ANALYSE D'IMPACT RELATIVE A LA PROTECTION DES DONNEES DE LA PLATEFORME TECHNOLOGIQUE 1.0 DU HEALTH DATA HUB

Mise à jour : 21/10/2020

Version communicable générée le 2 février 2021

1. Contexte

1.1. Vue d'ensemble

1.1.1. Quel est le traitement qui fait l'objet de l'étude ?

A la suite du rapport Villani sur l'intelligence artificielle rendu public le 28 mars 2018, le Président de la République a affirmé sa volonté de faire de la santé un des secteurs prioritaires pour le développement de l'intelligence artificielle en France. Deux actions majeures ont été annoncées : l'élargissement du système national de données de santé (SNDS) et la création d'un « Health Data Hub » pour faciliter l'accès aux données de santé et leur valorisation.

La loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé a concrétisé cette volonté en transformant l'Institut national des données de santé (INDS) en « Plateforme des données de santé » ou "Health Data Hub" (ci-après « HDH ») qui reprend les missions de l'Institut tout en les élargissant. Sa convention constitutive a été publiée par un arrêté du 29 novembre 2019.

La préfiguration du HDH a été réalisée conjointement par l'INDS et la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) du Ministère des Solidarités et de la Santé. Pendant cette période, la Drees était chargée du développement de la plateforme technologique qui sera mise à disposition des utilisateurs de données de santé dans un cadre précis à distinguer de la mise à disposition des données du "catalogue" qui ne sera possible qu'après parution des textes consacrés et autorisation par la CNIL du traitement (cf infra). L'INDS était chargé de la conception d'une offre d'accompagnement des utilisateurs, à la fois sur les plans humain et réglementaire. La préfiguration du HDH a donc impliqué les équipes du Ministère des Solidarités et de la Santé et celles de l'INDS, notamment leurs délégués à la protection des données respectifs.

Depuis le 30 novembre 2019, le groupement d'intérêt public Health Data Hub est officiellement créé et peut désormais assumer la responsabilité du développement de la plateforme technologique et l'ensemble des missions qui lui sont dévolues à l'article L.1462-1 du code de la santé publique. Le HDH reprend les missions de l'INDS et demeure ainsi le secrétariat unique par lequel transitent toutes les demandes d'accès des porteurs de projet à des bases de données de santé, hors recherches impliquant la personne humaine, et continue à assurer l'accompagnement des utilisateurs et de contribuer à la & mise en place de procédures simplifiées en accord avec la CNIL. Les missions nouvelles du HDH justifiant la construction d'une plateforme technologique sont les suivantes : la mise à disposition des données du Système national de données de santé (SNDS) ; la promotion de l'innovation ; l'accompagnement de porteurs de projets sélectionnés dans le cadre d'appels à projets et des responsables de données associés aux projets retenus ; et la réalisation de traitements de données pour le compte de tiers.

Cependant, les modalités de mise en œuvre de plusieurs de ces missions sont dépendantes des textes d'application de la loi et nécessitent des travaux avec le responsable de traitement du SNDS "historique" - la CNAM - et tous les responsables de traitement des bases sources du SNDS "élargi". Le traitement du SNDS serait sous la responsabilité de deux responsables de traitement : la CNAM et le HDH et porterait principalement sur la base principale couvrant l'ensemble de la population et composée d'abord des données mentionnées aux 1° à 4° de l'article L.1461-1 du code de la santé publique (le « SNDS historique ») qui sera progressivement complétée des autres données mentionnées à l'article L.1461-1 du code de la santé publique. Par ailleurs, un « catalogue » de différentes bases de données non exhaustives mais jugées prioritaires pour faire avancer les connaissances en santé sera aussi constitué.

La crise sanitaire actuelle liée à la pandémie de COVID-19 est venue accélérer les travaux du HDH en le poussant à préfigurer le "catalogue" en réunissant des données relatives à l'épidémie pour faciliter la recherche à son sujet. En effet, le gouvernement a demandé à Madame Stéphanie Combes, directrice du HDH et Monsieur Emmanuel Bacry, directeur scientifique du HDH, de piloter une mission visant à favoriser l'utilisation des données de santé comme outil participant à l'éclairage et la résolution de la crise sanitaire. Il apparaît que les données nécessaires au suivi de l'épidémie et à la compréhension de la pathologie sont dispersées et que les multiples initiatives lancées récemment se heurtent au manque de disponibilité des données. Sur le périmètre restreint des données permettant à ces initiatives d'aboutir dans les plus brefs délais à des solutions concrètes de management de la crise, les infrastructures du HDH ont été identifiées comme réceptacle temporaire de ces données et à ces seules fins.

La participation du HDH à l'effort national contre l'épidémie a été consacrée par un arrêté du 21 avril 2020 complétant l'arrêté du 23 mars 2020 prescrivant les mesures d'organisation et de fonctionnement du système de santé nécessaires pour faire face à l'épidémie de covid-19 dans le cadre de l'état d'urgence sanitaire. A la fin de l'état d'urgence sanitaire, fixée le 10 juillet 2020, un nouvel arrêté est venu prolonger le cadre d'intervention exceptionnel du HDH jusqu'à l'entrée en vigueur des dispositions prises en application de l'article 41 de la loi du 24 juillet 2019 susvisée et au plus tard le 30 octobre 2020. Un arrêté du 16 octobre 2020 a de nouveau modifié les dispositions de l'article 30 de l'arrêté du 10 juillet, considérant qu'il était nécessaire de prolonger les missions temporaires du HDH en raison de l'évolution de l'épidémie. L'échéance du 30 octobre est tombée de sorte que le HDH peut continuer de mettre à disposition des données liées à la Covid-19 jusqu'à l'entrée en vigueur des dispositions prises en application de la loi du 24 juillet 2019.

Néanmoins le HDH avait anticipé d'éventuelles difficultés dans le basculement vers le cadre de droit commun avant le 30 octobre 2020 et a soumis le 3 septembre 2020 une demande d'autorisation à la CNIL sur le fondement de l'article 66 de la loi Informatique et Libertés afin de conserver un entrepôt de données liées à la Covid-19. Le périmètre de cette demande est limité aux données que le HDH a pu collecter en vertu du cadre exceptionnel lié à l'urgence sanitaire et ses suites mais aussi toutes celles à venir qui seront utiles pour développer les connaissances sur une épidémie toujours active.

Si l'échéance du 30 octobre 2020 a été repoussée, le cadre réglementaire dans lequel le HDH évolue reste temporaire donc l'intérêt de conserver un entrepôt de données liées à la Covid-19 après la crise sanitaire reste vif. La demande d'autorisation déposée par le HDH reste pertinente et son instruction par la CNIL se poursuit.

La version 1.0 de la plateforme technologique est celle qui sera utilisée pour héberger l'entrepôt de données liées à la Covid-19 mais aussi les données des projets pilotes qui devaient préfigurer l'offre de service du HDH en dehors de tout contexte d'urgence. Une demande d'autorisation auprès de la CNIL n'était pas envisagée pour cette version 1.0 puisque les projets pilotes n'impliquent pas d'entrepôt de données et sont menés sous la responsabilité de traitement des porteurs de projets. Les circonstances font que la plateforme technologique 1.0 hébergera aussi l'entrepôt de données liées à la Covid-19 donc les conditions de sécurité de la plateforme technologique 1.0 seront examinées par la Commission à l'occasion de la demande d'autorisation pour l'entrepôt Covid.

Ainsi, la présente AIPD concerne non seulement les traitements qui seront réalisés pour l'entrepôt Covid mais aussi ceux nécessaires pour la mise à disposition d'environnements de travail sécurisés pour les projets pilotes.

La version 1.0 de la plateforme technologique prendra la suite de la version Covid-19 utilisée actuellement pour héberger les prémices de l'entrepôt constitué pendant la crise sanitaire. L'évolution principale entre la plateforme Covid-19 et la plateforme 1.0 réside dans l'automatisation de plusieurs processus, le renforcement de la sécurité, la possibilité de gérer des appariements de données et la mise en place d'une matrice d'exclusion qui préfigure ce que le HDH souhaite faire sur la version 1.1 de la plateforme pour faciliter l'exercice des droits des personnes.

