grille tarifaire dans les conventions accompagnant l'inscription au catalogue, le cas échéant après que la grille ait été approuvée par le comité stratégique des données de santé (CSDS).

- Le HDH pourra rétrocéder aux producteurs une part de la redevance correspondant à leurs propres coûts (coûts associés à l'inscription au catalogue et coûts d'aide aux utilisateurs lorsqu'elle leur incombe).

Introduction

La phase 1 de la mission a permis de clarifier les principes et règles de droit applicables à l'accès et la réutilisation des données du SNDS et la tarification de leur fourniture. Elle conclut notamment à la non-application d'un principe de gratuité en dehors des cas définis par les articles L. 1461-2 et L. 1461-5 du code de la santé publique et à l'application du régime juridique des redevances pour service rendu.

La phase 2 consiste à répondre aux questions opérationnelles qui se posent dans l'élaboration du régime de tarification, en vue de permettre au HDH de travailler ensuite à sa finalisation. La note s'appuie notamment sur une étude comparative auprès d'un certain nombre d'acteurs appliquant des redevances à la fourniture de données de santé ou d'autres données (cf. liste des entretiens en annexe).

L'analyse est centrée sur les bases de données du SNDS détenues par des organismes du secteur public au sens du règlement DGA et qui entrent donc dans son champ d'application. Au vu du projet d'arrêté relatif au catalogue du SNDS, cela paraît être le cas de la plus grande des bases dont l'inscription au catalogue est envisagée. Une partie est dédiée aux questions spécifiques posées par les bases détenues par des organismes privés.

1) Le règlement sur la gouvernance européenne des données, un texte dont l'application devrait être anticipée

La proposition de règlement sur la gouvernance européenne des données a fait l'objet d'un accord en trilogue (entre la Commission européenne, le Parlement européen et le Conseil) le 30 novembre 2021. Ce compromis a été confirmé par la commission compétente du Parlement européen (commission ITRE²⁸) le 8 décembre 2021. On peut donc s'attendre à une adoption définitive du règlement au cours du premier trimestre 2022 et à une entrée en vigueur avant juin 2023, l'article 35 prévoyant son application 15 mois après son entrée en vigueur, qui a lieu elle-même 20 jours après la publication du règlement au Journal officiel de l'Union européenne. Comme tout règlement, le règlement sur la gouvernance des données sera directement applicable en droit français, sans qu'il soit besoin de mesures de transposition.

Dès lors, il paraît souhaitable d'anticiper cette entrée en vigueur et de construire le cadre commun de tarification comme si le règlement était déjà applicable. Un cadre commun qui s'écarterait du règlement devrait être à nouveau modifié l'année suivante, ce qui serait une source d'instabilité pour les acteurs concernés (producteurs, HDH et réutilisateurs). Tout au plus peut-on envisager que le cadre commun, s'il est adopté au cours de l'année 2022, comporte une période transitoire permettant de maintenir temporairement certaines pratiques actuelles, afin de laisser le temps aux acteurs de se préparer à l'application des nouvelles règles. Il convient aussi de souligner que la tarification au coût marginal

²⁸ Industrie, recherche et énergie.

s'imposera indépendamment de l'inscription de la base de données au catalogue.

Le chapitre pertinent du règlement est le chapitre II, intitulé « Réutilisation de certaines catégories de données protégées détenues par des organismes du secteur public ». Les principales dispositions de ce chapitre intéressant le HDH sont les suivantes :

- L'article 3 dispose que ce chapitre s'applique à certaines catégories de données détenues par des organismes du secteur public, dont les données protégées pour des raisons de protection des données à caractère personnel. Les données de santé détenues par les administrations sont donc concernées²⁹.

- L'article 4 énonce un principe de prohibition des accords d'exclusivité portant sur ces données. De tels accords ne peuvent être conclus que s'ils sont nécessaires pour fournir un service d'intérêt général et la période d'exclusivité doit alors être limitée à 12 mois. Les accords d'exclusivité en cours au moment de l'entrée en vigueur du règlement ne peuvent être renouvelés et doivent s'achever au plus tard 30 mois après cette entrée en vigueur.

- L'article 5 est relatif aux conditions de réutilisation que peuvent définir les organismes du secteur public qui détiennent les données. Ces conditions doivent être non-discriminatoires, transparentes, proportionnées et objectivement justifiées au regard de la nature des données et du but de la réutilisation. L'article donne également une liste indicative des mesures pouvant être prises par l'organisme du secteur public pour préserver le caractère protégé des données : anonymisation ; accès dans le cadre d'un environnement de traitement sécurisé, qui peut se faire à distance où sur les lieux si l'accès à distance ne présente pas de garanties suffisantes. Si la réutilisation ne peut être autorisée que dans le respect des droits de propriété intellectuelle, l'article 5.7 dispose que « *les organismes du secteur public n'exercent pas le droit prévu à l'article 7, paragraphe 1, de la directive 96/9/CE pour le fabricant d'une base de données aux fins d'empêcher la réutilisation de données ou de limiter celle-ci au-delà des limites fixées par le présent règlement* ».

