

Paris,
20 juin 2024

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

Objet : mise en place d'un primo-accès pour les bases de données du catalogue du SNDS

Certains responsables de données souhaitent pouvoir bénéficier d'un primo-accès sur leurs données inscrites au Catalogue du SNDS afin de protéger leurs intérêts. Afin de pouvoir mettre en place ce primo-accès et de concilier à la fois les intérêts des responsables de données et les enjeux de réutilisation rapide des données, la période du primo-accès doit être limitée dans le temps, réservée aux responsables de données dont la demande serait considérée comme étant légitime, et mise en place de façon contractuelle.

I. La mise en place d'un primo-accès raisonnable semble justifiée pour encourager le partage des données tout en préservant l'intérêt de certains responsables de données

Le catalogue du SNDS est une collection de bases de données, construit de manière progressive avec des partenaires et sous l'égide du Comité stratégique des données de santé. Il se compose des bases d'intérêt prioritaires pour l'écosystème justifiant d'optimiser leur visibilité et de les partager *via* la Plateforme des données de santé (PDS). Les bases présentent au catalogue du SNDS peuvent être d'une grande diversité (cohortes, registres, entrepôts, bases administratives...) et peuvent être portées par différents types de responsables de données (des académiques, des start-ups, des industriels, des CHU...) dont les enjeux peuvent être différents. Par exemple, parmi les bases de données qui composent le catalogue du SNDS, certaines ont été créées et financées par l'Etat dans l'objectif de répondre à des missions de service public. C'est le cas des bases de données des registres de morbidité, des observatoires régionaux de santé, des agences régionales de santé ou bien encore de Santé Publique France.

Lors des nombreux échanges ayant eu lieu entre l'écosystème des responsables de bases de données et la PDS, une des demandes revenant fréquemment dans les discussions est celle de la mise en place d'une période durant laquelle leurs données chaînées aux données de la base principale du SNDS ne seraient accessibles qu'à eux. Malgré leur fort intérêt pour la mise au catalogue de leurs données, ils estiment que le risque de se créer de la concurrence (que d'autres porteurs de projets réalisent des projets de recherche avant eux) est trop important et peut nuire à leur activité et/ou à leur financement. A titre d'exemple, ce risque concurrentiel peut être présent :

- lorsque des équipes académiques de recherche trouvent des financements pour constituer une base de données (par exemple une cohorte) dans l'objectif de répondre à une question de recherche. Si cette cohorte est partagée dès sa constitution sans primo-accès, d'autres équipes de recherche, dotées de moyens humains et financiers plus importants pourraient répondre à la question de recherche avant l'équipe ayant créé la base de données. Cela mettrait en péril l'activité et les

financements de l'équipe académique productrice de ces données. Au-delà de cet aspect concurrentiel, cela reviendrait également à réaliser deux projets de recherche identiques sur les mêmes données et remettrait donc en question l'activité des organismes financeurs de projets si à chaque fois qu'ils attribuent des financements *via* par exemple un appel à projet, ils prennent le risque de ne pas voir le projet se réaliser *via* leur financement mais par d'autres équipes de recherche

- lorsque les bases de données font l'objet de financements privés. En effet, dans certains cas de figure les industriels financent des bases de données pour avoir l'exclusivité de l'exploitation de ces dernières. Si ces bases de données sont partagées sans primo-accès et donc sans période d'exclusivité pour les industriels, les industriels ne voudront plus financer ces bases de données. Ici aussi, la pérennité de la base de données est mise en danger.
- lorsque des responsables de données ont pour mission de produire des indicateurs épidémiologiques nationaux uniques et robustes afin de mieux piloter les politiques publiques. Si nous prenons l'exemple des registres de cancers, ces derniers ont dans leurs missions de service public la production d'indicateurs épidémiologiques. Si d'autres structures ayant plus de moyens financiers produisent ces mêmes indicateurs à des fréquences plus élevées, d'une part l'activité des registres est menacée, d'autre part, il y a un risque de se retrouver avec plusieurs taux d'incidence pour le cancer colorectal (par exemple) car la méthodologie ne serait pas la même.

Aussi, il est à noter qu'un des avantages de la mise au catalogue de leur base de données pour les responsables de données est de bénéficier de la réalisation par la PDS de l'appariement de leur base à celle de la base principale du SNDS. La mise en place d'un primo-accès permettrait donc aux responsables de données de pouvoir réaliser eux-mêmes des projets de recherche sur leur base chaînée à la base principale du SNDS et répondre ainsi à des nouvelles problématiques de recherche avant que d'autres équipes de recherche exploitent ces nouvelles données.

