

<p style="text-align: center;">Comité de sécurité de l'information</p> <p style="text-align: center;">Chambre sécurité sociale et santé</p>

CSI/CSSS/21/332

DÉLIBÉRATION N° 21/172 DU 5 OCTOBRE 2021 RELATIVE À LA COMMUNICATION DE DONNÉES À CARACTÈRE PERSONNEL RELATIVES À LA SANTÉ À PARTIR DE LA BASE DE DONNÉES HEALTHDATA COVID-19 II ET DE LA BASE DE DONNÉES ICAREDATA À L'UNIVERSITÉ D'ANVERS ET À L'INSTITUT DE MÉDECINE TROPICALE DANS LE CADRE DE LEUR PROJET DE RECHERCHE ID-COV

Le Comité de sécurité de l'information

Vu le Règlement (UE) n° 2016/679 du Parlement européen et du Conseil du 27 avril 2016 *relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE* (Règlement général relatif à la protection des données ou AVG);

Vu la loi du 30 juillet 2018 *relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel*;

Vu la loi du 15 janvier 1990 *relative à l'institution et à l'organisation d'une Banque-carrefour de la sécurité sociale*, notamment l'article 37;

Vu la loi du 3 décembre 2017 *relative à la création de l'Autorité de protection des données*, en particulier l'article 114, modifié par la loi du 25 mai 2018;

Vu la loi du 13 décembre 2006 *portant dispositions diverses en matière de santé*, en particulier l'article 42, § 2, 3°, modifié par la loi du 5 septembre 2018;

Vu la loi du 5 septembre 2018 *instituant le comité de sécurité de l'information et modifiant diverses lois concernant la mise en œuvre du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE*, notamment l'article 97;

Vu la loi du 21 août 2008 *relative à l'institution et à l'organisation de la plate-forme eHealth et portant dispositions diverses*;

Vu la demande de l'Université d'Anvers et de l'Institut de médecine tropicale;

Vu le rapport d'auditorat de la Plate-forme eHealth du 1^{er} septembre 2021;

Vu le rapport de monsieur Bart Viaene.

Émet, après délibération, la décision suivante, le 5 octobre 2021:

