PANTHÉON-ASSAS
UNIVERSITÉ
PARIS

BANQUE DES MÉMOIRES

Master de Droit du numérique
Dirigé par Jérome Passa
2024

L'anonymisation et la pseudonymisation : Clarification et évolution des notions

Jeanne Soubrié

Sous la direction de Lorraine Maisnier-Boché

MASTER 2 DROIT DU NUMERIQUE
UNIVERSITE PANTHEON ASSAS

UNIVERSITÉ PARIS II
PANTHÉON-ASSAS

Année universitaire 2023/2024

L'anonymisation et la pseudonymisation : Clarification et évolution des notions

Mémoire présenté pour l'obtention du Master II par **Jeanne Soubrié**

Sous la direction de **Maître Lorraine Maisnier-Boché**

Dans le cadre du M2 de Droit du numérique de Paris II Panthéon-Assas dirigé par **Maître Jérome Passa**

« La faculté n'entend donner aucune approbation ni improbation aux opinions émises dans ce mémoire, ces opinion doivent être considérés comme propres à leur auteur »

Remerciements

Je tiens à exprimer ma profonde gratitude à toutes les personnes qui ont contribué de manière significative à la réalisation de ce mémoire de recherche.

En premier lieu, je souhaite remercier ma directrice de mémoire, Lorraine Maisnier-Boché, pour son encadrement précieux et ses conseils avisés. Son expertise et son soutien ont été essentiels à l'aboutissement de ce travail.

Je remercie également mon père, pour son soutien émotionnel infaillible et ses relectures attentives. Sa patience et ses encouragements ont été une source de motivation continue tout au long de cette aventure.

Un grand merci à mon amie Lou pour les heures de travail que nous avons partagées en rédigeant nos mémoires respectifs. Nos séances de travail ont été à la fois productives et inspirantes, et je suis reconnaissante pour son soutien et sa camaraderie.

Egalement, je voudrais exprimer ma reconnaissance à mes grands-parents pour leur soutien affectif constant. Leur amour et leurs encouragements m'ont donné la force de persévérer dans les moments difficiles.

Je tiens à remercier mes professeurs et intervenants académiques pour leurs enseignements et leurs conseils tout au long de mon parcours. Leur expertise et leur dévouement ont été essentiels pour acquérir les connaissances nécessaires à la réalisation de ce mémoire. Tout particulièrement, je tiens à exprimer ma reconnaissance à mon directeur de master, Monsieur Jérome Passa. Mes collègues et camarades de classe méritent également toute ma gratitude pour leur soutien e† leurs discussions constructives.

Enfin, je remercie chaleureusement ma famille et mes amis pour leur soutien moral et logistique tout au long de cette aventure.

À tous, je vous adresse mes remerciements les plus sincères.

Jeanne

Principales abréviations

AES : Advanced Encryption Standard

Adresse IP : Adresse de Protocole Internet

CEPD : Contrôleur Européen de la Protection des Données

CJUE : Cour de Justice de l'Union Européenne

CNIL : Commission Nationale de l'Informatique et des Libertés

CRU : Conseil de Résolution Unique

G29 : Groupe de travail Article 29

ICO : Information Commissioner's Office

OCDE : Organisation de Coopération et de développement économique

Protocole HTTPS : Protocole Hyper Text Transfer Protocol Secure

RGPD : Règlement Général sur la Protection des Données

TAC : Technologie Améliorant la Confidentialité

TUE : Tribunal de l'Union Européenne

UE : Union Européen

VIN : Numéro d'Identification du Véhicule

VPN : Virtual Private Network

Sommaire

Introduction

« Data can be either useful or perfectly anonymous but never both. »[1]

Dans un monde de plus en plus connecté, la protection des données personnelles est devenue une préoccupation majeure pour les individus, les entreprises et les régulateurs. La transformation numérique et l'explosion des volumes de données collectées ont mis en lumière la nécessité de méthodes efficaces pour protéger la vie privée des individus et leurs données personnelles Les données personnelles sont définies comme toute information se rapportant à une personne physique identifiée ou identifiable, 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. Deux concepts clés émergent dans ce contexte : l'anonymisation et la pseudonymisation face aux données à caractère personnel.

Avec l'avènement du Règlement Général sur la Protection des Données (RGPD) en Europe, entré en vigueur en mai 2018, le paysage juridique en matière de protection des données a été profondément bouleversé en imposant des obligations strictes aux responsables de traitement et en renforçant les droits des individus sur leurs données personnelles. Selon son article 3, le règlement s'applique d'une part à toute entreprise ou organisation établie dans l'Union européenne (UE), quelle que soit la nationalité des personnes concernées par le traitement des données. Également, même si une entreprise ou une organisation n'est pas établie dans l'UE, si elle traite des données personnelles de personnes résidant dans l'UE dans le cadre de ses activités liées à l'offre de biens ou de services (payants ou gratuits) ou au suivi du comportement des individus se trouvant dans l'UE, elle est tenue de respecter les dispositions du RGPD.

Le Règlement établit plusieurs principes fondamentaux qui guident le traitement des données personnelles et garantissent la protection des droits et de la vie privée des individus. Les principaux concernent la licéité, la loyauté et la transparence, la limitation des finalités, la minimisation des données, l'exactitude des données, l'intégrité et la confidentialité, la responsabilité et la conformité et la limitation des données. Ce dernier principe suppose que données personnelles doivent être

[1] « Broken Promises of Privacy : Responding to te Surprising Failure of Anonymization », 2010, Paul Ohm, Law review 1701, site web Uclala W Review.

conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire aux finalités pour lesquelles elles sont traitées.

Au cœur des préoccupations en matière de protection des données et de ce principe de limitation se trouvent les notions de pseudonymisation et d'anonymisation, deux techniques visant à rendre les données moins identifiables, et donc moins sensibles aux atteintes à la vie privée.

La pseudonymisation est une technique de traitement des données personnelles qui consiste à remplacer les informations d'identification directes d'un individu par des identifiants ou des pseudonymes, de sorte que les données ne puissent plus être directement associées à une personne spécifique sans recourir à des informations supplémentaires. L'objectif de la pseudonymisation est de rendre les données moins identifiables tout en conservant leur utilité pour certaines finalités spécifiques, telles que la recherche, l'analyse ou d'autres activités légitimes. La pseudonymisation ne rend pas les données totalement non identifiables. Ainsi, le RGPD continu de s’appliquer. Cependant, elle réduit le risque de divulgation et de compromission des données en cas de violation de la sécurité, tout en permettant aux organisations de conserver une certaine capacité à utiliser et à exploiter ces données pour des activités autorisées.

L’anonymisation est une technique de traitement des données personnelles visant à rendre ces données totalement non identifiables de manière irréversible, de sorte qu'elles ne puissent plus être liées à une personne spécifique, directement ou indirectement. L'objectif de l'anonymisation est de supprimer complètement les données personnelles de sorte qu'elles ne puissent plus être utilisées pour identifier un individu, même en croisant les informations avec d'autres données disponibles. L'anonymisation altère les données de manière à ce qu'il soit impossible de relier ces données à un individu spécifique, même en utilisant des techniques avancées de recoupement ou d'analyse.

Les principes de pseudonymisation et d'anonymisation sont nés de la nécessité de concilier l'utilisation croissante des données avec le respect de la vie privée et des droits des individus. Ils offrent des méthodes efficaces pour réduire les risques associés à la collecte, au traitement et au stockage des données personnelles, tout en favorisant la conformité réglementaire et en renforçant la confiance dans les pratiques de gestion des données.

Dans quelle mesure la notion d'anonymisation permet-elle une impossible ré-identification des données personnelles, à l'inverse de la pseudonymisation, et a vocation à se clarifier et à évoluer avec l'émergence d'outils techniques innovants ?

Tout d'abord, nous examinerons les définitions essentielles de données personnelles, d'anonymisation et de pseudonymisation, ainsi que les enjeux juridiques, éthiques et sociétaux associés à ces concepts. Nous explorerons également le cadre législatif, en mettant en lumière les principaux textes européens régissant la protection des données. Puis, nous analyserons l'impact de la jurisprudence récente sur les définitions et l'application des notions de données personnelles, d'anonymisation et de pseudonymisation. Nous examinerons les décisions judiciaires clés et leur influence sur l'interprétation et l'évolution de ces concepts dans le contexte juridique actuel. Par la suite, nous explorerons les méthodes et les procédures alternatives émergentes visant à contourner l'application de la notion d'anonymisation. Notamment, nous étudierons les Technologies Améliorant la Confidentialité et les approches innovantes adoptées par des entreprises privées pour répondre aux défis croissants de protection des données tout en essayant de préserver leur utilité. Enfin, nous examinerons les tendances et les perspectives concernant l'évolution des définitions et de la compréhension des notions d'anonymisation et de pseudonymisation. Nous discuterons des facteurs qui pourraient influencer les changements futurs de ces concepts, tels que les avancées technologiques, les développements législatifs et les évolutions sociétales.

À travers cette étude, nous visons à fournir une compréhension approfondie des capacités et des limites de l'anonymisation par rapport à la pseudonymisation. Nous espérons démontrer que, bien que l'anonymisation offre une protection plus robuste contre la ré-identification, elle n'est pas sans défis techniques et pratiques. Par ailleurs, l'émergence d'outils techniques innovants promet de faire évoluer ces pratiques, offrant des solutions plus avancées pour protéger la confidentialité des données personnelles dans un paysage numérique en constante évolution.

Partie 1 : Compréhension des notions et analyse des décisions en la matière

Section 1 : Définitions, enjeux et cadre législatif : comprendre les fondements des notions de données personnelles, anonymisation et pseudonymisation

Afin d'identifier dans quelle mesure les notions de pseudonymisation et d'anonymisation ont vocation à se clarifier et à évoluer avec l'émergence de nouvelles techniques de ré-identification, il convient en premier lieu de définir les notions de données personnelles (I.), d'anonymisation (II.) et de pseudonymisation (III.).

I. L'élargissement de la notion de données personnelles

La notion de données personnelles a été définie précisément par le Règlement Européen pour la Protection des Données personnelles (A.) qui souligne l'importance de distinguer une approche spécifique pour les données sensibles (B.).

A. Une définition précise de la notion

Une donnée personnelle est une information relative à une personne physique identifiée ou identifiable[2]. Selon le Règlement général sur la protection des données (RGPD) de l'Union européenne, une personne physique identifiable est « une personne 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, mentale, économique, culturelle ou sociale ».[3]

Il faut noter que le concept d'identifiabilité est crucial dans la définition des données personnelles. En effet, une personne peut être identifiable même si son identification nécessite des informations supplémentaires qui ne sont pas directement liées à la donnée en question.

---

[2] Article 4 1) 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.

[3] *Ibid*.

Par exemple, une donnée telle que le numéro de téléphone d'une personne est directement liée à son identité et constitue donc une donnée personnelle. De même, une donnée telle que l'adresse IP d'un appareil numérique peut être considérée comme une donnée personnelle si elle peut être associée à une personne physique identifiable.

Si les données personnelles sont à traiter avec la plus grande des précautions, qu'en-est-il des données particulièrement sensibles ?

B. L'approche spécifique des données sensibles

Le RGPD reconnaît une catégorie de « données sensibles »[4]. Ces données font référence à des informations personnelles considérées comme étant particulièrement sensibles en raison de leur nature et du risque potentiel pour les droits et libertés des individus si ils sont atteints.

Elles sont définies ainsi : « les données à caractère personnel qui sont, par nature, particulièrement sensibles du point de vue des libertés et des droits fondamentaux méritent une protection spécifique, car le contexte dans lequel elles sont traitées pourrait engendrer des risques importants pour ces libertés et droits. »[5] Les données sensibles incluent notamment les données de santé, les données biométriques, les données génétiques, les données relatives à l'orientation sexuelle, les données relatives à l'origine raciale ou ethnique, les opinions politiques, les croyances religieuses ou philosophiques, et l'appartenance syndicale[6].

Le RGPD impose des restrictions spécifiques sur le traitement de ces données sensibles en raison de leur nature délicate. Leur traitement est généralement interdit, sauf dans des circonstances exceptionnelles où il existe une base légale spécifique, comme le consentement expresse de la personne concernée[7] ou lorsque le traitement est nécessaire à des fins médicales, de protection de la santé publique, ou pour la constatation, l'exercice ou la défense de droits en justice[8].

---

[4] Considérant 10 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.

[5] Considérant 51 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.

[6] « Donnée sensible », site web de la CNIL.

[7] « Conformité RGPD : comment recueillir le consentement des personnes ? », site web de la CNIL, 3 aout 2018.

[8] *Supra* note 6.

Il est crucial de faire particulièrement attention à ces données lors des processus d'anonymisation et de pseudonymisation. En effet, les données sensibles peuvent avoir un impact significatif sur la vie privée des individus si elles sont divulguées ou utilisées de manière inappropriée. Même après un processus d'anonymisation ou de pseudonymisation, il est important de minimiser le risque de divulgation involontaire ou d'utilisation abusive de ces données.

Alors, quel est le régime à appliquer si un organisme souhaite conserver des données sur une personne concernée mais que les finalités du traitement original ont été réalisées ?

II. Une stricte définition de la notion d'anonymisation

L'anonymisation est une notion qui a été définie précisément par le G29 (A.) qui dresse les techniques reconnues pour y parvenir (B.). Si l'anonymisation est appliquée correctement, le RGPD n'aura pas à s'appliquer (C.). Il pourra être possible d'illustrer une mauvaise application des critères du G29 par une décision rendue à l'encontre de Netflix (D.) et de la commune de Trente en Italie (E.).

A. Les incontournables critères du G29 pour définir la notion

« L'anonymisation est une technique consistant « à supprimer tout caractère identifiant à un ensemble de données » (…). L'anonymisation est donc marquée par le caractère irréversible de la perte du caractère identifiable d'individus. »[9]

Le G29 a contribué à établir des critères rigoureux pour évaluer l'anonymisation des données personnelles[10]. Le G29 est un organe consultatif indépendant composé des autorités de protection des données des États membres de l'Union européenne. Son nom fait référence à l'article 29 de la Directive 95/46/CE sur la protection des données[11]. Selon le G29, les données anonymisées doivent être irréversiblement dissociées de toute possibilité d'identifier les personnes concernées. Autrement dit, « Aucun seuil de risque de ré-identification acceptable n'est envisagé, mais plutôt une définition

[9] « Anonymisation ou pseudonymisation », site web de la Commission de contrôle des informations nominatives.

[10] « L'anonymisation des données personnelles », site web de la CNIL, 19 mai 2020.

[11] « G29 », site web de la CNIL.

sous la forme du risque zéro »[12]. L'anonymisation « est le résultat du traitement des données personnelles afin d'empêcher, de façon irréversible, toute identification »[13].

Pour assurer cette irréversibilité d'identification des données, « les responsables du traitement des données doivent tenir compte de plusieurs éléments, en prenant en considération l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre à des fins d'identification »[14]. Pour ainsi dire, l'anonymisation d'une donnée doit rendre l'identification impossible[15]. Pour qu'une donnée soit considérée comme anonyme, elle doit respecter trois critères principaux : l'individualisation (1.), la corrélation (2.) et l'inférence (3.). Il est important de noter qu'il est nécessaire de réévaluer périodiquement les nouvelles bases de données afin de s'assurer qu'aucun recoupement ne soit possible.

1. L'individualisation

Tout d'abord, la notion d'individualisation selon la CNIL concerne l'ensemble des éléments particuliers et spécifiques d'une personne qui permettent de l'identifier directement ou indirectement à partir des données qui la concernent.[16] Dans le contexte de l'anonymisation des données personnelles, l'individualisation est un critère important à considérer pour évaluer le risque de ré-identification des individus à partir des données anonymisées. En effet, même si les données ont été rendues anonymes, il est possible qu'elles conservent des caractéristiques suffisamment spécifiques pour permettre de relier ces données à une personne particulière.[17]

La CNIL insiste sur l'importance de prendre en compte le risque d'individualisation lors de l'anonymisation des données personnelles. Les techniques d'anonymisation doivent être conçues de

---

[12] « L'enjeu de l'anonymisation à l'heure du big data », Hélène Tanghe et Paul-Olivier Gibert, 25 janvier 2018, Revue Française des Affaires Sociales 2017/4, Cairn.

[13] *Ibid*.

[14] Groupe de travail « Article 29 » sur la Protection des données, 0829/14/FR WP216, Avis 05/2014 sur les Techniques d'anonymisation, adopté le 10 avril 2014.

[15] Conseil d'Etat, 10ème et 9ème chambres réunies, 8 février 2017, n° 393714, Dalloz.

[16] « Diffusion et réutilisation des informations publiques : « Open data » – données ouvertes – », Alexandre Lallet et Pearl Nguyen Duy, septembre 2020, Dalloz.

[17] « Protection des données personnelles : quelles garanties face a l'émergence des dispositifs innovants de sécurité ? », Romain Perray et Hélène Adda, 2019, Dalloz.

manière à réduire autant que possible ce risque, en supprimant ou en modifiant les éléments susceptibles de permettre l'identification des individus[18].

2. La corrélation

Le critère de corrélation est un aspect important à considérer lors de l'anonymisation des données personnelles. Il fait référence à la capacité de relier des données anonymisées à des individus spécifiques en les corrélant à d'autres ensembles de données disponibles.[19]

Lorsque des données sont anonymisées, leur objectif est de ne plus pouvoir être associées à une personne spécifique.[20] Cependant, même après anonymisation, il est possible que les données conservent des caractéristiques qui permettent de les relier à des individus, notamment lorsqu'elles sont croisées avec d'autres ensembles de données.

Le critère de corrélation prend en compte la possibilité de relier des données anonymisées à des individus en identifiant des similarités ou des relations entre ces données et d'autres ensembles de données accessibles. Par exemple, si des données anonymisées contiennent des caractéristiques spécifiques qui correspondent à des profils ou à des informations disponibles dans d'autres bases de données, cela peut permettre de les relier à des individus identifiables[21].

Pour évaluer le risque de corrélation, il est nécessaire de prendre en compte divers facteurs, tels que la nature et la sensibilité des données, la disponibilité d'autres ensembles de données et les techniques d'anonymisation appliquées[22]. Il est également important de considérer les éventuels liens entre les données anonymisées et d'autres sources d'information accessibles de manière publique ou privée.

---

[18] *Supra* note 14.

[19] *Supra* note 14.

