**AVIS DU CEPD SUR L’UTILISATION PAR L’ECDC DE DONNÉES PROVENANT DE LA PLATEFORME GISAID**
**(dossier 2021-0650)**

# INTRODUCTION

* Le 11 juin 2021, le CEPD a reçu une demande de consultation émanant du Centre européen de prévention et de contrôle des maladies (ECDC) concernant le traitement des données disponibles sur la plateforme GISAID. L’ECDC a fourni des informations supplémentaires en réponse à la demande du CEPD les 9 juillet, 23 août et 5 octobre 2021.
* Le CEPD émet le présent avis conformément à l’article 58, paragraphe 3, point c), du règlement (UE) 2018/1725[[1]](#footnote-1).

# CONTEXTE

## Mandat de l’ECDC et lutte contre la COVID-19

Dans le cadre de ses activités et de son mandat[[2]](#footnote-2), l’ECDC procède à une analyse séquentielle afin de suivre les variants du SARS-CoV-2 (ci-après la «COVID-19») dans le but d’évaluer leur incidence sur la pandémie, par exemple sur la transmissibilité du virus, une diminution éventuelle de l’efficacité des vaccins et des modifications dans la gravité de la maladie.

Conformément à l’article 3, paragraphe 1, du règlement fondateur de l’ECDC, le Centre a pour mission de **déceler, d’évaluer et de communiquer les menaces actuelles et émergentes que des maladies transmissibles représentent pour la santé**.En particulier, l’article 3, paragraphe 2, point a), l’article 5, paragraphe 1, l’article 10, paragraphe 1, et l’article 11, paragraphe 1, autorisent l’ECDC à recueillir, évaluer et diffuser les données scientifiques et techniques pertinentes, y compris à identifier les menaces émergentes pour la santé et à soutenir les activités en réseau des instances compétentes reconnues par les États membres par la gestion des réseaux de surveillance spécialisés et par la fourniture d’une expertise technique et scientifique. L’ECDC complète sa surveillance épidémiologique en intensifiant le séquençage complet du génome afin de mieux déceler et étudier les flambées épidémiques multinationales[[3]](#footnote-3).

S’agissant de la COVID-19 en particulier, le Conseil a explicitement chargé l’ECDC, au point 2, paragraphes 1, 3 et 4, de la [recommandation (UE) 2020/912 concernant la restriction temporaire des déplacements non essentiels vers l’UE](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02020H0912-20210621)[[4]](#footnote-4):

* d’évaluer les variants intéressants et les variants préoccupants (ci-après les «VOC»), sur la base des propriétés essentielles du virus [COVID-19] telles que la transmission, la gravité et la capacité à échapper à la réponse immunologique;
* de traiter les données concernant le «taux de dépistage», «taux de positivité des tests» et les «variants préoccupants et variants intéressants», sur la base des informations mises à la disposition de l’ECDC; et
* de publier et mettre régulièrement à jour une carte de la situation en ce qui concerne les variants préoccupants et les variants intéressants dans les pays tiers.

Afin de surveiller la prévalence des variants du SARS-CoV-2 et d’évaluer leur incidence épidémiologique, l’ECDC a besoin d’informations sur les séquences génétiques des variants de la COVID-19 signalés par les États membres (ci-après les «**données de séquençage**»).

En outre, la contribution scientifique de l’ECDC est la base de l’évaluation de la situation épidémiologique dans les pays tiers et des décisions politiques sur la levée des restrictions actuelles sur les déplacements non essentiels vers l’UE[[5]](#footnote-5).

Pour accomplir cette tâche, l’ECDC utilise notamment[[6]](#footnote-6) les données publiées par la **Global Initiative for Sharing Influenza Virus Data** (initiative mondiale de partage des données sur le virus de la grippe ou «initiative GISAID»).

Le traitement par l’ECDC des données relatives à la COVID-19 mises à disposition par la plateforme GISAID (voir les détails ci-dessous) fait l’objet du présent avis.

## GISAID

### L’initiative GISAID

Selon son site web[[7]](#footnote-7) et les informations supplémentaires fournies ultérieurement par l’ECDC[[8]](#footnote-8):

* la **plateforme GISAID** a été lancée lors de la 61e Assemblée mondiale de la santé en 2008. Créé en tant qu’alternative au modèle de partage dans le domaine public, le mécanisme de partage de GISAID a répondu aux préoccupations exprimées par les États membres en offrant une **base de données accessible au public et conçue par des scientifiques pour des scientifiques afin d’améliorer l’échange de données sur la grippe**. En 2013, la Commission européenne a reconnu GISAID en tant qu’organisme de recherche;
* l’Initiative promeut l’**échange rapide de données** sur tous les virus de la grippe et sur le coronavirus à l’origine de la COVID-19. Cela inclut la **séquence génétique** et les **données cliniques et épidémiologiques connexes associées aux virus humains**, ainsi que des données géographiques et spécifiques aux espèces associées aux virus aviaires et à d’autres virus animaux, en vue d’aider les chercheurs à comprendre comment les virus évoluent et se propagent lors des épidémies et des pandémies;
* s’agissant de sa **gouvernance**, l’Initiative GISAID bénéficie du soutien administratif d’une association allemande à but non lucratif (Freunde von GISAID e.V.), administrée et exploitée à des fins exclusivement caritatives, scientifiques et éducatives. Les activités de GISAID sont régies par plusieurs instances qui opèrent indépendamment les unes des autres. Le Conseil d’administration de Freunde von GISAID e.V., qui est chargé de superviser les activités, doit éviter au maximum les conflits d'intérêts potentiels en matière de sources de financement. La combinaison de fonds publics et privés provenant de diverses entités est censée assurer à GISAID une certaine indépendance vis-à-vis d’intérêts particuliers et nationaux;
* GISAID tient à jour une base de données mondiale sur les séquences génétiques de la grippe, la base de données EpiFluTM (ci-après la «base de données GISAID»). **L’accès libre** aux données de la base de données GISAID est **gratuit** pour toute personne qui accepte de s’identifier et de soutenir le mécanisme de partage GISAID régi par son [accord d’accès à la base de données](https://www.gisaid.org/registration/terms-of-use/)[[9]](#footnote-9). Tous les utilisateurs de bonne foi ayant un identifiant GISAID acceptent le principe de base de respecter un protocole scientifique, en reconnaissant les laboratoires d’origine qui fournissent les spécimens et les laboratoires ayant soumis l’échantillon qui produit la séquence et d’autres métadonnées, en assurant une exploitation loyale des résultats dérivés des données, et tous les utilisateurs conviennent que les données soumises à GISAID ne sont assorties d’aucune restriction, afin de promouvoir une collaboration entre chercheurs fondée sur un partage libre des données et le respect des droits et intérêts de tous.