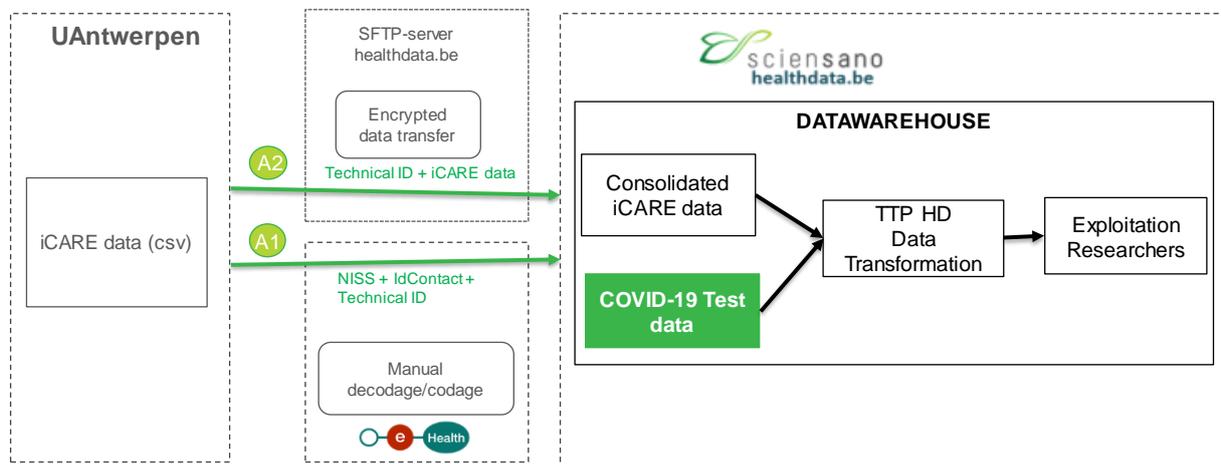
I. OBJET DE LA DEMANDE

1. Le département « *Family Medicine and Population Health* » (FAMPOP) de l'Université d'Anvers et le département « *Public Health* » de l'Institut de médecine tropicale ont introduit une demande visant à obtenir accès à des données à caractère personnel pseudonymisées relatives à la santé en provenance de la base de données Healthdata Covid-19 II de Sciensano et de la base de données iCAREdata de l'Université d'Anvers. Ces données seront utilisées dans le cadre de leur étude visant à développer des modèles susceptibles de prévoir la gravité de l'infection Covid-19 dans différents contextes cliniques (par exemple, postes de garde, hôpitaux). Ces modèles sont pertinents pour le triage des patients et pour l'organisation de la capacité des soins.
2. Le recours aux données de la base de données iCAREdata et de la base de données Healthdata Covid-19 II permet d'établir un cadre méthodologique pour l'harmonisation des données et pour le développement analytique d'un nouvel outil de prévision des risques. Cette prévision des risques permet de comprendre davantage le rôle de la comorbidité et d'autres facteurs dans la prévision de la progression de la maladie en ce qui concerne la Covid-19, et permet en d'autres termes de prédire sur la base de certains facteurs les chances de rétablissement pour un patient déterminé. Ce projet permettra de développer un outil unique de stratification des risques qui sera utile non seulement pour l'épidémie actuelle mais également pour des épidémies futures. Finalement, ce projet est considéré comme une validation de principe pour l'utilisation de données du monde réel dans la recherche ainsi qu'une validation de principe pour le développement méthodologique d'une prévision des risques susceptible d'être utilisée, moyennant certaines adaptations, dans d'autres pays ou d'autres systèmes de soins de santé chacun avec leur propre jeu de données unique.
3. Les symptômes rapportés de la Covid-19 proviennent de la base de données iCAREdata (UA), qui est gérée par le demandeur UA, département « *Family Medicine and Population Health* » (FAMPOP)¹. La base de données iCAREdata recueille des données de patients qui ont consulté les services de garde de médecine générale ou les services d'urgence. Cette base de données de recherche clinique a initialement été créée dans le cadre du projet iCAREdata, qui a été approuvé par le Comité par sa délibération 14/094 du 18 novembre 2014, modifiée le 28 juillet 2015, *relative à la communication de données à caractère personnel codées relatives à la santé par des postes de garde, les pharmacies affiliées à la Koninklijke Apothekersvereniging van Antwerpen (KAVA) et des services d'urgence au Centrum voor Huisartsengeneeskunde de l'université d'Anvers dans le cadre du projet iCAREdata.*

¹ La base de données du projet iCAREdata est conservée sur un serveur situé dans un local de serveurs du Campus Drie Eiken de l'université d'Anvers.

La banque de données HealthData COVID-19 II contient les résultats de tests relatifs à la COVID-19, tant les tests PCR que les résultats d'un éventuel CT-scan thoracique. Il s'agit des principales variables dépendantes qui confirment la présence de COVID-19 chez les patients et grâce au lien avec la base de données iCAREdata, ces variables permettront de déterminer les caractéristiques diagnostiques qui jouent un rôle important dans la COVID-19. La série de données iCAREdata ne contient pas ces données. Le couplage des deux séries de données est donc nécessaire pour pouvoir développer un modèle relationnel entre plaintes/symptômes et résultats.

4. Préalablement au couplage, les symptômes de l'ensemble de données iCAREdata sont uniformisés en quelque 40 codes de symptômes (codes ID-COV) à l'aide d'une méthode de datamining². Les consultations de patients avec des symptômes présumés de Covid-19 sont sélectionnées en vue du couplage avec la base de donnée HealthData Covid-19 II de Sciensano qui contient le résultat du test PCR pour ces consultations.
5. Une série de données a déjà été extraite de la base de données iCAREdata comprenant des enregistrements de consultations ayant eu lieu au cours de période épidémique de COVID-19 du 1/3/2020 au 30/11/2020. Cette série de données comprend quelque 23.000 patients présumés pour lesquels un prélèvement dans le cadre d'un test PCR a eu lieu. Un grand nombre d'enregistrements est nécessaire étant donné que pour toutes les consultations suspectes dans le cadre de Covid-19, il n'y a qu'un pourcentage limité de cas positifs. Ceci est caractéristique pour les consultations de la première ligne.
6. De manière schématique, les flux de données se dérouleront comme suit:



Les données iCAREdata sont transférées vers healthdata.be en 2 flux distincts :

² Les données initialement introduites dans iCAREdata sont notamment sous forme de texte libre et l'input est donc très variable et n'est pas directement utilisable dans un modèle. C'est pourquoi, préalablement au couplage, les symptômes dans l'ensemble de données iCAREdata sont uniformisés en une quarantaine de codes de symptômes à l'aide d'une méthode de datamining.