[20] *Supra* note 14.

[21] *Supra* note 14.

[22] « Des données à la responsabilité : de l'anonymisation à l'attaque par réidentification », François Viangalli, 1er août 2020, Revue Lamy droit de l'Immatériel n°173.

Autrement dit, « les données anonymes ne doivent pouvoir être ré-identifiées en les croisant avec d'autres jeux de données. Ainsi il doit être impossible de relier deux ensembles de données provenant de sources différentes concernant le même individu. »[23]

3. L'inférence

Enfin, le critère de l'inférence est un aspect crucial à prendre en compte lors de l'anonymisation des données personnelles. Il fait référence à la capacité de déduire des informations spécifiques sur des individus à partir des données anonymisées, même si ces informations ne sont pas directement fournies dans les données elles-mêmes[24].

L'inférence peut se produire lorsque des caractéristiques ou des schémas particuliers dans les données anonymisées permettent de tirer des conclusions sur des individus spécifiques. Par exemple, même si les données elles-mêmes sont anonymisées, des schémas de comportement ou des caractéristiques particulières peuvent permettre de déduire des informations sur les personnes concernées[25]. Pour évaluer le risque d'inférence, il est important de considérer les connaissances et les capacités des utilisateurs finaux qui pourraient tenter d'inférer des informations à partir des données anonymisées.

Pour minimiser les risque d'inférence, de corrélation et d'individualisation, il est recommandé d'appliquer des techniques d'anonymisation robustes et de mener une évaluation approfondie du contexte dans lequel les données seront utilisées. Ces trois critères ont notamment été appliqués par le Conseil d'Etat qui considère qu'une donnée « ne peut être regardée comme rendue anonyme que lorsque l'identification de la personne concernée, directement ou indirectement, devient impossible, que ce soit par le responsable du traitement ou par un tiers. Tel n'est pas le cas lorsqu'il demeure possible d'individualiser une personne ou de relier entre elles des données résultant de deux enregistrements qui la concernent. »[26]

---

[23] « Vos données sont-elles pseudonymisées ou anonymisées ? », 29 mars 2023, site web Association Pour la Sécurité des Systèmes d'Information de Santé.

[24] « Protection des données à caractère personnel : Précisions de la CNIL sur l'anonymisation de données personnelles », Lionel Costes, 19 mai 2020, site web de la CNIL.

[25] *Supra* note 14.

[26] *Supra* note 15.

Après avoir identifié les trois critères cumulatifs permettant de définir l'anonymisation, il convient de se demander comment appliquer en pratique l'anonymisation.

B. Les techniques fondamentales d'anonymisation

Il existe plusieurs techniques d'anonymisation des données, chacune ayant ses avantages et ses limites. Parmi les techniques couramment utilisées, on retrouve la généralisation (1.) et la randomisation (2.).

1. La généralisation

La généralisation est une technique d'anonymisation des données personnelles utilisée pour réduire le risque de ré-identification des individus tout en préservant leur utilité pour l'analyse ou d'autres traitements.[27] Cette technique consiste à regrouper ou à agréger des données afin de masquer les détails spécifiques tout en conservant les tendances ou les informations globales contenues dans les données.

Le processus de généralisation implique souvent la suppression ou la modification des détails précis des données tout en conservant des informations générales ou agrégées[28]. Par exemple, dans le cas des données géographiques, la généralisation peut consister à regrouper les adresses par quartier ou par région au lieu de conserver les adresses exactes des individus.

La généralisation peut être appliquée à différents types de données personnelles, y compris les données géographiques, temporelles, démographiques, etc. L'objectif est de réduire la spécificité des données tout en préservant leur utilité pour les analyses ou les traitements statistiques[29].

Cependant, il est important de noter que la généralisation peut entraîner une perte de précision des données. Plus les données sont généralisées, moins elles sont précises et détaillées. Il est donc nécessaire de trouver un équilibre entre la protection de la vie privée des individus et la pertinence des données pour les finalités prévues.

---

[27] *Supra* note 14.

[28] *Supra* note 14.

[29] « Introduction », Lionel Costes, 1er juin 2020, Revue Lamy Droit de l'Immatériel n°171.

2. La randomisation

La randomisation désigne la technique utilisée pour introduire du bruit aléatoire ou des modifications aléatoires dans les données afin de réduire le risque de ré-identification des individus. « La conséquence étant que les données ne sont plus liées à celle d'origine et nous altérons la véracité de la relation, tout en permettant un traitement statistique de l'ensemble du jeu de données. »[30]

L'objectif de la randomisation dans l'anonymisation des données est de perturber les schémas ou les caractéristiques des données personnelles de manière aléatoire, tout en préservant leur utilité pour l'analyse ou d'autres traitements. Cette technique vise à rendre plus difficile la corrélation des données anonymisées avec des individus spécifiques, tout en minimisant l'impact sur la qualité et la pertinence des données pour les finalités envisagées[31].

La randomisation contribue à renforcer le niveau de protection des données personnelles en introduisant un élément d'incertitude supplémentaire pour les personnes tentant de ré-identifier les individus à partir des données anonymisées[32].

Si le processus d'anonymisation est si complexe entre les critères à remplir et les techniques à appliquer, à quel régime sont soumises les données traitées de cette manière ?

C. La rationnelle non application du RGPD

« Les principes de la protection ne s'appliquent pas aux données rendues anonymes d'une manière telle que la personne concernée n'est plus identifiable »[33].

Lorsque les données personnelles sont anonymisées de manière irréversible, de sorte qu'elles ne peuvent plus être liées à une personne physique identifiable, le RGPD n'est plus applicable à ces

---

[30] « La protection des données personnelles, une obligation pour toutes les entreprises », 2019, Magazine « Sécuriser le traitement des traces numériques dans le cadre du RGPD : anonymisation et pseudonymisation », Cairn.

[31] *Supra* note 14.

[32] *Supra* note 14.

[33] Considérant 26 de la Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

données. En d'autres termes, une fois que les données ont été anonymisées de manière efficace et irréversible, elles ne sont plus considérées comme des données personnelles au sens du RGPD.

Ainsi, l'objectif de l'anonymisation est de réduire le risque de ré-identification des individus tout en préservant l'utilité des données pour les analyses ou les traitements ultérieurs. Lorsque les données sont anonymisées de manière efficace, les principes et les obligations du RGPD[34], tels que le consentement des personnes concernées, le droit à l'information, ou les obligations de sécurité des données, ne s'appliquent plus à ces données. Cela permet d'encourager les organismes à procéder à une véritable anonymisation qui leur permettra ainsi d'échapper au régime contraignant du RGPD.

Que se passe-t-il lorsque les données prétendument anonymisées sont partagées mais que les critères n'ont pas été appliqués correctement dans leur intégralité ?

D. Une illustration de l'échec d'une randomisation par Netflix

« Au cours des dernières années, les travaux sur la ré-identification ont démontré qu'en reliant des données supposées anonymes à des informations extérieures, un « attaquant » pouvait réussir à ré-identifier ces informations. »[35] Il peut arriver que le respect des règles en matière d'anonymisation ne garantisse pas pour autant la protection des données des personnes. Le G29 affirme que « les nombreux exemples d'anonymisation incomplète entraînent par la suite des effets négatifs, parfois irréparables, pour les personnes concernées »[36].

Cela s'est déjà produit notamment avec l'échec de Netflix en matière d'anonymisation lors d'un événement survenu en 2006.[37] Netflix a lancé un concours appelé le Netflix Prize dont l'objectif était d'encourager les développeurs et les chercheurs à améliorer les algorithmes de recommandation de films de Netflix en offrant un prix de 1 million de dollars à l'équipe ou à la personne qui parviendrait à améliorer le taux de précision des recommandations de 10%.[38]

---

[34] « RGPD : un premier bilan », 20 février 2019, Revue Petites affiches n°037, Lextenso.

[35] *Supra* note 12.

[36] *Supra* note 14.

[37] « Vie privée et big data », Jean-Charles Cointot et Yves Eychenne, 2014, Ouvrage *La Révolution du Big Data*, Cairn.

[38] « Les limites de l'anonymisation des données », 22 décembre 2019, site en ligne Malekal.

Netflix a fourni aux participants un ensemble de données contenant des évaluations anonymisées de films attribuées par les utilisateurs, ainsi que des informations sur les films eux-mêmes.[39] Pour garantir l'anonymat des utilisateurs, Netflix avait pris des mesures pour supprimer les informations directement identifiables telles que les noms d'utilisateurs et les adresses e-mail. En effet, « un bruit avait été ajouté dans la mesure où les évaluations avaient été légèrement augmentées ou diminuées. »[40]

Pour autant, malgré les efforts de Netflix pour anonymiser les données, des chercheurs ont démontré en 2007 qu'il était possible de ré-identifier certains utilisateurs en combinant les données du Netflix Prize avec d'autres ensembles de données disponibles publiquement.[41] En effet, en croisant les données de notation anonymisées de Netflix avec les évaluations publiques sur IMDb, il était possible de relier les évaluations anonymisées à des profils d'utilisateurs spécifiques : « il est apparu que 99 % des enregistrements des utilisateurs pouvaient être identifiés de manière unique dans l'ensemble de données en prenant comme critères de sélection huit évaluations et des dates comportant une marge d'erreur de 14 jours »[42].

Cet incident a révélé les limites de l'anonymisation des données et a mis en évidence les risques de ré-identification, même lorsque des mesures sont prises pour supprimer les informations directement identifiables.

De même, une décision a été rendue récemment contre la commune de Trente par l'Autorité italienne de protection des données concernant une insuffisance dans le processus d'anonymisation utilisé.

E. La récente décision de la commune de Trente en Italie concernant une anonymisation insuffisante

La commune de Trente en Italie a procédé à une collecte d'informations publiques auprès des citoyens sans leur demander leur accord préalable, mais en affirmant que l'ensemble des données

---

[39] « Robust De-anonymization of Large Sparse Datasets », A. Narayanan et Shmatikov, Ouvrage *IEEE Symposium on Security and Privacy,* 2008.

[40] *Supra* note 14.

[41] *Supra* note 23.

[42] *Supra* note 14.

avaient été anonymisées, ce qui lui permettait de se soustraire aux obligations du RGPD, notamment concernant la base légale[43].

Deux projets sont concernés : “Marvel” et “Protector”[44]. Le projet Marvel vise à améliorer la qualité de vie et les services aux citoyens, sans violer les limites éthiques et de confidentialité, de manière responsable pour l’intelligence artificielle. Pour cela, la ville de Trente utilise quatorze caméras de vidéosurveillance et des microphones afin d’analyser la collecte d’images et de sons dans l’espace public. La commune affirme que les données sont immédiatement anonymisées lors de la collecte grâce à un ordinateur qui substitue la voix des locuteurs et floutent leurs visages ainsi que toute plaque d’immatriculation présente sur les images. Ce processus est effectué avant l’analyse par les partenaires du projet.

Le projet Protector quant à lui, vise à améliorer la protection des lieux de culte en ville grâce à l’analyse des crimes de haine et des menaces terroristes, ainsi qu’à l’évaluation des mesures de sécurité et des réponses apportées par les forces de l’ordre dans de tels contextes. Pour cela, la municipalité s’appuie là encore sur des images des caméras de vidéosurveillance mais collecte également des messages haineux postés sur la plateforme X. La commune affirme que les données sont immédiatement anonymisées lors de la collecte grâce à ordinateur qui substitue le nom de l’utilisateur par un numéro. Ce processus est également effectué avant l’analyse par les partenaires du projet.

Pour la municipalité de Trente les données étaient ainsi complètement anonymisées, ce qui lui permettait de se soustraire aux obligations du RGPD. L’Autorité italienne de protection des données n’a pas été du même avis. En effet, cette dernière a condamné la commune de Trente pour non-respect des principes du RGPD à une amende de 50 000 euros.

Elle a distingué les deux projets. Concernant le projet Marvel, l’autorité a traité en deux temps l’usage des microphones et l’usage des caméras de vidéosurveillance. Tout d’abord, elle a considéré que la simple substitution de la voix de l’orateur n’est en aucun cas adaptée à l’anonymisation des données personnelles liées à une conversation. En effet, à partir du contenu des conversations, il est

[43] Article 6 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.

[44] Arrêté du 11 janvier 2024, 9977020, Garante Per La Protezione Dei Dati Personali.

possible d'identifier l'orateur, ses interlocuteurs ou les tiers mentionnés dans le discours. De même, la technique de floutage des visages ne peut pas être considérée comme appropriée pour assurer l'anonymisation effective des données, étant donné que les personnes concernées sont dans tous les cas potentiellement identifiables par d'autres caractéristiques physiques ou des éléments contextuels (tels que, par exemple, la corpulence, l'habillement, la position dans la scène filmée, des caractéristiques physiques particulières, etc.), des informations détenues par des tiers (comme, par exemple, des articles de presse relatifs à des événements d'actualité, informations fournies par les personnes dans la scène filmée, etc.), des informations qui peuvent être déduites, par exemple, de l'emplacement de la caméra (zones surplombant certains établissements commerciaux, cabinets médicaux ou écoles), et enfin, des informations relatives à l'itinéraire emprunté par une personne spécifique identifiée dans les images vidéo à travers les caractéristiques physiques et les éléments de contexte susmentionnés, étant donné la possibilité de suivre ses mouvements entre les différentes caméras installées[45].

Concernant le projet Protector, l'Autorité italienne de protection des données affirme que les noms des utilisateurs ayant inscrit des commentaires sur la plateforme Youtube n'ont pas été collectés ce qui permet, de fait, leur anonymisation. Cependant, les noms des utilisateurs ayant postés des messages haineux sur X ont été récoltés et échangés par un numéro unique. Ce processus ne permet pas de respecter le principe d'individualisation requis pour que l'anonymisation soit admise. Dans cette mesure, le simple échange des noms des utilisateurs par un numéro unique constitue seulement une pseudonymisation pour laquelle le RGPD doit s'appliquer.

Cette solution permet d'illustrer la complexité et la difficulté de parvenir à une véritable anonymisation des données selon les critères du G29. Cela a conduit à une sensibilisation accrue aux questions de confidentialité des données et à une prise de conscience des défis liés à la protection de la vie privée dans un environnement numérique où les données sont souvent interconnectées et accessibles. En conséquence, cela a incité les entreprises et les chercheurs à adopter des approches plus robustes pour protéger la confidentialité des données tout en préservant leur utilité pour les analyses et les traitements. Ainsi, à quel régime sont soumises les données qui ont fait l'objet de nombreuses mesures pour les rendre non directement identifiables ?

[45] *Ibid*.

III. Le flou juridique entourant la notion de pseudonymisation

Une définition de la pseudonymisation a été donnée par le RGPD (A.), mais cette dernière soulève certaines difficultés du fait de manque de précisions (B.). Différentes techniques sont envisageables afin de procéder à une pseudonymisation (C.).

A. La tentative de définition de la notion

La pseudonymisation est une technique de protection des données personnelles qui consiste à substituer les identifiants directs ou les données sensibles d'une personne par des identifiants artificiels ou des pseudonymes. Contrairement à l'anonymisation, la pseudonymisation ne rend pas les données totalement non identifiables, mais elle les rend moins directement liées à une personne physique identifiable[46]. La pseudonymisation suppose que « comme moins de personnes ont accès aux informations directement identifiantes, moins d'abus risquent de se produire »[47].

Le RGPD définit la pseudonymisation comme « 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 »[48].

L'objectif de la pseudonymisation est de réduire le risque de ré-identification des individus tout en permettant l'utilisation des données à des fins légitimes telles que la recherche, les analyses statistiques ou d'autres traitements. Les données pseudonymisées peuvent toujours être liées à une personne, mais cela nécessite l'utilisation d'informations supplémentaires qui sont conservées séparément et qui ne sont pas directement accessibles[49].

---

[46] « Recherche scientifique (hors santé) : Enjeux et avantages de l'anonymisation et de la pseudonymisation », 31 janvier 2022, site web de la CNIL.

[47] Chapitre 1 « Les composants de la notion de donnée à caractère personnel », Ouvrage *L'effectivité de la protection des personnes par le droit des données à caractère personnel*, S. Vergnolle, 13 septembre 2022, site web Stradalex Europe.

[48] Article 4 5) 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.

[49] *Supra* note 14.

La pseudonymisation implique souvent la substitution des identifiants directs (tels que le nom, l'adresse, le numéro de sécurité sociale) par des identifiants indirects ou des pseudonymes (tels que des numéros uniques générés aléatoirement)[50]. Ces pseudonymes permettent de dés-identifier les données tout en préservant leur utilité pour les traitements ou les analyses.

Il est important de noter que la pseudonymisation ne dispense pas les responsables de traitement des données de respecter les principes et les obligations du RGPD ou d'autres réglementations sur la protection des données[51]. Les données pseudonymisées restent des données personnelles et sont donc toujours soumises aux principes de licéité, loyauté, transparence, finalité, minimisation des données, exactitude, limitation de la conservation, intégrité et confidentialité.

B. Le manque de précisions concernant la définition

La définition précise de la pseudonymisation présente quelques faiblesses selon les autorités nationales et européennes en matière de protection des données. Ces faiblesses sont principalement liées à l'interprétation et à la mise en œuvre de la pseudonymisation dans le cadre de la protection des données personnelles.

Tout d'abord, les autorités nationales et européennes ont constaté qu'il y avait un manque de clarté dans les directives concernant la pseudonymisation[52]. Les lignes directrices fournies n'étaient pas toujours suffisamment détaillées pour permettre une application cohérente et efficace de la pseudonymisation dans différentes situations et contextes. En effet, « l'avis du groupe G29 pose une vision qui peut être perçue par des patriciens comme trop restrictive sans être suffisamment précise pour fournir des lignes directrices claires aux acteurs du marché de l'IA qui souhaitent intégrer ce type de technologies à leurs produits ou services. »[53] Ainsi, il a été souvent souligné qu'il était nécessaire de fournir des lignes directrices plus détaillées sur la manière de pseudonymiser efficacement les données personnelles tout en préservant leur utilité pour les traitements

[50] « 4054 - Quelles sont les données personnelles indirectement identifiantes ? », avril 2023, Guide Le Lamy droit du numérique.

[51] « Recherche scientifique et protection des données personnelles à l'ère du Big Data », Irene Olivan Garcia, 1er octobre 2019, Revue Lamy droit des affaires n°152.