### La base de données EpiCoV

En janvier 2020, au début de l’épidémie de COVID-19, GISAID a créé une nouvelle base de données baptisée **EpiCoVTM** (ci-après «EpiCoV»)[[10]](#footnote-10), sur le même modèle qu’EpiFlu[[11]](#footnote-11).

Les utilisateurs GISAID (y compris les autorités des États membres) d’EpiCoV remplissent les champs de données suivants[[12]](#footnote-12):

***Champs de données obligatoires:***

* Séquence génétique de l’échantillon du virus SARS-CoV-2
* Date de prélèvement de l’échantillon
* Date de soumission de l’échantillon à GISAID
* Pays de l’échantillon
* Laboratoire de prélèvement de l’échantillon
* Laboratoire ayant soumis les données à GISAID
* Technologie de séquençage utilisée pour analyser l’échantillon

***Champs de données facultatifs:***

* Âge de la personne sur laquelle l’échantillon a été isolé (ci-après le «patient»)
* Sexe du patient
* État clinique du patient (malade, pas malade)
* Informations générales sur la localisation du patient (par ex. ville, mais pas adresse professionnelle ou de résidence)
* État vaccinal du patient.

EpiCoV fonctionne comme une **base de données organisée**, ce qui signifie que l’équipe de conservation de GISAID examine chaque dossier soumis par un utilisateur qui devient accessible au public dans EpiCoV. Cette base de données permet la **collecte et le partage de données génomiques (de séquençage) sur les coronavirus et des informations connexes** (métadonnées), telles que des détails sur le spécimen dont le génome a été isolé et des caractéristiques essentielles des hôtes d’origine (animal, environnement, humain). Les métadonnées sont examinées afin d’**éviter que les données publiées dans EpiCoV puissent permettre de remonter jusqu’aux patients** sur lesquels l’échantillon clinique a été prélevé.

Pour ce faire, le processus de conservation comporte les contrôles suivants:

* le nom de l’entrée (nom du virus) ne peut contenir aucune information telle que le nom, prénom, adresse, identité de la personne, numéro de protocole interne de l’hôpital ou du laboratoire, etc.;
* les informations de localisation ne peuvent contenir aucune information détaillée sur l’adresse de résidence ou professionnelle de la personne. Seules des informations générales sont autorisées, comme la ville, de sorte que le lieu où le patient réside ou séjourne ne puisse pas être retrouvé;
* **seules des informations épidémiologiques générales** sont autorisées, à savoir l’âge et le sexe du patient sur lequel l’échantillon a été prélevé;
* vérifier que, parmi tous les champs des métadonnées, aucune donnée susceptible de permettre l’identification du patient n’ait été fournie par erreur, pour garantir la confidentialité et l’anonymat de chaque patient;
* pour éviter l’identification du patient source ou d’un groupe de patients, il n’est pas permis d’établir un lien entre les entrées de données et d’autres sources d’information, comme des articles de journaux ou d’autres sites web.

Les fournisseurs de données ont la possibilité de communiquer directement – presque en temps réel – avec l’équipe de conservation de GISAID. Ces communications permettent de trouver des solutions en temps utile en cas de problèmes liés à la soumission des données. Lorsque des données douteuses sont détectées, l’équipe de conservation en avise le fournisseur afin de dissiper les préoccupations et de permettre une mise à jour de l’entrée, si nécessaire. Cet échange de messages entre l’équipe de conservation et le fournisseur de données a lieu dans les heures qui suivent le chargement de l’entrée dans EpiCoVTM. Après avoir obtenu les confirmations nécessaires et les mises à jour du fournisseur de données, l’entrée est mise à la disposition du public dans EpiCoV.

## Traitement des données relatives à la COVID-19 extraites d’EpiCoV par l’ECDC

### Valeur ajoutée de la base de données EpiCoV de GISAID pour le travail de l’ECDC

Pour procéder à une analyse séquentielle en vue de suivre les variants de la COVID-19, l’ECDC insiste sur le fait que l’utilisation des données disponibles sur EpiCoV est **nécessaire sur le plan opérationnel**. Selon le Centre, dans le monde, GISAID est la principale entité à fournir un accès libre aux données génomiques sur les virus de la grippe et de la COVID-19. Les données de séquençage génétique et les données cliniques et épidémiologiques connexes associées au virus provenant du monde entier aident l’ECDC à mieux comprendre l’évolution et la propagation de la COVID-19.

L’ECDC ne téléverse aucune donnée dans EpiCoV. Le centre se limite à **collecter et à procéder au traitement ultérieur de données téléversées par d’autres entités** et, en particulier, par les États membres.

Les États membres ne sont pas tenus de fournir des informations à GISAID, mais l’ECDC les encourage fortement à le faire (voir ci-dessous)[[13]](#footnote-13).

En effet, l’ECDC estime que la meilleure méthode est que les États membres:

* signalent les séquences du virus à GISAID;
* signalent les cas à l’ECDC en utilisant le système de signalement fondé sur les cas (TESSy) et incluent l’identifiant GISAID dans les données communiquées de cette manière.

Selon l’ECDC, cette méthode de signalement permet l’analyse la plus approfondie des données, notamment la détection des événements individuels de transmission transfrontière, la détection de nouveaux variants et le suivi des tendances sur le plan géographique, l’analyse des percées en matière de vaccins et des réinfections, les changements de gravité de la maladie, etc.

L’ECDC est toutefois conscient que cela peut nécessiter trop de ressources ou que cela ne serait pas possible autrement dans les États membres. C’est la raison pour laquelle d’autres solutions existent, comme la transmission des séquences à GISAID et la communication d’informations agrégées sur les variants à l’ECDC. Ce modèle permet au Centre de suivre les tendances dans les différents pays et de détecter les nouveaux variants, mais il ne permet pas d’effectuer une analyse épidémiologique détaillée comme le permet le système de signalement fondé sur les cas.