- L'article 6 est relatif aux redevances que les organismes du secteur public peuvent percevoir pour autoriser la réutilisation des données. Celles-ci doivent être « *non discriminatoires, proportionnées et objectivement justifiées* » et ne pas restreindre la concurrence. Selon l'article 6.5, elles doivent être calculées « *sur la base des coûts liés au traitement des demandes de réutilisation* » : il s'agit donc d'une tarification au coût marginal, qui n'inclut pas les coûts de production. Il est précisé que les redevances doivent être limitées aux coûts induits par la reproduction, la fourniture et la diffusion des données, l'anonymisation ou les autres formes de préparation des données personnelles ou confidentielles, la maintenance de l'environnement de traitement sécurisé, les coûts d'acquisition des droits des tiers en dehors du secteur public et enfin les coûts d'assistance aux réutilisateurs cherchant à obtenir le consentement des personnes physiques ou des ayants droit. Les critères et la méthodologie de calcul des redevances doivent être publiés par les Etats membres. Les organismes du secteur public doivent favoriser les utilisations non commerciales, telles que celles relevant de la recherche scientifique, les PME et les

²⁹ Dans la version adoptée en trilogue, il été ajouté la mention « dans la mesure où ces données ne sont pas dans le champ d'application de la directive (UE) 2019/1024 (directive open data) ». L'article 1.2.h de cette directive prévoyant que celle-ci ne s'applique pas « *aux documents dont l'accès est exclu ou limité en application de règles d'accès pour des motifs de protection des données à caractère personnel* », cette directive ne s'applique pas aux données de santé détenues par les administrations et le règlement sur la gouvernance des données s'y applique donc.

start-ups et la société civile, et peuvent leur appliquer des tarifs réduits ou la gratuité. La liste des bénéficiaires et les critères de réduction sont rendus publics.

- Selon l'article 8a (numérotation provisoire), les demandes de réutilisation doivent être traitées par les organismes du secteur public dans un délai de deux mois, ce délai pouvant être allongé de trente jours pour les demandes particulièrement vastes et complexes. Il résulte des dispositions des articles 66 de la loi du 6 janvier 1978 et 90 du décret du 29 mai 2019³⁰ que le délai cumulé d'instruction par le CESREES et par la CNIL des demandes d'autorisation de traitement des données du SNDS peut aller jusqu'à 6 mois, ce qui excède les limites fixées par le règlement. Toutefois, l'article 1.2a du règlement prévoit qu'en cas de conflit avec le droit de l'Union ou le droit national relatif à la protection des données personnelles, la règle de protection des données personnelles doit prévaloir. Le délai d'instruction pouvant être considéré comme une garantie pour la protection des données personnelles, il ne semble pas qu'il y ait de contrariété avec le règlement. Il serait cependant utile d'aborder la question avec le ministère de la santé et la CNIL, pour vérifier que leurs analyses sont convergentes.

Au cours de l'entretien réalisé le 19 janvier, le SGAE a confirmé que la Commission préparait une proposition de règlement sur l'espace européen des données de santé, qui pourrait être adoptée par le collège des commissaires et ainsi rendue publique dans les prochains mois. S'agissant d'une proposition de règlement du Parlement européen et du Conseil, ce texte serait susceptible de déroger au règlement relatif à la gouvernance européenne des données. Il s'agit toutefois d'une perspective à la fois incertaine, faute d'informations plus précises sur le contenu de ce projet, et lointaine, en raison des délais de la procédure législative européenne. A ce stade, il convient donc de conduire l'analyse juridique en se fondant sur le règlement relatif à la gouvernance européenne des données, appelé à entrer en vigueur à brève échéance.

2) Le calcul de la redevance

2.1. Les dépenses pouvant être prises en compte

L'article 6.5 du règlement est construit autour de deux phrases :

- La première donne une définition générale des coûts sur la base desquels les redevances doivent être calculées, qui sont les coûts de traitement des demandes de réutilisation.
- La seconde, qui se présente comme limitative, énumère les catégories de coûts prises en compte : coûts induits par la reproduction, la mise à disposition et la diffusion des données ; coûts induits par l'anonymisation ou les autres formes de préparation des données personnelles ou confidentielles ; coûts induits par la maintenance de l'environnement de traitement sécurisé ; coûts d'acquisition des droits des tiers en dehors du secteur public ; coûts d'assistance aux réutilisateurs cherchant à obtenir le consentement des personnes physiques ou des ayants droit.

L'article 6.5 est proche dans sa rédaction de l'article 6.1 de la directive « open data » du 20 juin 2019, qui mentionnait aussi les coûts de reproduction, de mise à disposition et de diffusion des données, ainsi que les coûts d'anonymisation et de protection des informations

³⁰ Décret n° 2019-536 du 29 mai 2019 pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

confidentielles à caractère commercial. Il ne comporte en revanche aucune des exceptions qui figurait dans la directive, notamment celle en faveur des « *organismes du secteur public qui sont tenus de générer des recettes destinées à couvrir une part substantielle des coûts liés à l'accomplissement de leurs missions de service public* ». Pour ces organismes, l'article 6.4 autorisait, outre la couverture des coûts mentionnés à l'article 6.1, celle des coûts de « *collecte* » et de « *production* », ainsi qu'un « *retour sur investissement raisonnable* ». Il peut donc être déduit du rapprochement des deux textes que la couverture de ces deux catégories de coûts ainsi que la réalisation d'un bénéfice raisonnable sont exclus pour les données relevant du champ d'application du règlement DGA, y compris pour les organismes publics tenus de générer des recettes destinées à couvrir une part substantielle des coûts.

Si l'on applique les catégories de l'article 6.5 du règlement DGA au contexte de la mise à disposition des données du SNDS, les redevances pourraient notamment couvrir les coûts suivants :

- Les coûts supportés par le HDH pour les tâches liées à la mise à disposition des données : ceci inclut les coûts de l'interface technique, les coûts liés à l'aide aux utilisateurs concernant les procédures d'accès au SNDS et les explications données sur les bases de données (qui peuvent être assimilés à des coûts de mise à disposition) ;
- Les coûts supportés par d'autres organismes pour les tâches de mise à disposition : ceci inclut les coûts liés à l'aide aux utilisateurs lorsque celle-ci doit être fournie par les producteurs et les coûts d'appariement de plusieurs bases de données, cette opération étant effectuée par la CNAM ;
- Les coûts supportés par les producteurs pour l'inscription de la base au catalogue : ceci inclut les coûts de normalisation de la base permettant sa mise à disposition par le HDH et les coûts de mise à jour ; l'inscription au catalogue ayant pour seule finalité la mise à disposition des réutilisateurs, le coût des opérations nécessaires à cette inscription peut être pris en compte.