[52] Délibération n°2020-106 du 29 octobre 2020 portant avis sur un projet de décret relatif au système national des données de santé (demande d'avis n° 20011090), site web Légifrance.

[53] https://www.proquest.com/docview/2077057938/2966E9BC66CD4985PQ/4?accountid=162151

ultérieurs[54]. Ces lignes directrices devaient aborder des questions spécifiques telles que les techniques de pseudonymisation, les exigences de sécurité, et les bonnes pratiques à suivre.

Également, les divergences dans les interprétations de la pseudonymisation ont été relevées[55]. Les différentes autorités nationales, européennes et mondiales ont parfois des interprétations différentes de ce qu'implique exactement la pseudonymisation et de la manière dont elle doit être mise en œuvre dans la pratique. En effet, les autorités ont une approche plus au moins stricte de la notion de pseudonymisation. Par exemple, au Royaume-Uni, le Information Commissioner's Office (ICO) a dressé un guide pratique pour appliquer correctement la pseudonymisation. L'ICO précise « When assessing what pseudonymisation techniques to use and how to implement them you should take into account the type of attacker that may exist. »[56] En prenant en compte l'attractivité de la donnée en fonction de l'attaquant, interne ou externe, l'ICO propose d'adopter les seuils de pseudonymisation. Cette approche n'est pas celle adoptée en France qui se contente d'appliquer les mêmes techniques de pseudonymisation, quel que soit le type d'attaquant qui peut exister.

En réponse à ces faiblesses, les autorités nationales et européennes ont travaillé à élaborer des lignes directrices plus détaillées sur la pseudonymisation et ont encouragé les responsables de traitement des données à adopter des pratiques de pseudonymisation efficaces pour renforcer la protection de la vie privée des individus et assurer la conformité aux réglementations sur la protection des données.

Malgré ces faiblesses dans la définition théorique, les autorités ont-elles dressé des techniques pratiques pour parvenir à une pseudonymisation ?

C. Les principales techniques de pseudonymisation

Le RGPD dispose dans son considérant 32 que « compte tenu de l'état des connaissances, des coûts de mise en oeuvre et de la nature, de la portée, du contexte et des finalités du traitement ainsi que des risques, dont le degré de probabilité et de gravité varie, pour les droits et libertés des personnes

---

[54] *Ibid*.

[55] *Supra* note 50.

[56] « Chapter 3 : pseudonymisation : Draft anonymisation, pseudonymisation and privacy enhancing technologies guidances », février 2022, Information Commissioner's Office (ICO).

physiques, le responsable du traitement et le sous-traitant mettent en oeuvre les mesures techniques et organisationnelles appropriées afin de garantir un niveau de sécurité adapté au risque »[57].

La CNIL a établi la liste des principales techniques de pseudonymisation qui comprennent notamment[58] le chiffrement à clé secrète, le hachage, la tokenisation, le générateur de nombres aléatoires et le compteur. Le chiffrement à clé secrète est une méthode de chiffrement dans laquelle la même clé est utilisée à la fois pour le chiffrement et le déchiffrement des données. Cela signifie que la même clé, est partagé entre les parties autorisées pour chiffrer et déchiffrer les données. Comment cela fonctionne ? Tout d'abord, une clé de chiffrement est générée. Cette clé est une chaîne de bits qui peut avoir différentes longueurs, selon l'algorithme de chiffrement utilisé. Plus la clé est longue, plus elle est difficile à casser par des méthodes de force brute. Pour chiffrer des données, l'algorithme de chiffrement utilise la clé secrète pour transformer les données d'origine en une forme illisible et sécurisée, appelée texte chiffré[59]. Seules les personnes ou les systèmes qui possèdent la clé secrète peuvent déchiffrer ce texte chiffré pour récupérer les données d'origine. Pour déchiffrer les données, le même algorithme de chiffrement est utilisé, mais cette fois avec la clé secrète pour inverser le processus de chiffrement. Cela permet de restaurer les données d'origine à partir du texte chiffré, permettant ainsi aux destinataires autorisés d'accéder aux informations[60].

La technique du hachage est une méthode qui transforme les données en une chaîne de caractères alphanumériques de longueur fixe, appelée empreinte ou hash[61]. Cette empreinte est unique pour chaque ensemble de données d'entrée, mais il est difficile de retrouver les données d'origine à partir de l'empreinte. Pour une même entrée, le même algorithme de hachage produira toujours la même sortie. Cela garantit la cohérence et la fiabilité du processus de hachage. Idéalement, chaque ensemble de données d'entrée devrait produire un cash unique. Cependant, étant donné que la sortie du hachage est souvent de taille fixe, il est possible que deux ensembles de données différents

---

[57] Considérant 32 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.

[58] *Supra* note 46.

[59] « Techniques et meilleures techniques de pseudonymisation : Recommandations sur l'usage des technologies conformément aux dispositions en matière de protection des données et de respect de la vie privée », novembre 2019, European Union Agency For Cybersecurity (ENISA).

[60] Notes de cours de l'étudiante Jeanne Soubrié, « Droit et chaînes de blocs », printemps 2023, Québec, Université Laval.

[61] « La pseudonymisation des données à caractère personnel : une mesure de sécurité à déployer ? », Laure Landes-Gronowski et Marie Miliotis, 6 novembre 2019, site web Agil'it.

produisent le même haché, ce qu'on appelle une collision. Les bons algorithmes de hachage sont conçus pour minimiser les risques de collision.

La tokenisation remplace les données sensibles par des jetons ou des identifiants uniques, généralement stockés dans une table de correspondance[62]. Ces jetons n'ont aucune signification en dehors du contexte de la table de correspondance, ce qui rend difficile la ré-identification des individus à partir des données pseudonymisées. Les tokens sont généralement aléatoires et non prédictibles, ce qui les rend difficiles à deviner ou à reconstruire pour quelqu'un qui ne dispose pas de l'accès au référentiel sécurisé. Cela garantit la confidentialité des données, même en cas d'accès non autorisé à la base de données ou de compromission du système.

Le générateur de nombres aléatoires consiste à remplacer les données sensibles par des valeurs aléatoires ou des pseudonymes générés de manière aléatoire[63]. Ces pseudonymes ne sont pas directement liés aux individus, ce qui réduit le risque de ré-identification. La qualité d'un générateur de nombres aléatoires est évaluée en fonction de plusieurs critères, notamment l'indépendance des nombres générés, l'uniformité de leur distribution, leur périodicité (ou leur absence), leur reproductibilité et leur résistance aux tests statistiques. Les générateurs de haute qualité sont capables de produire des séquences de nombres qui semblent être aléatoires selon une variété de critères statistiques, ce qui les rend adaptés à un large éventail d'applications sensibles.

Le compteur peut être utilisé pour générer des identifiants uniques ou des pseudonymes pour les données personnelles[64]. Plutôt que d'utiliser des informations directement identifiables comme les noms, les adresses ou les numéros de sécurité sociale, les données sont associées à des pseudonymes générés par un compteur. Un compteur est utilisé pour générer des séquences d'identifiants uniques ou de pseudonymes. Chaque fois qu'une nouvelle entrée de données est pseudonymisée, le compteur est incrémenté pour produire un nouvel identifiant. Les pseudonymes générés par le compteur sont ensuite associés aux données personnelles correspondantes. Par exemple, un pseudonyme généré peut être associé à un utilisateur dans une base de données. Les pseudonymes générés par le compteur peuvent être utilisés dans les analyses de données ou les échanges de données entre organisations. Cela permet de protéger la confidentialité des données en

---

[62] « Pseudonymisation des données : principes, techniques et bonnes pratiques », 7 février 2023, site web Vaadata.

[63] « Introduction aux générateurs de nombres aléatoires », site web Ofcm.

[64] *Supra* note 46.

ne divulguant pas les informations personnelles directement identifiables. En utilisant un compteur pour générer des pseudonymes, il est possible de garantir l'unicité des identifiants tout en conservant la cohérence des données[65]. Cela permet de suivre les mêmes individus ou entités à travers différentes sources de données sans révéler leur identité réelle.

Cette liste de techniques non limitatives peuvent être utilisées seules ou combinées ensemble pour pseudonymiser efficacement les données personnelles tout en préservant leur utilité pour les traitements ou les analyses ultérieurs.

Ainsi, comment ces notions d'anonymisation et de pseudonymisation sont-elles appréciées par les juridictions françaises et européennes ? Dans quelle mesure ces notions s'appliquent-elles conformément aux règles édictées par le G29 ? Ces règles laissent-elles la place à une interprétation in concreto par les juridictions ?

Section 2 : L'évolution des notions avec la jurisprudence récente

Comme nous venons de le voir, les notions de donnée personnelles, pseudonymisation et anonymisation ont été définies par le G29. Cependant, ces notions laissent place à quelques interprétations propres aux Cours. C'est pour cela qu'il convient d'analyser certaines de ces décisions notamment la décision Breyer rendue en octobre 2016 qui reconnaît les adresses IP comme données personnelles (I.) ; la décision Gesamtverband Autoteile-Handel de la CJUE rendue en novembre 2023 qui intègre sous condition le Numéro d'Immatriculation du Véhicule dans la notion de données personnelles (II.) ; et la décision du CRU contre le CEPD rendue en avril 2023 qui procède à une remise en cause de l'appréciation de l'anonymisation (III.).

I. La décision Breyer d'octobre 2016 : l'insertion des adresses IP dynamiques dans la notion de données personnelles

La Cour de Justice de l'Union Européenne (CJUE) a rendu une décision retentissante le 19 octobre 2016 permettant de qualifier les adresses de protocole internet (« adresses IP ») de données personnelles. Avant toute chose, il sera nécessaire de comprendre dans quelle contexte cette décision a été rendue (A.), avant d'analyser les deux questions préjudicielles qui ont été posées à la

[65] « La pseudonymisation des données personnelles dans le cadre du RGPD », 20 décembre 2022, site web Le blog data.

Cour de Justice de l'Union Européenne. La première question préjudicielle visait à savoir si les adresses IP dynamiques constituent des données à caractère personnel au sens de la Directive 95/46/CE (B.). La deuxième question préjudicielle visait à savoir si un fournisseur de services de médias en ligne peut collecter et utiliser les données personnelles des internautes sans leur consentement afin de garantir le bon fonctionnement du site (C.).

A. Une nécessaire mise en contexte

Tout d'abord, il est important de noter que cette décision a été rendue au visa de l'article 2 sous a) et de l'article 7 sous f) de la directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données[66]. Cette directive a été abrogée le 25 mai 2018 mais constituait auparavant le texte de référence en matière de protection des données personnelles. C'est aujourd'hui le RGPD qui la remplace. Malgré cela, cette décision reste pertinente dans sa substance quant à la qualification accordée aux adresses IP dynamiques.

Un recours a été mené par Patrick Breyer, un militant allemand pour les droits numériques et la protection de la vie privée contre la République fédérale d'Allemagne[67]. Patrick Breyer contestait le fait qu'en vertu de la loi allemande sur les médias en ligne, le Telemediengesetz, les sites internet pouvaient conserver « les termes entrés dans les champs de recherche, la date et l'heure de la consultation, le volume des données transférées, la constatation du succès de la consultation et l'adresse IP de l'ordinateur à partir duquel la consultation a été effectuée. »[68]

Suite au rejet du recours de Patrick Breyer en première instance et à la réformation partielle de la décision en deuxième instance, la République fédérale d'Allemagne et Patrick Breyer ont formé un recours devant la Cour fédérale de justice allemande, le Bundesgerichtshof. Pour rendre sa décision, le Bundesgerichtshof a saisi la CJUE de deux questions préjudicielles.

---

[66] « Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données », Journal officiel n° L 281 du 23/11/1995 p. 0031 - 0050.

[67] Cour de justice de l'Union européenne, 2ème Chambre, 19 octobre 2016, n° C-582/14, Lamyline.

[68] *Ibid*.

B. La confortante qualification de l'adresse IP comme donnée à caractère personnel sous conditions

Avant toute chose, il faut procéder à la distinction entre les adresses IP statique et dynamique. L'adresse IP statique est une adresse IP qui est assignée de manière permanente à un dispositif sur un réseau[69]. Elle ne change pas et reste la même chaque fois que le dispositif se connecte au réseau. Elle est généralement configurée manuellement par un administrateur réseau ou assignée de manière fixe par un fournisseur de services internet (ISP). L'adresse IP dynamique change à chaque nouvelle connexion à l'internet et ne permet pas, par elle-même, d'identifier directement un utilisateur spécifique[70].

La question préjudicielle posée à la CJUE était alors de savoir si une adresse IP dynamique pouvait être considérée comme une donnée à caractère personnel dans le contexte où une partie tierce (comme un fournisseur d'accès à internet) détient des informations supplémentaires qui, combinées à l'adresse IP dynamique, permettent l'identification d'un utilisateur[71].

La CJUE a conclu que les adresses IP dynamiques constituent des données à caractère personnel au sens de la Directive 95/46/CE, sous certaines conditions. Tout d'abord, la CJUE a jugé qu'une adresse IP dynamique peut être considérée comme une donnée à caractère personnel si le responsable du traitement (par exemple, l'opérateur du site web) dispose de moyens légaux permettant de combiner cette adresse IP avec des informations supplémentaires détenues par un tiers (comme le fournisseur d'accès à internet) afin d'identifier l'utilisateur. D'autre part, la Cour a précisé que l'identification doit être réalisable par des moyens « que le responsable du traitement ou une autre personne est raisonnablement susceptible d'utiliser »[72] pour identifier l'utilisateur. Cela implique une évaluation concrète des moyens dont dispose le responsable du traitement.

Cette qualification découle du fait que les adresses IP peuvent être utilisées pour identifier de manière indirecte un individu, surtout lorsque ces données sont combinées avec d'autres informations détenues par un fournisseur de services Internet.

---

[69] « Document de la Commission COM (2002) 96 final du 21 février 2002 : Communication de la Commission au Conseil et au Parlement européen - L'internet nouvelle génération : priorités d'actions dans la migration vers le nouveau protocole internet IPv6 », Dalloz.

[70] « Chapitre 131 - Données de connexion », Livre 1 *Les données à caractère personnel*, Christiane Féral-Schuhl, 2020-2021, Dalloz.

[71] *Supra* note 67.

[72] *Supra* note 67.

Ainsi, la CJUE a considéré que la conservation des adresses IP dynamiques et d'autres données de communication électronique devait respecter les principes de proportionnalité et de nécessité, et ne pouvait être justifiée que dans le cadre de la lutte contre la criminalité grave, tout en tenant compte des droits fondamentaux à la vie privée et à la protection des données des individus[73]. Cette décision a des implications importantes pour la protection des données, car elle élargit la définition de ce qui peut être considéré comme une donnée personnelle, en prenant en compte non seulement les informations directement identifiables mais aussi celles qui peuvent permettre une identification indirecte via des informations supplémentaires.

C. La possibilité troublante par le fournisseur de services de médias en ligne de collecter les données personnelles des utilisateurs sans leur consentement

Concernant la question de savoir si les fournisseurs de services de médias en ligne pouvaient collecter des données personnelles des utilisateurs sans leur consentement, la CJUE a répondu en se référant aux conditions établies par la Directive 95/46/CE pour le traitement des données personnelles. Selon l'article 7(f) de cette directive, le traitement des données personnelles est licite si « le traitement est nécessaire aux fins des intérêts légitimes poursuivis par le responsable du traitement ou par le ou les tiers auxquels les données sont communiquées, à condition que ne prévalent pas les intérêts ou les libertés et droits fondamentaux de la personne concernée. »[74]

Tout d'abord, la CJUE a reconnu que la collecte et l'utilisation des adresses IP dynamiques par un fournisseur de services de médias en ligne peuvent être justifiées par un intérêt légitime, notamment pour assurer la sécurité et le bon fonctionnement de ses services, protéger ses infrastructures contre les cyberattaques, et permettre la traçabilité en cas de tentative d'accès non autorisé[75]. Egalement, la Cour a souligné qu'il est nécessaire de mettre en balance les intérêts légitimes du fournisseur de services et les droits fondamentaux des utilisateurs, notamment leur droit à la vie privée et à la protection des données personnelles[76]. Cette évaluation doit être faite au cas par cas. La CJUE a conclu qu'un fournisseur de services de médias en ligne peut, sous certaines conditions, collecter et utiliser les données personnelles des internautes sans leur consentement préalable. Cependant, cette

---

[73] *Supra* note 50.

[74] *Supra* note 46.

[75] *Supra* note 22.

[76] *Supra* note 67.

collecte doit être strictement nécessaire pour les intérêts légitimes du fournisseur et ne doit pas porter atteinte de manière disproportionnée aux droits et libertés fondamentaux des utilisateurs[77].

Ces conclusions de la CJUE ont des implications importantes pour la protection des données personnelles et la gestion des adresses IP dynamiques par les fournisseurs de services en ligne, en renforçant la nécessité d'une évaluation équilibrée entre les intérêts légitimes des entreprises et les droits des utilisateurs.

Est ce que le VIN, numéro d'identification du véhicule, constitue également une donnée personnelle ?

II. La décision Gesamtverband Autoteile-Handel de la CJUE de novembre 2023 : l'intégration sous condition du VIN dans la notion de données personnelles

Cette décision rendue par la CJUE qui reconnaît le Numéro d'Identification du Véhicule (VIN) comme une donnée personnelle (A.) a été fortement contestée à cause de son manque de clarté (B.).

A. L'étonnante reconnaissance du VIN comme donnée personnelle indirecte

La décision de la Cour de Justice de l'Union Européenne le 9 novembre 2023 porte sur la qualification du VIN comme donnée personnelle[78].

En l'espèce, le litige opposait l'association professionnelle allemande du commerce de gros de pièces automobiles Gesamtverband Autoteile-Handel eV à la société Scania CV AB, un constructeur de véhicules suédois. Le règlement allemand 2018/858 « requires car manufacturers to make accessible to independent operators, including repairers, spare parts distributors and publishers of technical information, the information necessary for the repair and maintenance of the vehicules they manufacture. »[79] L'association a saisi un tribunal allemand, affirmant que les informations fournies par le constructeur de poids lourds Scania à ses membres ne respectent ni la

[77] « CJUE : les adresses IP « dynamiques » sont des données personnelles au sens du droit de l'Union », Elisabeth Autier, 8 novembre 2016, Dalloz actualités.

[78] Décision Cour de justice de l'Union européenne, 3e chambre, 9 novembre 2023, n° C-319/22, Dalloz.