- Le flux A1 contient le NISS du patient, le 'IdContact' et un ID technique (un ID technique unique par enregistrement). Les données sont transférées par l'Université d'Anvers, via eHealth pour le décodage/codage manuel du NISS et d'IdContact, vers le datawarehouse de healthdata.be (HD-DWH).
- Le flux A2 contient d'autres données iCAREdata avec le même ID technique. Etant donné que cette série de données ne contient pas de données qui doivent être pseudonymisées, celle-ci sera transmise directement de l'Université d'Anvers vers le DWH Healthdata, soit via SFTP, soit via une autre méthode de transfert permettant le cryptage.

L'ID technique est crypté par l'expéditeur, tandis que le NISS et l'IdContact sont dépseudonymisés/pseudonymisés par eHealth en tant que tiers de confiance. Après réception, la plateforme healthdata.be consolidera les 2 flux de données distincts sur la base de l'ID technique. Après la consolidation et le contrôle de qualité technique, l'ID technique sera immédiatement et définitivement supprimé de l'infrastructure healthdata.be. La plateforme healthdata.be conserve un log de ces processus techniques.

Après consolidation des deux séries de données sur la base d'un même pseudo-ID, healthdata.be interviendra comme tiers de confiance (Trusted Third Party = TTP HD) pour transformer certaines variables afin de réduire les risques potentiels de réidentification. Les détails de cette transformation (calculs de dates exactes) sont précisés au point 19 de cette délibération. L'environnement distinct du datawarehouse healthdata.be, dans lequel ces transformations auront lieu, est uniquement accessible aux collaborateurs du TTP HD.

Après le processus de transformation de données, les variables autorisées sont transférées vers un environnement d'exploitation dans le cadre du projet de recherche. Les chercheurs concernés de l'Université d'Anvers et de l'Institut de médecine tropicale obtiendront accès à cet environnement d'exploitation au sein du Datawarehouse Healthdata où ils pourront réaliser leurs analyses de données.

Les différents environnements du datawarehouse healthdata.be sont équipés du logiciel GUARDIUM permettant de surveiller et de protéger les activités de la base de données en temps réel. Ce logiciel d'audit permet de créer des loggings qui enregistrent de manière permanente tout utilisateur, tout accès, toute période d'accès, toute activité et tout résultat de cette activité et les rendent consultables. Pour toute demande d'accès, un avis est demandé au Data Protection Officer de Sciensano au moyen de la Standard Operating Procedure. Cette personne vérifie, sur la base des rôles au sein de Sciensano (p.ex. collaborateur du service TTP) et du contenu de la délibération de la chambre sécurité sociale et santé du Comité de sécurité de l'information (p.ex. en ce qui concerne les chercheurs autorisés), si les droits d'accès peuvent être accordés. L'accès a lieu sur la base d'une authentification à deux facteurs.

7. Les données à caractère personnel pseudonymisées suivantes sont demandées:

A partir de la base de données iCAREdata :

- Numéro NISS (pseudonymisé) (INSZ_Cod) ;
- Numéro unique pour l'identification du contact (idContact) ;
- Code technique (Technical ID) ;
- Diagnostic (DiagnTekst) ;
- Code diagnostic (DiagnCod) ;
- Motif du contact (ContactRedenTekst) ;
- Code pour motif du contact (ContactRedenCod) ;
- Code unique pour l'identification du poste de garde (idWachtpost) ;
- Type de contact: conseil par téléphone, consultation, visite à domicile (ConType) ;
- Eventuel renvoi vers un tiers (ContVerwerk) ;
- Année de naissance du patient (GebJaar) ;
- Sexe du patient ;
- Date du contact (ContDatumStart) (mois, année);
- Codes ID-Cov³.