En revanche, la couverture des coûts supportés en amont par les producteurs (coûts de collective, coûts de création de la base de données) et a fortiori la réalisation d'un bénéfice seraient contraires à l'article 6.5 du règlement.

2.2. Le mode d'évaluation : forfaitaire ou analytique ?

Comme l'explique l'étude du Conseil d'Etat de 2002, la jurisprudence administrative admet une certaine « forfaitisation » des redevances, ce qui signifie que la personne publique n'a pas nécessairement à justifier que le montant des redevances est identique à l'euro près aux charges que celles-ci ont vocation à couvrir. Le juge administratif n'exerce qu'un contrôle de l'erreur manifeste d'appréciation sur le montant de la redevance (cf. notamment CE, 19 mars 2010, *Syndicat des compagnies aériennes autonomes (SCARA) et Fédération nationale de l'aviation marchande (FNAM)*, n° 305047, Tab. ; 27 octobre 2016, *SEMRAP*, n° 383501, Tab.).

Le règlement s'avère cependant plus exigeant, l'article 6.6 prévoyant que les Etats membres définissent les critères et la méthodologie de calcul et la publient à l'avance et que l'organisme du secteur public doit publier une description des principales catégories de coûts et des règles utilisées pour la répartition des coûts.

2.3. L'impossibilité de valoriser les droits de propriété intellectuelle sur les bases de données

L'article L. 321-3 du code des relations entre le public et l'administration (CRPA) dispose : « *Sous réserve de droits de propriété intellectuelle détenus par des tiers, les droits des administrations mentionnées au premier alinéa de l'article L. 300-2 du présent code, au titre des articles L. 342-1 et L. 342-2 du code de la propriété intellectuelle, ne peuvent faire obstacle à la réutilisation du contenu des bases de données que ces administrations publient en application du 3° de l'article L. 312-1-1 du présent code.* ». Cet article ne s'applique qu'aux bases de données rendues publiques, ce qui correspond s'agissant du SNDS aux données mentionnées par l'article L. 1461-2 du code de la santé publique.

L'article 5.7 du règlement DGA a une portée plus large. Il s'applique à l'ensemble des données couvertes par le règlement, et donc à toutes les bases de données du SNDS détenues par des organismes du secteur public. S'il subordonne la réutilisation au respect des droits de propriété intellectuelle, il interdit aux organismes du secteur public de faire valoir les droits prévus par l'article 7 de la directive du 11 mars 1996 concernant la protection juridique des bases de données : ceci paraît valable à la fois pour les droits de l'organisme qui détient les données dont la réutilisation est demandée et pour ceux d'un organisme du secteur public tiers. Ainsi si le HDH détient une base de données sur laquelle un autre organisme du secteur public (le producteur) détient des droits, cet organisme tiers ne pourra les faire valoir pour restreindre la réutilisation.

L'article 5.7 ne mentionne que le droit dit *sui generis* prévu par l'article 7 de la directive du 11 mars 1996 et non le droit d'auteur qui peut exister sur certaines bases de données qui « *par le choix ou la disposition des matières, constituent une création intellectuelle propre à leur auteur* » (article 3 de la directive). Mais il est probable que peu de bases de données relevant du SNDS puissent être couvertes par un tel droit, qui implique une originalité dans la construction-même de la base.

Ceci limite considérablement la portée, s'agissant du SNDS, de la mention à l'article 6.5, parmi les coûts pouvant être pris en compte, des coûts « d'acquisition des droits des tiers » car il s'agit seulement des tiers qui ne sont pas des organismes du secteur public.

2.4. La valorisation de services complémentaires

Le chapitre II du règlement DGA s'applique à la réutilisation des données. S'il encadre la tarification des opérations associées à la mise à disposition, telles que les tâches d'anonymisation, le fonctionnement de l'interface technique ou l'aide aux utilisateurs, il ne traite pas des services qui peuvent être fournis par ailleurs, notamment en procédant à une analyse des données à la demande d'une tierce personne. De tels services peuvent donc être tarifés librement par le HDH et par les producteurs de bases de données du SNDS, dans le respect de la jurisprudence du Conseil d'Etat sur les redevances pour service rendu (cf. la note d'analyse juridique de la phase 1).

Toutefois, ces services peuvent également être fournis par d'autres acteurs, tels que des bureaux d'études ou des organismes de recherche, travaillant à partir des données mises à disposition par le HDH et les producteurs. Le HDH et les producteurs doivent donc veiller à ce que leurs pratiques de tarification ne fassent pas obstacle à l'activité concurrentielle de ces opérateurs économiques (CE, 29 juillet 2002, *Société CEGEDIM*, n° 200886, Rec.), en

excluant notamment toute pratique de vente liée imposant aux réutilisateurs de recourir aux services internes d'analyses des données.

3) Les cas d'application de la gratuité

On peut relever tout d'abord que l'article 1^{er} de la loi du 7 octobre 2016 pour une République numérique, qui interdit la perception d'une redevance pour l'échange d'informations publiques entre administrations, n'est pas applicable au SNDS, dont les données ne constituent pas des informations publiques. Les seuls cas dans lesquels la loi impose la gratuité sont donc ceux prévus par les articles L. 1461-2 et L. 1461-5 du code de la santé publique. L'article L. 1461-2 ne pose pas de difficultés pour identifier les jeux de données concernées, puisqu'il s'agit des bases de données anonymes (statistiques agrégées ou données individuelles constituées de telle sorte que l'identification, directe ou indirecte, des personnes concernées y est impossible), mises en « open data » par le HDH. Les catégories mentionnées par l'article L. 1461-5 soulèvent en revanche quelques questions, d'autant plus que les travaux parlementaires n'éclairent pas les termes choisis et qu'aucune circulaire n'est venue exposer leur interprétation par l'administration.