L’ECDC a donc publié un «**Guide pour un suivi représentatif et ciblé du SARS-CoV-2 génomique**» à l’intention des États membres de l’UE/EEE en mai 2021[[14]](#footnote-14). En ce qui concerne le partage de données et le signalement, le guide prévoit ce qui suit (p. 11):

*«Le partage de données et le signalement sont décrits dans le rapport intitulé “Méthodes de détection et d’identification des variants du SARS-CoV-2”, publié par l’ECDC et le Bureau régional de l’OMS pour l’Europe le 3 mars 2021 et le Guide technique de l’ECDC “Séquençage du SARS-CoV-2: première mise à jour”, publié le 18 janvier 2021.*

*Les cas détectés de SARS-CoV-2 doivent être communiqués chaque semaine au système européen de surveillance (TESSy). La détection de nouveaux VOC ou de foyers de VOC circulant actuellement doit être immédiatement signalée en recourant au système d’alerte précoce et de réaction, tandis que la détection de VOC doit être signalée à TESSy chaque semaine.*

*En outre,* ***il est recommandé que les séquences de SARS-CoV-2 soient soumises à GISAID (***[***https://www.gisaid.org***](https://www.gisaid.org)***), la plateforme européenne de données sur la COVID-19 (***[***https://www.covid19dataportal.org***](https://www.covid19dataportal.org)***)[[15]](#footnote-15)****, ou à d’autres bases de données publiques en temps utile (idéalement une ou deux semaines après le prélèvement de l’échantillon).* ***Les données brutes[[16]](#footnote-16) peuvent également être déposées sur la plateforme de données relatives à la COVID-19. Les États membres ne doivent pas téléverser des informations ou des données permettant de relier la séquence de virus soumis à un patient particulier dans une base de données nationale****[[17]](#footnote-17), telles que des identifiants personnels (nom, date de naissance, code d’identification unique, par exemple) ou des données démographiques granulaires (âge du patient, par exemple), etc. Il incombe à l’État membre de veiller à ce que les données de séquençage soient soumises conformément à toutes les règles et dispositions réglementaires pertinentes».*

*Il incombe également aux États membres qui signalent des détections de déterminer si le virus est un variant figurant dans la liste des variants, quelle que soit la méthode utilisée pour la détection ou l’identification. Des variables relatives au signalement de VOC ont été mises en œuvre dans les types de dossiers TESSy fondés sur les cas (NCOV) et dans les types agrégés (NCOVAGGR), où la séquence ID à 12 chiffres du RAPPORT TECHNIQUE DE L’ECDC “Guide de suivi représentatif et ciblé du SARS-CoV-2 génomique" (par exemple* ***les identifiants de GISAID[[18]](#footnote-18)****) doit également être mentionnée. Les numéros d’accès ENA/SRA désignant des données de séquence brutes peuvent également être introduits dans TESSy. Il est également important de partager les lectures de séquençage brutes parce que cela permet des approches bio-informatiques uniformes pour parvenir à un génome consensus, ce qui permet d’éviter des erreurs potentielles dans la génération de séquences consensus, tout en permettant également d’autres analyses (par exemple, de variants minoritaires). Toutes les données épidémiologiques disponibles, y compris la gravité et le pays d’infection probable, devraient être fournies si les données sont soumises en utilisant le type de dossier fondé sur les cas (NCOV). (...)».*

### Identifiant de dossier national attribué par les États membres et code d’identification secondaire attribué par GISAID

Selon l’ECDC[[19]](#footnote-19), les États membres traitent des données relatives à la santé selon une approche fondée sur les cas (données relevant d’un cas particulier de la maladie COVID-19) en utilisant un **identifiant de dossier national**.Cet identifiant de dossier national est attribué par les autorités sanitaires des États membres conformément à leur droit national. Les États membres utilisent les données relatives à la santé fondées sur les cas pour leurs propres analyses. Dans ce contexte, ces données sont soumises aux dispositions du RGPD[[20]](#footnote-20). L’identifiant de dossier national poursuit les principaux objectifs suivants: s’assurer qu’il n’y a pas de double entrée du même virus, être en mesure de mettre à jour les dossiers existants avec de nouvelles informations et permettre aux utilisateurs de demander des informations complémentaires au fournisseur de données en cas d’études impliquant plus d’un laboratoire. Comme indiqué plus haut, cet identifiant de dossier national n’est ni le numéro de passeport ni le numéro de la carte d’identité nationale.

En ce qui concerne en particulier l’analyse par l’ECDC des données de séquençage, l’identifiant de dossier national, qui figure dans le rapport fondé sur les cas mis à la disposition de l’ECDC par l’intermédiaire de **TESSy**, vise à garantir que les données de séquençage analysées sont exactes. Cela revêt une importance capitale pour pouvoir donner suite aux préoccupations éventuelles de l’État membre ainsi que pour faire en sorte que les résultats scientifiques de l’ECDC soient conformes aux méthodes et aux normes scientifiques établies.

L’ECDC n’intervient pas dans l’attribution de l’identifiant du dossier national. **L’ECDC** ne détient pas la clé de l’identifiant de dossier national et **n’a** donc **pas les moyens de retrouver l’identité** du patient dont l’échantillon a été isolé.

Le **code d’identification secondaire** (ou identifiant GISAID), qui est associé à un échantillon particulier de séquences de COVID-19, est créé de manière aléatoire par la plateforme GISAID lorsque les données sont téléversées par les fournisseurs de données (États membres et autres). Le code d’identification de l’échantillon ne contient aucun élément en lien avec le patient dont l’échantillon a été isolé (voir la description du processus de conservation ci-dessus).

L’ECDC a également indiqué que TESSy comporte un champ facultatif où les pays peuvent mentionner l’identifiant GISAID correspondant pour chaque cas. Ce champ a été utilisé par quelques États membres. Dans certains cas, les États membres utilisent toujours le même identifiant pour les signalements dans TESSy et dans GISAID.