A partir de la base de données Healthdata Covid-19 II

Prescription test laboratoire Covid-19 :

- Numéro NISS (pseudonymisé) ;
- Date du contact (EncounterStartDateTime) convertie en nombre de jours entre la date du contact (ContDatumStart) et la date à laquelle le contact a eu lieu (EncounterStartDateTime);⁴
- Date d'apparition des symptômes (ProblemStartDate) convertie en nombre de jours par rapport à la date du contact (ContDatumStart)⁵
- Date d'apparition des symptômes (ProblemStartDate) convertie en nombre de jours par rapport à la date à laquelle le contact a eu lieu (EncounterStartDateTime) ;
- Le motif de la prescription de test (TestPrescribedReason) ;
- Le lieu de prélèvement (CollectionLocation) ;
- Numéro d'identification du matériel prélevé (SpecimenId1, SpecimenId2 et SpecimenId3) ;
- Description du matériel prélevé (SpecimenMaterial1, SpecimenMaterial2 et SpecimenMaterial3).

Résultat du test de laboratoire Covid-19 :

- Numéro NISS (pseudonymisé) ;
- Numéro d'identification du matériel prélevé (SpecimenId1, SpecimenId2 et SpecimenId3) ;
- Le résultat du test (TestResult1, TestResult2 et TestResult3)⁶ ;

³ Codes développés sur la base d'une procédure de datamining dans une phase antérieure (avant le couplage) du projet ID-COV.

⁴ La date à laquelle le contact a eu lieu (EncounterStartDateTime) est remplacée par un champ calculé, à savoir le nombre de jours entre la date du contact (ContDatumStart en provenance de la base de données iCAREdata) et EncounterStartDateTime (COVID DB II, indication temporelle non nécessaire)

⁵ La date d'apparition des symptômes est remplacée par un champ calculé, à savoir le nombre de jours entre la date du contact (iCAREdata) et ProblemStartDate (COVID DB II). Un champ calculé supplémentaire est ajouté, à savoir le nombre de jours entre EncounterStartDate (COVID DB II) et ProblemStartDate (COVID D II).

⁶ Il s'agit d'une liste de codes comprenant 4 valeurs : (260373001=Déteecté ; 419984006=Non concluant ; 260415000=Non déteecté ; 125154007=Spécimen inadéquat pour analyse)

- Description du matériel prélevé (SpecimenMaterial1, SpecimenMaterial2 et SpecimenMaterial3).

8. Le Comité prend acte du fait que l'Institutional Review Board de l'Institut de médecine tropicale et le Comité d'éthique de l'Universitair Ziekenhuis Antwerpen ont rendu des avis positifs concernant cette étude.

I. COMPÉTENCE

9. En vertu de l'article 42, § 2, 3° de la loi du 13 décembre 2006 portant dispositions diverses en matière de santé, la chambre sécurité sociale et santé du Comité de sécurité de l'information est compétente pour rendre une autorisation de principe concernant toute communication de données à caractère personnel relatives à la santé.

10. Conformément à l'article 10, § 3, de l'accord de coopération du 25 août 2020, les données à caractère personnel telles que communiquées et conservées dans la Base de données Healthdata Covid-19 II, ne peuvent être transmises à des tiers aux fins stipulées à l'article 3, § 1^{er}, 4° qu'après la délibération, visée à l'article 11, de la Chambre sécurité sociale et santé du Comité de sécurité de l'information.

11. Compte tenu de ce qui précède, le Comité estime qu'il peut se prononcer sur la communication de données à caractère personnel relatives à la santé, telle que décrite dans la demande d'autorisation.

II. EXAMEN

A. ADMISSIBILITÉ

12. Le traitement de données à caractère personnel est uniquement autorisé pour des finalités déterminées, explicites et légitimes et le traitement de données à caractère personnel relatives à la santé est en principe interdit⁷.

13. L'interdiction du traitement de données à caractère personnel relatives à la santé ne s'applique pas lorsque 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.