D'un point de vue réglementaire, le décret n° 2021-848 du 29 juin 2021 relatif au traitement de données à caractère personnel dénommé « système national des données de santé » prévoit de mettre à disposition les données dans un délai raisonnable : *"Article R 1461-2 III. – La Plateforme des données de santé et la Caisse nationale de l'assurance maladie mettent des données à disposition des organismes responsables des traitements prévus aux 1o et 2o du I de l'article L. 1461-3 dans un délai raisonnable."* Par conséquent, la formulation actuelle des textes ne s'oppose pas à la mise en place d'un primo-accès dont on pourrait juger qu'il est de durée raisonnable.

Au regard de ces enjeux et des demandes de l'écosystème des responsables de données de santé et compte-tenu des textes applicables, il semble bénéfique de proposer un primo-accès pour certaines bases de données du catalogue du SNDS. Néanmoins, afin de respecter le cadre réglementaire et d'accélérer la recherche en santé, **il est nécessaire de cadrer ce primo-accès ainsi que son périmètre d'application (acteurs concernés, usages pour lesquels elle s'applique, etc.).**

II. Un tel primo-accès pourrait s'appliquer aux responsables de données dont les activités principales sont issues de l'exploitation scientifique de leur base de données, pour un délai de 6 à 12 mois précisé dans la convention de partenariat avec la PDS, à l'exception des utilisations à des fins de mission de service public.

Au regard des éléments précédents, un primo-accès respectant les caractéristiques suivantes pourrait être mis en place pour les bases de données du catalogue du SNDS.

- Bases de données concernées

Les bases de données pouvant bénéficier d'un primo-accès sont celles faisant l'objet d'un risque concurrentiel c'est-à-dire :

- les cohortes financées par des fonds publics ou des fonds mixtes (publics et privés) ;
- les cohortes, registres ou EDS financés uniquement par des fonds privés ;
- les bases de données créées pour répondre à des missions de service public.

- Périmètre (usages concernés)

Il s'agit de refuser l'accès aux données d'une base du catalogue pendant sa période de primo-accès pour toutes demandes provenant d'un porteur de projet ne faisant pas partie de l'organisme responsable de données à l'exception des demandes d'accès émanant de la puissance publique que les agences en charge des sujets traitent en fonction de leur périmètre de responsabilité.

- Durée du primo-accès et modalité d'application

La durée du primo-accès accordée pourra être de 6 à 12 mois conformément aux pratiques observées au niveau international et en fonction de la durée des activités menacées des responsables de données.

En effet, que ce soit au niveau international ou plus spécifiquement au niveau des HDAB, nous pouvons constater que peu de pays mettent en place un primo-accès sur les données avant de les mettre à disposition (cf. Annexe 1). Lorsqu'il y a un primo-accès, celui-ci peut aller de 3 à 18 mois en fonction des pays avec une majorité autour de 12 mois (Norvège, Belgique, Etats-Unis, Croatie) et se fait généralement sur demande faisant l'objet d'une évaluation.

Deux cas de figure peuvent être imaginés pour cette durée de primo-accès :

- elle pourrait être fixée dans la convention qui lie le responsable de données et la PDS pour être applicable de manière fixe à tous les usages concernés, par exemple dans le cas des cohortes ;
- elle pourrait faire l'objet d'une évaluation projet par projet et trouver à s'appliquer selon le cas, avec une durée maximale fixée dans la convention, par exemple pour les responsables de données ayant des missions de service public.

Dans ce second cas, le conseil scientifique et éthique de la base de données est habilité à déterminer si le primo-accès est applicable ou non mais doit le justifier.

- Prise en compte du primo-accès dans le processus de sélection des bases de données au catalogue du SNDS

La mise en place d'un primo-accès sur une base de données du catalogue du SNDS implique :

- que cette demande ait été mentionnée par le responsable de données dans sa lettre d'intention (cf. Annexe 2) ;
- que le sujet ait été discuté avec la PDS notamment dans l'optique de préciser si cette demande est légitime et que les exceptions ont été comprises par le responsable de données. Cette discussion est aussi l'occasion de préciser la durée du primo-accès et les modalités de mise en oeuvre retenues (cf. supra)¹ ;

¹ Le point est ajouté à la check list du processus de sélection des bases du catalogue du SNDS jointe en annexe 3

- lorsqu'il exprime son avis sur l'inclusion de la base au catalogue du SNDS, le comité stratégique prend en compte les aspects relatifs au primo-accès (applicabilité, durée) ;
- en cas d'inscription de la base au catalogue du SNDS, le contrat qui lie le responsable de données et la PDS précisera la durée du primo-accès et les modalités d'application, et rappellera les exceptions.