Concernant les « *traitements de données concernant la santé demandés par l'autorité publique* », il s'agit notamment des traitements requis par les études qu'exigent les autorités sanitaires en lien avec les autorisations des produits de santé. Il serait utile d'identifier avec les agences concernées (ANSM, HAS, CNAM) les principaux types d'études et de traitement susceptibles d'être couverts par ces dispositions et les documents qui pourraient être requis des réutilisateurs pour s'assurer que leur demande en relève.

Concernant les « *recherches réalisées exclusivement pour les besoins de services publics administratifs* », cette catégorie recouvre en premier lieu les demandes entrant dans le cadre de « l'accès permanent » prévu par le I-2^o de l'article L. 1461-3. Il ne semble cependant pas qu'elle doive s'y limiter. Les universités, les établissements de santé et les établissements publics à caractère scientifique et technologique (EPST) tels que le CNRS et l'INSERM sont des services publics administratifs et l'article L. 1461-5 interdit donc de soumettre à une redevance leurs demandes de réutilisation des données du SNDS.

4) La différenciation des tarifs

4.1. Les cas de différenciation imposés par le règlement DGA

L'article 6.4 du règlement DGA impose d'appliquer des tarifs réduits ou la gratuité dans les cas suivants :

- Les demandes de réutilisation des PME : comme le suggère le considérant n° 2 du règlement, il y a lieu de retenir la définition européenne des PME (Recommandation de la Commission du 6 mai 2003 concernant la définition des micro, petites et moyennes entreprises), qui est aussi celle du décret n° 2008-1354 du 18 décembre 2008 relatif aux critères permettant de déterminer la catégorie d'appartenance d'une entreprise pour les besoins de l'analyse statistique et économique : il s'agit des entreprises qui emploient moins de 250 personnes et qui ont un chiffre d'affaires annuel n'excédant pas 50 millions d'euros ou un total de bilan n'excédant pas 43 millions d'euros ;
- Les demandes des startups : le règlement ne définit pas ce terme et il n'existe pas non plus

de définition transversale des startups en droit de l'Union européenne ; si les startups sont le plus souvent des PME³¹ au sens de la définition ci-dessus, leur croissance rapide peut les conduire à dépasser ces seuils ;

- Les demandes de la société civile : là encore cette notion n'est pas définie ; il pourrait être envisagé d'appliquer un tarif réduit ou la gratuité aux organismes à but non lucratif n'intervenant pas dans un secteur concurrentiel³² (celles qui sont exonérées d'impôt sur les sociétés, de TVA et des autres impôts commerciaux en application de la jurisprudence *Association Jeune France* du Conseil d'Etat (CE, Sect., 1^{er} octobre 1999, n° 170289, Rec.)) ;
- Les demandes des établissements d'éducation : cette catégorie semble en grande part incluse dans celle des « *recherches réalisées exclusivement pour les besoins de services publics administratifs* », pour laquelle l'article L. 1461-5 impose déjà la gratuité.

4.2. Les cas de différenciation autorisés par la jurisprudence administrative

Selon la jurisprudence du Conseil d'Etat, « *la fixation de tarifs différents applicables, pour un même service rendu, à diverses catégories d'usagers d'un service public implique, à moins qu'elle ne soit la conséquence nécessaire d'une loi, qu'il existe entre les usagers des différences de situation appréciables ou que cette mesure soit justifiée par une nécessité d'intérêt général en rapport avec les conditions d'exploitation du service* » (CE, Sect., 10 mai 1974, *Sieur Denoyez et Sieur Chorques*, n° 88032, Rec. ; 28 février 1996, *Etablissement public du musée du Louvre*, n° 163528, Tab.)

Dans la précédente note, il avait été envisagé d'appliquer des tarifs plus élevés aux réutilisateurs établis dans un Etat non membre de l'Union européenne. Cependant, compte tenu du plafonnement au coût marginal imposé par le règlement DGA, il n'existe en réalité guère de marge de manœuvre en ce sens (sauf à appliquer un tarif inférieur au coût marginal à l'ensemble des réutilisateurs établis dans l'Union européenne).

L'éventualité d'un tarif réduit pour les membres du HDH était également mentionnée. Toutefois, un certain nombre de ces membres (Etat, CNAM, CCMSA, Union nationale des régimes spéciaux, HAS, ANSM, Santé publique France, InCa, ABM, CNSA) bénéficient d'un accès permanent et gratuit. S'agissant des autres membres, qui sont notamment des fédérations d'organismes complémentaires, d'offeurs de soins ou d'industriels, ainsi que certains établissements publics de santé, il n'est pas évident qu'il existe par rapport à d'autres acteurs « *des différences de situation appréciables* » : on peut par exemple douter que l'AP-HP, qui est membre du HDH, soit dans une situation différente au regard du service public de l'accès aux données de santé d'un établissement public de santé non membre ou même d'une clinique privée.

5) Les règles applicables aux bases de données produites par des organismes du secteur privé

5.1. L'identification des organismes du secteur privé

³¹ Le régime fiscal favorable des jeunes entreprises innovantes est d'ailleurs réservé aux PME (article 44 *sexies*-0 A du code général des impôts).

³² Cette seconde condition apparaît nécessaire, car des organismes à but non lucratif telles que les mutuelles ou les institutions de prévoyance peuvent être en concurrence avec des organismes à but lucratif, de sorte qu'une différence de tarif pourrait être injustifiée.

On qualifie ici d'organismes du secteur privé qui ne sont pas des organismes du secteur public au sens du règlement DGA, et auxquels celui-ci n'est donc pas applicable. L'article 2 du règlement range parmi les organismes du secteur public l'Etat, les collectivités locales et les « *organismes de droit public* », notion qui recouvre elle-même les organismes combinant trois caractéristiques : ils ont été créés pour satisfaire spécifiquement des besoins d'intérêt général ayant un caractère autre qu'industriel ou commercial ; ils sont dotés de la personnalité juridique ; ils sont soit financés majoritairement par des organismes du secteur public, soit soumis dans leur gestion à un contrôle de ces organismes, soit dirigés par un organe composé au moins pour moitié de représentants du secteur public. Cette définition est identique à celle de l'article 2 de la directive du 20 juin 2019 sur l'open data.