[79] « Obligation on car makers to communicate vehicle ID numbers is compatible with GDPR », Thomson Reuters, Westlaw Edge UK.

forme ni le contenu requis par les obligations légales de ce règlement[80]. Le tribunal a saisit la Cour de Justice de l'Union Européenne pour répondre à trois questions préjudicielles.

La question posée à la CJUE qui nous intéresse consistait à déterminer si les VINs devaient être traités comme des données personnelles, nécessitant des protections supplémentaires en vertu du RGPD[81].

La Cour de justice de l'Union européenne va clarifier l'inclusion du VIN dans la notion de données personnelles. La Cour a jugé que le VIN, lorsqu'il est associé à des informations permettant l'identification d'une personne physique, constitue une donnée personnelle[82]. La décision repose sur le fait que le VIN peut indirectement identifier une personne en combinant le numéro avec d'autres informations disponibles (par exemple, les bases de données des propriétaires de véhicules). C'est à cette condition seulement que le VIN sera considéré comme une donnée personnelle. En effet, la Cour assure que le VIN sera considéré comme tel dans la mesure seulement « où celui qui y a accès pourrait disposer de moyens lui permettant de l'utiliser pour identifier le propriétaire du véhicule auquel il se rapporte ou la personne pouvant disposer de ce véhicule à un titre juridique autre que celui de propriétaire. »[83] Cela signifie que le VIN constitue une donnée personnelle uniquement si la personne qui y a accès peut ré-identifier le propriétaire grâce à la base de données permettant la corrélation. En effet, « La CJUE a jugé que le VIN, étant un simple code alphanumérique, ne constituait pas en soi une donnée personnelle. Toutefois, cette évaluation pourrait changer si le certificat d'enregistrement était également disponible et qu'une personne physique y était inscrite. »

Cette intégration sous condition renforce la protection des données personnelles en garantissant que des informations comme le VIN, qui peuvent identifier indirectement une personne, soient traitées conformément aux exigences de confidentialité et de protection des données. Cela a pour conséquence que les constructeurs doivent traiter les VIN avec les mêmes précautions que les autres données personnelles, garantissant leur traitement conforme aux normes de protection des données.

---

[80] *Ibid*.

[81] *Supra* note 78.

[82] *Supra* note 78.

[83] *Supra* note 78.

Ainsi, cette décision marque une étape importante dans l'extension de la notion de données personnelles, intégrant des éléments techniques comme les numéros de véhicules lorsqu'ils sont associés à des informations permettant l'identification des individus, renforçant ainsi la protection de la vie privée.

Cependant, cette décision présente de nombreuses limites.

B. Les limites de cette décision par le manque de clarté

Bien que cette décision apparaisse protectrice au regard du RGPD en adoptant une vision large de la notion de donnée personnelle, la CJUE ne précise pas réellement dans quelles mesures la condition peut s'appliquer. Bien que le VIN soit reconnu comme une donnée personnelle lorsqu'il permet d'identifier une personne physique, la décision n'explicite pas clairement quels types de données ou contextes spécifiques justifient cette association. La CJUE précise seulement que le VIN « acquiert ce caractère (personnel) à l'égard de quiconque dispose raisonnablement de moyens permettant de l'associer à une personne déterminée. »[84] Cependant, cette notion de raisonnabilité n'est détaillée ni dans la décision, ni dans le RGPD (ce qui fera l'objet d'une explication complète à la partie 2, section 2). Il apparait probable que le critère de « raisonnabilité » soit interprété de manière large par les Cours pour s'assurer que le VIN soit bien considéré comme une donnée personnelle. Ainsi, pour les entreprises, surtout les constructeurs automobiles et les opérateurs indépendants, la mise en œuvre de cette décision peut s'avérer complexe. L'absence de directives précises sur la manière de traiter les VINs en tant que données personnelles peut mener à des interprétations variées et à une application incohérente de la protection des données.

De fait, la décision peut nécessiter des changements significatifs dans les systèmes et processus existants pour le traitement des données. Les entreprises doivent non seulement adapter leurs politiques de confidentialité mais aussi leurs systèmes techniques pour s'assurer que les VINs sont traités conformément aux réglementations sur les données personnelles, ce qui peut être coûteux.

Enfin, « la CJUE ne précise pas non plus clairement si l'évaluation de la présence de données personnelles doit être (uniquement) effectuée du point de vue de la partie qui détient les données

[84] *Supra* note 78.

(« approche relative ») ou également du point de vue de tiers. (« approche objective »). »[85] On retrouvera justement ce manque de clarté quant au point de vue adopté dans la décision du CRU contre le CEPD d'avril 2023.

III. La décision CRU c. CEPD d'avril 2023 : une remise en cause de l'appréciation de l'anonymisation

La décision rendue par le Contrôleur Européen de la Protection des Données (CEPD) contre le Conseil de Résolution Unique (CRU) n'avait pas fait de vague car elle affirmait simplement que les données personnelles sont considérées comme pseudonymisées lorsqu'il est possible de les recouper par la suite, afin de ré-identifier les individus, grâce à une base de données qui regroupe tous les identifiants, position couramment adoptée (A.). Cependant, le Tribunal de l'Union Européenne (TUE) a condamné cette première décision en adoptant une nouvelle définition de l'anonymisation (B.). Cette décision intéressante laisse place à de nouveaux questionnements et présente certaines limites (C.).

A. La décision attendue du CRU par le CEPD

C'est le 26 avril 2023 que le Tribunal de l'Union Européenne a rendu une décision majeure concernant les notions de pseudonymisation et de données personnelles[86]. Le litige opposait d'une part le Conseil de Résolution Unique (CRU), une agence de l'Union européenne établie pour gérer la résolution ordonnée des défaillances bancaires au sein de l'Union bancaire européenne[87] et d'autre part, le Contrôleur Européen de la Protection des Données (CEPD) qui joue un rôle crucial dans la supervision et la protection des données personnelles au sein des institutions, organes et agences de l'Union européenne[88].

En l'espèce, le CRU traitait les informations liées aux actionnaires et créanciers de la Banco Popular Español et souhaitait procéder à l'évaluation des actifs de la banque[89]. Pour procéder à cela, le CRU a mandaté le cabinet Deloitte et lui a transféré des commentaires émanant d'actionnaires de

---

[85] « Decoding Data? ECJ's verdict on Vehicle Identification Numbers as personal data », Lennart Schübler, 9 novembre 2023, site web Bird&bird.

[86] Décision du Tribunal de l'Union Européenne, 8e chambre élargie, 26 avril 2023, T-557/20, Lexbase.

[87] « Single Resolution Mechanism », European Council.

[88] Page d'accueil du sire web European Data Protection Supervisor.

[89] *Supra* note 86.

la banque en supprimant les identifiants directs. Pour cela, un code alphanumérique a été appliqué aux commentaires. Un code alphanumérique est une séquence de caractères qui peut inclure à la fois des lettres (de l'alphabet) et des chiffres[90]. L'ensemble des codes alphanumériques et des informations s'y rattachant étaient stockées dans une base de données à laquelle Deloitte n'avait pas accès. Cependant, cinq actionnaires ont introduit une plainte contre le CRU auprès du CEPD.

En l'espèce, le CEPD reprochait au CRU de ne pas avoir alerté les personnes concernées que leurs données avaient été communiquées au cabinet Deloitte[91]. Le CRU arguait que les données avaient été anonymisées lors du transfert ce qui le dispensait d'informer les personnes concernées du transfert de leurs données. Ainsi, le CEPD reprochait au CRU de ne pas avoir respecté ses obligations en matière de protection des données personnelles en n'ayant pas réellement anonymisé les données et en procédant seulement à une pseudonymisation[92].

Le CRU a estimé, en se positionnant à sa place, que les données transférées à Deloitte ne pouvaient pas être ré-identifiées car le cabinet ne disposait pas de la base de données. En effet, « pour le CRU, aucune donnée à caractère personnel n'était communiquée à Deloitte qui n'avait eu accès qu'à des réponses à des questionnaires auxquelles étaient attribués des codes alphanumériques »[93]. Cependant, le CEPD a considéré que cela n'était pas suffisant et que le CRU devait également se positionner à la place de Deloitte afin d'identifier si le cabinet était en capacité de procéder à la ré-identification des données. Le CEPD a estimé que cette absence de mesures adéquates augmentait le risque de ré-identification des individus concernés et que les données devaient alors êtres considérées comme seulement pseudonymisées. Quelle a été la décision du Tribunal de l'Union européenne ?

B. L'étonnante condamnation du Tribunal de l'Union Européenne en rappel à la décision Breyer

Pour rendre sa décision, le Tribunal de l'Union Européenne va raisonner en deux temps.

---

[90] Délibération de la CNIL n° 2008-005 du 10 janvier 2008 Objet : portant autorisation unique de mise en oeuvre par les entreprises ou organismes exploitants de médicaments de traitements automatisés de données a caractère personnel relatifs a la gestion des données de santé recueillies dans le cadre de la pharmacovigilance des médicaments postérieurement a leur mise sur le marché.

[91] *Supra* note 86.

[92] « Anonymisation through separation : what recent cases teach us about the EU's anonymisation standards », Lore Leitner, Gabe Maldoff et Mickey Lee, site web Westlaw Edge UK.

[93] « Notion de « donnée à caractère personnel » : les précisions du TUE », Lionel Costes, 2 mai 2023, Le Lamy de Droit de l'Immatériel.

Tout d'abord, concernant la notion de données personnelles[94], le TUE reproche au CEPD de ne pas avoir correctement qualifié la notion de données personnelles dans les faits de l'espèce. En effet, il a considéré que « toute opinion personnelle constituait une donnée à caractère personnel. Il a également admis ne pas avoir examiné le contenu des commentaires produits par les réclamants lors de la phase de consultation. »[95] Cependant, « pour le Tribunal, cet examen n'aurait pas dû se faire in abstracto sur la base de cette présomption, mais in concreto, et ce, en prenant en compte le contenu, la finalité et l'effet du traitement des informations transmises a Deloitte »[96]. Ainsi, le TUE affirme que pour considérer une opinion personnelle comme une donnée personnelle, il faut analyser la situation de l'espèce et ne pas associer automatiquement une opinion personnelle à une donnée à caractère personnel. En effet, le CEPD « s'est appuyé sur une forme de présomption de la nature personnelle des données »[97] qui n'avait pas lieu d'être selon le TUE.

Dans un second temps, le TUE s'est fondé sur le considérant 16 du Règlement (UE) 2018/1725[98] (similaire au considérant 26 du RGPD) qui suppose que « 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 »[99]. C'est pour cela que le Tribunal affirme que pour qu'une donnée soit considérée comme une donnée à caractère personnel, il n'est « pas requis que toutes les informations permettant d'identifier la personne concernée doivent se trouver entre les mains d'une seule personne. »[100] Le Tribunal affirme ainsi qu'il faut seulement analyser le risque ré-identification au regard de la personne qui détient les données initialement, non pas au regard du nouveau détenteur.

---

[94] « Décision CRU c/ Deloitte : Une nouvelle pierre a l'édifice de la notion de donnée a caractère personnel », Lorette Dubois, 1er juin 2023, Dalloz actualités.

[95] *Supra* note 86.

[96] *Ibid*.

[97] « Droit des données personnelles », Lorraine Maisnier-Boché, 9 septembre 2023, Communication Commerce Electronique n°9.

[98] 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) n° 45/2001 et la décision n° 1247/2002/CE.

[99] *Ibid*.

[100] *Supra* note 86.

En effet, dans sa décision, « le TUE juge que le CEPD aurait dû rechercher si les auteurs des informations transmises à Deloitte étaient directement identifiables par Deloitte ou si Deloitte disposait de moyens légaux et réalisables en pratique lui permettant d'accéder aux informations supplémentaires nécessaires à la réidentification des auteurs des commentaires. »[101] En l'absence de cette analyse, le TUE a annulé la décision du CEPD qui avait considéré que les données transmise à Deloitte étaient seulement pseudonymisées. Le TUE insiste sur le fait qu'il faut prendre en compte « 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 »[102]. En l'espèce, le TUE a considéré que les données transmises à Deloitte ne pouvaient pas être ré-identifiées par Deloitte et constituaient ainsi des données anonymisées. En effet, « Deloitte n'aurait pas eu et n'aurait toujours pas de moyens légaux d'accéder aux informations supplémentaires et d'identification. »[103] Cette approche in concreto permet d'adapter la notion d'anonymisation en fonction des faits de l'espèce. En effet, « le Tribunal invite à une approche relative de la donnée personnelle et (considère) qu'une donnée pourrait désormais être considérée comme anonymisée au sein d'un organisme et pseudonymisée ou personnelle au sein d'un autre. »[104] Mais, quelles sont les limites d'une telle décision ?

C. Les multiples limites d'une telle décision

La décision T-557/20 de la Cour de justice de l'Union européenne (CJUE) présente plusieurs limites et points de tension, en particulier en ce qui concerne la protection des données personnelles et l'application de l'anonymisation.

Tout d'abord, en l'espèce, Deloitte n'est pas soumis aux obligations du RGPD pour les données récoltées, car elles ne permettent pas, en l'état, d'identifier des individus spécifiques. Cette situation soulève des préoccupations sur le niveau de protection des données personnelles lorsqu'elles sont traitées par des tiers. En effet, en appliquant cette décision, « le destinataire des informations ne sera, lui, pas soumis à la LPD (ou au RGPD), puisque les données ne seront pas pour lui des

---

[101] « Des données personnelles pseudonymisées transférées à un tiers deviennent-elles anonymes ? », Alexandre Jotterand, 13 juin 2023, site web SwissPrivacy.

[102] *Supra* note 98.

[103] *Supra* note 86.

[104] *Supra* note 97.

données personnelles »[105]. De cette manière, Deloitte ne sera pas tenu responsable en cas de fuite de données. Cela pose évidemment des questionnements sur les responsabilités qui pourront être engagées en cas de fuite de données chez le destinataire. De ce fait, les personnes concernées pourraient voir leurs droits limités, car les protections offertes par le RGPD ne s'appliquent pas pleinement si les données sont considérées comme anonymisées et non comme des données personnelles. Les individus perdent alors certaines garanties de protection des données, telles que le droit d'accès, de rectification ou d'effacement.

Egalement, même si Deloitte ne peut pas directement ré-identifier les données, il existe toujours un risque que des informations supplémentaires puissent être combinées à ces données prétendument anonymisées, permettant une ré-identification indirecte. La décision ne traite pas en profondeur de ce scénario, laissant une certaine zone d'incertitude[106]. Cela pourrait arriver notamment « si la situation factuelle change, par exemple parce que le destinataire transmet à son tour les données (qui sont anonymes pour lui) à des tiers, ou les publie sur internet, l'analyse juridique change également et des données qui ont été anonymes pendant un moment peuvent soudainement redevenir personnelles. »[107] Dans ce cas, les données en la possession de Deloitte initialement considérées comme anonymisées mais non personnelles pourraient être ré-identifiées, ce qui remettrait en cause la qualification initiale.

Enfin, la décision reflète la complexité des réglementations en matière de protection des données, où l'application des règles du RGPD dépend fortement de l'interprétation et de la mise en œuvre des techniques d'anonymisation. Cette complexité peut entraîner des incohérences dans la protection des données et des difficultés pour les entités comme le CRU et Deloitte à naviguer dans ces obligations.

Ainsi, la décision de la CJUE présente des limites importantes en termes de clarté et de mise en œuvre des mesures de protection des données. Elle souligne la nécessité d'un équilibre entre la transparence et la protection des données personnelles, tout en laissant des zones d'ombre sur les exigences précises de conformité sous le RGPD en laissant place à une approche in concreto[108]. Ces

---

[105] *Supra* note 101.

[106] *Supra* note 94.

[107] *Supra* note 105.

[108] « Nouvel éclaircissement de la notion de données personnelles », Jerôme Lasserre Capdeville, 2023, Dalloz IP/IT 2023. 540.

ambiguïtés peuvent poser des défis pour les institutions cherchant à respecter les réglementations tout en protégeant efficacement les données personnelles. Cette décision est alors extrêmement contestable ce qui explique l'appel qui a d'ores et déjà été formé.

Au vu du manque de clarté des définitions de données personnelles, pseudonymisation et anonymisation, quelles sont les évolutions à venir ? Quelles sont les alternatives qui pourront être utilisées pour recourir à des notions mieux définies ?

Partie 2 : Les alternatives à l'anonymisation et la pseudonymisation et les évolutions à venir

Section 1 : L'adoption de procédures alternatives à l'anonymisation

Afin de contourner la notion d'anonymisation, il est possible de recourir aux Technologies Améliorant la Confidentialité (I.). Il pourra être intéressant de se pencher spécifiquement sur l'exemple des données synthétiques (II.). Enfin, il sera nécessaire d'étudier la pertinence des outils développés par des entreprises privées pour anonymiser les données (III.).

I. Les Technologies Améliorant la Confidentialité (TAC) pour limiter la récolte de données personnelles

Les TAC sont des technologies adoptées pour protéger davantage les données personnelles afin de ne pas recourir directement au processus d'anonymisation (A.). Il en existe une multitude et il convient de s'intéresser aux principales (B.).

A. L'utilisation des TAC pour contourner l'anonymisation

Les Technologies Améliorant la Confidentialité (TAC) (également appelées « Privacy Enhancing Technology » (PET)) sont un ensemble d'outils, de méthodes et de pratiques conçus pour protéger les données personnelles et la vie privée des individus[109]. Ces technologies visent à minimiser la collecte de données personnelles, à garantir la confidentialité des données des utilisateurs, et à prévenir les abus liés à la surveillance et à la collecte de données[110]. Les TAC protègent les informations personnelles identifiables contre les accès non autorisés et les fuites de données. Elles garantissent que les données des individus restent confidentielles et sécurisées. Autrement dit, « l'objectif principal et l'objectif de la technologie d'amélioration de la confidentialité est de sécuriser les données personnelles des utilisateurs et de garder leurs informations secrètes. Les utilisateurs peuvent désormais contrôler la manière dont leurs données sont utilisées (…) lorsque leurs données sont envoyées ou utilisées par des entreprises en ligne, des marchands ou d'autres utilisateurs. »[111]

---

[109] « PETS controls matrix : a systematic approach for assessing online and mobile privacy tools », 20 décembre 2016, ENISA.