Les principales caractéristiques cadrant la possibilité d'un primo-accès sur les données des bases du catalogue sont résumées en Annexe 4.

Liste des Annexes

- Annexe 1 - Benchmark international et des HDAB sur les primo-accès des données
- Annexe 2 - Template lettre d'intention pour inscrire sa base à l'arrêté ministériel "catalogue"
- Annexe 3 - Check-list du processus de sélection des bases du Catalogue du SNDS
- Annexe 4 - Principales caractéristiques cadrant la possibilité d'un primo-accès sur les données des bases du catalogue

Annexe 1 - Benchmark international et des HDAB sur les primo-accès des données

Benchmark international

Le tableau ci-dessous permet de comparer les pratiques de primo-accès concernant les principales structures d'accès internationales hors HDAB. Ce benchmark a été réalisé à partir des éléments recueillis lors d'entretiens réalisés auprès de ces structures en 2019 et 2024.

Structures	Pays	Périmètre des données concernés	Cadre du primo-accès	Durée du primo-accès
UK Biobank ⁽¹⁾	Royaume-Uni	Données issues de questionnaires et examens, données d'imagerie, données d'accéléromètre, données issues du séquençage de l'exome entier	Industriels du consortium travaillant sur la cohorte	9 mois
Clinical Practice Research Datalink (National Health Service) ⁽²⁾	Royaume-Uni	Données de soins primaires anglaises, potentiellement enrichies par des données hospitalières	Pas de primo-accès	n/a
SAIL Databank ⁽¹⁾	Pays de galles	Données de sources et de types multiples dont les soins primaires, les données hospitalières, des données de dépistage et des données démographiques	Pas de primo-accès	n/a
Vivly (données essais cliniques) ⁽²⁾	US	Données issues d'essais cliniques du monde entier (études observationnelles et interventionnelles)	Possibilité de demander un primo-accès d'une durée limitée pour les producteurs de données.	~12-18 mois
Health Verity ⁽¹⁾	US	Données administratives, dossiers patients, données hospitalières, données des pharmacies et laboratoires, etc.	Définition des règles par le producteur de données	n/a
Framingham Heart Study (FHS) ⁽²⁾	US	Données cliniques, génétiques et d'imagerie ainsi que des échantillons	Possibilité pour les producteurs de données de demander un primo-accès d'une durée limitée	3-12 mois

Insights (essais cliniques) ⁽²⁾	US	Données cliniques et génomiques ainsi que des échantillons	Partage des données après la publication des résultats de l'essai	n/a
Regenstrief ⁽²⁾	US	Données de soin	Pas de primoaccès	n/a
National Cancer Institute's Genomic Data Commons/The Cancer Genome Atlas (TCGA) ⁽²⁾	US	Données génétiques des cancers et des caractéristiques cliniques des patients ayant un cancer	Pas primoaccès ou durée très courte	non précisé

(1) Information recueillie en 2024 ; (2) Information recueillie en 2019

Benchmark des HDABs

Le tableau ci-dessous permet de comparer les pratiques de primoaccès concernant les principales structures d'accès internationales HDAB déjà établies ou en cours de constitution. Ce benchmark a été réalisé à partir des éléments recueillis lors d'entretiens réalisés auprès de ces structures en 2019 et 2024.

Structure	Pays	Périmètre des données concernés	Cadre du primoaccès	Durée du primoaccès
DHDA ⁽¹⁾	Danemark	Registres nationaux Données de la Biobanque danoise Données cliniques Données socio-économiques	Pas de primoaccès	n/a
NIPH ⁽²⁾	Norvège	Données des 17 registres nationaux (entre autres données hospitalières, données de naissances, registres de cancers, registres régionaux de qualité médicale) et des cohortes de recherche	Possibilité pour les chercheurs de demander un primoaccès d'une durée limitée mais ces demandes sont appréciées au cas par cas	~12 mois
Findata ⁽¹⁾	Finlande	Registres médicaux, bases médico-administratives, données socio-économiques (retraites, allocations, logement, famille etc.)	Pas de primoaccès	n/a
Croatian Institute of Public Health (CIPH) ⁽¹⁾	Croatie	Données de soins et issues d'enquêtes	Primoaccès avec condition de publication préliminaire	12 mois