14. A la lumière de ce qui précède, le Comité est par conséquent d'avis qu'il existe un fondement admissible pour le traitement des données à caractère personnel pseudonymisées relatives à la santé envisagé.

⁷ Art. 9, point 1 RGPD.

III. FINALITÉ

15. Conformément à l’art. 5, b) du RGPD, le traitement de données à caractère personnel est uniquement autorisé pour des finalités déterminées, explicites et légitimes.
16. L'objectif de l'étude est de développer un modèle de prédiction sur la base des symptômes rapportés afin de faire des prédictions sur la présence ou non d'une infection Covid (résultat test PCR).
17. Au vu de cet objectif, le Comité considère que le traitement des données à caractère personnel envisagé poursuit bien des finalités déterminées, explicites et légitimes.

IV. PROPORTIONNALITÉ

18. Conformément à l’art. 5, b) et c) du RGPD, les données à caractère personnel doivent être adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont obtenues et pour lesquelles elles sont traitées ultérieurement.
19. L’utilisation de la base de données iCAREdata est motivée comme suit :
 - Le **numéro NISS pseudonymisé** et le **code technique** : nécessaires pour le couplage des séries de données. L’ID technique est crypté par l’expéditeur, tandis que les numéros NISS sont pseudonymisés par le TTP eHealth. Après réception, la plateforme healthdata.be consolidera les 2 flux de données distincts (voir le point 6) sur la base de l’ID technique. Après la consolidation et le contrôle de qualité technique, l’ID technique sera immédiatement et définitivement supprimé de l'infrastructure healthdata.be.
 - Le **numéro unique pour l’identification du contact** : une même personne peut avoir plusieurs IdContact. Dans le cadre du modèle de prédiction, les chercheurs doivent pouvoir disposer de variables de chaque IdContact distinct et les mettre en rapport avec une même personne. Ceci permet de tenir compte du facteur temps. Les IdContact sont pseudonymisés par le TTP eHealth.
 - **Diagnostic** (thésaurus) : il s’agit du diagnostic posé par le médecin à la fin de la consultation. Il s’agit d’une interprétation professionnelle et sans doute d’un meilleur prédicteur du résultat (infection à la Covid-19) que la variable *motif du contact*. Le thésaurus est une liste standardisée de possibilités qui permet une comparaison avec d’autres bases de données.
 - **Code diagnostic** (thésaurus) : il s’agit du code généré pour le diagnostic posé par le médecin. C’est un code standardisé qui peut être lu par une machine. Tout comme « *code motif du contact* », il s’agit d’une variable de contrôle.
 - **Motif du contact** (thésaurus) : il s’agit de la raison pour laquelle le patient demande une aide médicale. Cette variable est pertinente pour l’étude du pronostic. Certaines plaintes donneront lieu à un déroulement plus ou moins grave. Il s’agit d’informations fournies par le patient et donc d’informations différentes du diagnostic qui est une

interprétation du médecin. Il s'agit de variables distinctes qui sont introduites séparément dans le modèle.