[110] « Technologies d'amélioration de la confidentialité – Un survol des outils et des techniques », rapport préparé par la Division de l'analyse de la technologie du Commissariat à la protection de la vie privée du Canada, novembre 2017, site web du Commissariat à la Protection de la vie privée du Canada.

[111] « Guide des technologies améliorant la confidentialité (TAC) », Monideepa Mrinal Ro, 25 mai 2022, site web Manage Engine Blog.

Ces technologies aident les organisations à se conformer aux régulations sur la protection des données, telles que le RGPD. Elles permettent aux entreprises de respecter les exigences légales concernant la collecte, le traitement et le stockage des données personnelles. En sécurisant les données et en limitant l'accès aux informations sensibles les TAC réduisent le risque de violations de données, ce qui protège les entreprises contre les coûts élevés associés aux amendes, aux pertes de réputation et aux litiges[112].

L'utilisation de TAC renforce la confiance des utilisateurs envers les services et les entreprises, car ils savent que leurs données personnelles sont protégées et traitées de manière sécurisée. Cette confiance accrue peut améliorer la fidélité des clients et renforcer les relations commerciales. Les TAC permettent de mettre en place des contrôles d'accès stricts et des mécanismes d'audit pour surveiller qui peut accéder aux données et comment elles sont utilisées[113]. Cela aide à prévenir les accès non autorisés et à détecter les activités suspectes. Il existe une grande diversité de TAC.

B. La multitude de technologies existantes améliorant la confidentialité

Il existe une multitude de TAC, développés au fil des années. Il convient de s'intéresser succinctement aux technologies les plus utilisées.

Tout d'abord, il existe les virtuals private network (VPN) qui créent une connexion sécurisée et chiffrée entre l'utilisateur et un serveur distant, masquant ainsi l'adresse IP de l'utilisateur et protégeant ses données contre l'interception[114]. Cela est particulièrement utile sur les réseaux Wi-Fi publics. En effet, « l'idée d'un VPN est avant tout de renforcer la sécurité de sa connexion a un réseau. Il permet également de limiter la récupération de nos données personnelles et éventuellement de débloquer des contenus dans certaines régions où la censure règne. »[115] En masquant l'adresse IP et en chiffrant les données, un VPN augmente l'anonymat en ligne, rendant plus difficile pour les tiers de suivre les activités ou de créer un profil détaillé des habitudes de navigation d'une personne concernée.

---

[112] *Supra* note 110.

[113] « Privacy by Design : An Overview of Privacy Enhancing Technologies », 26 novembre 2008, Entreprise Privacy Group, site web DSP u Toronto.

[114] « Régime du Réseau privé virtuel avocat », 1er mai 2012, Revue Lamy Droit de l'Immatériel n°82.

[115] « L'autodétermination informationnelle a l'épreuve des évolutions technologiques », Pierre Bordais, 2024, Dalloz IP/IT 2024. 72.

Egalement, Tor[116] est un réseau décentralisé qui anonymise le trafic Internet en le faisant passer par plusieurs nœuds (ou "relais") avant d'atteindre sa destination[117]. Chaque nœud ne connaît que l'origine et la destination immédiates du trafic, ce qui rend très difficile de retracer l'activité jusqu'à l'utilisateur. Le réseau Tor est conçu pour résister à l'analyse de trafic, rendant difficile pour les attaquants l'analyse du volume et de la fréquence du trafic pour en déduire des informations sur les activités en ligne des individus[118]. Certains auteurs considèrent que « pour ce qui est de la navigation sur les darknets, l'utilisation des logiciels comme le fameux Tor constitue aussi un dispositif d'anonymisation d'adresses IP. »[119]

Le protocole Hyper Text Transfer Protocol Secure (HTTPS) sécurise les communications entre le navigateur de l'utilisateur et les sites web en chiffrant les données échangées[120]. Lors d'une connexion à un site via HTTPS, le serveur web présente un certificat numérique au navigateur[121]. Ce certificat est délivré par une autorité de certification de confiance et vérifie l'identité du site web, assurant que la communication s'effectue avec le site légitime et non avec un imposteur. De plus, HTTPS assure l'intégrité des données en utilisant des sommes de contrôle et des signatures numériques. Cela garantit que les données transmises n'ont pas été modifiées ou altérées pendant le transit[122].

Enfin, les données synthétiques peuvent être considérées comme des TAC, notamment en ce qui concerne la confidentialité et la sécurité des données. Il convient de les étudier plus en détails.

II. Un cas d'usage des TAC : les données synthétiques comme substitut aux données anonymisées ?

---

[116] « Tout ce qu'il faut savoir sur le navigateur Tor », J. M. Porup, IDG NS (adapté par Jean Elyan), 20 octobre 2019, site web Le monde informatique.

[117] « Le dark web », mis à jour en novembre 2023, Le Lamy droit pénal des affaires.

[118] « Tor en pratique », David A. Levastre-Bodoule Sosso, 2021, Dalloz IP/IT 2021. 89.

[119] « Le clair-obscur autour du délit d'administration d'une plateforme en ligne pour permettre une transaction illicite », Stefan Trifkovic, 28 février 2024, Lexbase.

[120] *Ibid*.

[121] « 4051 - Comment identifier les données ou informations, les traitements et les personnes soumis à la réglementation relative à la protection des données personnelles ? », mis à jour en avril 2023, Le Lamy droit du numérique.

[122] *Ibid*.

Comme nous venons de le voir, les Technologies Améliorant la Confidentialité permettent une approche alternative aux critères de l'anonymisation adoptés par la CNIL. Il est possible d'illustrer cela avec l'utilisation des données synthétiques qui présentent certains avantages (A.) mais également de nombreux inconvénients (B.). Les Etats européens se sont prononcés sur la question de l'utilité de ces données synthétiques (C.).

A. Le principe des données synthétiques et leurs précieux avantages

Les données synthétiques sont des données artificiellement générées qui imitent les caractéristiques statistiques et les structures des données réelles[123], mais qui ne correspondent à aucune donnée individuelle réelle.[124] Elles sont utilisées dans divers domaines, notamment dans la recherche, l'analyse de données et l'apprentissage automatique, pour protéger la confidentialité des données réelles tout en permettant des analyses et des tests.[125] Les données synthétiques peuvent être générées à partir de modèles statistiques, d'algorithmes de génération aléatoire ou d'autres méthodes, et sont souvent utilisées lorsque l'accès aux données réelles est limité ou réglementé.[126]

Le Conseil de l'Europe en 2020 est venu définir la notion de données synthétiques en s'appuyant sur une définition de l'Organisation de Coopération et de développement économique (OCDE) : « Les données synthétiques sont générées à partir d'un modèle construit sur des données réelles. Elles devraient être représentatives des données réelle originales. »[127] Egalement, la Commission Européenne a récemment précisé que ces données synthétiques permettaient « l'augmentation de données qui implique des techniques permettant de générer artificiellement des points de données supplémentaires a partir de données existantes. »[128]

---

[123] « IPEN 2021 on Synthetic Data - Keynote Speech by Wojciech Wiewiórowski », Wojciech Wiewiórowski, 18 juin 2021, site web European Data Protection Supervisor.

[124] « Synthetic Data », Robert Riemann, site web European Data Protection Supervisor.

[125] « Synthetic Data and Privacy : Experiences Implementing Data Synthesis in a Global Life Sciences Company », Stephen Bamford, 16 juin 2021, Janssen Research & Development, site web EDPS Europa.

[126] « IPEN 2021 on Synthetic Data - Introduction by Thomas Zerdick », Thomas Zerdick, 18 juin 2021, European Data Protection Supervisor.

[127] Conseil de l'Europe, « Intelligence artificielle et protection des données », rapport sur l'intelligence artificielle, 2020.

[128] Règlement d'exécution (UE) n° 2023/1507 de la Commission, 20 juill. 2023, établissant les spécifications techniques des besoins en données ainsi que les délais pour la soumission des métadonnées et des rapports sur la qualité pour le thème « Utilisation des TIC et commerce électronique » pour l'année de référence 2024, conformément au règlement (UE) n° 2019/2152 du Parlement européen et du Conseil, Dalloz.

Les données synthétiques offrent plusieurs avantages par rapport au RGPD. En effet, elles assurent la confidentialité des données en remplaçant les informations sensibles par des données générées de manière artificielle, réduisant ainsi le risque de divulgation d'informations personnelles[129]. Egalement, les données synthétiques permettent aux chercheurs et aux analystes d'accéder à des ensembles de données représentatifs tout en évitant les restrictions associées à l'utilisation de données réelles, telles que les consentements requis et les restrictions d'accès[130]. Les données synthétiques peuvent être utilisées pour préserver l'intégrité des données réelles en permettant l'analyse et le partage d'informations sans compromettre la confidentialité des individus concernés[131].

Par ailleurs, les données synthétiques présentent d'autres avantages qui ne sont pas attachés au RGPD. Par exemple, les données synthétiques « helps training machine learning algorithms that need an immense amount of labeled training data, which can be costly or come with data usage restrictions. Moreover, manufacturers can use synthetic data for software testing and quality assurance. Synthetic data can help companies and researchers build data repositories needed to train and even pre-train machine learning models, a technique referred to as transfer learning. »[132] Ainsi, les données synthétiques permettent notamment de produire des algorithmes à moindre coût et participent à la recherche et au développement grâce à cet ensemble de données généré artificiellement.

Cependant, bien que les données synthétiques, qui ne se fondent sur aucune donnée individuelle réelle, présentent de nombreux avantages, elles posent également certaines difficultés qui les empêchent d'être totalement efficaces.

B. Les nombreux risques à l'utilisation de cette alternative

Les données synthétiques présentent de nombreux risques notamment une perte de représentativité en ne capturant pas fidèlement toutes les nuances et les variations présentes dans les données réelles, ce qui peut entraîner conclusions biaisées dans les analyses et les modèles. Cette perte de

---

[129] *Supra* note 124.

[130] « IPEN 2021 on Synthetic Data », Omar Ali Fidal, 18 juin 2021, European Data Protection Supervisor.

[131] *Supra* note 124.

[132] *Supra* note 124.

représentativité peut s'illustrer également par le fait que « la distribution choisie peut également être trop simpliste au vu des tendances mineures présentes dans l'ensemble source »[133]. Ainsi, même si l'objectif de limiter la ré-identification des données personnelles est louable, utiliser des données synthétiques en n'utilisant que partiellement des informations utiles[134] pourrait devenir d'autant plus préjudiciable au regard de la mauvaise représentativité des données.

En effet, si les données synthétiques sont générées à partir de modèles ou d'algorithmes biaisés, cela peut entraîner la propagation de biais dans les analyses et les décisions qui en découlent, amplifiant ainsi les préjugés existants. En effet, « la distribution choisie peut résulter d'hypothèses erronées, conduisant à une modélisation approximative des données »[135]. Lors de la création de données synthétiques, certaines informations pertinentes peuvent être perdues ou mal représentées, ce qui peut compromettre la qualité et la pertinence des analyses et des prédictions qui en découlent.

Egalement, l'une des difficultés majeures concernant l'utilisation de données synthétiques comme données anonymisées réside dans le fait que « ces données pourraient permettre à un tiers malveillant d'en déduire des informations confidentielles à propos des personnes dont les données ont été utilisées pour synthétiser les données. »[136] C'est pour cela qu'affirmer que les données synthétiques sont assimilables à des données anonymisées est sans doute abusif. En effet, bien que seules des données artificiellement générées soient utilisées, celles-ci reflètent des tendances sur des données qui ont obligatoirement été récoltées à un moment donné. En effet, « si le caractère aléatoire dans la génération de données synthétiques n'est pas suffisamment paramétré, il serait utile que l'on puisse parvenir à ré-identifier une ou plusieurs personnes physiques. »[137]

Ainsi, même si le recoupement apparaît difficile à envisager, les pirates informatiques sont de plus en plus performants et parviennent facilement à établir des corrélations et des inférences au sens des critères de la CNIL[138]. En effet, il ne serait pas juste d'affirmer que les données synthétiques

---

[133] « Données synthétiques - Et l'Homme créa les données à son image 2/2 », Alexis Léautier, 17 août 2022, Laboratoire d'Innovation Numérique de la CNIL (LNIC).

[134] *Ibid*.

[135] *Supra* note 133.

[136] *Supra* note 133.

[137] « La génération de données synthétiques par des systèmes d'intelligence artificielle, une nouvelle méthode de protection et de valorisation des données », Bertrand Cassar, 2024, Dalloz IP/IT 2024.282.

[138] *Supra* note 14.

empêchent automatiquement tout risque de ré-identification. Ce point de vue est contrebalancé par de nombreux auteurs[139] qui estiment à l'inverse que les données synthétiques offrent une plus grande protection que les données anonymisées.[140] Cependant, il apparait que le développement d'outils techniques innovants permet en pratique un recoupement des données.

Il convient de conclure que, bien que les données synthétiques soient conçues pour préserver la confidentialité des données personnelles, il existe toujours un risque résiduel de ré-identification ou de divulgation non autorisée, surtout si les techniques de génération ne sont pas correctement appliquées ou si les données sont combinées avec d'autres sources d'information. Ainsi, cette alternative d'utilisation des données synthétiques est intéressante au regard d'une application juste du RGDP afin de limiter les risques de fuites de données personnelles. Cependant, il apparait difficile de qualifier les données synthétiques comme de véritables données anonymisées, supprimant tout risque de ré-identification. Il faudrait pouvoir disposer d'une véritable définition de la notion de données synthétiques, avec des critères précis et rigides, comme le G29 avec la notion d'anonymisation, afin de pouvoir s'assurer d'un seuil de fiabilité suffisant des données synthétisées.

D'autres Etats européens se sont prononcés sur la question et s'interrogent sur les apports que pourraient avoir à long terme l'utilisation de données synthétiques.

C. L'utilisation pratique des données synthétiques par les Etats

Au Royaume-Uni, l'Information Commissioner's Office (ICO) s'est prononcé sur la question des données synthétiques. Après avoir défini la notion, il considère que « to the extent that synthetic data cannot be related to identified or identifiable living individuals, it is not personal data and therefore data protection obligations do not apply when you process it. »[141] Cette vision des choses n'est pas celle adoptée en Union Européenne puisque la réglementation sur la protection des données continue à s'appliquer aux données synthétiques. L'ICO considère tout de même qu'en cas de risque de ré-identification des données au stade de la transformation des données réelles en données synthétiques, il sera important d'appliquer les protections nécessaires.[142] Mais, il reste

---

[139] *Supra* note 137.

[140] *Supra* note 125.

[141] « How should we assess security and data minimisation in AI ? », Information Commissioner's Office (ICO), site web ICO ORG.

[142] *Ibid*.

intéressant de prendre en compte le point de vue de cette autorité qui estime que lorsque les données sont considérées comme synthétiques, le risque de ré-identification disparaît.

En Norvège, l'Autorité de protection des données a encouragé l'utilisation de données synthétiques pour limiter le risque d'intrusion à la vie privée.[143] En effet, l'Autorité a condamné à une amende la Confédération norvégienne du sport en 2021 pour avoir effectué des tests avec des données personnelles sans prendre en compte les impacts potentiels et les principes du RGPD.[144] Elle a ainsi affirmé que « the testing could have been achieved with processing of synthetic data. »[145] Ainsi, l'Autorité norvégienne prend conscience que l'utilisation de données synthétiques permettrait de mieux répondre aux exigences du RGPD.

L'Autorité de données fédérale allemande affirme quant à elle que « also research in the field of synthetic data shows enormous promise, and more funding should be funnelled into this area. »[146] En effet, afin de trouver des alternatives à l'anonymisation, l'Autorité allemande cherche à développer davantage de moyens afin de mieux définir et comprendre plus précisément les enjeux de la notion de données synthétiques et l'apport véritable qu'elle pourrait avoir dans un monde où les fuites de données et les ré-identifications sont de plus en plus récurrentes.

Après avoir identifié dans quelle mesure les données synthétiques sont intéressantes du point de vue du RGPD mais souvent erronées et ne permettant pas automatiquement une véritable anonymisation, il convient de s'interroger sur les nouvelles techniques d'anonymisation que prétendent adopter certaines entreprises.

III. L'adoption de présumées techniques d'anonymisation innovantes pour contourner la définition originelle

Plusieurs entreprises privées ont décidé de développer leurs propres outils d'anonymisation. Anonos a notamment créé l'outil Anonymeter accueilli favorablement par la CNIL (A.). Egalement, Sarus a

---

[143] « Norwegian Confederation of Sport fined for inadequate testing », 2021, site web Datalsynet.

[144] *Ibid*.

[145] *Supra* note 143.

[146] « Opinion of the Data Ethics Commission », Daten Ethik Commission, site web BFDI BUND.

commercialisé un outil se basant sur la confidentialité différentielle (B.). Enfin, Aircloak s'est basé sur une approche dynamique pour développer un outil censé garantir la protection des données (C.).

A. Anonymeter, un outil prometteur par Anonos pour une protection des données par une approche avec les données synthétiques

Anonos est une entreprise spécialisée dans les technologies de protection des données et de la confidentialité[147]. Elle développe des solutions pour anonymiser et pseudonymiser les données, permettant aux entreprises de se conformer aux régulations sur la protection des données tout en exploitant les données de manière sécurisée pour l'analyse et le traitement. Leurs outils permettent de minimiser les risques liés à la confidentialité des données tout en maximisant leur utilité. Récemment, Anonos a même été nommé comme « le leader de l'innovation technologique dans le secteur britannique des services de désidentification des données des patients »[148] par le prix Frost & Sullivan Technology Innovation Leader. Depuis novembre 2022, Anonos a acquis Statice, une entreprise spécialisée dans les logiciels de données synthétiques.[149]

Depuis sa création en 2012, Anonos a développé de nombreux outils afin d'optimiser la protection des données personnelles tels que Data Embassy, Prompt Protector, Test Data Management, Variant Twins Et Anonymeter[150]. L'outil Variant Twins permet de pseudonymiser des données personnelles[151]. En effet, cet outil permet de créer un pseudonyme unique à attribuer à différents moments « to replace data values for various purposes so they cannot be correlated without access to a Master Look-Up Table that is kept separately and securely »[152]. Cette technologie est supposée répondre aux critères apposés par le G29 concernant la pseudonymisation et devrait répondre aux besoins des entreprises. Certains considèrent que ces Variant Twins « rendent légaux et éthiques l'utilisation, le partage, le traitement et l'analyse des données, et tirent parti de la valeur des

---

147 Page d'accueil du site web Anonos.

148 « Anonos nommé par Frost & Sullivan comme le leader de l'innovation technologique dans le secteur britannique des services de désidentification des données des patients », Businesswire, 12 janvier 2023, site web Silicon.