Si l'on se réfère au projet d'arrêté relatif aux bases de données alimentant la base principale et le catalogue du système national des données de santé, la grande majorité des producteurs sont sans nul doute des organismes du secteur public : c'est le cas du ministère de la santé, de l'ensemble des établissements publics administratifs (ARS, INSERM, Santé Publique France), des établissements publics de santé et des universités.

S'agissant de la société Unicancer, il s'agit d'un groupement de coopération sanitaire de moyens (GCSM) composé des centres de lutte contre le cancer (CLCC) et de la fédération hospitalière Unicancer. Les CLCC sont quant à eux des établissements de santé de droit privé à but non lucratif. Les CLCC remplissent bien les critères de l'organisme de droit public au sens du règlement DGA (créés dans un but d'intérêt général, dotés de la personnalité morale et financés par l'assurance-maladie³³), de sorte qu'il en va de même pour la société Unicancer.

S'agissant du Centre régional de coordination des dépistages des cancers en Occitanie, davantage d'informations seraient nécessaires concernant cet organisme.

Le HDH a signalé la présence d'une base dans le projet de catalogue d'une base de données de la société Roche. Il s'agit bien sûr d'un organisme du secteur privé.

5.2. Les conséquences de la qualification d'organisme du secteur privé

Dans un schéma d'accès direct auprès du producteur pour la réutilisation des données, celui-ci est tout à fait libre de sa politique tarifaire : il n'est soumis ni au règlement DGA ni à la jurisprudence administrative sur les redevances pour service rendu.

En revanche, dès lors que la base est inscrite au catalogue du SNDS, lorsque l'accès se fait auprès du HDH, elle devient une base de données détenue par un organisme du secteur public entrant dans le champ d'application du règlement DGA et soumise à ses principes. Ceci crée un problème de discordance tarifaire et pourrait inciter les organismes privés contrôlant des bases de données relevant du SNDS à s'opposer à leur inscription au catalogue. Ce sujet pourrait être traité au moyen de dispositions d'harmonisation du cadre commun de tarification (cf. infra).

Du point de vue du HDH, la seule spécificité induite par l'origine de la base de données dans

³³ La question de savoir si les organismes de sécurité sociale sont des organismes de droit public au sens de la directive open data ne semble pas avoir été déjà tranchée par la jurisprudence. La réponse ne fait cependant guère de doute, dès lors que les organismes de sécurité sociale sont créés par la loi dans un but d'intérêt général, sont dotés de la personnalité morale et sont contrôlés par l'Etat.

le secteur privé tient à la possibilité, dans le cadre de l'article 6.5 du règlement DGA, de refacturer les coûts d'acquisition des droits de propriété intellectuelle auprès de ce dernier.

6) L'architecture du cadre commun de tarification

Pour mémoire, dans l'état actuel du droit, chaque producteur d'une base de données relevant du SNDS est compétent pour fixer la redevance applicable à l'accès à cette base. Lorsqu'une base de données est inscrite au catalogue, le HDH devient responsable de sa mise à disposition (III-2° de l'article R. 1461-3 du code de la santé publique) et son conseil d'administration est alors compétent pour fixer les tarifs de ce service (articles 3 et 9.4 de la convention constitutive). Toutefois, la mise à disposition pouvant toujours être effectuée directement par le producteur (IV de l'article R. 1461-2), le tarif applicable est alors celui défini par le producteur.

Ceci pose un double problème de cohérence :

- Une même base inscrite au catalogue peut faire l'objet de deux régimes tarifaires distincts selon que l'accès s'opère auprès du HDH ou du producteur ;
- Indépendamment de l'inscription au catalogue, des choix très différents³⁴ peuvent être effectués par les producteurs concernant les coûts à prendre en compte pour calculer la redevance, les cas de gratuité et les critères de modulation des tarifs ; il en résulte un problème de lisibilité de l'objet commun qu'est le SNDS.

6.1. Le contenu du cadre commun : la question du degré d'harmonisation

Schématiquement, trois degrés d'harmonisation peuvent être envisagés.

L'harmonisation minimale : le maintien d'une tarification distincte pour chaque base du SNDS

Dans cette option, on impose seulement que le cadre tarifaire pour une base de données inscrite au catalogue soit le même, que l'accès se fasse par l'intermédiaire du HDH ou directement auprès du producteur.

L'harmonisation complète : un régime tarifaire uniforme

Cette option implique la mise en place d'un cadre tarifaire unique, tant le tarif que ses critères de étant identiques quelle que soit la base du catalogue dont l'accès est sollicité et indépendamment du guichet d'accès (HDH ou producteur). Ce tarif uniforme pourrait être imposé y compris aux organismes du secteur privé contrôlant des bases relevant du SNDS.

Le maintien d'une tarification par base mais selon des critères communs

Dans cette option intermédiaire, chaque producteur demeure compétent pour définir le cadre tarifaire d'accès à sa base de données, mais ce cadre doit être défini en respectant des critères communs. Ce type de schéma « décentralisé » se retrouve dans nombre de textes législatifs et réglementaires concernant différents types de redevances pour service rendu (cf. exemples en annexe).

En l'absence du règlement DGA, les redevances pourraient couvrir les coûts de production et

³⁴ Qui peuvent s'expliquer par divers facteurs tels que les possibilités de financement dont dispose par ailleurs le producteur, la notoriété de sa base de données, le profil des réutilisateurs potentiels, etc.

tenir compte de la valeur économique de la prestation pour le bénéficiaire. Ces coûts étant variables selon la base concernée, le schéma d'harmonisation minimale et le schéma décentralisé pourraient se justifier. Toutefois, dès lors que le règlement DGA impose la tarification au coût marginal, le schéma d'harmonisation complète paraît s'imposer, car les coûts de mise à disposition dépendront peu de la base recherchée. Par ailleurs, il est possible que pour une même base, ces coûts soient différents pour le HDH et pour le producteur, qui n'utilisent pas nécessairement les mêmes moyens techniques ; toutefois, pour des raisons de cohérence, il est souhaitable qu'un tarif unique soit appliqué indépendamment du guichet d'accès dès lors que la base est inscrite au catalogue.