- **Code motif du contact** (thésaurus) : Il s'agit du code généré pour le motif du contact sur la base du terme de thésaurus introduit. C'est un code standardisé qui peut être lu par une machine. Tout comme « *code diagnostic* », il s'agit d'une variable de contrôle.
- **Code unique pour l'identification du poste de garde** : étant donné qu'il y a un regroupement de données, il est important de contrôler ceci. Contrôler signifie qu'un facteur de correction est introduit dans l'analyse statistique pour corriger l'effet du regroupement (observations d'un même poste de garde) sur les relations entre prédicteur et variable dépendante à examiner. Le regroupement signifie que les données d'un même poste de garde se ressemblent plus que les données de différents postes de garde. Ceci a une influence sur la valeur des prédictions. C'est pourquoi une correction statistique doit être effectuée. Pour réaliser cette correction, il est nécessaire de savoir combien d'enregistrements proviennent de quel poste de garde.
- **Type de contact** (conseil par téléphone, consultation, visite à domicile) : ceci est nécessaire afin de comprendre pourquoi des champs sont laissés à blanc pour certains IdContact. Par exemple, en cas de conseil par téléphone, il n'y a objectivement pas de champs ou seulement de manière limitée. Cette information est donc nécessaire pour l'interprétation des symptômes (absents) et constitue donc une variable de contrôle pour le modèle.
- **Eventuel renvoi vers un tiers** : Il s'agit de la partie du traitement mis en œuvre (par exemple renvoi vers un spécialiste) et donc d'une étape dans le processus entre la présentation de la plainte et le résultat (infection Covid-19 oui/non). Etant donné que le but de l'étude est de prédire le résultat sur la base de plusieurs indicateurs dans le processus, il s'agit d'informations dont les chercheurs ont besoin pour développer un modèle optimal. Il s'agit de l'indication selon laquelle le patient est renvoyé pour des examens supplémentaires ou une hospitalisation.
- **Année de naissance du patient** : il apparaît de la littérature que l'âge est un prédicteur possible pour le résultat d'une infection à la Covid-19.
- **Sexe du patient** : il apparaît de la littérature que le sexe est un prédicteur possible pour le résultat d'une infection à la Covid-19.
- **Date du contact**: le temps entre la date du contact et la date du diagnostic est un prédicteur possible du résultat. Un terme souvent utilisé est retard diagnostique. Le retard diagnostique est un indicateur qui sera testé dans le modèle de prédiction. Les chercheurs n'obtiennent cependant pas accès à la date exacte, mais uniquement à l'année et au mois. Healthdata.be effectuera cette agrégation en tant que tiers de confiance.
- **Codes ID-Cov** : il s'agit des codes développés à l'aide de procédures de datamining dans une phase antérieure (pré-couplage). Ces codes contiennent des informations sur les plaintes et symptômes. Nécessaires pour pouvoir identifier les patients Covid-19 potentiels et importants pour comprendre le tableau clinique des patients Covid-19 présumés. Il y aura environ 40 catégories.

L'utilisation des données de la base de données Healthdata Covid-19 II est motivée comme suit :

- Le **numéro NISS pseudonymisé** : nécessaire pour relier les bases de données entre elles.
- **Date à laquelle le contact a eu lieu** : étant donné que plusieurs contacts sont possibles, tout comme plusieurs tests, il est nécessaire de relier la date de test correcte à la date de contact exacte avec le prestataire de soins. Ceci est nécessaire pour l'analyse de survie, dans le cadre de laquelle le temps est un élément important. Les chercheurs n'auront cependant pas accès à la date exacte. Healthdata.be effectuera un calcul en tant que tiers de confiance. Sur la base de ce calcul, les chercheurs obtiendront uniquement un nombre positif ou négatif, c'est-à-dire le nombre de jours entre le contact auprès du poste de garde dans la base de données iCAREdata (date 0) et le contact mentionné dans la base de données Healthdata Covid-19.
- **Date d'apparition des symptômes** : il s'agit d'une variable pertinente pour déterminer la durée des plaintes. Les chercheurs n'auront cependant pas accès à la date exacte. Healthdata.be effectuera un calcul en tant que tiers de confiance. Sur la base de ce calcul, les chercheurs obtiendront uniquement un nombre positif ou négatif, c'est-à-dire le nombre de jours entre le contact auprès du poste de garde et la date d'apparition des symptômes. ContDatumStart.
- **Motif de la prescription de test** : ceci est pertinent afin de distinguer les personnes qui sont testées en raison de symptômes des personnes qui sont testées pour d'autres raisons.
- **Lieu de prélèvement** : ceci fournit des renseignements sur l'endroit où le prélèvement a été effectué et fournit des informations partielles sur le lieu où se trouve le patient (par exemple un hôpital, prélèvement décentralisé ou centre de soins résidentiels). Il est important d'opérer une distinction entre les patients hospitalisés et non-hospitalisés.
- **Numéro d'identification du matériel prélevé** : A titre de référence pour permettre la demande d'information auprès de l'organisation source. Dans un contexte transmurale, ce numéro est constitué d'un numéro d'échantillon y compris l'identification de l'organisation émettrice, de sorte à être unique en dehors des murs de l'organisation. Ceci est demandé à la fois pour les *prescriptions de test laboratoire Covid-19* et pour les *résultats de test laboratoire Covid-19*, de sorte à pouvoir les mettre en rapport.
- **Description du matériel prélevé** : Le matériel de prélèvement a une influence sur la sensibilité du test. Il est donc important de pouvoir interpréter la fiabilité du résultat. Dans certains cas, un patient peut subir plusieurs tests Covid-19 avec prélèvement de différents matériels corporels. Dans ces cas, tous les tests et matériels sont nécessaires pour l'étude. Ceci est demandé à la fois pour les *prescriptions de test laboratoire Covid-19* et pour les *résultats de test laboratoire Covid-19* en vue de contrôles croisés.
- **Le résultat du test** : ceci est nécessaire pour le développement du modèle de prédiction. Il s'agit d'une variable dépendante importante, à savoir la confirmation en cas de présomption de Covid-19.