Afin de sécuriser les relations avec les différentes parties prenantes, que ce soit dans le cadre de l'entrepôt Covid ou des projets pilotes, le HDH signe des conventions, en amont, avec les sources des données et, en aval, avec les utilisateurs des données. Ainsi, une convention de transfert de données doit être signée entre le HDH et chaque responsable de traitement des données sources afin d'encadrer spécifiquement cette remontée des données. Il s'agit notamment d'en préciser le périmètre, la fréquence de mise à jour et la sécurisation du transfert. Du côté des utilisateurs, le HDH doit s'assurer que tout responsable de traitement d'un projet d'utilisation des données est dûment autorisé à solliciter l'accès et une convention de mise à disposition des données doit être conclue. En outre, chaque utilisateur habilité doit s'engager individuellement à respecter les conditions générales d'utilisation de la plateforme technologique 1.0 avant d'y accéder. L'utilisateur projet ne pourra traiter que le périmètre des données visé par son autorisation CNIL préparé par l'opérateur et mis à disposition au sein de l'espace projet qui lui est réservé.

Il est à noter que côté utilisateurs chaque responsable de traitement doit réaliser également une démarche d'homologation couvrant l'espace projet qui lui est confié. Cette démarche doit inclure une AIPD, sur le périmètre de son projet, décrivant précisément les données qu'il souhaite utiliser et les traitements envisagés. Elle conduit chaque responsable de traitement à mettre en place les mesures de sécurité, de gouvernance, d'organisation et des mesures techniques appropriées à son projet, en complément des mesures déjà prévues par le HDH.

Afin d'atteindre l'objectif de mise en œuvre de l'entrepôt de données de santé liées à la COVID-19 et des projets pilotes sur la plateforme technologique 1.0, objets de cette AIPD, plusieurs traitements de données à caractère personnel sont impliqués dont la responsabilité incombe au HDH :

- la collecte et la conservation de données de santé pseudonymisées liées à la COVID-19 ;
- la mise à disposition des données de santé relatives à la COVID-19 à des projets (incluant de possibles appariements de données) par le biais de la plateforme technologique 1.0 ;
- la mise à disposition de données de santé pseudonymisées aux projets pilotes ;
- la gestion des comptes ayant des droits sur la plateforme technologique 1.0 ;
- la traçabilité de l'activité des comptes ayant des droits sur la plateforme technologique 1.0.

Concernant la mise à disposition de données de santé, que ce soit pour l'entrepôt Covid ou pour les projets pilotes, la finalité est de préparer et mettre à la disposition des projets, dans leur espace projet, des jeux de données de santé pseudonymisées issues de différentes sources afin qu'elles soient utilisées de manière adaptée aux besoins du projet et conformément au cadre réglementaire d'accès aux données personnelles de santé. Les données de santé pseudonymisées seront traitées sur la plateforme technologique 1.0 :

- d'une part, par les utilisateurs habilités pour chaque projet autorisé (dit « utilisateurs projet externes », qui seront des chercheurs, des employés de start-ups, etc.) ;
- d'autre part, par des agents du HDH :
    - Les "opérateurs données", qui préparent les données pour les projets, valident leur contenu, réalisent des audits sur les données ingérées / exportées par les utilisateurs projet mais qui n'ont pas les droits pour faire transiter les données d'un espace à un autre ;
    - Les "opérateurs projet", qui préparent les espaces projets, font transiter les données entre les espaces, attribuent les droits et accès aux comptes des utilisateurs projet mais qui ne peuvent pas réaliser de traitement sur les données.

Il est à noter que les traitements de données de santé pseudonymisées réalisés au sein d'un espace projet relèvent de la responsabilité de traitement du porteur du projet et sont donc décrits dans l'AIPD réalisée dans le cadre de la démarche d'homologation de ce projet.

Concernant la gestion des comptes, la finalité est de permettre la gestion des accès à la plateforme technologique 1.0 (authentification, gestion des droits, etc.). Des données à caractère personnel sont nécessaires pour la création du compte et la gestion des accès et des identités sur la plateforme technologique 1.0. Elles ne sont accessibles et utilisées que par les agents du HDH chargés de cette tâche :

- Les "opérateurs plateforme", qui créent les comptes ;
- Le superopérateur, qui attribue les droits aux comptes des opérateurs ;
- Les opérateurs projet qui attribuent les droits aux comptes des utilisateurs projet et ferment les comptes des utilisateurs selon la procédure de gestion des identités et des habilitations ;

Concernant la traçabilité de l'activité des comptes, la finalité principale est de suivre *a posteriori* les actions des comptes sur la plateforme technologique dans le cadre d'une démarche de sécurité. Les traces sont consultées et exploitées par les opérateurs sécurité du HDH. Elles sont stockées au sein d'un puits de traces et envoyées à un tiers-archiveur de confiance pour les sceller et garantir leur force probante.

Plus précisément, les traces sont consultées et utilisées dans les cas de figure suivants :

- Contrôle automatique par des algorithmes pour détecter des comportements anormaux ;
- Contrôle aléatoire effectué par des opérateurs sécurité dans le cadre de leur tâche de maintien en condition de sécurité de la plateforme technologique ;
- Recherches spécifiques pour une investigation en cas d'incident sur la plateforme technologique 1.0 ;
- Élaboration de statistiques d'utilisation de la plateforme technologique 1.0 dans le cadre d'une démarche d'amélioration continue. Ceci est considéré comme une mesure de sécurité car l'objectif est de maintenir et adapter les mesures de sécurité en fonction de l'utilisation qui en est faite.

La plateforme technologique 1.0 repose sur les offres de Microsoft Azure, une solution d'hébergement dans le Cloud disposant de la certification « Hébergeur de données de santé » https://esante.gouv.fr/labels-certifications/hds/liste-des-herbergeurs-certifies) :

- Une offre « Infrastructure as a Service », qui met à disposition une virtualisation de l'infrastructure physique permettant de stocker les données et applications, de fournir de la puissance de calcul et de mettre en réseau la plateforme technologique 1.0.
- Une offre « Platform as a Service », qui propose un ensemble de services « logiciels » (tels que des bases de données, des outils d'analyse de traces, etc.) entièrement managés par Microsoft pouvant être intégrés dans la plateforme technologique 1.0.

1.1.2. Quelles sont les responsabilités liées aux traitements

Pour la suite de l'AIPD, les différents traitements de données qui relèvent de la responsabilité du HDH seront désignés de la manière suivante :

Pour l'entrepôt Covid :

- traitement 1 : Collecte et conservation de données de santé pseudonymisées dans l'entrepôt COVID
- traitement 2 : Mise à disposition de données de santé de l'entrepôt COVID

Pour les projets pilotes :

- traitement 3 : Mise à disposition de données de santé pseudonymisées pour les projets pilotes

Pour la gestion des utilisateurs :

- traitement 4 : Gestion des comptes des utilisateurs de la plateforme technologique 1.0
- traitement 5 : Collecte de la trace des activités des comptes utilisateur

Traitement 1- Collecte et conservation de données de santé pseudonymisées dans l'entrepôt COVID:

Responsable de traitement : HDH

Sous-traitants : Microsoft

Destinataires : Les opérateurs sont destinataires des données de santé pseudonymisées relatives à l'épidémie afin d'assurer la réception et la conservation des données dans des conditions opérationnelles et de sécurité.

Traitement 2 - Mise à disposition de données de santé de l'entrepôt COVID :

Responsable de traitement : HDH

Sous-traitants : Microsoft

Destinataires :

- les utilisateurs projets sont destinataires des données de santé pseudonymisées afin de conduire les projets relatifs à l'épidémie ;
- les opérateurs sont destinataires des données de santé pseudonymisées pour la préparation des données aux projets.

Traitement 3 - Mise à disposition de données de santé pseudonymisées pour les projets pilotes :

Responsable de traitement : HDH

Sous-traitants : Microsoft

Destinataires :

- les utilisateurs projets sont destinataires des données de santé pseudonymisées afin de conduire les projets pilotes ;
- les opérateurs sont destinataires des données de santé pseudonymisées pour la préparation des données aux

projets

Traitement 4 - Gestion des comptes des utilisateurs de la plateforme technologique 1.0 :

Responsable de traitement : HDH

Sous-traitants : Microsoft

Destinataires : Les opérateurs sont destinataires des données des utilisateurs afin de maintenir la plateforme technologique 1.0 en condition opérationnelle et de sécurité.

Traitement 5 - Collecte de la trace des activités des comptes utilisateur :

Responsable de traitement : HDH

Sous-traitants : Microsoft, Tiers archiveur de confiance

Destinataires : Les opérateurs sont destinataires des données d'activité afin de maintenir la plateforme technologique 1.0 en condition opérationnelle et de sécurité.