### Catégories de données recueillies dans EpiCoV et traitées ultérieurement par l’ECDC

Tous les utilisateurs agréés d’EpiCoV ont accès à l’ensemble des données, à savoir tant les **données de séquençage** (la séquence génomique du virus) que les **métadonnées** (par ex., tranche d'âge, sexe, état clinique comme indiqué plus haut) stockées dans la base de données. Ils n’ont pas accès à l’identifiant de dossier national attribué par les États membres, parce qu’il ne peut pas être introduit dans EpiCoV (voir la note de GISAID sur les garanties concernant la vie privée des patients et la confidentialité). Le code d’identification secondaire aléatoire de l’échantillon est, quant à lui, accessible à tous les utilisateurs agréés d’EpiCoV.

En outre, GISAID fournit à l’ECDC des **données agrégées** sur la prévalence de la COVID-19. Ces données agrégées sont fondées sur les données de séquençage stockées dans EpiCoV. Elles sont utilisées à des fins de surveillance épidémiologique. En particulier, GISAID transmet à l’ECDC des mises à jour instantanées en temps réel des données agrégées sur la prévalence des variants de la COVID-19 dans l’UE/EEE. Ces données sont destinées à être utilisées dans les rapports hebdomadaires de l’ECDC et figurent dans le [tableau de bord de suivi sur la page web de l’ECDC](https://www.ecdc.europa.eu/en/covid-19/situation-updates/variants-dashboard). Étant donné que les données agrégées sont entièrement anonymisées, aucune mesure de sécurité spécifique ne s’applique; elles sont transférées à l’ECDC au moyen d’une interface de programmation sécurisée (API).

Le nombre de membres du personnel scientifique de l’ECDC ayant le statut d’utilisateur agréé est limité. Ces membres du personnel peuvent se connecter à EpiCoV avec leurs identifiants. Ils accèdent à EpiCoV à partir de leur ordinateur professionnel. Ce traitement est soumis à la politique de sécurité de l’information de l’ECDC (ECDC/IP/63) et aux mesures de sécurité TIC standard mises en place pour protéger l’infrastructure TIC du Centre, notamment des pare-feu, une protection antivirus, le chiffrement des données et des mots de passe, lorsque cela est jugé nécessaire. Des utilisateurs non agréés ne peuvent pas consulter les données extraites d’EpiCoV qui font l’objet d’un traitement ultérieur par l’ECDC et l’accès en dehors de l’ECDC n’est possible qu’en utilisant une connexion VPN sécurisée de l’ECDC. Les données sont régulièrement sauvegardées et un logiciel de sécurité est maintenu à jour. Les locaux de l’ECDC ne sont pas accessibles au public et sont sécurisés conformément aux normes de sécurité de l’UE.

L’ECDC ne dispose pas d’une politique de conservation pour les données extraites d’EpiCoV et traitées ultérieurement.

### Questions de l’ECDC

Considérant les informations susmentionnées, l’ECDC souhaite savoir:

* si les données disponibles sur EpiCoV sont considérées comme des données à caractère personnel;
* dans l’affirmative, de quelle nature est le rapport entre l’ECDC et les États membres en ce qui concerne l’utilisation de GISAID; et
* dans l’affirmative, comment l’ECDC devrait traiter les données en interne pour se conformer au règlement.

# ANALYSE JURIDIQUE ET RECOMMANDATIONS

Le présent avis du CEPD porte sur les opérations de traitement effectuées par l’ECDC en ce qui concerne **la collecte et l’utilisation ultérieure de données provenant d’EpiCoV**.Dans ce cadre, le CEPD présente ci-dessous son analyse et ses recommandations.

## Les données traitées par l’ECDC sont-elles des données à caractère personnel?

On entend par «données à caractère personnel» toute information concernant une personne physique identifiée ou identifiable (ci-après la «personne concernée»). Une personne physique identifiable est une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu’un nom, un numéro d’identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale[[21]](#footnote-21).

On entend par «pseudonymisation» le traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que:

* ces informations supplémentaires soient conservées séparément et
* soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable.

Selon le considérant 16 du règlement, les «données à caractère personnel qui ont fait l’objet d’une pseudonymisation et qui pourraient être attribuées à une personne physique par le recours à des informations supplémentaires devraient être considérées comme des informations concernant une personne physique identifiable». **Pour déterminer si une personne physique est identifiable, il convient de prendre en considération l’ensemble des moyens raisonnablement susceptibles d’être utilisés par le responsable du traitement ou par toute autre personne pour identifier la personne physique directement ou indirectement, tels que le ciblage[[22]](#footnote-22).** Pour établir si des moyens sont raisonnablement susceptibles d’être utilisés pour identifier une personne physique, il convient de prendre en considération l’ensemble des facteurs objectifs, tels que le coût de l’identification et le temps nécessaire à celle-ci, en tenant compte des technologies disponibles au moment du traitement et de l’évolution de celles-ci».

En conséquence, les données à caractère personnel qui ont fait l’objet d’une pseudonymisation devraient être considérées comme des informations sur une personne physique identifiable, parce que les informations supplémentaires – telles qu’un code ou un numéro – permettent d’identifier la personne[[23]](#footnote-23). Il est important de relever qu’à la lumière de la définition des données pseudonymes, il est indifférent que les entités qui traitent les données pseudonymes et celles qui détiennent les informations supplémentaires ne soient pas les mêmes. Le fait qu’il s’agisse d’entités différentes ne rend pas anonymes les données pseudonymes. La pseudonymisation est considérée comme une garantie supplémentaire qui devrait être appliquée dans le cadre de la recherche scientifique afin d’assurer le respect du principe de minimisation des données[[24]](#footnote-24).

Dans le même temps, les données à caractère personnel ne sont considérées comme telles que lorsque la personne concernée est identifiable en prenant en considération l’ensemble des moyens raisonnablement susceptibles d’être utilisés par le responsable du traitement ou par toute autre personne pour identifier la personne physique directement ou indirectement (considérant 16 du règlement). De même, dans l’arrêt Breyer (C‑582/14), la Cour de justice déclare que des données peuvent être considérées comme ayant un caractère personnel par rapport à une entité particulière lorsque celle-ci dispose de moyens juridiques lui permettant d’identifier une personne concernée en combinant des données supplémentaires provenant d’une autre source. L’arrêt Breyer n’implique pas que la classification en données anonymes dépende uniquement du fait que des informations supplémentaires peuvent être obtenues «légalement» par le responsable du traitement ou par un tiers. Cet arrêt semble seulement citer en exemple le fait que, lorsque la loi elle-même prévoit la possibilité d’accéder à des informations supplémentaires aux fins de la ré-identification d’une personne concernée, il est objectivement clair qu’un «moyen raisonnablement susceptible d’être utilisé» existe pour ré-identifier la personne concernée. Par conséquent, chaque fois qu’il est possible de ré-identifier la personne concernée, on ne saurait dire que les informations sont anonymes.