20. Les symptômes rapportés proviennent de la base de données iCAREdata (UA)⁸ et la base de données HealthData Covid-19 II contient les résultats de tests Covid-19, tant les tests PCR que les résultats de CT-scans thoraciques éventuels. En couplant les bases de données, les variables demandées permettront de déterminer quels caractéristiques diagnostiques jouent un rôle important dans la Covid-19. Le lien entre les deux jeux de données est donc nécessaire pour le développement d'un modèle relationnel entre plaintes/symptômes et résultats.
21. Le principe de proportionnalité implique que le traitement doit en principe être réalisé au moyen de données anonymes, ce qui est le cas en l'espèce. Cependant, si la finalité ne peut être réalisée au moyen de données anonymes, des données à caractère personnel pseudonymisées peuvent être traitées. Vu la nécessité de réaliser une analyse détaillée sur la base de ces données, les chercheurs ont besoin d'avoir accès à des données pseudonymisées afin d'être en mesure de réaliser ces analyses qu'il ne serait pas possible de réaliser à l'aide de données anonymes. Cette finalité justifie donc le traitement de données à caractère personnel pseudonymisées.
22. Les données à caractère personnel ne peuvent être conservées sous une forme permettant l'identification des personnes concernées que pendant la durée nécessaire à la réalisation des finalités pour lesquelles les données à caractère personnel sont traitées.
23. A compter du 1^{er} octobre 2021⁹, la documentation pertinente de l'étude sera conservée pendant 5 ans sur la plateforme healthdata.be. Dans ce délai, il est possible que des informations complémentaires concernant les données soient demandées aux chercheurs, qui devront y répondre. Après 5 ans et au plus tard le 1^{er} janvier 2027, les données seront intégralement supprimées. Seuls les services mentionnés qui ont été communiqués au Comité auront accès aux fichiers de données ID-Cov.
24. Le Comité constate qu'il n'est pas nécessaire pour les chercheurs que la Plate-forme eHealth conserve le lien entre la série de données pseudonymisées et l'identité des personnes.
25. Le Comité estime que le traitement de ces données à caractère est en principe adéquat, pertinent et non excessif.

V. TRANSPARENCE

26. Conformément à l'art. 12 du RGPD, le responsable du traitement doit prendre des mesures appropriées pour fournir toute information en ce qui concerne le traitement à la personne concernée d'une façon concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples. Les informations sont fournies par écrit ou par d'autres moyens y compris, lorsque c'est approprié, par voie électronique.

⁸ Le jeu de données iCAREdata contient des données des postes de garde et des centres de triage. Il s'agit de l'endroit primaire où de nombreuses personnes présumées infectées à la Covid-19 ont été examinées et ces données sont dès lors les plus complètes pour identifier les plaintes et symptômes de personnes présumées infectées à la Covid-19.

⁹ Il est prévu que le rapportage final soit terminé pour le 1^{er} octobre 2022.

27. Le responsable du traitement est dispensé de cette obligation lorsque la communication de telles informations se révèle impossible ou exigerait des efforts disproportionnés, en particulier pour le traitement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques sous réserve des conditions et garanties visées à l'article 89, paragraphe 1, du RGPD.¹⁰
28. L'Université d'Anvers et l'Institut de médecine tropicale ne sont pas en mesure de recontacter les patients, puisque les données de contact ne sont pas communiquées. Ce serait également une charge pour les patients.
29. Le Comité estime par conséquent que la demande répond aux exigences de transparence.