1.1.3. Quels sont les référentiels applicables ?

Les référentiels applicables au traitement sont :

- Politique de sécurité des systèmes d'information de l'Etat du 17 juillet 2014 ;
- Arrêté du 1er octobre 2015 portant approbation de la politique de sécurité des systèmes d'information pour les ministères chargés des affaires sociales (PSSI MCAS) ;
- Référentiel général de sécurité V2 du 13 juin 2014 ;
- Arrêté du 22 mars 2017 relatif au référentiel de sécurité du SNDS (qui renvoie au RGS, à la PGSSI-Santé et à la PSSI-MCAS) ;
- Loi n°2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé.

1.2. Données, processus et supports

1.2.1. Quelles sont les données traitées ? Description des données, destinataires et durées de conservation

Traitement 1- Collecte et conservation de données de santé pseudonymisées dans l'entrepôt COVID :

- **Données :** Les données concernées sont les données de santé pseudonymisées liées à la COVID-19 dans la continuité de ce qui était autorisé dans le cadre des arrêtés du 21 avril puis du 10 juillet 2020. Il s'agit :
  - des données issues du système national des données de santé mentionné à l'article L. 1461-1 du même code ;
  - des données de pharmacie ;
  - des données de prise en charge en ville telles que des diagnostics ou des données déclaratives de symptômes issues d'applications mobiles de santé et d'outils de télésuivi, télésurveillance ou télémédecine ;
  - des résultats d'examens biologiques réalisés dans les laboratoires hospitaliers et les laboratoires de biologie médicale de ville ;
  - des données relatives aux urgences collectées par l'Agence nationale de santé publique dans le cadre du réseau de surveillance coordonnée des urgences ;
  - des données relatives aux appels recueillis au niveau des SAMU ;
  - des données relatives à l'activité et à la consommation de soins dans les établissements ou services médico-sociaux et dans les EHPAD ;
  - des enquêtes réalisées auprès des personnes pour évaluer leur vécu ;
  - des données cliniques telles que imagerie, pharmacie, biologie, virologie, compte-rendu médicaux de cohortes de patients pris en charge dans des centres de santé en vue de leur agrégation.

Les pseudonymes fournis par le responsable de données ne sont pas conservés et les données sont associées aux identifiants suivants : Partie non communicable en raison du secret lié à la sécurité des systèmes d'information.

Les mécanismes de gestion des pseudonymes et des identifiants sont décrits au paragraphe "3.1.1.9 Pseudonymisation".

- **Destinataires :** Les destinataires de ces données sont les opérateurs appelés « opérateurs projet » qui déplacent les jeux de données d'un espace de la plateforme à un autre et les opérateurs appelés « opérateurs données » qui les préparent pour un projet.

- **Durée de conservation :** Les données sont conservées sur la plateforme technologique 1.0 en plusieurs endroits distincts : au sein de l'espace opérateur et au sein des espaces projet. Au sein de l'espace opérateur, les données sont conservées pendant une durée de 10 ans dans la mesure où, selon toute vraisemblance, la Covid-19 est destinée à rester active encore plusieurs années.
  Les pseudonymes fournis par les responsables de données ne sont pas conservés sur la plateforme technologique1.0 .

Traitement 2- Mise à disposition de données de santé de l'entrepôt COVID :

- **Données** : Les données sont mises à disposition au sein d'un espace projet, associées à un identifiant projet

- **Destinataires** : Les destinataires de ces données sont les opérateurs appelés « opérateurs projet » qui déplacent les jeux de données d'un espace de la plateforme à un autre, les opérateurs appelés « opérateurs données » qui les préparent pour un projet, et les « utilisateurs projet » qui traitent les données pour mener leurs travaux.

- **Durée de conservation** : Au sein d'un espace projet, la durée de conservation est fondée sur le temps nécessaire à la conduite du projet. Cette durée de conservation diffère donc d'un projet à un autre, elle est déterminée par le responsable de traitement du projet et fait partie des éléments indiqués dans le dossier de demande d'autorisation.

Traitement 3 - Mise à disposition de données de santé pseudonymisées pour les projets pilotes :

- **Données** : Les données concernées sont les données de santé pseudonymisées nécessaires à la réalisation de projets soumis à une procédure d'accès aux données sous contrôle de la CNIL. Il s'agit notamment de données issues du SNDS. Un pseudonyme généré aléatoirement est attribué au sein de l'espace opérateur par le HDH pour remplacer les identifiants fournis par les responsables de données fournissant des jeux de données. Étant entendu que les responsables de données ne fournissent pas au HDH des données directement nominatives.

- **Destinataires :** Les destinataires de ces données sont les opérateurs appelés « opérateurs projet » qui déplacent les jeux de données d'un espace de la plateforme à un autre, les opérateurs appelés « opérateurs données » qui les préparent pour un projet, et les « utilisateurs projet » qui traitent les données pour mener leurs travaux.

- **Durée de conservation** : Les données sont conservées sur la plateforme technologique 1.0 en deux endroits distincts : au sein de l'espace projet et au sein de l'espace opérateur.

  - ❖ Au sein de l'espace projet, la durée de conservation est fondée sur le temps nécessaire à la conduite du projet. Cette durée de conservation diffère donc d'un projet à un autre, et elle est déterminée par le responsable de traitement du projet et fait partie des éléments indiqués dans le dossier de demande d'autorisation.
  - ❖ Au sein de l'espace opérateur, les données sont conservées à des fins fonctionnelles pendant une durée équivalente à celle de l'espace projet correspondant. En outre, à des fins de sécurité et pour la durée de conservation est fondée sur celle des traces systèmes et de sécurité, soit 12 mois glissants, afin de laisser la possibilité au HDH de mener des contrôles a posteriori sur toutes les opérations préalables à la mise à disposition des données au sein des espaces projet, les données sont conservées sur une période de 12 mois glissants comme les traces systèmes et de sécurité.

  La correspondance entre l'identifiant attribué au sein de l'espace opérateur par le HDH et les pseudonymes utilisés en dehors de la plateforme technologique 1.0 n'est pas conservée sauf pour les projets nécessitant une mise à jour des extractions de données, pour la durée strictement nécessaire à la mise à jour.

Traitement 4 - Gestion des comptes des utilisateurs de la plateforme technologique 1.0 :

- **Données** : Les données concernées sont les données à caractère personnel des utilisateurs de la plateforme technologique 1.0, à savoir leur nom, prénom, données de contact professionnelles (mail, adresse postale, numéro de téléphone), fonction, organisme de rattachement et identifiant du générateur de jeton logiciel.

- **Destinataires** : Les destinataires de ces données sont les opérateurs de la plateforme technologique 1.0 du HDH pour la gestion des accès et des privilèges. Plus particulièrement, les opérateurs appelés « opérateurs plateforme » créent les comptes, les opérateurs appelés « opérateurs projet » attribuent les droits des utilisateurs projet tandis que le superopérateur attribue les droits des opérateurs (opérateurs plateforme, opérateurs sécurité, opérateurs projet, opérateurs données). Les personnes intervenant dans le cadre d'audits et les autorités d'enregistrement (i.e. la personne désignée pour habiliter des personnes de son organisme à accéder à la plateforme technologique 1.0) peuvent également, sur demande spécifique et légitime de leur part, être amenés à traiter les données à caractère personnel des utilisateurs de la plateforme technologique 1.0.

- **Durée de conservation** : Ces données seront conservées pendant toute la durée d'existence du compte utilisateur (i.e. pendant la durée d'utilisation puis pendant la durée de mise en quarantaine du compte, cette dernière durant douze mois, afin d'assurer la cohérence des traces collectées en cas de contentieux) puis sont placées en archives intermédiaires pendant 5 ans. Une revue trimestrielle des comptes des utilisateurs projet sera réalisée par les opérateurs projet, en collaboration avec les autorités d'enregistrement, et une revue trimestrielle des comptes des opérateurs sera réalisée par le superopérateur.

Traitement 5 - Collecte de la trace des activités des comptes utilisateur :

- **Données** : Les données concernées sont les traces de l'activité des utilisateurs sur la plateforme technologique 1.0, à savoir les traces système, les traces de sécurité et les traces applicatives liées à la connexion sur la plateforme (e.g.

date, IP), à l'utilisation des services de la plateforme (e.g., nom de compte, nom de la ressource, requêtes, résultats renvoyés, messages d'erreur) et à la déconnexion (e.g., nom de compte, date).

- **Destinataires** : Les destinataires de ces données sont les opérateurs sécurité pour le maintien en condition de sécurité, les opérateurs plateforme pour le maintien en condition opérationnelle, le tiers archiveur pour le scellement des traces, les personnes intervenant dans le cadre des audits sur demande spécifique et légitime de leur part, et les autorités d'enregistrement sur leur demande, pour les utilisateurs de leur organisme.