Dans le cas d’espèce, où les données sont pseudonymisées, les seules entités qui pourraient avoir la clé pour potentiellement relier l’identifiant du dossier attribué par GISAID à une personne concernée sont les entités qui téléversent ces données (les États membres qui sont des responsables conjoints du traitement et relèvent du RGPD).

Compte tenu de l’arrêt Breyer, l’ECDC souligne que les données téléversées par les États membres (responsables conjoints du traitement avec l’ECDC) pourraient n’être des données à caractère personnel que par rapport aux entités de téléversement (à savoir les États membres eux-mêmes, qui sont les seules entités capables de ré-identifier les personnes concernées). L’ECDC a indiqué que GISAID et tout autre utilisateur ayant accès aux données sur la plateforme GISAID n’ont aucun moyen légal de ré-identifier les personnes concernées et qu’il n’existe pas de moyens raisonnablement susceptibles d’être utilisés qui permettraient cette ré-identification.

À cet égard, le CEPD souligne que l’arrêt Breyer a été rendu avant l’entrée en vigueur du règlement, qui contient désormais la définition de la «pseudonymisation», encore précisée au considérant 16, qui couvre une interprétation large des données pseudonymisées.

En l’espèce, TESSy comporte un champ facultatif où les États membres peuvent inclure l’identifiant GISAID[[25]](#footnote-25). En outre, certains États membres utilisent toujours le même identifiant pour les rapports TESSy et GISAID dans EpiCoV. Par conséquent, l’ECDC est en mesure, à tout le moins dans certains cas, d’établir un lien entre les données obtenues de GISAID (EpiCoV) et les données provenant de TESSy. Ces dernières comprennent l’identifiant de dossier national, qui est une donnée pseudonymisée. Dès lors, bien que l’ECDC n’ait pas accès à la clé de l’identifiant de dossier national, les données collectées par l’ECDC dans EpiCoV devraient être considérées comme des données pseudonymisées à caractère personnel, puisque les personnes concernées restent identifiables indirectement[[26]](#footnote-26), à tout le moins dans certains cas (à savoir lorsqu’il est possible d’établir un lien entre les données stockées dans TESSy et celles stockées dans EpiCoV). Le CEPD comprend bien que l’ECDC ne s’intéresse pas à la ré-identification des personnes concernées en combinant les données de TESSy et de GISAID. Néanmoins, on ne saurait nier que le Centre, en raison de son étroite coopération avec les États membres et du fait que certains États membres utilisent exactement le même identifiant lorsqu’ils communiquent des données à GISAID et à TESSy, est, dans l’absolu, en mesure de procéder à cette ré-identification.

Eu égard aux observations qui précèdent, les échantillons particuliers de séquences de SARS-CoV-2 téléversées dans EpiCoVTM et les informations connexes sont donc des données à caractère personnel qui ont fait l’objet d’une pseudonymisation par GISAID et sont traitées par l’ECDC conformément au considérant 16 et à l’article 3 du règlement.

## Rapport entre l’ECDC et les États membres

L’ECDC considère qu’il agit comme responsable conjoint du traitement aux côtés des États membres en ce qui concerne le traitement des données à caractère personnel provenant de GISAID, tout en soulignant qu'il n’existe pas d’autre solution que la plateforme GISAID et qu’il se trouve dans une situation «à prendre ou à laisser» vis-à-vis de cette dernière.

Premièrement, il importe de définir et d’expliquer la notion de responsable du traitement au sens du règlement. Le responsable du traitement est l'institution ou organe de l'UE, la direction générale ou toute autre entité organisationnelle qui, seule ou conjointement avec d'autres, détermine les finalités et les moyens du traitement de données à caractère personnel[[27]](#footnote-27).

Le responsable du traitement décide pourquoi et comment les données seront traitées.

Le responsable du traitement a un contrôle réel et une influence déterminante sur le traitement et un rôle prépondérant dans celui-ci. Le responsable du traitement peut donc provoquer ou mettre fin à un traitement, tirer profit d’un traitement ou avoir un intérêt dans un traitement[[28]](#footnote-28).

Pour déterminer si une entité est un responsable du traitement ou non au sens du règlement, il convient de prendre en considération un certain nombre d’éléments spécifiques[[29]](#footnote-29).

En particulier, le responsable du traitement décide seul des éléments essentiels d’un traitement, à savoir

* sa base juridique,
* sa ou ses finalités,
* les données à caractère personnel traitées,
* auprès de quelles personnes elles seront collectées et traitées,
* si les données traitées seront divulguées et à qui,
* combien de temps les données à caractère personnel seront conservées (durée de conservation),
* comment les personnes concernées peuvent exercer leurs droits et comment le responsable du traitement répondra à leurs demandes,
* si le responsable du traitement jouit d’une autonomie et d’une indépendance dans le cadre des tâches qui lui incombent,
* si le responsable du traitement choisit un sous-traitant pour effectuer le traitement ou une partie de celui-ci pour son compte, même si le sous-traitant choisi met en œuvre des mesures techniques et organisationnelles spécifiques (éléments non essentiels du traitement).

Si l’entité répond positivement à la majorité des éléments susvisés, elle est un responsable du traitement au sens du règlement.

En l’espèce, l’ECDC extrait des catégories particulières de données de la plateforme GISAID pour accomplir ses missions, telles qu’elles sont exposées plus haut, à la page 2, conformément à son règlement fondateur. L’ECDC détermine la finalité du traitement des données de GISAID ainsi que la plupart des moyens essentiels du traitement, à savoir quelles données seront traitées pour la finalité spécifique, provenant de quelles catégories de personnes concernées, les destinataires, la politique de conservation et son autonomie dans l’accomplissement de ses tâches. À la lumière de ces éléments, l’ECDC exerce un contrôle réel et une influence déterminante sur les données qu’il traite en les extrayant de GISAID et il est donc un **responsable du traitement** au sens de l’article 3, paragraphe 8, du règlement.