VI. MESURES DE SÉCURITÉ

30. Conformément à l'article 5, f) du RGPD, le demandeur doit prendre toutes les mesures techniques et organisationnelles appropriées pour protéger les données à caractère personnel. Ces mesures doivent garantir un niveau de protection adéquat, compte tenu, d'une part, de l'état de la technique en la matière et des frais qu'entraîne l'application de ces mesures et, d'autre part, de la nature des données à protéger et des risques potentiels.
31. Pour garantir la confidentialité et la sécurité du traitement de données, tout organisme qui conserve, traite ou communique des données à caractère personnel est tenu de prendre des mesures dans les onze domaines d'action suivants liés à la sécurité de l'information: politique de sécurité; désignation d'un conseiller en sécurité de l'information; organisation et aspects humains de la sécurité (engagement de confidentialité du personnel, information et formations régulières du personnel sur le thème de la protection de la vie privée et sur les règles de sécurité); sécurité physique et de l'environnement; sécurisation des réseaux; sécurisation logique des accès et des réseaux; journalisation, traçage et analyse des accès; surveillance, revue et maintenance; système de gestion des incidents de sécurité et de la continuité (systèmes de tolérance de panne, de sauvegarde, ...); documentation.
32. Le Comité constate que l'échange de données mis en œuvre fait appel à l'infrastructure de la plateforme healthdata.be. Les échanges s'effectuent conformément aux dispositions de la délibération n° 15/009 du 17 février 2015, dernièrement modifiée le 3 mars 2020, relative à la méthode générique d'échange de données à caractère personnel codées et non codées relatives à la santé, dans le cadre de healthdata.be et healthstat.be.
33. Conformément à l'article 9, point 3, du RGPD, le traitement de données à caractère personnel relatives à la santé peut uniquement être effectué sous la surveillance et la responsabilité d'un professionnel des soins de santé. Le Comité constate que le traitement de données à caractère personnel s'effectuera sous la surveillance et la responsabilité de trois professionnels des soins de santé, à savoir respectivement au sein de l'Institut de médecine tropicale, au sein de l'Université d'Anvers et au sein de healthdata.be.

¹⁰ Art. 14, alinéa 5, b) du RGPD.

34. Le Comité rappelle que lors du traitement de données à caractère personnel, le professionnel des soins de santé ainsi que ses préposés ou mandataires sont soumis au secret.
35. Le Comité constate que l'Université d'Anvers, l'Institut de médecine tropicale et Healthdata ont chacun désigné un délégué à la protection des données.
36. Il est interdit d'entreprendre toute action visant à convertir les données à caractère personnel pseudonymisées qui ont été communiquées en données à caractère personnel non pseudonymisées.
37. Le Comité rappelle que dans le cadre du traitement à des fins scientifiques, le responsable du traitement doit respecter les dispositions du titre 4 de la loi du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel.
38. Le Comité attire explicitement l'attention sur les dispositions du Titre 6. de la loi du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel, qui prévoit des sanctions administratives et pénales sévères dans le chef du responsable du traitement et des sous-traitants pour la violation des conditions prévues dans le RGPD et la loi du 30 juillet 2018 précitée.
39. Le Comité constate enfin qu'une analyse de risque « small cell » (SCRA) est prévue et que celle-ci sera réalisée par P-95. Le Comité souhaite être informé du résultat.

Par ces motifs,

la chambre sécurité sociale et santé du comité de sécurité de l'information

conclut que la communication des données à caractère personnel, telle que décrite dans la présente délibération, est autorisée moyennant le respect des mesures de protection des données qui ont été définies, en particulier les mesures en matière de limitation de la finalité, de minimisation des données, de limitation de la durée de conservation des données et de sécurité de l'information et sous réserve de la réalisation d'une « Small Cell Risk Analysis » (SCRA) par P-95 afin d'exclure toute possibilité de réidentification des intéressés

Bart VIAENE
Président

Le siège de la chambre sécurité sociale et santé du comité de sécurité de l'information est établi dans les bureaux de la Banque Carrefour de la sécurité sociale, à l'adresse suivante: Quai de Willebroeck 38 - 1000 Bruxelles.