- **Durée de conservation** : Les traces systèmes et de sécurité sont conservées sur une période de douze mois glissants, les traces applicatives sont conservées sur une période de vingt-quatre mois glissants.

1.2.2. Quel est le cycle de vie des données ?

Partie non communicable en raison du secret lié à la sécurité des systèmes d'information.

1.2.3. Quels sont les supports de données ?

Traitement 1- Collecte et conservation de données de santé pseudonymisées dans l'entrepôt COVID :

Les données de santé pseudonymisées reçues sont stockées uniquement dans la plateforme technologique 1.0. Cette plateforme est hébergée dans les centres de données de Microsoft Azure situés en France dans la région "France central" (région parisienne).

L'ensemble des flux de données sont réalisés sur des canaux sécurisés : sur un réseau privé virtuel entre le responsable de données et la plateforme technologique 1.0 pour l'ingestion des données, sur le « backbone » (réseau interne de Microsoft, coupé d'Internet) de Microsoft pour la circulation entre les différents espaces de la plateforme technologique 1.0.

Traitement 2- Mise à disposition de données de santé de l'entrepôt COVID

Les données de santé pseudonymisées reçues sont stockées uniquement dans la plateforme technologique 1.0. Cette plateforme est hébergée dans les centres de données de Microsoft Azure situés en France dans la région "France central" (région parisienne).

L'ensemble des flux de données sont réalisés sur des canaux sécurisés : sur un réseau privé virtuel entre le responsable de données et la plateforme technologique 1.0 pour l'ingestion des données, sur le « backbone » (réseau interne de Microsoft, coupé d'Internet) de Microsoft pour la circulation entre les différents espaces de la plateforme technologique 1.0 .

Traitement 3 - Mise à disposition de données de santé pseudonymisées pour les projets pilotes :

Les données de santé pseudonymisées reçues sont stockées uniquement dans la plateforme technologique 1.0. Cette plateforme est hébergée dans les centres de données de Microsoft Azure situés en France dans la région "France central" (région parisienne).

L'ensemble des flux de données sont réalisés sur des canaux sécurisés : sur un réseau privé virtuel entre le responsable de données et la plateforme technologique 1.0 pour l'ingestion des données, sur le « backbone » (réseau interne de Microsoft, coupé d'Internet) de Microsoft pour la circulation entre les différents espaces de la plateforme technologique 1.0.

Traitement 4 - Gestion des comptes des utilisateurs

Comme pour les données de santé pseudonymisées, les données personnelles des utilisateurs de la plateforme sont stockées sur la plateforme technologique 1.0.

Traitement 5 - Collecte de la trace des activités des utilisateurs

Les traces sont stockées à la fois dans le puits de traces de la plateforme technologique 1.0 et sur les serveurs du tiers archiveur de confiance, localisés en France. La transmission des traces scellées se fait de manière journalière via l'interface HTTPS de la solution, avec l'émission d'un conteneur chiffré.

Contexte	Acceptable / A corriger ?	Mesures correctives
Vue d'ensemble	Acceptable	
Données, processus et support	Acceptable	

2. Principes fondamentaux

2.1. Mesures garantissant la proportionnalité et la nécessité du traitement

2.1.1. Finalités et fondements

Le HDH est responsable de l'opération de la plateforme technologique 1.0 qui repose sur trois fonctionnalités impliquant un traitement de données à caractère personnel :

- L'exposition de jeux de données conformes au besoin et au cadre réglementaire pour les mettre à disposition des équipes projet dans leur espace projet ;
- La gestion des droits et des accès des comptes des utilisateurs de la plateforme technologique 1.0 ;
- La traçabilité de l'activité des utilisateurs de la plateforme technologique 1.0.

Dans le cadre de l'entrepôt Covid-19, à la différence des projets pilotes, le HDH est aussi responsable de traitement de la collecte et de la conservation des données liées à la COVID-19.

Le HDH a été créé par la loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé et ses missions sont définies à l'article L.1462-1 du code de la santé publique. Elles comprennent en particulier :

- l'accompagnement de porteurs de projets et des responsables de données associés aux projets ;
- la réalisation pour le compte de tiers des opérations nécessaires à un traitement de données issues du SNDS pour lequel ce tiers a obtenu une autorisation ;
- la réunion, l'organisation et la mise à disposition des données du SNDS
- ainsi que la promotion de l'innovation dans l'utilisation des données de santé.

En outre, l'article L. 1461-1 du même code impose que l'accès aux données du SNDS s'effectue dans des conditions assurant « la confidentialité et l'intégrité des données et la traçabilité des accès et des autres traitements ».

Ces dispositions de droit français (code de la santé publique) instituent une obligation pour le HDH de traiter des données à caractère personnel relatives aux citoyens, d'une part, et aux utilisateurs de la plateforme technologique 1.0, d'autre part. Les finalités du traitement sont clairement définies et il ne fait pas de doute que c'est le HDH qui est visé par ces dispositions.

En outre, au début de la crise sanitaire liée à la Covid-19, le gouvernement a demandé à Madame Stéphanie Combes, directrice du HDH et Monsieur Emmanuel Bacry, directeur scientifique du HDH , de piloter une mission visant à favoriser l'utilisation des données de santé comme outil participant à l'éclairage et la résolution de la crise sanitaire. Il apparaissait en effet que les données nécessaires au suivi de l'épidémie et à la compréhension de la pathologie étaient dispersées et que de multiples initiatives se heurtaient au manque de disponibilité des données. A la fin de l'état d'urgence sanitaire, la Direction générale de la santé a renouvelé son souhait que le HDH continue à favoriser les projets de recherche sur la Covid-19 et à collecter des données utiles pour mieux connaître la maladie et gérer les suites de la crise sanitaire.

Compte tenu de ses missions et de la volonté des pouvoirs publics de favoriser la recherche sur l'épidémie, le HDH a rassemblé les données utiles pour la lutte contre la COVID-19 et souhaite continuer à en collecter et à en donner l'accès à tous les utilisateurs autorisés, soit parce qu'ils interviennent dans la gestion des suites de la crise sanitaire, soit à des fins de recherche dans l'objectif d'améliorer les connaissances sur la COVID-19.

Les traitements de données à caractère personnel couverts par la présente AIPD relèvent d'une obligation légale en vertu des articles L.1461-1 et L.1462-1 du code de la santé publique pour ce qui concerne l'opération de la plateforme technologique 1.0 . Ils relèvent d'une mission d'intérêt public pour ce qui concerne la constitution d'un entrepôt de données liées à la COVID-19.

Le fondement juridique du traitement des données au sens du RGPD est donc l'article 6-1-c (« le traitement est nécessaire au respect d'une obligation légale à laquelle le responsable du traitement est soumis ») et, s'agissant particulièrement des données de santé, l'exception de l'article 9-2-j est mobilisée (« le traitement est nécessaire à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques, conformément à l'article 89, paragraphe 1, sur la base du droit de l'Union ou du droit d'un État membre qui doit être proportionné à l'objectif poursuivi, respecter l'essence du droit à la protection des données et prévoir des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée »).

Acceptable / Améliorable / A corriger	Commentaires d'évaluation	Mesures correctives
Acceptable		

2.1.2. Les données collectées sont-elles adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées (minimisation des données) ?

Traitement 1- Collecte et conservation de données de santé pseudonymisées dans l'entrepôt COVID :

Les données concernées sont les données de santé pseudonymisées liées à la COVID-19 dans la continuité de ce qui était autorisé dans le cadre des arrêtés du 21 avril puis du 10 juillet 2020. Il s'agit :

- des données issues du système national des données de santé mentionné à l'article L. 1461-1 du même code ;
- des données de pharmacie ;
- des données de prise en charge en ville telles que des diagnostics ou des données déclaratives de symptômes issues d'applications mobiles de santé et d'outils de télésuivi, télésurveillance ou télémédecine ;
- des résultats d'examens biologiques réalisés dans les laboratoires hospitaliers et les laboratoires de biologie médicale de ville ;
- des données relatives aux urgences collectées par l'Agence nationale de santé publique dans le cadre du réseau de surveillance coordonnée des urgences ;
- des données relatives aux appels recueillis au niveau des SAMU ;
- des données relatives à l'activité et à la consommation de soins dans les établissements ou services médico-sociaux et dans les EHPAD ;
- des enquêtes réalisées auprès des personnes pour évaluer leur vécu ;
- des données cliniques telles que imagerie, pharmacie, biologie, virologie, compte-rendu médicaux de cohortes de patients pris en charge dans des centres de santé en vue de leur agrégation.