Belgian Health Data ⁽²⁾	Belgique	Données de soins primaires, données hospitalières, de laboratoires et d'autres spécialités médicales	Possibilité pour les chercheurs de demander un primoaccès d'une durée limitée selon les règles en vigueur pour les données concernées	~12 mois
Health Data Lab ⁽¹⁾	Allemagne	Données de soin	Pas de primoaccès	n/a

(1) Information recueillie en 2024 ; (2) Information recueillie en 2019

Annexe 2 - Template Lettre d'intention pour inscrire sa base à l'arrêté ministériel "catalogue"

Expéditeur
[A compléter]
A l'attention de
Madame Stéphanie COMBES
Directrice du Health Data Hub (Plateforme de données de santé)
9, rue Georges Pitard 75015 Paris

[Lieux], le [date]

Objet : Lettre d'intention de partage de données au sein du catalogue du Health Data Hub

Affaire suivie par : [A compléter]

Madame la Directrice,

En tant que responsable de données, nous sommes conscients que nous avons la chance de disposer en France d'un patrimoine de données de santé probablement sans équivalent dans le monde. Ce patrimoine est le fruit d'un travail remarquable conduit par celles et ceux qui comme nous, au niveau de l'Assurance Maladie, des établissements de santé ou des instituts de recherche notamment, conçoivent les bases, collectent les données, les regroupent et œuvrent au quotidien pour garantir leur qualité et la pertinence scientifique des usages qui en sont faits.

Créé en fin 2019 après un an d'incubation au ministère de la santé, le Health Data Hub réunit les données de santé jugées prioritaires au sein de son catalogue et les met à disposition sous format déidentifié à des porteurs de projet d'intérêt public, principes que nous soutenons. Nous avons également pris acte que le catalogue du Health Data Hub est documenté et construit de manière progressive en partenariat avec les responsables de la collecte des données et porté par un arrêté pris après avis de la CNIL et décision du comité stratégique du Système National des Données de Santé qui en fixe sa composition.

En sa qualité [statut social de l'organisme], [nom du responsable de données] souhaite vous informer de son intention de proposer [nom de la base de données] en partage au sein du catalogue de données du Health Data Hub, et ce en accord avec son responsable scientifique [nom du responsable scientifique].

[Nom de la base de données] est [type de base : cohorte, registre, entrepôt de données...] comportant des données [cliniques, d'imagerie, biologiques, génomiques...]. Les données concernent environ [nombre de personnes] personnes depuis [année]. La couverture géographique de la base est [à compléter]. La base de données sera mise à jour tous les [nombre de mois] mois.

Ses thèmes portent sur [à compléter].

[Nom du responsable de données] voit dans la répllication de la base de données au sein de la plateforme du Health Data Hub une opportunité stratégique pour d'une part, améliorer la

qualité des soins et l'accompagnement des patients, en particulier si elle est chaînée au SNDS historique ou à une autre base ; d'autre part, donner à cette base la visibilité qu'elle mérite en accord avec le plan science ouverte porté par la Ministre de l'Enseignement Supérieur, de la Recherche et de l'Innovation, auquel il souscrit pleinement. *[Ce paragraphe peut être à compléter selon les spécificités de la base. Donnés à titre d'exemples : développer des collaborations scientifiques, proposer une base présentant une innovation pour la collecte et/ou le traitement et/ou l'utilisation finale des données par rapport à l'état actuel de l'art ; développer des projets de recherche nécessitant une forte puissance de calcul, une importante capacité de stockage ; proposer des données mise en qualité (contrôle qualité, format interopérable...), représentative à l'échelle régionale/nationale ; faciliter l'appariement entre différentes bases de données ; rendre plus visible et accessible la base pour des porteurs de projet (ex : start-up)...]*

[Nom du responsable de données] confirme par ailleurs qu'il a conscience que les modalités du partage sont portées par la convention de partenariat qui a été partagée et qui reprend notamment les principes de valorisation scientifique et économique, ainsi que de gouvernance d'accès.

[Nom du responsable de données] souhaite qu'un primo-accès soit mis en place sur sa base de données pendant une durée de [x mois] pour la raison suivante : [à compléter selon les spécificités de la base].