149 Page d'accueil du site web Statice.

150 « Will You Be Ready for DORA in January 2025? How Anonos Variant Twins Revolutionize Cybersecurity Under the EU Digital Operational Resilience Act », Gary LaFever, 16 janvier 2024, Linkedin.

151 « How Variant Twins Work », site web Anonos.

152 *Ibid*.

données tout en protégeant les droits des individus. »[153], selon Maeirah Ashaie, consultante chez Frost&Sullivan.

Enfin, l'outil le plus reconnu d'Anonos est Anonymeter. Cet outil a une grande importance car il a été reconnu par la CNIL comme respectant les trois critères d'individualisation, de corrélation et d'inférence[154]. La CNIL en a profité pour « saluer l'approche adoptée par Statice (acquis par Anonos) »[155] et l'« accueille favorablement »[156]. Cette lettre envoyée par la CNIL le 13 février 2023 permet de reconnaître à Anonos une légitimité européenne[157].

En quoi consiste cet outil ? Anonymeter est un outil d'évaluation de la confidentialité des données qui permet de mesurer et de vérifier le niveau d'anonymisation des ensembles de données, garantissant que les données respectent les normes et réglementations de confidentialité en vigueur[158]. L'outil utilise une approche basée sur des attaques pour évaluer trois types de risques principaux : l'identification unique (« singling out »), la liaison (« linkability ») et l'inférence (« inference »)[159]. L'outil utilise des attaques simulées pour formuler des hypothèses sur les enregistrements dans le jeu de données original en se basant sur les données synthétiques. Pour l'évaluation, Anonymeter compare les hypothèses de l'attaquant avec les valeurs réelles des données originales pour mesurer le succès de l'attaquant. L'estimation du risque permet de répéter l'attaque sur un jeu de données de contrôle (qui n'a pas été utilisé pour générer les données synthétiques). Anonymeter distingue entre les informations obtenues grâce aux données synthétiques et les véritables fuites de confidentialité. Le résultat est un rapport détaillé qui résume les risques de confidentialité et fournit des recommandations pour les atténuer, aidant ainsi les entreprises à déterminer si les données peuvent être considérées comme anonymes et conformes aux réglementations.

---

[153] *Supra* note 48.

[154] Lettre de la CNIL à Mr Ali Fdal par Bertrand Pailhès, 13 février 2023, N/réf : AGE/BPS/CS231015.

[155] *Ibid*.

[156] *Supra* note 154.

[157] *Supra* note 154.

[158] « Ensuring Synthetic Data Privacy », site web Anonos.

[159] *Ibid*.

Cependant, après avoir vanté les mérites de la solution Anonymeter développée par Statice, la CNIL a affirmé que « bien qu'Anonymeter semble être un outil prometteur pour l'évaluation des risques résiduels de ré-identification pour les ensembles de données synthétiques, le service de l'expertise technologique souligne que d'autres indicateurs issus de la littérature scientifique pourraient utilement lui être associés. » [160] En effet, dans la deuxième partie de sa lettre, la CNIL dresse une liste des points que Anonos pourrait améliorer afin d'assurer un outil encore plus puissant. La CNIL conclu en affirmant « qu'Anonymeter est un outil prometteur et pertinent dans le contexte de la protection des données personnelles. Il recommande donc aux aux chercheurs et institutions de le tester dans divers contextes, afin de confirmer son utilité et fiabilité, et d'affiner les critères d'évaluation et les seuils acceptables »[161].

Peut-on penser que qu'Anonos serait sur le point de développer un outil qui pourrait être adopté par les autorités de protection européennes ? Une chose est sûre, Anonymeter présente aujourd'hui une technologie qui apparaît fiable et protectrice des données personnelles. L'entreprise Sarus en fait-elle autant ?

B. Un outil innovant par Sarus pour une protection des données par une approche sur la confidentialité différentielle

Sarus est une startup parisienne spécialisée dans la protection de la vie privée pour l'analyse et l'intelligence artificielle[162]. Elle propose une couche de confidentialité permettant aux data scientists et analystes de travailler sur des données sensibles sans les voir directement, grâce à des technologies comme la confidentialité différentielle et des données synthétiques, constituant des Technologies Améliorant la Concurrence (TAC)[163]. Leur solution aide les entreprises à utiliser pleinement leurs données tout en respectant les réglementations de confidentialité et en minimisant les risques de violation de données.

Le fondateur et CEO de Sarus a décidé de développer son propre outil de protection des données après avoir fait le constat que « Data security is probably more important than ever. There is more and more awareness about the risks, including the risk that information that is released today may

[160] *Supra* note 154.

[161] *Supra* note 154.

[162] Page d'accueil du site web Sarus.

[163] « Differential Privacy », site web Sarus.

be used in the future to carry out cyber attacks. Even encrypted data can no longer be shared lightly with the advent of store-now-decrypt-later attacks. »[164] Cet outil de confidentialité différentielle « permet l'exploitation statistique de données individuelles agrégées sans compromettre la vie privée des individus concernés. »[165] Il applique des techniques avancées, comme l'ajout de bruit statistique aux données, pour garantir que les informations personnelles ne puissent pas être ré-identifiées[166]. En effet, « Differential Privacy is based on the addition of statistical noise to computation results. The noise introduces a level of uncertainty limiting how much information about one individual may be revealed. Such noise shall be sufficient to hide the effect of one single individual, but not excessive, to keep the result accurate. »[167] Cela permet de protéger la vie privée des individus tout en permettant des analyses précises et fiables sur les données.

Il est important de noter que cette notion de confidentialité différentielle a été développée bien auparavant par Cynthia Dwork qui a introduit et formalisé ce concept en 2006[168]. Dwork a développé les fondements théoriques qui démontrent comment la confidentialité différentielle peut être appliquée à divers types d'analyses et de requêtes sur les bases de données[169]. Ses travaux ont établi des critères précis pour mesurer la quantité de bruit à ajouter, en fonction du niveau de confidentialité souhaité et du type de requête.

Pour développer et commercialiser cet outil d'analyse de données, Sarus a levé deux millions d'euros auprès de Serena et XAnge[170]. Cette levée de fonds visait à renforcer le développement de leur technologie et à étendre leur portée sur le marché, en particulier pour les secteurs nécessitant une protection rigoureuse des données, comme la finance et la santé.

Pour illustrer l'intérêt de cet outil, il est possible d'utiliser un exemple concret. Prenons une ville qui souhaite déterminer les emplacements optimaux pour de nouveaux arrêts de bus afin de couvrir

---

[164] « Maxime Agostini's interview on Cybernews », 27 juillet 2022, The Sarus Blog, site web Sarus.

[165] « La confidentialité différentielle au service de la privacy », 20 décembre 2019, site web In Cyber News.

[166] « Quand l'IA garantit la confidentialité des données », 17 octobre 2023, site web French Impact Technologie IOT.

[167] *Supra* note 163.

[168] « Calibrating noise to sensitivity in private data analysis », Cynthia Dwork, Frank McSherry, Kobbi Nissim et Adam Smith Ouvrage *Theory of cryptography*, 2006.

[169] Ouvrage *The Algorithmic Foundations of Differential Privacy*, Cynthia Dwork, 2014.

[170] « Sarus lève 2 ME pour accélérer l'innovation digitale », Claude Leguilloux, 22 avril 2020, site web boursier.com.

au mieux les besoins de la population[171]. Pour ce faire, elle a besoin d'analyser des données de géolocalisation sensibles, telles que les lieux d'habitation et de travail des habitants. Les données de géolocalisation des résidents, incluant les adresses domiciliaires et professionnelles, sont collectées de manière anonyme dans un système centralisé. L'outil développé par Sarus applique la confidentialité différentielle aux données collectées. Cela se fait en générant plusieurs ensembles de données (avec des cartes) légèrement modifiés, où certains individus sont aléatoirement ajoutés ou retirés. Ce processus ajoute du bruit statistique pour protéger les informations personnelles tout en permettant des analyses significatives. L'outil génère une série de cartes représentant les emplacements potentiels pour les arrêts de bus, en fonction des données bruitées. Pour chaque carte, l'emplacement optimal de l'arrêt de bus est calculé. Si, pour chaque carte, l'arrêt de bus est placé au même endroit ou à des endroits très proches, cela indique que la solution est robuste et indépendante de la présence ou de l'absence de tout individu spécifique. La ville utilise les résultats pour déterminer les emplacements optimaux pour les nouveaux arrêts de bus. Les emplacements identifiés de manière cohérente sur toutes les cartes générées sont choisis, garantissant ainsi que la décision est basée sur des données robustes et non sur des informations individuelles spécifiques. L'analyse montre que, sur 100 cartes générées, l'arrêt de bus optimal est placé au même endroit dans 95% des cas. Cela signifie que l'emplacement choisi est indépendant de l'ajout ou du retrait de données spécifiques d'individus.

Ainsi, cet exemple permet d'affirmer que « la confidentialité différentielle offre une garantie forte de maintien de l'anonymat en s'appliquant à un algorithme et non à un résultat »[172]. Cependant, cet outil développé par Sarus, bien que très intéressant, n'a reçu pour le moment aucune certification ou validation par une autorité de protection des données européennes. Il est alors important de garder certaines réserves quant à son effectivité concernant une véritable anonymisation telle qu'elle est spécifiées dans les critères du G29.

C. Aircloak Diffix, un outil novateur par Aircloack pour une protection des données par une approche dynamique

Aircloak est une entreprise spécialisée dans la protection de la vie privée des données. Fondée en 2014 en Allemagne, cette société propose des solutions de confidentialité pour les entreprises,

---

[171] Exemple inspiré de « Sarus lève deux millions d'euros pour son outil d'analyse de données respectueux de la vie privée », Alice Vitard, 21 avril 2020, site web Usine digitale.

[172] *Supra* note 165.

permettant de traiter et d'analyser des données sensibles tout en garantissant la confidentialité des informations personnelles. Cette entreprise procède à la distinction entre donnée dynamique et donnée statique[173]. Elle estime que les données statiques sont des données qui ne changent pas ou changent très peu après leur création initiale[174]. Elles incluent des ensembles de données archivés, des enregistrements historiques, et des bases de données qui ne sont pas régulièrement mises à jour. Les données dynamiques en revanche sont des données qui changent fréquemment ou qui sont constamment mises à jour[175]. Cela peut inclure des flux de données en temps réel, tels que les transactions financières, les données des capteurs IoT, ou les interactions utilisateur sur un site web.

C'est sur la base des données dynamiques que Aircloak a développé "Aircloak Diffix", une technologie de protection des données[176]. Diffix permet aux entreprises d'exploiter les données tout en assurant que les informations personnelles des individus restent anonymes. Cette technologie combine des techniques de protection de la vie privée, telles que l'ajout de bruit statistique, pour permettre l'analyse des données sans compromettre la confidentialité. Plus précisément, « Aircloak gives an analyst access to all the underlying data, and dynamically tailors the anonymization to the specific query and the data requested. The system understands what data is sensitive under which circumstances, freeing the analyst from error-prone manual configuration. »[177]

Comment cet outil fonctionne en pratique ?[178] Le Aircloak Diffix fonctionne en cinq étapes. Les données sensibles sont importées dans le système Aircloak Diffix. Ces données restent dans un environnement sécurisé. Puis, les administrateurs définissent des règles de confidentialité spécifiques, telles que les niveaux de bruit à ajouter et les types de requêtes autorisées. Par la suite, les utilisateurs peuvent interroger les données en utilisant des outils d'analyse ou des requêtes SQL standard (instructions utilisées pour interagir avec des bases de données relationnelles)[179]. Lorsqu'une requête est soumise, Aircloak Diffix intervient pour anonymiser les résultats. Avant de retourner les résultats, Aircloak Diffix ajoute du bruit statistique et applique des filtres pour

---

[173] « Diffix and the Fundamental Law of Information Recovery », Paul Francis, 11 août 2020, site web Aircloak.

[174] « Anonymization with Aircloak : how it works », site web Aircloak.

[175] *Ibid*.

[176] « Aircloak Insights », site web Aircloak.

[177] *Supra* note 174.

[178] « Aircloak Anonymization », juin 2017, site web Aircloak.

[179] « Cloud santé - Anonymisation dynamique », site web Euris.

masquer les informations pouvant mener à une ré-identification. Par exemple, si une requête vise des groupes trop petits qui pourraient révéler des individus, Aircloak Diffix ajuste les résultats pour éviter cela. En effet, Aircloak Diffix « marginally altering the values produced from the query. Either the dimensions or the aggregates can be altered. Altering the aggregates themselves is usually done by adding small amounts of statistical noise »[180]. Enfin, les résultats de la requête sont renvoyés à l'utilisateur, anonymisés de manière à préserver la confidentialité des données tout en fournissant des informations utiles pour l'analyse.

Ainsi, cet outil de protection des données des utilisateurs, avant même la récolte de données, est intéressant au regard du RGPD. Cependant, cet outil n'a pas reçu, comme Sarus, de certification par une Autorité de protection des données européennes. Alors, est-il vraiment possible de parler d'un outil permettant une anonymisation des données comme l'entend le G29 ? Dans quelle mesure est-il possible d'avoir confiance en ces entreprises qui font la promotion d'outils soi-disant nécessaires et efficaces pour parvenir à une véritable anonymisation ? Faut-il se fier à ces nouveaux outils émergents ? Quoi qu'il en soit, ce qu'il convient de garder à l'esprit, ce sont les trois critères qui selon le G29 permettent une anonymisation : inférence, corrélation et individualisation.

Dans quelle mesure cette notion d'anonymisation a-t-elle vocation à évoluer ? Est-ce-que les entreprises sauront s'adapter à temps face au Big data en créant de nouvelles techniques de chiffrage ?

Section 2 : Les définitions de l'anonymisation et de la pseudonymisation ont-t-elles vocation à évoluer ?

« Il est à noter que le risque d'identification peut augmenter avec le temps et dépend aussi des progrès des technologies de l'information et des communications. Les règles juridiques doivent donc, le cas échéant, être formulées d'une manière technologiquement neutre et tenir compte, dans l'idéal, des capacités d'évolution des technologies de l'information. »[181] Cette citation, tirée du site de la CNIL, permet de souligner l'un des enjeux majeurs de la notion d'anonymisation : l'évolution et le développement de technologies numériques augmentant le risque de ré-identification.

[180] *Supra* note 174.

[181] *Supra* note 14.

Pour terminer cette réflexion sur la clarification et l'évolution de la notion d'anonymisation, il s'agira de traiter de manière prospective les limites de la notion d'anonymisation (I.), avant de présenter les nouveaux moyens de chiffrement qui sont envisageables pour y remédier (II.). Il sera alors possible de conclure en abordant la question de la modernisation des législations françaises et européennes qui peuvent être envisagées (III.).

I. Les importantes limites à l'anonymisation

Bien que les critères d'inférence, d'individualisation et de corrélation aient déjà été fixés par le G29, le RGPD tente d'encadrer au mieux la notion d'anonymisation afin de l'adapter in concreto en fonction de « l'ensemble des moyens raisonnablement susceptibles d'être utilisés »[182]. Cependant, cette notion de raisonnabilité n'a pas réellement était définie, ce qui contribue à créer un flou juridique important (A.). Par ailleurs, à l'heure du big data, une autre limite à l'anonymisation réside dans la mise à disposition de données massives qui permettent un recoupement des données facilement (B.).

A. Une caractérisation imprécise de la notion de raisonnabilité

Le considérant 26 du RGPD dispose que « 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 »[183]. Cependant, la notion de « raisonnabilité » dans la définition de l'anonymisation peut donner lieu à une caractérisation peu précise, car elle dépend souvent du contexte, des normes et des exigences spécifiques de chaque situation. La « raisonnabilité » fait référence à l'évaluation subjective de ce qui est considéré comme adéquat, approprié ou acceptable dans un contexte donné.

Le considérant précise par la suite que « 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

[182] Considérant 26 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.

[183] *Ibid*.

l'évolution de celles-ci. »[184] Cette formulation reste relativement peu précise et « laisse transparaître un certain aléa sur la qualification d'un jeu de données anonymes. »[185]

En effet, plusieurs auteurs soulignent l'importance de prendre en compte les capacités réelles des acteurs du traitement des données lors de l'évaluation du caractère raisonnable des mesures de protection des données. Adopter un standard abstrait déconnecté des capacités réelles peut conduire à une qualification quasi automatique des données en tant que données à caractère personnel, sans tenir compte du niveau réel de risque de réidentification.

« L'analyse du caractère raisonnable des moyens susceptibles d'être utilisés pour réidentifier une personne diverge drastiquement entre les responsables du traitement. Par exemple, les services de renseignement – tels que la National Security Agency (NSA) ou la Direction générale de la Sécurité extérieure (DGSE) – ou les grandes entreprises privées – telles que Criteo, Facebook ou Google – n'ont absolument pas les mêmes capacités de ré-identification que des plus petites structures telles que le bureau de police local ou des entreprises comme Carrefour ou OVH. En adoptant un standard abstrait, c'est-à-dire sans lien avec les capacités réelles du responsable du traitement ou de son sous-traitant, le législateur pose un principe de qualification quasi automatique des données en données à caractère personnel. »[186]

Cette citation soulève un point crucial : les différents acteurs ne possèdent pas tous les mêmes capacités et les mêmes outils pour ré-identifier des données. Elle met en évidence les disparités dans les ressources et les compétences techniques disponibles pour effectuer la ré-identification des personnes à partir des données anonymisées.

D'une part, les grandes organisations, telles que les services de renseignement ou les géants de la technologie disposent souvent de vastes ressources et de technologies avancées pour analyser les données et potentiellement ré-identifier les individus. Leurs capacités techniques et leurs budgets importants leur permettent d'exploiter des méthodes sophistiquées de traitement des données, y compris l'utilisation de l'intelligence artificielle et de l'apprentissage automatique pour identifier les schémas et les corrélations dans les données.

---

[184] *Supra* note 182.

[185] *Supra* note 12.

[186] *Supra* note 47.