Une convention de transfert de données est signée entre le HDH et chaque responsable de traitement des données sources afin d'encadrer spécifiquement cette remontée des données. Il s'agira notamment d'en préciser le périmètre, la fréquence de mise à jour et la sécurisation du transfert.

Traitement 2 - Mise à disposition de données de santé de l'entrepôt COVID

Quel que soit le projet, les formalités applicables doivent être accomplies et la pertinence des données est vérifiée à cette occasion. La plateforme technologique 1.0 n'héberge que les données ainsi validées lors d'un processus d'accès réglementaire.

De surcroît, une équipe projet n'a accès qu'aux données strictement nécessaires à son projet, elle n'a en aucun cas accès aux données stockées dans la plateforme technologique 1.0 pour d'autres projets. Ceci est assuré techniquement par l'architecture de la plateforme technologique 1.0 et opérationnellement par les opérateurs.

Traitement 3 - Mise à disposition de données de santé pseudonymisées pour les projets pilotes

Quel que soit le projet, les formalités applicables doivent être accomplies et la pertinence des données est vérifiée à cette occasion. La plateforme technologique 1.0 n'héberge que les données ainsi validées lors d'un processus d'accès réglementaire.

De surcroît, une équipe projet n'a accès qu'aux données strictement nécessaires à son projet, elle n'a en aucun cas accès aux données stockées dans la plateforme technologique 1.0 pour d'autres projets. Ceci est assuré techniquement par l'architecture de la plateforme technologique 1.0 et opérationnellement par les opérateurs.

Traitement 4 - Gestion des comptes des utilisateurs

En vertu de sa mission de mise à disposition des données prévue par l'article L. 1462-1 du code de la santé publique et afin de respecter l'obligation légale posée par l'article L. 1461-1 du même code selon laquelle l'accès aux données du SNDS s'effectue dans des conditions assurant « *la confidentialité et l'intégrité des données et la traçabilité des accès et des autres traitements* », le HDH doit traiter les données à caractère personnel relatives aux utilisateurs.

Les données collectées pour la gestion des comptes sont circonscrites au strict nécessaire tel que décrit dans la réponse à la question 1.2.1 Quelles sont les données traitées?

Traitement 5 - Collecte de la trace des activités des utilisateurs

Les traces collectées sont nécessaires au maintien en condition de sécurité de la plateforme technologique 1.0.

Les traces système, les traces de sécurité et les traces applicatives sont liées à la connexion sur la plateforme (e.g., date, IP), à l'utilisation des services de la plateforme (e.g., nom de compte, nom de la ressource, requêtes, résultats renvoyés, messages d'erreur) et à la déconnexion (e.g., nom de compte, date).

Il est prévu que ces données permettent également l'élaboration de statistiques dans une démarche d'amélioration continue.

Acceptable / Améliorable / A corriger	**Commentaires d'évaluation**	**Mesures correctives**
Acceptable	- La pertinence des données est déterminée en amont par le responsable de traitement du projet et vérifiée lors de la procédure d'accès aux données sous le contrôle de la CNIL - La grande diversité et l'important volume de traces collectées sont justifiés par les textes applicables.	

2.1.3. Les données sont-elles exactes et tenues à jour ?

Traitement 1- Collecte et conservation de données de santé pseudonymisées dans l'entrepôt COVID

La convention de transfert de données signée entre le HDH et chaque responsable de données doit notamment préciser le périmètre, la fréquence de mise à jour et la sécurisation du transfert des données. La qualité des données relève de la responsabilité des responsables de données tels que la CNAM, des établissements de santé ou des laboratoires de biologie médicale.

Traitement 2 - Mise à disposition de données de santé de l'entrepôt COVID

S'agissant des mêmes données de santé que pour le traitement 1, leur qualité et leur fréquence de mise à jour relèvent du même mécanisme qui s'appuie sur la responsabilité des responsables de données.

Traitement 3- Mise à disposition de données de santé pseudonymisées pour les projets pilotes

La convention de transfert de données signée entre le HDH et chaque responsable de données doit notamment préciser le périmètre, la fréquence de mise à jour et la sécurisation du transfert des données. La qualité des données de santé relève de la responsabilité des responsables de données tels que la CNAM, des hôpitaux ou des entreprises privées.

Traitement 4 - Gestion des comptes des utilisateurs

Chaque personne qui a la responsabilité de désigner les utilisateurs au sein d'un projet (l'autorité d'enregistrement) doit fournir à l'opérateur projet qui lui est attribué une liste d'utilisateurs habilités pour son projet. Pour chacun des utilisateurs habilités l'opérateur projet s'assure de la complétude des informations nécessaires à la création du compte telles que décrites dans la réponse à la question 1.2.1 Quelles sont les données traitées ? Par l'intermédiaire de l'autorité d'enregistrement, les utilisateurs ont ensuite le droit et le devoir de tenir à jour ou de demander toute modification de leurs données à caractère personnel.

Traitement 5 - Collecte de la trace des activités des utilisateurs

Les traces sont centralisées sur la plateforme technologique 1.0 dans un puits de traces et envoyées de manière journalière par conteneur sécurisé à un tiers archiveur de confiance puis scellées afin de garantir leur intégrité. Le puits de trace et le scellement ne permettent aucune modification des traces.

Acceptable / Améliorable / A corriger	Commentaires d'évaluation	Mesures correctives
Acceptable		

2.1.4. Quelle est la durée de conservation des données ?

Traitement 1 - Collecte et conservation de données de santé pseudonymisées de l'entrepôt COVID :

Les données sont conservées sur la plateforme technologique 1.0 au sein de l'espace opérateur pendant une durée de 10 ans.

Les pseudonymes fournis par les responsables de données ne sont pas conservés sur la plateforme technologique 1.0.

Traitement 2 - Mise à disposition de données de santé de l'entrepôt COVID

Au sein de l'espace projet, la durée de conservation est fondée sur le temps nécessaire à la conduite du projet. Cette durée de conservation diffère donc d'un projet à un autre, elle est déterminée par le responsable de traitement du projet et fait partie des éléments indiqués dans le dossier de demande d'autorisation.

Traitement 3 - Mise à disposition de données de santé pseudonymisées pour les projets pilotes

Les données sont conservées sur la plateforme technologique 1.0 en deux endroits distincts : au sein de l'espace projet et au sein de l'espace opérateur.

- Au sein de l'espace projet, la durée de conservation est fondée sur le temps nécessaire à la conduite du projet. Cette durée de conservation diffère donc d'un projet à un autre, elle est déterminée par le responsable de traitement du projet et fait partie des éléments indiqués dans le dossier de demande d'autorisation.
- Au sein de l'espace opérateur, les données sont conservées à des fins fonctionnelles pendant une durée équivalente à celle de l'espace projet correspondant.

La correspondance entre l'identifiant attribué au sein de l'espace opérateur et les pseudonymes utilisés en dehors de la plateforme technologique 1.0 n'est pas conservée sauf pour les projets nécessitant une mise à jour des extractions de données, pour la durée strictement nécessaire à la mise à jour.

Traitement 4 - Gestion des comptes des utilisateurs

Les données nécessaires à la gestion des comptes utilisateurs sont conservées durant toute la durée d'utilisation du compte par son propriétaire puis en base active pendant douze mois après la mise en quarantaine du compte associé et cinq ans en archive intermédiaire.

Les comptes, centralisant les données personnelles des utilisateurs de la plateforme, doivent être conservés pour assurer la cohérence des traces qu'ils ont générées. La durée de conservation de ces informations en base active est donc la même que la durée de conservation maximale des traces associées.

Les archives intermédiaires présentent un intérêt administratif, notamment en cas de contentieux, justifiant ainsi de les conserver.

Les opérateurs plateforme sont chargés de la suppression des comptes des utilisateurs de la plateforme à la fin de la durée de mise en quarantaine, le superopérateur est chargé de la suppression des comptes des opérateurs de la plateforme à la fin de la durée de mise en quarantaine.

Traitement 5 - Collecte de la trace des activités des utilisateurs

Les données sont stockées au sein d'un puits de traces, et conservées chez un tiers archiveur de confiance sur une période de douze mois glissants pour les traces liées aux événements systèmes et événements de sécurité et sur une période de vingt-quatre mois glissants pour les traces applicatives.

Ces données sont utilisées à des fins de sécurité de la plateforme technologique 1.0 , que ce soit pour détecter des incidents et mener des investigations *a posteriori*, ou détecter d'éventuelles anomalies et améliorer de manière continue les mesures de sécurité.