L'équipe du *[nom du responsable scientifique]* se tient à disposition de la direction afin de déterminer les modalités techniques, l'accompagnement du Health Data Hub et le calendrier de la mise à disposition des données.

Je vous prie de croire, Madame la Directrice, en l'expression de ma considération distinguée.

Prénom Nom, Fonction

Signature

Copies :

[A compléter]

Annexe 3 - Check-list des éléments clés abordés avec le responsable de données en amont du comité stratégique des données de santé

Notice de complétion

- Checklist visant à être complétée par le secrétariat du Comité stratégique des données de santé lors de la candidature d'une base de données à son inscription au catalogue de la Plateforme de données de santé
- Destinataires : les membres du comité stratégique des données de Santé

Informations générales

- Base : <XXX>
- Date de complétion : <JJ/MM/AAAA>

ACCORD DE PRINCIPE DES RESPONSABLES DE DONNEES

OUI NON

Le responsable de données² a donné son accord de principe pour la mise au catalogue de sa base de données : une lettre d'intention a été signée par les personnes habilitées à le faire.

--	--

PROTECTION DE LA VIE PRIVÉE, SECURITE ET CONFIDENTIALITE DES DONNEES

Note aux membres du GT catalogue : cette partie pourra être enrichie/adaptée au regard des retours finaux que la CNIL fera pour les premières bases du catalogue.

OUI NON

La base dispose d'une autorisation réglementaire à jour.

--	--

² Responsable de données : organisme qui est à l'origine de la collecte des données dans le cadre de ses propres activités et sous sa responsabilité.

La durée de conservation des données est compatible avec la mise au catalogue de la base de données

Le responsable des données est en capacité d'informer individuellement la personne concernée ou à défaut est en capacité d'argumenter une demande de dérogation à l'information individuelle.

Le responsable des données est en capacité technique de mettre en place un canal de transfert selon le mode nominal proposé par la Plateforme de données de santé ou à défaut un mode alternatif répondant aux contraintes de sécurité du référentiel SNDS.

Le responsable de données est en capacité d'exercer les droits des personnes concernées.

ENRICHISSEMENT DE LA BASE PAR DES DONNÉES DE LA BASE PRINCIPALE DU SNDS (critère non bloquant)

OUI NON

La base de données est chainable aux données de la base principale
Si oui,

- appariement via le NIR certifié
- appariement via une reconstitution du NIR
- appariement via des variables communes entre la base principale du SNDS et la base de données

MODALITÉS DE PARTAGE DES DONNÉES

OUI NON

Il est mentionné dans la lettre d'intention du responsable de données que :

- la convention de partenariat entre le responsable de données et la Plateforme des données de santé a été partagée et que le responsable de données est aligné sur ses principes ;
- les principes de valorisation scientifique ont été partagés entre les parties³ (collaborations scientifiques, règles de signatures, modalités de

³ Le HDH, et le Responsable de Données sont désignés collectivement par le terme « Parties ».

mise en place de collaborations entre le demandeur et le responsable de données) ;

- les principes de valorisation économique ont été partagés entre les parties ;
- les principes des modalités d'accès/gouvernance ont été discutés entre les parties.

Les sujets prioritaires du responsable de données définis dans le cadre des missions publiques propres aux organismes publics (exemple : indicateurs épidémiologiques pour les registres de morbidité) et la période de primoaccès permettant de réaliser ces sujets prioritaires ont été discutés entre les parties.

Annexe 4 - Principales caractéristiques cadrant la possibilité d'un primo-accès sur les bases de données du catalogue

Bases de données ou organismes concernés	Financements	Cas d'usage concernés par le primo-accès	Exceptions	Durée du primo-accès
- Cohorte	- Publics - Publics et privés	- Uniquement pour les cas d'usage ayant la même finalité que celle déposée par l'équipe ayant produit la base de données (évaluation au cas par cas réalisée par le responsable de données)	- Demande émanant de la puissance publique	Entre 6 et 12 mois
- Cohortes/Registres /Entrepôts de données	- Privés uniquement	- Tous les cas d'usage	- Demande émanant de la puissance publique - Demande émanant des organismes privés financeurs	Entre 6 et 12 mois
- Bases de données créées pour répondre à des missions de service public	- Publics	- Uniquement pour des cas d'usage répondant aux mêmes missions que celles du responsable de données concerné (évaluation au cas par cas réalisée par le responsable de données)	- Demande émanant de la puissance publique	Entre 6 et 12 mois