D'autre part, les petites entreprises ou les organisations moins puissantes peuvent avoir des capacités limitées pour ré-identifier les individus à partir de données pseudonymisées. Leurs ressources techniques et financières plus modestes peuvent limiter leur capacité à mener des analyses approfondies ou à utiliser des technologies avancées de traitement des données.

En conséquence, il est crucial pour les régulateurs et les législateurs de reconnaître ces disparités et d'adopter des normes et des réglementations qui tiennent compte des capacités techniques et des ressources des différentes organisations. Cela pourrait inclure des exigences différenciées en matière de protection des données en fonction de la taille, de la nature et des capacités techniques des acteurs du traitement des données, afin de garantir une protection adéquate de la vie privée tout en permettant l'innovation et le développement économique.

B. Les Big Data ou l'augmentation conséquente de la récolte de données personnelles

« À l'heure des données massives, toute donnée peut un jour ou l'autre devenir porteuse d'un risque de ré-identification. »[187] Les big data, c'est-à-dire les données massives, peuvent poser de nouveaux défis à l'anonymisation. En effet, elles comprennent souvent une multitude de variables et de points de données. Il est possible de définir les big data comme « l'internet des objets, l'ouverture des données et leur accessibilité, l'interconnexion généralisée transformant les internautes en fournisseurs de données personnelles et d'une manière générale les évolutions technologiques liées aux capacités computationnelles qui traitent l'information sur toute sa chaîne de transformation »[188].

Ce nouveau phénomène signifie que même si les données sont aujourd'hui anonymisées, il devient de plus en plus aisé de les relier à des individus en combinant plusieurs sources de données ou en utilisant des techniques d'analyse sophistiquées. Autrement dit, les big data permettent souvent de découvrir des corrélations et des modèles complexes entre les différentes variables. Même si les données sont anonymisées, les modèles statistiques peuvent être utilisés pour identifier des tendances spécifiques et potentiellement révéler des informations sur les individus. En effet, « les chercheurs en informatique ont, depuis longtemps, montré qu'on pouvait retrouver l'identité des

---

[187] *Supra* note 12.

[188] « Introduction », Évelyne Broudoux et Ghislaine Chartron, Ouvrage *Big Data - Open Data : Quelles valeurs ? Quels enjeux ?*, 2015, Cairn.

personnes en croisant des données anonymisées avec des informations publiquement disponibles. »[189]

Également, les progrès dans les techniques d'analyse de données, telles que l'apprentissage automatique et l'intelligence artificielle, permettent d'extraire des informations significatives à partir de données brutes. Ces techniques peuvent contourner l'anonymisation en identifiant des schémas ou des caractéristiques uniques dans les données qui peuvent être utilisés pour identifier des individus.

Par ailleurs, dans de nombreux cas, les données anonymisées peuvent être conservées pendant de longues périodes de temps. Or, avec le temps, de nouvelles techniques et de nouveaux outils peuvent être développés pour briser l'anonymisation, compromettant ainsi la confidentialité des données. En effet, même si les données sont anonymisées au départ, les avancées technologiques peuvent permettre la ré-identification des individus en croisant des données avec d'autres ensembles de données disponibles publiquement. Ainsi, « la difficulté tient à ce que l'anonymisation en amont de la donnée par celui qui la collecte et la traite ne garantit nullement que, par recoupement, il ne soit néanmoins pas possible, en aval, de ré-identifier la personne en question. »[190]

Pour ces raisons, bien que l'anonymisation soit une technique importante pour protéger la vie privée des individus, elle ne peut pas toujours garantir l'anonymat absolu dans le contexte des big data. « L'hypothèse d'un risque zéro peut donc s'avérer préjudiciable à la notion de données anonymisées voire même encourager le traitement de données à caractère personnel plutôt que le traitement de données anonymisées. »[191]

C'est notamment pour cela que malgré les efforts réalisés par la CNIL pour encadrer la notion d'anonymisation, les évolutions techniques remettent continuellement en cause les critères établis. En effet, « l'état de l'art des techniques d'anonymisation ne permet pas d'assurer une

---

[189] « Au delà des big data. Les sciences sociales et la multiplication des données numériques », Étienne Ollion et Julien Boelaert, Ouvrage *Sociologie* (Vol. 6), mars 2015, Cairn.

[190] « L'extension du domaine de la donnée », Valérie-Laure Benabou, Ouvrage *Legicom* (N° 59), février 2017, Cairn.

[191] *Supra* note 12.

anonymisation fiable à 100 %. Il existe toujours des risques résiduels qui compromettent l'anonymisation strictement technique. »[192]

Afin de palier les limites liées à la définition donnée par le RGPD et à l'accumulation de données, l'utilisation de moyens de chiffrement pourrait-il contribuer à améliorer l'efficacité de l'anonymisation et de la pseudonymisation ?

II. La création de nouveaux moyens de chiffrement au profit de l'anonymisation et de la pseudonymisation ?

Afin de faire face aux big data et aux puissants moyens de ré-identification dont disposent les acteurs, la solution réside dans le développement de moyens de chiffrement beaucoup plus puissants. Pour cela, les Etats Unis ont développé l'Advanced Encryption Standard, réputé comme le plus puissant du monde grâce à un chiffrement symétrique (A.). Une solution de chiffrement asymétrique est également envisagé par la Blockchain afin d'empêcher tout risque de ré-identification (B.).

A. Le chiffrement AES ou l'algorithme n'ayant connu aucune attaque

L'Advanced Encryption Standard (AES), en français « Standard de chiffrement avancé », est un algorithme de chiffrement symétrique largement utilisé dans le monde entier pour sécuriser des données sensibles[193]. Contrairement au chiffrement asymétrique, où une paire de clés est utilisée (une pour le chiffrement et une pour le déchiffrement), l'AES utilise une seule clé pour à la fois chiffrer et déchiffrer les données[194].

L'AES est réputé pour sa robustesse et sa sécurité. Il est approuvé par le gouvernement des États-Unis pour le chiffrement des informations sensibles et est utilisé par de nombreuses organisations à travers le monde pour protéger leurs données.[195] L'AES prend en charge trois tailles de clé

---

[192] *Supra* note 12.

[193] « Le chiffrement, ou l'apport de la cryptologie à la sécurisation du stockage, de la transmission et du traitement des données », Louis Goubin, Annale des Mines - Réalités industrielles 2022/3, août 2022, Cairn.

[194] Cour de cassation, Assemblée plénière 7 novembre 2022, n° 21-83.146, Lexbase.

[195] « Qu'est-ce qu'une paire de clefs privée/publique ? », Jean-Guillaume Dumas, Pascal Lafourcade, Etienne Roudeix, Ariane Tichit et Sébastien Varrette, Ouvrage *Les NFT en 40 questions,* 2022, Cairn.

différentes : 128 bits, 192 bits et 256 bits[196]. Les données sont traitées par blocs de 128 bits. La longueur de la clé choisie détermine le niveau de sécurité offert par l'algorithme[197].

L'AES est conçu pour être rapide et efficace, ce qui le rend adapté à une utilisation dans une large gamme d'applications, y compris les systèmes embarqués, les applications web et les protocoles de sécurité réseau[198]. L'AES est résistant à de nombreuses attaques cryptographiques connues, y compris les attaques par force brute et les attaques différentielles[199]. En effet, « Jusqu'à présent, il n'existe aucune attaque réelle connue qui permettrait à une personne n'ayant pas connaissance de la clé de lire les données chiffrées par l'AES, évidemment si ce dernier est correctement mis en œuvre.[200] » Cependant, il est important de noter que la sécurité de l'AES dépend en grande partie de la longueur et de la qualité de la clé utilisée[201].

L'AES est reconnu comme une norme internationale par l'Institut national des standards et de la technologie (NIST) des États-Unis, ainsi que par d'autres organisations de normalisation à travers le monde[202]. Cet algorithme puissant pourrait être utilisé en France et dans l'Union européenne afin de renforcer la confidentialité et assurer une anonymisation des données.

Cependant, le chiffrement AES ne remplace pas les valeurs des données par d'autres valeurs, ni ne supprime les liens entre les données et les individus. Il rend simplement les données illisibles sans la clé appropriée[203]. Pour l'anonymisation, les données doivent être modifiées de manière à ce qu'elles ne puissent plus être associées à des individus spécifiques, ce qu'AES ne fait pas. De plus, le chiffrement AES est réversible, ce qui signifie que, avec la clé de déchiffrement, on peut retrouver les données originales. Cela permet une réversibilité des données car les données peuvent toujours être déchiffrées.

---

[196] Décision n° 2008/616/JAI du 23 juin 2008 concernant la mise en oeuvre de la décision 2008/615/JAI relative à l'approfondissement de la coopération transfrontalière notamment en vue de lutter contre le terrorisme et la criminalité transfrontalière, Lamyline.

[197] « Fort Knox pour vos données : comment le chiffrement AES sécurise votre entreprise », site web Veritas.

[198] Décision CNIL, 5 juillet 2021, n° 2012-219, Dalloz.

[199] *Ibid*.

[200] « Sécurité absolue : le chiffrement AES-128 d'Ajax Systems est désormais certifié par le NIST », 9 avril 2024, site web Ajax.

[201] *Supra* note 199.

[202] « Advanced Encryption Standard (AES) », site web Iot Industriel.

[203] « Recommendation for Block Cipher Modes of Operation : Methods and Techniques », Morris Dworkin, 2001, National Institute of Standards and Technology.

Ainsi, l'AES pourrait être utilisé en complément pour sécuriser les données anonymisées contre tout accès non autorisé. En effet, sa robustesse face aux attaques démontre la puissance de son fonctionnement et il serait intéressant de le mettre au profit de l'anonymisation en complément d'autres mécanismes de sécurité ou en adaptant la réversibilité et la suppression des liens entre les données et les individus afin de créer un algorithme infaillible pour assurer l'anonymisation.

Par ailleurs, l'utilisation du chiffrement AES pourrait être très utile pour procéder à une pseudonymisation. En effet, le chiffrement AES peut être utilisé pour transformer des identifiants personnels (comme les numéros de sécurité sociale, les adresses e-mail, etc.) en pseudonymes sécurisés[204]. Par exemple, en chiffrant un identifiant avec une clé secrète, on obtient une valeur pseudonymisée qui peut remplacer l'identifiant original dans les bases de données. Grâce à la réversibilité des données, il serait possible de ré-identifier les individus si nécessaire, par exemple pour des audits ou des enquêtes. Enfin, en chiffrant les identifiants avec AES, on assure que même si les pseudonymes sont exposés, ils ne peuvent pas être directement associés aux identifiants originaux sans la clé de déchiffrement. Cela ajoute une couche de sécurité supplémentaire. Ainsi, le chiffrement AES peut jouer un rôle crucial dans la pseudonymisation en générant des pseudonymes sécurisés et en protégeant les identifiants personnels. Cependant, pour assurer une sécurité maximale, il est essentiel de gérer rigoureusement les clés de chiffrement.

Est ce que l'utilisation d'un chiffrement asymétrique, tel que celui employé sur la blockchain, pourrait être pertinent à utiliser dans le cadre d'une anonymisation des données ?

B. Le fonctionnement inédit de la blockchain avec le chiffrement asymétrique

La blockchain est une technologie de registre distribué qui permet de stocker des enregistrements de manière transparente, sécurisée et décentralisée[205]. Elle utilise à la fois le chiffrement symétrique et asymétrique pour sécuriser les données et garantir l'intégrité du système[206].

---

[204] *Ibid*.

[205] « Du Bitcoin aux DAO : les fondations techniques de la blockchain », Jean-Noël Colin, Ouvrage *Les blockchains et les smart contracts à l'épreuve du droit*, 30 octobre 2020.

[206] « Comprendre les grands principes de la cryptologie et du chiffrement », 24 octobre 2016, site web de la CNIL.

Il sera intéressant de se pencher plus spécifiquement sur le chiffrement asymétrique qui est généralement utilisé pour assurer l'authenticité et l'intégrité des transactions dans une blockchain. Chaque participant de la blockchain possède une paire de clés : une clé publique et une clé privée[207]. La clé publique est partagée avec les autres participants, tandis que la clé privée est gardée secrète. Lorsqu'un participant souhaite effectuer une transaction, il signe numériquement la transaction à l'aide de sa clé privée[208]. Cette signature est vérifiée par les autres participants à l'aide de la clé publique correspondante pour s'assurer que la transaction est authentique et non altérée.[209] En effet, « Seules les personnes disposant de la clé de déchiffrement peuvent connaître la vraie donnée. C'est un moyen très efficace de lutte contre la fuite, l'interception et le vol des données. »[210]

L'utilisation de la clé privée pour signer numériquement des données peut aider à garantir l'authenticité des informations sans révéler la clé privée elle-même. En effet, l'utilisation de clés publiques dans les transactions permet de pseudonymiser les parties prenantes. Au lieu de révéler leur identité réelle, les utilisateurs peuvent être représentés par leurs clés publiques.[211] Cela peut contribuer à améliorer l'anonymisation en fournissant une assurance que les transactions ou les données proviennent d'une entité authentique sans divulguer d'informations personnelles.

Le chiffrement asymétrique offre un niveau supplémentaire de sécurité aux transactions en ligne. Même si les clés publiques sont disponibles publiquement, il est extrêmement difficile de déchiffrer les données sans la clé privée correspondante.[212] Cela réduit considérablement les risques de fraude et de piratage, améliorant ainsi la sécurité des transactions en ligne. Grâce à ce fonctionnement, la blockchain parvient à garantir la confidentialité, l'authenticité et l'intégrité des données et des transactions stockées sur la chaîne, tout en offrant un environnement sécurisé et décentralisé.[213]

---

[207] *Supra* note 60.

[208] « La cryptomonnaie », journaliste Binh An Vu Van, réalisatrice Jeannita Richard, épisode de Radio Canada du 31 mars 2019.

[209] « La chaîne de blocs Ethereum veut faire sa révolution verte », 12 septembre 2022, Agence France Presse, site web Radio Canada.

[210] « Le RGPD : un atout ou un frein pour la cybersécurité ? », Adel Jomni, 2019, Dalloz IP/IT 2019. 352.

[211] « Le droit à l'épreuve des technologies », Reza Moradinejad, Ouvrage *Comment la chaîne de blocs va transformer le droit,* Charlaine Bouchard, 2020, p. 115-134.

[212] « Chiffrement asymétrique : définition, architecture et utilisation », site web Okta.

[213] Notes du cours de l'étudiante Jeanne Soubrié « Protection et sécurité des données personnelles à l'ère du numérique », printemps 2023, Pierre-Luc Déziel, Québec, Université Laval.

Il serait alors possible d'envisager d'utiliser ce chiffrement asymétrique dans un autre contexte que celui de la blockchain. En effet, les entreprises pourraient avoir recours à ce mécanisme de chiffrement asymétrique dans le cadre de leurs relations commerciales, notamment avec leurs sous-traitants. Par exemple, les entreprises utiliseraient ce mécanisme en chiffrant les données avec la clé publique du sous-traitant. De cette manière, les entreprises peuvent s'assurer que seules les personnes disposant de la clé privée correspondante peuvent accéder et déchiffrer les données. Cela réduit les risques de divulgation non autorisée des informations pendant le transfert.

Egalement, une fois que les données sont entre les mains des sous-traitants, le chiffrement asymétrique peut être utilisé pour garantir leur stockage sécurisé. Les sous-traitants peuvent stocker les données chiffrées et ne peuvent les déchiffrer que lorsqu'ils en ont besoin, en utilisant leur clé privée. Cela protège les données sensibles contre les accès non autorisés, même en cas de compromission du système de stockage.

De plus, le chiffrement asymétrique peut être utilisé pour gérer les accès et les autorisations aux données des sous-traitants de manière sécurisée. Les entreprises peuvent attribuer des clés publiques spécifiques à chaque sous-traitant et définir des autorisations en fonction de ces clés. Par exemple, certaines clés peuvent permettre un accès en lecture seule, tandis que d'autres peuvent permettre la modification des données. Cela permet de morceler les accès en permettant un contrôle en cascade pour déterminer qui peut accéder et manipuler les données.

Ainsi, bien que ce fonctionnement de chiffrement asymétrique soit utilisé majoritairement au sein de la blockchain, il serait intéressant de l'extirper de son cadre initial afin de l'appliquer de manière plus généralisée aux relations commerciales entre un responsable de traitement et ses sous traitants pour procéder à une pseudonymisation d'autant plus sécurisée.

De quelle manière les législations française et européenne évoluent-elles afin de s'adapter aux nouvelles techniques de ré-identification ? Quels sont véritablement les progrès qui pourraient être apportés dans les textes pour sécuriser et rendre plus fiable la notion d'anonymisation ?

III. La nécessaire mise à jour de la législation française et européenne

Après avoir analysé les difficultés auxquelles fait face la notion d'anonymisation, il est intéressant d'envisager les mises à jour législatives qui permettraient de réduire au mieux les risques de ré-identification des données personnelles. Pour cela, le Royaume-Uni s'illustre comme précurseur en ne prenant pas seulement en compte les caractéristiques techniques et en appliquant une protection in concreto (A.). Alors, afin d'assurer une juste et cohérente application de la notion d'anonymisation, serait-il envisageable d'adopter la théorie des seuils de risques variables ? (B.)

A. L'approche du Royaume-Uni avec l'attractivité de la donnée

Au Royaume-Uni, avant l'entrée en vigueur du RGPD, un code de bonnes pratiques a été publié en 2012 par le Information Commissioner's Office (ICO) sur l'anonymisation des données[214]. Evidemment, bien que le code de bonnes pratiques sur l'anonymisation des données soit toujours en vigueur, les organisations sont tenues de le compléter par d'autres ressources et de tenir compte des développements récents dans le domaine de la protection des données pour garantir une conformité adéquate.

Ce code prend en compte l'attractivité de la donnée pour déterminer à quel point elle doit être protégée : « In reality though, some types of data will be more attractive to a motivated intruder than others – and more consequential for individuals. »[215] Egalement, leur analyse porte sur les conséquences d'un vol ou d'une perte de données en fonction de sa sensibilité, du contexte dans lequel elle a été divulguée et de l'identité de la personne concernée[216].

En effet, le code de bonnes pratiques dispose que, certaines données pourraient être recherchées plus spécifiquement pour des raisons particulières et il convient de les analyser.[217] Par exemple, certaines données pourraient être récoltées plus précisément dans le but de causer de l'embarras, du tort ou de nuire à la réputation d'une personne ou d'une organisation. Cela pourrait inclure des informations embarrassantes ou compromettantes qui pourraient être utilisées pour faire chanter la victime. Egalement, il est probable que les données sur des personnalités publiques, telles que les politiciens, les célébrités ou les personnalités médiatiques, soient davantage recherchées pour

---

[214] « Anonymisation : managing data protection risk code of practice », Information Commissioner's Office (ICO).