Les données sont supprimées de la plateforme technologique 1.0 et des zones de stockage du tiers archiveur de confiance à la fin de leur durée de conservation.

Acceptable / Améliorable / A corriger	Commentaires d'évaluation	Mesures correctives
Acceptable		

2.2. Mesures protectrices des droits des personnes concernées

2.2.1. Comment les personnes concernées sont-elles informées à propos du traitement ?

La présente AIPD porte non seulement sur la collecte et la conservation de données liées à la COVID-19 mais aussi sur la mise à disposition de données pseudonymisées pour les projets pilotes ainsi que sur l'opération de la plateforme technologique 1.0 du HDH sur ce périmètre.

Deux types de population sont concernés par cette information, d'une part, les citoyens dont les données de santé sont traitées et d'autre part, les utilisateurs de la plateforme technologique 1.0 concernés.

Au niveau du HDH, dans la même logique qu'une obligation de transparence sur tous les projets menés avec des données du SNDS est prévue en application de l'article L.1461-3 du code de la santé publique, le HDH tient un répertoire public accessible sur son site internet où la liste et les caractéristiques de tous les projets liés à la Covid-19 et de tous les projets pilotes sont recensés. Ainsi les citoyens connaissent les sources des données et les responsables de traitement de chaque projet.

En outre, le HDH tiendra la liste de tous les contacts (DPD ou contact opérationnel selon le choix des acteurs) auprès de qui les personnes pourront exercer leurs droits et les accompagne dans cet exercice selon les modalités précisées dans un document intitulé "Procédure de réponse à l'exercice des droits relatifs aux données personnelles".

Concernant les fonctionnalités de gestion des comptes utilisateurs et de traçabilité de leurs activités, les utilisateurs sont

informés lors de leur circuit d'arrivée sur la plateforme technologique notamment grâce à la lecture et la signature des conditions générales d'utilisation. Ces CGU sont accessibles librement par les utilisateurs au sein de leur espace projet et détaillent les modalités d'exercice des droits relatifs à leurs données à caractère personnel.

Acceptable / Améliorable / A corriger	**Commentaires d'évaluation**	**Mesures correctives**
Acceptable		

2.2.2. Si applicable, comment le consentement des personnes concernées est-il obtenu ?

Ce point n'est pas applicable à la plateforme technologique 1.0 :

- Concernant les données de santé pseudonymisées relatives à la COVID-19 et aux projets pilotes : la réutilisation des données des personnes concernées à des fins de recherche dans le domaine de la santé n'est pas soumise à leur consentement préalable mais à une simple information.

- Concernant les données collectées et utilisées pour la gestion des comptes des utilisateurs : le traitement est fondé sur le respect d'une obligation légale et les données sont recueillies directement auprès des utilisateurs.

- Concernant les traces de l'activité des comptes utilisateurs : la collecte des traces pour répondre au besoin de la sécurité de la plateforme répond à une obligation légale du HDH.

Acceptable / Améliorable / A corriger	**Commentaires d'évaluation**	**Mesures correctives**
Acceptable		

2.2.3. Comment les personnes concernées peuvent-elles exercer leurs droits d'accès et droit à la portabilité ?

Les traitements de données liés à l'entrepôt Covid et aux projets pilotes n'étant fondés ni sur le consentement de la personne ni sur l'exécution d'un contrat, les conditions d'applicabilité du droit à la portabilité prévu par l'article 20 du RGPD ne sont pas réunies.

Concernant le droit d'accès aux données de santé, dans la mesure où celles-ci sont pseudonymisées et que la correspondance entre les identifiants sur la plateforme technologique 1.0 et l'identité des personnes n'est pas conservée par le HDH , remonter à l'identité des personnes afin de confirmer que des données à caractère personnel les concernant sont ou ne sont pas traitées représenterait des efforts disproportionnés. Par conséquent, le HDH ne peut pas permettre l'accès auxdites données à caractère personnel. C'est la raison pour laquelle, dans le cadre des traitements de la plateforme 1.0, si un citoyen exerce son droit d'accès auprès du DPD du HDH, ce dernier ne peut que les inviter à contacter les responsables de données (DPD ou contact opérationnel selon le choix du responsable de données) qui disposent des identifiants directs des personnes et peuvent confirmer s'ils ont communiqué des données au HDH. En outre, seuls les responsables de données sont capables, en raison de leur expertise et de leur connaissance des données, de fournir aux personnes les données dans un format compréhensible et aisément accessible, selon l'exigence posée par l'article 12 du RGPD.

En tout état de cause, les coordonnées des correspondants à contacter pour exercer les droits sont aussi disponibles sur le site internet du HDH dans le répertoire public tel que décrit dans la réponse à la question *2.2.1 Comment les personnes concernées sont-elles informées à propos du traitement ?*

Les modalités détaillées de réponse à l'exercice des droits des citoyens sont précisées dans un document intitulé "Procédure de réponse à l'exercice des droits relatifs aux données personnelles".

Concernant le droit d'accès des utilisateurs à leurs données personnelles, le délégué à la protection des données du HDH est le point de contact des utilisateurs de la plateforme qui souhaitent exercer leurs droits. Les coordonnées du DPD sont indiquées dans les CGU de la plateforme technologique 1.0 et disponibles sur demande.

Acceptable / Améliorable / A corriger	**Commentaires d'évaluation**	**Mesures correctives**
Acceptable		

2.2.4. Comment les personnes concernées peuvent-elles exercer leurs droits de rectification et droit à l'effacement (droit à l'oubli) ?

La mise en œuvre de l'entrepôt est fondée sur une mission d'intérêt public, alors que l'opération de la plateforme technologique 1.0 du HDH est fondée sur une obligation légale. Ces deux fondements juridiques écartent l'application du droit à l'effacement, conformément à l'article 17-3-b) et c) du RGPD. En ce qui concerne les traitements réalisés dans les espaces projets de la plateforme technologique 1.0 qui relèvent des responsables de traitement des projets, le droit à l'effacement s'applique dans les conditions décrites dans l'AIPD du projet concerné.

Concernant le droit à la rectification des données de santé, pour les mêmes raisons justifiant que le HDH n'est pas en mesure de confirmer que des données à caractère personnel d'une personne identifiée sont ou ne sont pas traitées sur la plateforme technologique 1.0, il n'a pas non plus la possibilité de permettre leur rectification. En outre, l'application du droit de rectification soulève un problème de fond pour le HDH qui n'apparaît pas légitime à rectifier des données dont il n'est pas le responsable de traitement pour la collecte originelle et sur lesquelles ils ne disposent pas d'une proximité avec les personnes concernées ni d'une expertise métier. Si une rectification des données est pertinente, elle le sera avant tout au niveau des sources des données donc chez les responsables de données et toute rectification sera répercutée chez le HDH lors d'une mise à jour des données. En cas d'exercice du droit de rectification par les citoyens auprès du DPD du HDH, il est donc prévu, comme mentionné précédemment, que ce dernier les invite à contacter les responsables de données (DPD ou contact opérationnel selon le choix du responsable de données) qui disposent éventuellement des identifiants directs des personnes et peuvent donc donner suite à une demande d'effacement ou de rectification sur leur périmètre de responsabilité de traitement. En tout état de cause, comme mentionné précédemment, les coordonnées des correspondants à contacter pour exercer les droits sont disponibles sur le site internet du HDH dans le répertoire public tel que décrit dans la réponse à la question 2.2.1 Comment les personnes concernées sont-elles informées à propos du traitement ?

Les modalités détaillées de réponse à l'exercice des droits des citoyens sont précisées dans un document intitulé "Procédure de réponse à l'exercice des droits relatifs aux données personnelles".

Concernant les données des utilisateurs, ces derniers peuvent contacter leur autorité d'enregistrement qui fera le lien avec un opérateur projet pour rectifier les données nécessaires à la gestion de leur compte utilisateur. Le traitement de ces données étant fondé sur une obligation légale, les utilisateurs ne peuvent toutefois pas demander leur effacement avant la fin de la période de conservation fixée à deux ans après la fermeture du compte.

Enfin, les traces d'activité ne peuvent pas être rectifiées ni effacées avant la fin de la durée de conservation définie du fait des contraintes de sécurité de la plateforme.