Il est demandé si l’ECDC et les États membres peuvent être qualifiés de responsables conjoints du traitement. Il y a responsabilité conjointe du traitement lorsque deux ou plusieurs responsables du traitement et un ou plusieurs responsables du traitement autres que les institutions et organes de l’Union déterminent conjointement les finalités et les moyens du traitement; ils sont les responsables conjoints du traitement[[30]](#footnote-30).

Les États membres introduisent volontairement des données dans GISAID, qu’ils collectent et utilisent pour leurs propres analyses conformément à leur cadre juridique, notamment le RGPD. L’ECDC extrait des données de GISAID pour ses propres évaluations et publications. Le Centre a publié sur son site web un rapport technique contenant des orientations à l’intention des États membres. Bien que tant le Centre que les États membres traitent des données à des fins similaires, ils bénéficient séparément des avantages de GISAID. Le rapport de l’ECDC ne contient que des orientations et des recommandations en vue de parvenir à la meilleure coopération possible, mais ce n’est pas un instrument contraignant pour les États membres et il ne saurait donc être considéré comme exerçant une influence déterminante sur la finalité et les éléments essentiels des moyens.

Ces éléments conduisent à la conclusion que les deux entités sont autonomes dans la réalisation de leurs analyses respectives des données et qu’elles ne déterminent pas conjointement les moyens essentiels du traitement.

Par conséquent, l’ECDC et les États membres ne sauraient être considérés comme des responsables conjoints du traitement au sens de l’article 28 du règlement, mais plutôt comme des **responsables du traitement distincts**.

## Rapport entre l’ECDC et GISAID

L’ECDC souligne qu’il n’a pas besoin de donner d’instructions spécifiques à GISAID en tant que sous-traitant de données à caractère personnel pas plus qu’il ne doit se préoccuper de savoir qui a accès aux données. Le Centre négocie actuellement un accord afin de formaliser sa relation avec GISAID.

Il importe de définir et d’expliquer la notion de sous-traitant au sens du règlement. Un sous-traitant est la personne physique ou morale, l’autorité publique, le service ou un autre organisme qui traite des données à caractère personnel pour le compte et sur instruction du responsable du traitement[[31]](#footnote-31).

Pour déterminer si une entité est ou non un sous-traitant, il convient de passer en revue la liste ci-dessous.

Un sous-traitant ne devrait pas décider des éléments essentiels d’un traitement, à savoir:

* sa base juridique,
* sa finalité,
* quelles données à caractère personnel devraient être collectées et auprès de qui,
* si les données traitées doivent être divulguées et à qui,
* la durée de conservation des données collectées.

Le sous-traitant devrait uniquement servir l’intérêt du responsable du traitement dans l’exécution de la ou des tâches qui lui sont assignées, peut jouir d’une certaine indépendance et peut formuler des suggestions sur des questions relevant de son expertise et sur la manière dont il exécutera ses tâches pour se conformer aux instructions du responsable du traitement (par exemple, quel équipement utiliser, quelles mesures de sécurité prendre).

Toutefois, la décision finale appartient au responsable du traitement après qu’il a été pleinement informé des éléments, des risques et des mesures proposés par le sous-traitant. Le responsable du traitement et le sous-traitant devraient mettre ces décisions à exécution dans le cadre d’un contrat ou de tout autre acte juridique ou instrument contraignant.

En l’espèce, l’ECDC ne donne aucune instruction à GISAID, mais bénéficie d’un accès gratuit et libre à ses services. Le fait que le Centre ait accepté des conditions spécifiques de GISAID pour pouvoir utiliser les données ne fait pas automatiquement de GISAID un sous-traitant. À la lumière de la liste susvisée, un sous-traitant ne devrait pas décider des moyens essentiels d’un traitement, à savoir la base juridique du traitement des données utilisée, sa finalité, auprès de quelles personnes concernées les données sont collectées, etc. Au contraire, l’ECDC a une influence déterminante sur le choix de la base juridique, les finalités et la manière dont les données seront traitées après leur extraction de GISAID ainsi que sur leurs destinataires.

Eu égard aux observations qui précèdent, l’ECDC et GISAID sont des **responsables du traitement distincts**, qui déterminent des finalités et des moyens différents pour le traitement des données, mais sont soumis aux obligations découlant du règlement et du RGPD, respectivement.

## Obligations de l’ECDC en tant que responsable du traitement des données collectées dans EpiCoV

L’ECDC a mentionné qu’«EpiCoV est accessible aux membres accrédités de la communauté scientifique (y compris les boursiers et les chercheurs de l’UE/EEE et de pays tiers). En tant qu’utilisateurs agréés à titre personnel, les membres du personnel de l’ECDC peuvent consulter toutes les données stockées sur EpiCoV et les exploiter pour les publications scientifiques du Centre (par ex. des évaluations des risques). L’ECDC devrait donc donner des instructions claires à son personnel quant à la manière de traiter les données à caractère personnel aux fins de cette opération de traitement.

L’ECDC a également déclaré que «... lorsque des entités autres que des États membres téléversent des données à caractère personnel sans tenir compte des recommandations de l’ECDC, nous ne considérons pas que le Centre soit le responsable du traitement, étant donné qu’il n’est pas en mesure de déterminer pleinement la finalité et les moyens de ces traitements ultérieurs».

Cette situation fait référence au traitement de données d’EpiCoV par d’autres entités que l’ECDC pour leurs propres finalités et n’exonère pas le Centre de sa responsabilité au titre du règlement. En effet, eu égard à ce qui précède, l’ECDC est le responsable de sa propre opération de traitement de données, à savoir les données qu’il collecte dans EpiCoV et traite ultérieurement dans le cadre de son mandat.