6.2. L'instrument juridique du cadre commun

Sauf lorsqu'elles affectent la libre administration des collectivités territoriales, les règles relatives aux redevances pour service rendu ne relèvent pas du domaine de la loi mais du domaine réglementaire. Pour les redevances perçues par l'Etat, l'article 4 de la loi organique du 1^{er} août 2001 relative aux lois de finances (LOLF) prévoit qu'elles sont établies et perçues sur la base de décrets en Conseil d'Etat pris sur le rapport du ministre chargé des finances et du ministre intéressé, qui doivent être ratifiés dans la plus prochaine loi de finances. Le décret peut cependant se borner à préciser l'objet de la redevance et les prestations donnant lieu à rémunération et renvoyer à l'arrêté la définition des règles de calcul du tarif. Pour les redevances perçues par des établissements publics ou des groupements d'intérêt public, la fixation de la redevance relève de l'organe compétent de la personne publique (le plus souvent le conseil d'administration) ; toutefois, un décret, pris sur le fondement du pouvoir réglementaire autonome de l'article 37 de la Constitution, pourrait définir des règles communes s'imposant à ces organes.

Dès lors que l'on s'inscrit dans le schéma d'harmonisation complète, le conseil d'administration du HDH est compétent pour fixer les tarifs applicables à l'ensemble des bases inscrites au catalogue. Ces tarifs ne s'imposeront pas en revanche aux producteurs lorsque l'accès s'effectue directement auprès d'eux ; il convient donc d'identifier un instrument juridique garantissant la cohérence. Trois options sont envisageables :

- Inscire cette règle dans un décret, qui prévoirait que l'inscription au catalogue impose au producteur d'appliquer les tarifs définis par le HDH lorsque l'accès s'effectue directement auprès de lui ;
- Inscire la même règle dans la convention conclue entre le HDH et le producteur à l'occasion de l'inscription de la base au catalogue ;
- S'en tenir à une recommandation en ce sens du comité stratégique des données de santé (CSDS), ce qui entre dans ses attributions, l'article 1^{er} de l'arrêté du 29 juin 2021 portant création de ce comité disposant qu'il est chargé d'émettre « *des recommandations afin de favoriser le partage des données relevant du système national des données de santé, en prenant en particulier en compte ses aspects financiers et juridiques* ».

S'agissant des organismes du secteur privé, les tarifs qu'ils appliquent ne sont pas des redevances pour service rendu ; ils relèvent de la liberté des prix conformément à l'article L. 410-2 du code de commerce. Cet article permet de réglementer les prix par décret en Conseil d'Etat pris après avis du Conseil de la concurrence, dans les secteurs « *où la concurrence par les prix est limitée en raison soit de situations de monopole ou de difficultés*

durables d'approvisionnement » : il n'est toutefois pas évident que ces conditions soient réunies pour l'accès aux bases de données du SNDS. La seule voie envisageable pour imposer aux organismes du secteur privé d'appliquer les tarifs du HDH serait donc une disposition législative, inscrite dans le code de la santé publique.

6.3. Les modalités de rétrocession

Dès lors qu'une base est inscrite au catalogue, le HDH est compétent pour percevoir la redevance puisque c'est lui qui assure la mise à disposition des données. Il est toutefois envisageable qu'une partie de la redevance perçue soit rétrocédée au producteur. Cette faculté est prévue par l'article 3 de la convention constitutive du HDH, qui stipule que la Plateforme « peut tarifier des services en lien avec ses missions et reverser une partie de ses recettes aux partenaires tels que les producteurs de données qui ont contribué à leur réalisation ».

L'article 107 de la loi de simplification et d'amélioration de la qualité du droit (dite « loi Warsmann ») dispose certes que « *le groupement d'intérêt public ne donne pas lieu au partage de bénéfices* » et que « *les excédents annuels de la gestion ne peuvent qu'être utilisés à des fins correspondant à l'objet du groupement ou mis en réserve* ». Il n'existe pas de jurisprudence à ce jour concernant ces dispositions. Lues de manière littérale, elles interdisent seulement que le bénéfice constaté en fin d'exercice soit distribué aux membres. A supposer qu'elles doivent être lues comme s'appliquant à toute opération de rétrocession de redevance en cours d'exercice, dès lors que la tarification est limitée à la couverture des coûts de mise à disposition, elle ne génère aucun bénéfice pour le HDH et la rétrocession au profit d'un membre du GIP ne peut donc être qualifiée de partage de bénéfice.

La rétrocession devrait porter sur les coûts de mise à disposition supportés par le producteur : coûts d'inscription au catalogue, coûts d'aide aux utilisateurs lorsque celle-ci fait intervenir les moyens du producteur.

Les modalités de la rétrocession pourraient être fixées par la convention conclue entre le HDH et le producteur à l'occasion de l'inscription de la base au catalogue. Son principe pourrait être prévu par décret.

6.4. La place de la négociation dans la fixation des redevances

Plusieurs des producteurs interrogés ont fait état de pratiques de négociation du prix de la mise à disposition de leur base de données, permettant d'obtenir un meilleur tarif selon le profil du réutilisateur et notamment selon ses moyens financiers. Il n'est pas évident qu'une telle pratique soit compatible avec le régime juridique des redevances pour service rendu. Selon une étude du Conseil d'Etat³⁵, « la compétence de l'autorité administrative s'exerce en principe de manière unilatérale », même si les redevances sont assimilables à des prix sur le plan économique. Nous n'avons pas identifié dans la jurisprudence administrative de précédent dans lequel la redevance aurait été fixée par contrat³⁶. La négociation de gré à gré

³⁵ *Redevances pour service rendu et redevances pour occupation du domaine public*, octobre 2002.