Acceptable / Améliorable / A corriger	**Commentaires d'évaluation**	**Mesures correctives**

Acceptable		

2.2.5. Comment les personnes concernées peuvent-elles exercer leurs droits de limitation et droit d'opposition ?

Les modalités détaillées de l'exercice du droit d'opposition sont précisées dans un document intitulé "Procédure de réponse à l'exercice des droits relatifs aux données personnelles".

a) Concernant l'entrepôt Covid :

Le HDH favorise l'exercice du droit d'opposition de deux manières :

- D'abord en amont de l'arrivée des données sur la plateforme technologique 1.0, il est à souligner que les projets nécessitent souvent une phase de préparation des données assez conséquente chez les responsables de données. Pendant cette préparation, les responsables de données peuvent sans difficultés donner suite à une demande de droit d'opposition et ainsi retenir les données correspondantes avant transmission au HDH. En facilitant la diffusion de l'information auprès des personnes concernées, grâce à son répertoire public (tel que décrit dans la réponse à la question 2.2.1 Comment les personnes concernées sont-elles informées à propos du traitement ?), le HDH permet aux personnes concernées d'exercer plus facilement leur droit d'opposition auprès des responsables de données.
- Ensuite, le HDH met en place, avec l'aide des responsables de données, une logique de matrice d'exclusion. Lorsque le HDH reçoit du responsable de données un lot d'identifiants à exclure, l'opérateur projet lance la procédure d'exclusion. Celle-ci consiste à lancer l'automate de génération des identifiants de stockage correspondants puis à lancer l'automate d'ajout à la matrice d'exclusion. Cette matrice d'exclusion permet de ne plus partager les données attenantes lors des futures extractions de données sollicitées par les porteurs de projets. La matrice d'exclusion est conservée au sein de l'espace opérateur dans la même zone que les données fournies par le responsable de données.

b) Concernant les projets pilotes :

S'agissant des projets dont la responsabilité incombe aux porteurs des projets pilotes, dans la mesure où ils ne sont pas fondés sur une obligation légale, les patients peuvent naturellement s'y opposer. Pour les mêmes raisons justifiant que le HDH n'est pas en mesure de confirmer que des données à caractère personnel d'une personne identifiée sont ou ne sont pas traitées sur la plateforme technologique 1.0, il n'a pas non plus la possibilité de donner suite à une demande d'opposition. Conformément à l'article 11 du RGPD, le HDH ne conserve pas les pseudonymes d'origine, ces derniers étant remplacés par des nouveaux identifiants projet générés aléatoirement et en aucun cas dérivés du pseudonyme d'origine chez le responsable des données ou du NIR des personnes. Ainsi, une même personne qui serait concernée par plusieurs projets pilotes se verrait attribuer autant de pseudonymes différents qu'il y a de projets pilotes.

Néanmoins, il existe une situation où le HDH peut favoriser l'exercice du droit d'opposition vis-à-vis des projets pilotes : si un projet peut justifier de l'utilité de conserver la correspondance entre l'identifiant projet généré aléatoirement et le pseudonyme d'origine, par exemple en cas de mise à jour nécessaire des jeux de données mis à disposition, alors le HDH propose de conserver une table de correspondance entre ces pseudonymes. Cette dernière est conservée pendant la durée nécessaire à la ou aux mises à jour d'un projet et stockée sur l'espace opérateur, protégée par un chiffrement avec une clé dédiée. Son accès nécessite l'action conjointe d'un opérateur projet et d'un opérateur donnée. Dans cette hypothèse, l'exercice du droit d'opposition sera possible mais nécessitera toujours l'intervention des responsables de données qui connaissent l'identité directe des personnes. En pratique, à chaque mise à jour des données : soit le responsable met à jour la nouvelle livraison de données mais également les anciennes, soit il fournit les identifiants d'origine correspondant à toutes les personnes qui ont exercé leur droit d'opposition depuis le précédent envoi, dans la logique de la matrice d'exclusion. Dans ce deuxième cas, l'opérateur données du HDH reçoit cette liste d'identifiants et, après action conjointe avec un opérateur projet, met à jour les données dans l'espace projet sans les données des personnes qui ont exprimé leur opposition.

Dès lors, si les conditions décrites ci-dessus sont réunies pour l'entrepôt Covid ou les projets pilotes, en cas d'exercice du droit d'opposition par les citoyens auprès du DPD du HDH, ce dernier les invite à contacter les responsables de données (DPD ou contact opérationnel selon le choix du responsable de données) qui disposent des identifiants directs des personnes et peuvent donc donner suite à leur exercice du droit d'opposition. En tout état de cause, les coordonnées des correspondants à contacter pour exercer les droits sont disponibles sur le site internet du HDH dans le répertoire public.

Enfin, les utilisateurs ne peuvent pas s'opposer aux traitements de leurs données qui sont nécessaires à la gestion de leur compte utilisateur ni aux traitements de leurs traces d'activité du fait de l'obligation légale de maintenir la plateforme technologique 1.0 en condition de sécurité. Ils peuvent néanmoins exercer leur droit à la limitation auprès du délégué à la protection des données du HDH.

Acceptable / Améliorable / A corriger	**Commentaires d'évaluation**	**Mesures correctives**
Acceptable		

2.2.6. Les obligations des sous-traitants sont-elles clairement définies et contractualisées ?

Deux sous-traitants interviennent sur la plateforme technologique 1.0 :

Microsoft :

La préfiguration du HDH a été menée par la DREES qui a lancé les travaux de développement de la plateforme technologique 0.9 et eu recours aux services de Microsoft dans le cadre d'un marché UGAP. Une convention de cession de la plateforme technologique a été signée entre l'Etat et le HDH qui en est donc désormais propriétaire.

Les discussions contractuelles entre Microsoft et le HDH ont abouti à la signature d'un contrat le 15 avril 2020.

À la suite de l'ordonnance du Conseil d'Etat du 19 juin 2020 et de divers échanges avec la CNIL, le HDH a ouvert des négociations avec Microsoft pour préciser par voie d'avenant les garanties encadrant le stockage et le traitement des données de santé en Union Européenne ainsi que le non-accès des ingénieurs Microsoft aux données de santé en clair. Ces avenants ont été conclus le 7 juillet et le 3 septembre 2020.

A la suite d'une seconde ordonnance du Conseil d'Etat en date du 13 octobre 2020, de nouvelles négociations ont été ouvertes avec Microsoft pour apporter des garanties supplémentaires sur l'absence de transferts de données de santé hors UE ainsi que préciser le droit applicable en cas de divulgation des données.

Tiers archiveur de confiance :

Des garanties sont apportées par le tiers archiveur sur la sécurité physique de ses centres de données. Les données envoyées sont scellées et ne sont pas consultables par le sous-traitant.

Acceptable / Améliorable / A corriger	**Commentaires d'évaluation**	**Mesures correctives**
Acceptable		

2.2.7. En cas de transfert de données en dehors de l'Union européenne, les données sont-elles protégées de manière équivalente ?

Conformément au référentiel de sécurité du Système National des Données de Santé (SNDS), les données de santé de l'entrepôt COVID stockées sur la plateforme technologique 1.0 sont hébergées en Union Européenne. Plus précisément, les données sont stockées dans les centres de données Microsoft situés en France dans la région "France central" (région parisienne), certifiés "Hébergeur de données de santé". Il est de la responsabilité du HDH de spécifier la zone géographique des ressources déployées et Microsoft s'engage contractuellement à ne pas stocker ni traiter les données en dehors de cette zone géographique. Le HDH met en œuvre des restrictions techniques et des contrôles pour s'assurer que les ressources déployées sont effectivement hébergées en Union Européenne.

En effet, Microsoft ne s'engage pas contractuellement à ce que des administrateurs situés en dehors de l'Union Européenne n'interviennent pas à des fins de maintenance, de résolution d'incident ou de support. Interdire une telle possibilité d'accès empêcherait Microsoft d'honorer ses engagements de niveau de service (99,9% de disponibilité mensuelle garantis pour la plupart des Services en Ligne). Le support 24 x 7 de « suivi du soleil » nécessite de pouvoir faire appel à des équipes de support réparties sur les différentes plaques géographiques pour être en mesure de réagir rapidement à la résolution de situations critiques.

En conséquence, des mesures de contrôle sont mises en place, du côté de Microsoft et du côté du HDH pour faire respecter l'interdiction d'accès aux données. Les ingénieurs Microsoft, notamment les employés à temps plein et les sous-traitants / fournisseurs, n'ont pas accès par défaut aux données des clients. Pour la majorité des cas de support, l'accès aux données clients n'est pas nécessaire. Dans les rares cas, sur requête du client, où un accès serait nécessaire pour la résolution d'un incident, un système de disposition d'accès privilégié « juste-à-temps » (« Just-in-Time ») est déployé. Il est à noter que les données concernées sont des données d'utilisation de la plateforme, telles que des "logs" ou "traces", et non pas les données de santé. Les procédures sont auditées et validées dans le cadre d'un audit produit par un cabinet d'audit tiers indépendant.