Toutefois, considérant que l’ECDC n’est pas en mesure de toujours identifier les personnes concernées, l’article 12 du règlement s’applique. Si l’ECDC est en mesure de démontrer qu’il ne peut pas identifier la personne concernée, par exemple s’il existe une demande de droit d’accès, le Centre n’a pas besoin de procéder à la ré-identification des personnes concernées pour répondre à cette demande. Par conséquent, conformément à cet article, l’ECDC peut être exempté de certaines obligations incombant au responsable du traitement, telles que celles décrites aux articles 17 à 22 du règlement.

En tout état de cause, dans le cadre de son rôle en tant que responsable du traitement, l’ECDC est tenu de se conformer au règlement et de le prouver, même lorsque le traitement de données à caractère personnel ne nécessite pas l’identification des personnes concernées.

Recommandation nº 1 : L’ECDC devrait informer les personnes concernées de cette opération de traitement en publiant un avis général relatif à la protection des données sur son site web, qui mentionnerait uniquement le traitement de données pseudonymisées, sans identification des personnes concernées.

## Recommandations concernant le traitement de données par l’ECDC

### Durée de conservation

L’ECDC n’a pas publié de politique de conservation des données en ce qui concerne le traitement des données de séquençage des virus SARS-CoV-2 et des données à caractère personnel connexes.

Compte tenu du principe de limitation de la conservation[[32]](#footnote-32), le responsable du traitement doit adopter une durée maximale de conservation des données à caractère personnel collectées qui est nécessaire au regard de la finalité pour laquelle elles sont traitées. Les données à caractère personnel peuvent être stockées plus longtemps lorsqu’elles sont traitées à des fins scientifiques conformément à l’article 13, sous réserve de la mise en œuvre de mesures techniques et organisationnelles appropriées[[33]](#footnote-33). L’ECDC traite des données pseudonymisées, ce qui constitue déjà une garantie appropriée pour les droits et libertés des patients.

**Recommandation nº 2 :** L’ECDC devrait fixer une durée de conservation maximale pour les données qu’il traite en tenant compte de la finalité pour laquelle elles ont été utilisées et de la nécessité de les conserver pour une durée (x), conformément à l’article 4, paragraphe 1, point e), du règlement.

### Mesures de sécurité techniques et organisationnelles

Les données à caractère personnel doivent être traitées conformément aux principes de la protection des données énoncés à l’article 4 du règlement afin de garantir la qualité des données. En particulier, le principe d’exactitude [article 4, point d), du règlement] est primordial pour ces opérations de traitement afin de garantir avec une certitude raisonnable que les données sont exactes et tenues à jour.

Conformément à l’article 33 du règlement, en tant que responsable du traitement, l’ECDC devrait mettre en œuvre des mesures techniques et organisationnelles appropriées afin de garantir un niveau de sécurité adapté au risque, tout en cherchant à réduire ces risques pour les droits et libertés des patients.

Comme indiqué à l’article 12, paragraphe 1, du règlement, si les finalités pour lesquelles un responsable du traitement traite des données à caractère personnel n’imposent pas ou n’imposent plus au responsable du traitement d’identifier une personne concernée, celui-ci n’est pas tenu de conserver, d’obtenir ou de traiter des informations supplémentaires pour identifier la personne concernée.

Eu égard au fait que les données sont pseudonymisées et que l’ECDC ne possède a priori pas la clé de pseudonymisation, le CEPD reconnaît que le Centre est libre de prendre les mesures techniques et de sécurité appropriées, qui peuvent être abaissées en fonction des risques pour les personnes concernées.

Néanmoins, en tant que responsable de ce traitement, l’ECDC doit mettre en œuvre des mesures de sécurité pour s’assurer que les données à caractère personnel sont traitées conformément au règlement, notamment en appliquant la protection des données dès la conception et par défaut.

L’ECDC a publié des lignes directrices à l’intention des États membres sur la manière de téléverser des données dans GISAID afin que le Centre puisse les utiliser de manière optimale ultérieurement pour ses propres finalités.

**Recommandation nº 3 :** L’ECDC devrait informer les membres de son personnel qui sont des utilisateurs de GISAID de la finalité de ces opérations de traitement ainsi que de leurs conditions. En outre, l’ECDC devrait recommander que le personnel ne prenne aucune mesure pour obtenir des informations supplémentaires susceptibles de ré-identifier les personnes concernées.

**Recommandation nº 4:** Au vu de l’évolution constante des technologies, y compris la possibilité de ré-identifier des personnes concernées en se fondant sur les informations disponibles dans EpiCoV, l’ECDC devrait régulièrement évaluer le niveau de sécurité de cette opération de traitement, compte tenu de ses risques particuliers.

# CONCLUSION

Le CEPD a analysé le traitement de données à caractère personnel relatif à la collecte et au traitement ultérieur de données relatives à la COVID-19 mises à la disposition du public par GISAID, à savoir la notion de données à caractère personnel, le rapport entre l’ECDC et GISAID, le rapport entre l’ECDC et les États membres et le rôle de l’ECDC en tant que responsable du traitement.

Le CEPD a formulé plusieurs recommandations afin de garantir la conformité du traitement avec le règlement, notamment en ce qui concerne la durée de conservation des données et les mesures de sécurité techniques et organisationnelles.

À la lumière du principe de responsabilité, le CEPD attend de l’ECDC qu’il **mette en œuvre les recommandations susmentionnées, et décide donc de clôturer le dossier**.