³⁶ Deux décisions peuvent néanmoins être mentionnées. S'agissant de l'exercice d'une activité libérale à l'hôpital public par certains praticiens hospitaliers, qui donne lieu au versement d'une redevance, il est jugé que celle-ci n'est pas déterminé par le contrat conclu par le praticien avec l'hôpital et qui détermine les modalités de son activité libérale (CE, 29 mai 2009, *Syndicat national de chirurgie*

avec chaque client ne garantit pas le respect du principe d'égalité, qui interdit de traiter de façon différente, sans motif d'intérêt général, deux personnes se trouvant dans la même situation. En outre, les dispositions de l'article 6.6 du règlement DGA sur la transparence de la méthodologie de calcul de la redevance imposent à tout le moins qu'une négociation se déroule dans un cadre public et connu des réutilisateurs potentiels.

En tout état de cause, la limitation des tarifs à la couverture des coûts de mise à disposition paraît limiter fortement l'intérêt d'une négociation.

plastique reconstructrice et esthétique et autre et syndicat national de défense de l'exercice libéral de la médecine, n° 318071, Tab.). S'agissant de la fourniture de données publiques, une personne publique peut fixer par contrat les droits de propriété intellectuelle qu'elle perçoit à l'occasion de la commercialisation de ces données (CE, Ass., 10 juillet 1996, *Société Direct Mail Promotion et autres*, n° 168702, Rec.), mais il s'agit d'une rémunération de nature différente de la redevance pour service rendu, qui s'ajoute à cette dernière.

Annexe : Exemples de redevances dont la fixation est décentralisée

Les droits de port : les droits de port sont les redevances perçues par les gestionnaires des ports (grands ports maritimes, ports autonomes ou collectivité territoriale responsable) à raison des opérations commerciales ou des séjours des navires et de leurs équipages effectués dans le port. Les dispositions réglementaires du code des transports définissent les composantes des droits de port (redevance sur le navire, redevance de stationnement, etc), le mode de calcul de chacune d'entre elles (par exemple, la redevance sur le navire est calculée en fonction de son volume, de ses dimensions et de son tirant d'eau) et les possibilités de modulation et d'abattements. Les droits de port sont fixés selon ces règles par le gestionnaire du port (conseil d'administration du port autonome, directoire du grand port maritime ou collectivité territoriale responsable). Leur présentation doit respecter un « cadre type uniforme » fixé par arrêté conjoint du ministre chargé des finances et du ministre chargé des ports maritimes.

Les redevances aéroportuaires : il s'agit des redevances perçues pour les services publics aéroportuaires rendus sur les aérodromes ouverts à la circulation aérienne publique. Elles sont dues par les usagers de l'aérodrome (les compagnies aériennes) et versées à l'exploitant de celui-ci. Le code de l'aviation civile définit les composantes de ces redevances (redevance d'atterrissage, redevance de stationnement et redevance par passager), le mode de calcul de chacune d'entre elles (par exemple, la redevance d'atterrissage est fonction de la masse maximale certifiée au décollage de l'aéronef), les dépenses pouvant être prises en compte (dépenses correspondant à des investissements déjà réalisés, dépenses engagées et, dans certains cas, dépenses futures) et les possibilités de modulation (destinées par exemple à réduire ou à compenser les atteintes à l'environnement). Les redevances sont fixées par l'exploitant de l'aérodrome ou la collectivité territoriale responsable.

Les redevances d'eau potable et d'assainissement : ces redevances sont fixées par les communes ou les intercommunalités compétentes, dans des conditions fixées par le code général des collectivités territoriales. Celui-ci définit les dépenses que ces redevances ont vocation à couvrir (charges consécutives aux investissements, au fonctionnement et aux renouvellements nécessaires à la fourniture des services, ainsi que les charges et les impositions de toute nature afférentes à leur exécution) et le mode de calcul (pour la redevance d'eau potable, le montant est calculé en fonction du volume et peut comprendre une part fixe, qui est fonction notamment du nombre de logements desservis par le réseau).

Annexe 4 - Modèle de financement pour les centres de ressources biologiques

L'appellation centre de ressources biologiques (CRB) est apparue en 1999 à l'issue du travail de l'OCDE puis a été entérinée en 2001 par le Ministre chargé de la recherche. Les premières banques d'échantillons, en particulier les banques d'échantillons humains bien que très précieuses pour la recherche médicale, ont malheureusement été souvent constituées et gérées dans de mauvaises conditions, faute de moyens, de contrôle et de formation spécifique. Ainsi, une dynamique obligatoire d'autorisation de ces infrastructures a été introduite par l'article L 1243-4 du Code de la santé publique.

Depuis 2016, les CRB des établissements de santé assurant la conservation des échantillons biologiques humains sont éligibles à un financement encadré par le Ministère de la santé : la dotation MIG D04. Cette dotation vise à financer les missions de préparation, conservation et mise à disposition des ressources biologiques avec pour objectif de pourvoir des programmes de recherche avec des éléments du corps humain pour in fine développer de nouvelles approches diagnostiques et thérapeutiques. Les missions relatives à la prise en charge des échantillons biologiques humains (sang, urines, sperme, biopsies...) assurées par les CRB comportent :

- 1) La réception des échantillons biologiques à partir de leur site de prélèvement. Cette étape permet de vérifier la qualité pré-analytique de l'échantillon (prélèvement et acheminement effectués dans de bonnes conditions) et de l'enregistrer au niveau du CRB (catégorisation et anonymisation de l'échantillon).
- 2) La préparation de l'échantillon biologique. Cette étape consiste à transformer l'échantillon afin qu'il puisse être analysé dans de bonnes conditions par la suite (centrifugation des échantillons sanguins, inclusions des biopsies en bloc paraffine ...).
- 3) Le stockage de l'échantillon. Cette étape permet de conserver l'échantillon de manière optimale en attendant sa mise à disposition pour analyse. La conservation peut être effectuée à froid (4°C ou congélation) ou à température ambiante. Quel que soit le mode de conservation, le CRB contrôle régulièrement, à l'aide de relevés automatiques, les courbes de température.
- 4) La mise à disposition des échantillons à des fins de recherche. Les échantillons conservés au CRB sont indexés dans un catalogue en ligne disponible pour la communauté scientifique. Les laboratoires qui souhaitent disposer d'un échantillon biologique doivent en faire la demande écrite au CRB via un bon de sortie. Une fois la demande vérifiée et validée par le CRB, celui-ci s'assure que l'échantillon est correctement sorti et acheminé vers le laboratoire.