En outre, dans le cas du HDH, un contrôle supplémentaire est activé : il s'agit de la fonctionnalité Customer Lockbox. La Customer Lockbox, fournie avec un enregistrement d'audit complet, permet à un client de Microsoft d'approuver ou refuser l'accès aux données. Par défaut, le HDH refusera tout accès aux données par les ingénieurs Microsoft.

La combinaison de ces mesures techniques et organisationnelles, et de la protection en confidentialité de tous les échanges sur les réseaux entre la plateforme et les utilisateurs par l'utilisation de protocoles chiffrant, conformes au Référentiel Général de Sécurité, permettent d'éviter le transfert de données de santé hors Union Européenne. Le stockage et le traitement des données de santé au sein de l'Union Européenne s'est traduit juridiquement dans un avenant au contrat entre Microsoft et le HDH au début du mois de septembre 2020.

Cependant, il est à relever que les données relatives aux utilisateurs de la plateforme peuvent être collectées par Microsoft et transférées en dehors de l'Union européenne, notamment dans le cadre de leur authentification avant accès à la plateforme technologique 1.0.

A la suite de l'ordonnance du Conseil d'Etat du 19 juin 2020, il est décidé d'informer de l'existence de transferts de données hors Union Européenne de la manière suivante :

"La plateforme technologique du HDH est hébergée dans les centres de données Microsoft situés en Union Européenne, certifiés « Hébergeur de données de santé ». Compte tenu du contrat passé avec son sous-traitant et du fonctionnement des opérations d'administration de la plateforme technologique, il est possible que des données techniques d'usage de la plateforme (qui ne révèlent aucune information de santé) soient transférées vers des administrateurs situés en dehors de l'Union Européenne. Ces transferts de données sont encadrés par les clauses contractuelles types adoptées par la Commission Européenne dont une copie peut être obtenue auprès du Délégué à la protection des données du HDH."

Par ailleurs un arrêt de la CJUE du 16 juillet 2020 a invalidé le Privacy Shield au motif que « *la primauté des exigences relatives à la sécurité nationale, à l'intérêt public et au respect de la législation américaine, rendant ainsi possibles des ingérences dans les droits fondamentaux des personnes dont les données sont transférées vers ce pays* ». Le contrat entre le HDH et Microsoft s'appuie sur les clauses contractuelles types dont la validité a été confirmée par la Cour dans ce même arrêt. Néanmoins, dans la mesure où la Cour a tout de même pris soin de rappeler que les personnes concernées par des données transférées doivent bénéficier d'un niveau de protection équivalent à celui garanti au sein de l'UE et que cette protection doit être évaluée, au-delà des clauses contractuelles, en tenant compte du cadre juridique d'un éventuel accès par les autorités publiques du pays tiers, la vérification du niveau de protection des données des utilisateurs de la plateforme technologique concernés par un éventuel transfert de données hors UE est en cours d'instruction par le HDH et Microsoft.

A la suite d'une seconde ordonnance du Conseil d'Etat en date du 13 octobre 2020, de nouvelles négociations ont été ouvertes avec Microsoft pour apporter des garanties supplémentaires sur l'absence de transferts de données de santé hors UE ainsi que préciser le droit applicable en cas de divulgation des données.

Acceptable / Améliorable / A corriger	**Commentaires d'évaluation**	**Mesures correctives**
Acceptable		

3. Etude des risques liés à la sécurité des données

3.1. Mesures existantes ou prévues

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

3.2. Accès illégitime à des données

Quels pourraient être les principaux impacts sur les personnes concernées si le risque se produisait ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Quelles sont les principales menaces qui pourraient permettre la réalisation du risque ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Quelles sources de risques pourraient en être à l'origine ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Quelles sont les mesures initiales, parmi celles identifiées, qui contribuent à traiter le risque ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Comment estimez-vous la gravité du risque, notamment en fonction des impacts potentiels et des mesures prévues ?

Maximale.

Comment estimez-vous la vraisemblance du risque, notamment au regard des menaces, des sources de risques et des mesures prévues ?

Limitée.

3.3. Modification non désirée de données

Quels pourraient être les principaux impacts sur les personnes concernées si le risque se produisait ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Quelles sont les principales menaces qui pourraient permettre la réalisation du risque ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Quelles sources de risques pourraient-elles en être à l'origine ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Quelles sont les mesures initiales, parmi celles identifiées, qui contribuent à traiter le risque ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Comment estimez-vous la gravité du risque, notamment en fonction des impacts potentiels et des mesures prévues ?

Importante.

Comment estimez-vous la vraisemblance du risque, notamment au regard des menaces, des sources de risques et des mesures prévues ?

Limitée.

3.4. Disparition de données

Quels pourraient être les principaux impacts sur les personnes concernées si le risque se produisait ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Quelles sont les principales menaces qui pourraient permettre la réalisation du risque ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Quelles sources de risques pourraient-elles en être à l'origine ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Quelles sont les mesures initiales, parmi celles identifiées, qui contribuent à traiter le risque ?

Partie non communicable conformément à la jurisprudence de la CADA (Conseil 20183041 - séance du 08/11/2018) en raison de la protection de la sécurité des systèmes d'information des administrations.

Comment estimez-vous la gravité du risque, notamment en fonction des impacts potentiels et des mesures prévues ?

Importante.

Comment estimez-vous la vraisemblance du risque, notamment au regard des menaces, des sources de risques et des mesures prévues ?

Négligeable.

4. Formalisation de la validation

Avis du Délégué à la Protection des Données

L'analyse d'impact présente de manière précise tant les aspects techniques qu'opérationnels des traitements liés à l'entrepôt COVID et aux projets pilotes, i.e. aussi bien les traitements des données concernées que l'opération de la plateforme technologique version 1.0.

La réalisation de cette AIPD a été menée dans un contexte particulièrement contraint en raison de la crise sanitaire actuelle, du contentieux devant le Conseil d'Etat au sujet des éventuels transferts de données hors Union européenne et des travaux en cours pour basculer vers le cadre de droit commun.

Mon opinion en tant que DPD est que la plateforme technologique est néanmoins maîtrisée et que la version 1.0 destinée aux projets pilotes est en mesure d'accueillir convenablement un entrepôt de données liées à la Covid-19. Même si un tel entrepôt n'était pas envisagé, son utilité en réponse à la crise sanitaire et la gestion de ses suites est démontrée au regard des enjeux à mobiliser les données de santé dans ce contexte. La conservation de l'entrepôt à plus long terme à des fins de recherche apparaît aussi utile pour l'amélioration des connaissances scientifiques. Je considère que les mesures mises en œuvre réduisent les risques résiduels à un niveau acceptable et formule donc un avis favorable à la mise en œuvre de la plateforme technologique 1.0.

Néanmoins il conviendra de continuer à investiguer les conséquences de l'arrêt C-311/18 de la CJUE du 16 juillet 2020, dit « Schrems II », sur le choix du prestataire d'hébergement de la plateforme technologique. En particulier, conformément à l'ordonnance du Conseil d'Etat du 13 octobre 2020, un nouvel avenant doit être conclu avec le prestataire afin d'apporter plus de garanties sur l'absence de transferts de données de santé en dehors de l'Union européenne. Il faudra également continuer de rechercher, en vertu de l'article 28 du RGPD, la mise en œuvre des mesures techniques et organisationnelles appropriées pour garantir au mieux la protection des droits des personnes concernées.

Date et signature : 21 octobre 2020, Thomas Duong

Décision du responsable de traitement

Le 21 octobre 2020, Mme Stéphanie Combes, directrice de la Plateforme des données de santé et autorité qualifiée pour la sécurité des systèmes d'information, valide l'Analyse d'impact relative à la protection des données (AIPD) de la plateforme technologique version 1.0 du HDH.

Le traitement doit permettre de poursuivre l'hébergement de données de santé provenant de différentes sources en vue de faciliter leur mise à disposition des projets pilotes du HDH, d'une part, et des équipes de recherche mobilisées dans la lutte contre l'épidémie Covid-19 et ses conséquences, d'autre part.

La manière dont il est prévu de mettre en œuvre les mesures à la fois juridiques et techniques et de traiter les risques est en effet jugée acceptable au regard de ces enjeux. La mise en œuvre du plan d'actions devra être démontrée ainsi que l'amélioration continue de l'AIPD.

Date et signature : 21 octobre 2020, Stéphanie Combes