Fait à Bruxelles, le 30 novembre 2021

***[signature électronique]***

Wojciech Rafał WIEWIÓROWSKI

1. Règlement (UE) 2018/1725 du Parlement européen et du Conseil du 23 octobre 2018 relatif à la protection des personnes physiques à l’égard du traitement des données à caractère personnel par les institutions, organes et organismes de l’Union et à la libre circulation de ces données, et abrogeant le règlement (CE) nº 45/2001 et la décision nº 1247/2002/CE (JO L 295 du 21.11.2018, p. 39). [↑](#footnote-ref-1)
2. Règlement (CE) nº 851/2004 du Parlement européen et du Conseil du 21 avril 2004 instituant un Centre européen de prévention et de contrôle des maladies (JO L 142 du 30.4.2004, p. 1), (ci-après le «règlement fondateur de l’ECDC»). [↑](#footnote-ref-2)
3. Voir ECDC (en anglais), [Document de programmation unique 2021-2023, page 45)](https://www.ecdc.europa.eu/sites/default/files/documents/Single-Programming-Document-2021-2023.pdf). [↑](#footnote-ref-3)
4. Council Recommendation (EU) 2020/912 of 30 June 2020 on the temporary restriction on non-essential travel into the EU and the possible lifting of such restriction. [↑](#footnote-ref-4)
5. Point 1, paragraphe 2, de la recommandation (UE) 2020/912. [↑](#footnote-ref-5)
6. L’ECDC et les États membres utilisent également le système européen de surveillance (TESSy) pour l’échange d’informations aux fins de la prévention des «maladies transmissibles» qui sont pertinentes pour l’Union, telles que la tuberculose, la rougeole, le SARS, le H1N1, etc. Voir avis du CEPD du 3 septembre 2010: https://edps.europa.eu/sites/default/files/publication/10-09-03\_tessy\_ecdc\_en.pdf [↑](#footnote-ref-6)
7. <https://www.gisaid.org/> [↑](#footnote-ref-7)
8. À la suite de la demande de consultation, le CEPD et l’ECDC ont organisé une vidéoconférence entre leurs équipes le 1er juillet 2021. Le CEPD a ensuite posé des questions supplémentaires (courrier électronique du 2 juillet 2021) et l’ECDC a fourni des informations et des documents supplémentaires (courrier électronique du 8 juillet 2021 et appel téléphonique du 23 juillet 2021). [↑](#footnote-ref-8)
9. <https://www.gisaid.org/registration/terms-of-use/> [↑](#footnote-ref-9)
10. Cf. note présentée par l’ECDC avec sa demande de consultation («GISAID EpiCoV – Note sur les garanties pour la vie privée des patients et la confidentialité»). [↑](#footnote-ref-10)
11. Selon l’ECDC, l’intention de GISAID est que, en l’absence de conditions générales spécifiques pour EpiCoV, les conditions générales publiées concernant l’utilisation d’EpiFlu s’appliquent également à l’utilisation d’EpiCoV, à tout le moins par analogie (cf. courrier électronique de l’ECDC du 23 août 2021). [↑](#footnote-ref-11)
12. Cf. explication fournie par l’ECDC dans son courrier électronique adressé au CEPD le 8 juillet 2021. [↑](#footnote-ref-12)
13. Voir les réponses de l’ECDC aux questions du CEPD par courrier électronique du 8 juillet 2021. [↑](#footnote-ref-13)
14. Le guide est disponible sur le site web de l’ECDC: [rapport technique du 3 mai 2021 sur le «Guide pour un suivi représentatif et ciblé du SARS-CoV-2 génomique»](https://www.ecdc.europa.eu/en/publications-data/guidance-representative-and-targeted-genomic-sars-cov-2-monitoring) [↑](#footnote-ref-14)
15. Caractères gras ajoutés. [↑](#footnote-ref-15)
16. L’ECDC considère les données brutes comme une lecture brute générée par le matériel de laboratoire de séquençage et une analyse de données primaires, comprenant entre des milliers et des millions de courtes séquences non organisées, comme indiqué dans la réponse du 23 août 2021. [↑](#footnote-ref-16)
17. Caractères gras ajoutés. [↑](#footnote-ref-17)
18. Caractères gras ajoutés. [↑](#footnote-ref-18)
19. Voir les réponses de l’ECDC aux questions du CEPD par courrier électronique du 8 juillet 2021. [↑](#footnote-ref-19)
20. 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 (règlement général sur la protection des données ou RGPD). [↑](#footnote-ref-20)
21. Article 3, paragraphe 1, du règlement. [↑](#footnote-ref-21)
22. Caractères gras ajoutés. [↑](#footnote-ref-22)
23. Considérant 16 du règlement. [↑](#footnote-ref-23)
24. Article 13 du règlement. Voir aussi paragraphe 44, page 11, du document du Comité européen de la protection des données sur la réponse à la demande de clarification de la Commission européenne relative à l’application cohérente du RGPD, axé sur la recherche médicale (2 février 2021):
<https://edpb.europa.eu/sites/default/files/files/file1/edpb_replvec_questionnaireresearch_final.pdf> [↑](#footnote-ref-24)
25. Le CEPD a émis un avis de contrôle préalable sur TESSy en 2010, dans lequel il considère que les données traitées par TESSy sont des données à caractère personnel

(<https://edps.europa.eu/sites/default/files/publication/10-09-03_tessy_ecdc_en.pdf>). [↑](#footnote-ref-25)
26. Voir considérant 16 du règlement. [↑](#footnote-ref-26)
27. Article 3, paragraphe 8, du règlement. [↑](#footnote-ref-27)
28. Pour en savoir plus sur la notion de responsable du traitement, pages 7 à 12 des lignes directrices du CEPD concernant les notions de responsable du traitement, de sous-traitant et de responsabilité conjointe du traitement dans le cadre du règlement:
<https://edps.europa.eu/sites/default/files/publication/19-11-07_edps_guidelines_on_controller_processor_and_jc_reg_2018_1725_en.pdf> [↑](#footnote-ref-28)
29. Idem, page 13 des lignes directrices du CEPD. [↑](#footnote-ref-29)
30. Article 28, paragraphe 1, du règlement. Voici quelques exemples de responsables conjoints du traitement existants au sein des institutions de l’UE: l’outil partagé en ligne des RH de l’UE, la base de données d’experts de l’UE, les portails des projets de recherche des agences exécutives, les plateformes sur les maladies rares avec les États membres, etc. Voir également pages 22 à 26 des lignes directrices du CEPD:
<https://edps.europa.eu/sites/default/files/publication/19-11-07_edps_guidelines_on_controller_processor_and_jc_reg_2018_1725_en.pdf> [↑](#footnote-ref-30)
31. Article 3, paragraphe 12, du règlement. Voir pages 15 à 18 des lignes directrices du CEPD:
<https://edps.europa.eu/sites/default/files/publication/19-11-07_edps_guidelines_on_controller_processor_and_jc_reg_2018_1725_en.pdf> [↑](#footnote-ref-31)
32. Article 4, paragraphe 2, du règlement. [↑](#footnote-ref-32)
33. Article 4, paragraphe 1, point e), du règlement. [↑](#footnote-ref-33)