En 2022, 51 établissements bénéficient de la dotation MIG D04 qui s'élève à 26,6 M€. Cette dotation est composée d'une part fixe (150 000 € par CRB) et d'une part variable (323 000 € en moyenne par CRB). La répartition de cette part variable repose sur 5 groupes d'indicateurs d'activité permettant de mesurer un volume d'échantillons tout au long de leur prise en charge par le CRB, ces 5 groupes d'indicateurs comprennent :

- 1) le nombre de réceptions (tout groupe d'échantillons d'un même patient reçu conjointement) et le nombre de requalifications (échantillon reçu mais utilisé à une fin différente de celle initialement programmée)

- 2) le nombre de préparations à partir de ces échantillons : préparations simples (centrifugation, aliquotage) et complexes (transformation d'échantillons, par exemple inclusion d'une biopsie en bloc paraffine)
- 3) le nombre de nouveaux échantillons stockés au CRB (en distinguant ceux conservés à température ambiante de ceux conservés au froid)
- 4) le nombre total d'échantillons conservés au CRB (en distinguant ceux conservés à température ambiante de ceux conservés au froid)
- 5) le nombre d'échantillons mis à disposition (comprend le nombre d'envois indépendamment du nombre d'échantillons et le nombre d'échantillons sortis à des fins de recherche)

Annexe 5 - Lettre de mission du groupe de travail

Paris,
18 mai 2022

Lettre de mission du Groupe de travail "Financement" du Comité stratégique des données de santé

L'arrêté du 29 juin 2021 a créé le comité stratégique des données de santé (comité stratégique) auprès du ministre chargé de la santé afin de lui apporter des éléments d'orientation et de décision relatifs à la mise en œuvre et au développement du système national des données de santé.

Ce comité stratégique est notamment chargé d'émettre des recommandations afin d'améliorer la collecte et de favoriser le partage des données relevant du SNDS, en prenant en particulier en compte ses aspects financiers et juridiques ainsi que de conduire et diffuser une réflexion prospective sur le cadre juridique relatif à la collecte et au partage des données.

Il apparaît aujourd'hui que les problématiques liées au financement constituent aujourd'hui un frein majeur à la collecte et au partage des données de santé.

En effet, la complexité d'accès aux financements - majoritairement publics - pour constituer des bases de données de santé stratégiques ; l'absence de financement pérenne et la reconnaissance insuffisante des efforts de collecte entraînent aujourd'hui un réflexe prioritaire qui limite la constitution d'un patrimoine utile de données et son partage.

Cette situation conduit également les acteurs à nouer des partenariats public-privé et à proposer des services commerciaux afin d'assurer un niveau minimal de revenus issus de la tarification. Ces services d'analyse de données tarifés font nécessairement obstacle à une ouverture large des données.

Ce constat se retrouve pour différentes typologies de bases de données : entrepôts d'établissement de santé, cohortes, registres, etc.

Le comité stratégique a décidé de se saisir de ce sujet en mettant en place un groupe de travail dédié au financement des bases de données de santé (GT Financement).

Le GT financement s'attachera à établir un état des lieux sur les mécanismes actuels de financement (moyens et vecteurs) des bases de données de santé revêtant un caractère stratégique. La diversité des pratiques devra être illustrée.

Il est également attendu que le GT puisse établir des propositions permettant de renforcer la pérennité des financements afin de pouvoir alimenter et actualiser les bases dans la durée. Ces propositions devront être chiffrées et illustrées quant aux vecteurs mobilisables (Sigaps, Merri...).

Le GT s'attachera également à proposer une distinction entre ce qui lui semble relever d'une ouverture et d'une gratuité d'accès d'une part, de ce qui pourrait générer une tarification permettant d'enclencher une rétribution vertueuse pour les producteurs.

Enfin, le GT pourra faire des propositions d'évolution de la réglementation actuelle afin de simplifier les mécanismes de financement et de tarification pour les rendre plus accessibles, plus incitatifs et ainsi accélérer la collecte et le partage des données de santé.

Ces analyses devront être articulées avec les travaux actuellement menés au niveau européen relatif à la gouvernance des données (DGA) et qui visent à définir de nouvelles règles visant à favoriser le partage des données au sein de l'UE en vue de stimuler l'innovation dans le numérique ; créer de nouvelles règles européennes sur la neutralité des marchés de données ; faciliter la réutilisation de certaines données détenues dans le secteur public. Une attention particulière sera portée à la création du futur espace de données de santé européen.

De manière très opérationnelle, et compte tenu de l'appel à projets visant à financer le déploiement des entrepôts de données de santé hospitaliers qui se lance dans le cadre de la stratégie d'accélération de la santé numérique, il apparaît indispensable de traiter en priorité le sujet des entrepôts de données de santé hospitaliers. Le diagnostic et les propositions sur ce sujet sont attendus dans le courant de l'été 2022.

Pour les autres types de bases stratégiques de données de recherche, il serait utile, pendant la période où on traite des EDS, de documenter les modalités de financement de ces différents types de bases de données de recherche. Ces informations permettront la poursuite des travaux sur le financement de la collecte et du partage afin de faire émerger des modèles systémisables.

Les travaux relatifs à la tarification et aux évolutions réglementaires sont attendus d'ici à la fin de l'année 2022.