[215] *Ibid*.

[216] *Supra* note 12.

[217] *Supra* note 215.

révéler des informations sensationnelles ou scandaleuses qui pourraient intéresser les médias ou le public. Notamment, les données concernant une campagne politique ou une activité militante contre une organisation ou une personne spécifique ont plus de chances de faire l'objet de fuites. Cela pourrait inclure la divulgation d'informations compromettantes pour discréditer une organisation ou une figure politique. Enfin, les données relatives à des événements locaux, à des faits divers dans une région sont susceptibles d'être piratées pour de simples raisons de curiosité, pour avoir davantage d'information sur des événements récents.

Tous ces exemples permettent d'illustrer que lors de piratages, de fuites, de vols de données, il est nécessaire d'analyser l'aspect sociologique afin d'identifier les raisons de cette attaque malveillante. En effet, différentes catégories de données peuvent présenter un intérêt accru pour les intrus motivés en raison de leur potentiel à être utilisées à des fins malveillantes, lucratives ou personnelles. Cela pourrait permettre de mieux anticiper les données susceptibles de faire l'objet de cyberattaques afin d'appliquer des seuils d'anonymisation et de pseudonymisation particulièrement élevés pour ces données spécifiques.

L'ensemble de ces éléments permet d'ajuster les notions en fonction de différents facteurs et de rendre les processus techniques plus justes et efficaces. Il pourrait être envisageable d'adopter cet ensemble de facteurs en Union européenne afin d'adapter les notions en fonction du contexte.

Ainsi, qu'entrainerait en pratique la prise en compte de nouveaux éléments dans la définition des notions d'anonymisation et de pseudonymisation ?

B. L'adoption de seuils de risques variables pour une meilleure application des notions

La théorie des seuils de risques variables est une approche qui reconnaît que le risque d'identification de données sensibles peut varier en fonction de divers facteurs, tels que le contexte, la sensibilité des données et les techniques de désidentification utilisées. En effet, « du droit français, l'anonymat a un caractère absolu et strict. Il n'envisage pas la mise en balance coût/ avantage d'une éventuelle levée de l'anonymat tel que le permet la directive 95/46/CE »[218]. Cela aboutit à ce que « les données pour lesquelles l'identification des individus requiert la mise en œuvre de moyens déraisonnables sont considérées comme nominatives par la loi française, mais

[218] *Supra* note 12.

anonymes selon la directive européenne »[219]. Cette théorie vise alors à établir des seuils de risques acceptables pour l'anonymisation des données, en tenant compte de ces facteurs variables.

Les données peuvent être considérées comme plus ou moins sensibles en fonction du contexte dans lequel elles sont utilisées et des informations qu'elles contiennent. Par exemple, les informations médicales et financières sont généralement considérées comme très sensibles, tandis que d'autres types de données peuvent être moins sensibles[220].

Par ailleurs, l'évaluation des risques d'identification des données est une étape importante dans le processus d'anonymisation. Cela implique d'identifier les risques potentiels d'identification des individus dans un ensemble de données et de déterminer si ces risques sont acceptables ou non dans un contexte donné.

Enfin, en fonction des risques identifiés, des mesures de protection supplémentaires peuvent être nécessaires pour garantir la sécurité et la confidentialité des données. Cela peut inclure la mise en œuvre de contrôles d'accès, de politiques de sécurité des données et d'autres mesures visant à réduire les risques d'identification des individus.

En résumé, la théorie des seuils de risques variables reconnaît que la protection de la vie privée des individus dans un ensemble de données peut nécessiter une approche flexible qui prend en compte les divers facteurs qui influent sur le risque d'identification des données sensibles. « Les responsables du traitement des données devraient concentrer leur attention sur les moyens concrets qui seraient nécessaires pour inverser la technique d'anonymisation, notamment en termes de coût et de savoir-faire requis pour mettre en œuvre ces moyens, et sur l'appréciation de leur probabilité et de leur gravité. »[221] En établissant des seuils de risques acceptables et en mettant en œuvre des mesures de protection appropriées, il est possible de minimiser les risques d'identification tout en permettant une utilisation efficace et sécurisée des données.

À mesure que la technologie continue de progresser, il est crucial de rester vigilant face aux nouvelles menaces et opportunités. Les recherches futures devraient se concentrer sur le

---

[219] « Données de santé : anonymat et risque de ré-identification », juillet 2015, n°64, Dossiers Solidarité et Santé.

[220] *Supra* note 14.

[221] *Supra* note 14.

développement de méthodes d'anonymisation plus robustes, capables de résister aux techniques de ré-identification émergentes. De plus, une collaboration étroite entre les chercheurs, les régulateurs et les praticiens est essentielle pour assurer une mise en œuvre efficace des meilleures pratiques et le respect des normes éthiques.

L'anonymisation et la pseudonymisation sont des outils complémentaires dans l'arsenal de la protection des données. Alors que l'anonymisation offre un niveau de protection supérieur en rendant les données véritablement anonymes, la pseudonymisation apporte une flexibilité nécessaire pour de nombreuses applications pratiques. L'émergence d'outils techniques innovants est prometteuse pour l'évolution de ces pratiques, mais il est impératif de continuer à innover et à adapter les stratégies de protection des données pour répondre aux défis sans cesse renouvelés du paysage numérique. En fin de compte, l'objectif est de garantir la confidentialité des données personnelles tout en permettant leur utilisation responsable et bénéfique.

Bibliographie

Règlements Européens et directives :

« Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données », Journal officiel n° L 281 du 23/11/1995 p. 0031 - 0050.

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.

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) n° 45/2001 et la décision n° 1247/2002/CE.

Règlement d'exécution (UE) n° 2023/1507 de la Commission, 20 juill. 2023, établissant les spécifications techniques des besoins en données ainsi que les délais pour la soumission des métadonnées et des rapports sur la qualité pour le thème « Utilisation des TIC et commerce électronique » pour l'année de référence 2024, conformément au règlement (UE) n° 2019/2152 du Parlement européen et du Conseil.

Jurisprudences :

Cour de justice de l'Union européenne, 2ème Chambre, 19 octobre 2016, nº C-582/14, Lamy.

Conseil d'Etat, 10ème et 9ème chambres réunies, 8 février 2017, n° 393714, Dalloz.

Décision du Tribunal de l'Union Européenne, 8e chambre élargie, 26 avril 2023, T-557/20, Dalloz.

Décision Cour de justice de l'Union européenne, 3e chambre, 9 novembre 2023, n° C-319/22, Dalloz.

Cour de cassation, Assemblée plénière 7 novembre 2022, n° 21-83.146, Lexbase.

Documents du European Data Protection Supervisor :

« IPEN 2021 on Synthetic Data - Introduction by Thomas Zerdick », Thomas Zerdick, 18 juin 2021, European Data Protection Supervisor.
« IPEN 2021 on Synthetic Data - Keynote Speech by Wojciech Wiewiórowski », Wojciech Wiewiórowski, 18 juin 2021, site web European Data Protection Supervisor.
« IPEN 2021 on Synthetic Data », Omar Ali Fidal, 18 juin 2021, European Data Protection Supervisor.
Page d'accueil du sire web European Data Protection Supervisor.
« Synthetic Data and Privacy : Experiences Implementing Data Synthesis in a Global Life Sciences Company », Stephen Bamford, 16 juin 2021, Janssen Research & Development, site web European Data Protection Supervisor.
« Synthetic Data », Robert Riemann, site web European Data Protection Supervisor.

Documents du Conseil de l'Europe :

« Intelligence artificielle et protection des données », rapport sur l'intelligence artificielle, Conseil de l'Europe, 2020.
« Single Resolution Mechanism », Conseil de l'Europe.

Documentations de la CNIL et G29 :

« Comprendre les grands principes de la cryptologie et du chiffrement », 24 octobre 2016, site web de la CNIL.
« Conformité RGPD : comment recueillir le consentement des personnes ? », 3 aout 2018, site web de la CNIL.
« Donnée sensible », site web de la CNIL.
« Données synthétiques - Et l'Homme créa les données à son image 2/2 », Alexis Léautier, 17 août 2022, Laboratoire d'Innovation Numérique de la CNIL (LNIC).
« G29 », site web de la CNIL.
« L'anonymisation des données personnelles », 19 mai 2020, site web de la CNIL.
« Protection des données à caractère personnel : Précisions de la CNIL sur l'anonymisation de données personnelles », Lionel Costes, 19 mai 2020, site web de la CNIL.
« Recherche scientifique (hors santé) : Enjeux et avantages de l'anonymisation et de la pseudonymisation », 31 janvier 2022, site web de la CNIL.

Décision CNIL, 5 juillet 2021, n° 2012-219, Dalloz.
Délibération de la CNIL n° 2008-005 du 10 janvier 2008 Objet : portant autorisation unique de mise en oeuvre par les entreprises ou organismes exploitants de médicaments de traitements automatisés de données a caractère personnel relatifs a la gestion des données de santé recueillies dans le cadre de la pharmacovigilance des médicaments postérieurement a leur mise sur le marché.
Groupe de travail « Article 29 » sur la Protection des données, 0829/14/FR WP216, Avis 05/2014 sur les Techniques d'anonymisation, adopté le 10 avril 2014.
Lettre de la CNIL à Mr Ali Fdal par Bertrand Pailhès, 13 février 2023, N/réf : AGE/BPS/CS231015.

Documentations officielles :

« Document de la Commission COM (2002) 96 final du 21 février 2002 : Communication de la Commission au Conseil et au Parlement européen - L'internet nouvelle génération : priorités d'actions dans la migration vers le nouveau protocole internet IPv6 », Dalloz.
Décision nº 2008/616/JAI du 23 juin 2008 concernant la mise en oeuvre de la décision 2008/615/ JAI relative à l'approfondissement de la coopération transfrontalière notamment en vue de lutter contre le terrorisme et la criminalité transfrontalière, Lamyline.
Délibération n° 2020-106 du 29 octobre 2020 portant avis sur un projet de décret relatif au système national des données de santé (demande d'avis n° 20011090), site web Légifrance.

Documents d'autorités étrangères :

« Anonymisation : managing data protection risk code of practice », Information Commissioner's Office (ICO).
Arrêté du 11 janvier 2024, 9977020, Garante Per La Protezione Dei Dati Personali.
« Chapter 3 : pseudonymisation : Draft anonymisation, pseudonymisation and privacy enhancing technologies guidances », février 2022, Information Commissioner's Office (ICO).
« How should we assess security and data minimisation in AI ? », Information Commissioner's Office (ICO), site web ICO ORG.
« Opinion of the Data Ethics Commission », Daten Ethik Commission, site web BFDI BUND.

Ouvrages :

« Au delà des big data. Les sciences sociales et la multiplication des données numériques », Étienne Ollion et Julien Boelaert, Ouvrage Sociologie (Vol. 6), mars 2015, Cairn.
« Calibrating noise to sensitivity in private data analysis », Cynthia Dwork, Frank McSherry, Kobbi Nissim et Adam Smith, Ouvrage Theory of cryptography, 2006.
« Chapitre I : Les composantes de la notion de donnée à caractère personnel », S Vergnolle, Ouvrage L'effectivité de la protection des personnes par le droit des données à caractère personnel, 13 septembre 2022, Strada Lex Europe.
« Chapitre 131 - Données de connexion », Livre 1 Les données à caractère personnel, Christiane Féral-Schuhl, 2020-2021, Dalloz.
« Du Bitcoin aux DAO : les fondations techniques de la blockchain », Jean-Noël Colin, Ouvrage Les blockchains et les smart contracts à l'épreuve du droit, 30 octobre 2020.
« L'extension du domaine de la donnée », Valérie-Laure Benabou, Ouvrage Legicom (N° 59), février 2017, Cairn.
« Le droit à l'épreuve des technologies », Reza Moradinejad, Ouvrage Comment la chaîne de blocs va transformer le droit, Charlaine Bouchard, 2020.
Ouvrage The Algorithmic Foundations of Differential Privacy, Cynthia Dwork, 2014.
« Qu'est-ce qu'une paire de clefs privée/publique ? », Jean-Guillaume Dumas, Pascal Lafourcade, Etienne Roudeix, Ariane Tichit et Sébastien Varrette, Ouvrage Les NFT en 40 questions, 2022, Cairn.
« Robust De-anonymization of Large Sparse Datasets », A. Narayanan et Shmatikov, Ouvrage IEEE Symposium on Security and Privacy, 2008.« Introduction », Évelyne Broudoux et Ghislaine Chartron, Ouvrage Big Data - Open Data : Quelles valeurs ? Quels enjeux ?, 2015, Cairn.
« Vie privée et big data », Jean-Charles Cointot et Yves Eychenne, Ouvrage La Révolution du Big Data, 2014, Cairn.

Revues et doctrine :

« Anonymisation through separation : what recent cases teach us about the EU's anonymisation standards », Lore Leitner, Gabe Maldoff et Mickey Lee, site web Westlaw Edge UK.
« Broken Promises of Privacy : Responding to te Surprising Failure of Anonymization », Paul Ohm, Law review 1701, 2010, site web Uclala W Review.

« CJUE : les adresses IP « dynamiques » sont des données personnelles au sens du droit de l'Union », Elisabeth Autier, 8 novembre 2016, Dalloz actualités.
« Décision CRU c/ Deloitte : Une nouvelle pierre a l'édifice de la notion de donnée a caractère personnel », Lorette Dubois, 1er juin 2023, Dalloz actualités.
« Des données à la responsabilité : de l'anonymisation à l'attaque par réidentification », François Viangalli, 1er août 2020, Revue Lamy droit de l'Immatériel n°173.
« Des données personnelles pseudonymisées transférées à un tiers deviennent-elles anonymes ? », Alexandre Jotterand, 13 juin 2023, site web SwissPrivacy.
« Diffusion et réutilisation des informations publiques : « Open data » – données ouvertes – », Alexandre Lallet et Pearl Nguyen Duy, septembre 2020, Dalloz.
« Droit des données personnelles », Lorraine Maisnier-Boché, 9 septembre 2023, Communication Commerce Electronique n°9.
« Introduction », Lionel Costes, 1er juin 2020, Revue Lamy Droit de l'Immatériel n°171.
« La génération de données synthétiques par des systèmes d'intelligence artificielle, une nouvelle méthode de protection et de valorisation des données », Bertrand Cassar, 2024, Dalloz IP/IT 2024.282.
« La protection des données personnelles, une obligation pour toutes les entreprises », 2019, Magazine « Sécuriser le traitement des traces numériques dans le cadre du RGPD : anonymisation et pseudonymisation », Cairn.
« L'autodétermination informationnelle a l'épreuve des évolutions technologiques », Pierre Bordais, 2024, Dalloz IP/IT 2024. 72.
« Le chiffrement, ou l'apport de la cryptologie à la sécurisation du stockage, de la transmission et du traitement des données », Louis Goubin, Annale des Mines - Réalités industrielles 2022/3, août 2022, Cairn.
« Le clair-obscur autour du délit d'administration d'une plateforme en ligne pour permettre une transaction illicite », Stefan Trifkovic, 28 février 2024, Lexbase.
« Le dark web », mis à jour en novembre 2023, Le Lamy droit pénal des affaires.
« L'enjeu de l'anonymisation à l'heure du big data », Hélène Tanghe et Paul-Olivier Gibert, 25 janvier 2018, Revue Française des Affaires Sociales 2017/4, Cairn.
« Le RGPD : un atout ou un frein pour la cybersécurité ? », Adel Jomni, 2019, Dalloz IP/IT 2019. 352.
« Notion de « donnée à caractère personnel » : les précisions du TUE », Lionel Costes, 2 mai 2023, Le Lamy de Droit de l'Immatériel.

« Nouvel éclaircissement de la notion de données personnelles », Jerôme Lasserre Capdeville, 2023, Dalloz IP/IT 2023. 540.
« Obligation on car makers to communicate vehicle ID numbers is compatible with GDPR », Thomson Reuters, Westlaw Edge UK.
« Protection des données personnelles : quelles garanties face a l'émergence des dispositifs innovants de sécurité ? », Romain Perray et Hélène Adda, 2019, Dalloz.
« Recherche scientifique et protection des données personnelles à l'ère du Big Data », Irene Olivan Garcia, 1er octobre 2019, Revue Lamy droit des affaires n°152.
« Régime du Réseau privé virtuel avocat », 1er mai 2012, Revue Lamy Droit de l'Immatériel n°82.
« RGPD : un premier bilan », 20 février 2019, Revue Petites affiches n°037, La base Lextenso.
« Tor en pratique », David A. Levastre-Bodoule Sosso, 2021, Dalloz IP/IT 2021. 89.

« 4051 - Comment identifier les données ou informations, les traitements et les personnes soumis à la réglementation relative à la protection des données personnelles ? », mis à jour en avril 2023, Le Lamy droit du numérique.
« 4054 - Quelles sont les données personnelles indirectement identifiantes ? », avril 2023, Guide Le Lamy droit du numérique.

Documents de l'ENISA :

« PETS controls matrix : a systematic approach for assessing online and mobile privacy tools », 20 décembre 2016, ENISA.
« Techniques et meilleures techniques de pseudonymisation : Recommandations sur l'usage des technologies conformément aux dispositions en matière de protection des données et de respect de la vie privée », novembre 2019, European Union Agency For Cybersecurity (ENISA).

Documents de Anonos, Sarus et Aircloak :

« Aircloak Anonymization », juin 2017, site web Aircloak.
« Aircloak Insights », site web Aircloak.
« Anonymization with Aircloak: how it works », site web Aircloak.
« Differential Privacy », site web Sarus.

« Diffix and the Fundamental Law of Information Recovery », Paul Francis, 11 août 2020, site web Aircloak.
« Ensuring Synthetic Data Privacy », site web Anonos.
« How Variant Twins Work », site web Anonos.
« Maxime Agostini's interview on Cybernews », 27 juillet 2022, The Sarus Blog, site web Sarus.
Page d'accueil du site web Anonos.
Page d’accueil du site web Sarus.
Page d’accueil du site web Statice.

Articles de presse :

« Advanced Encryption Standard (AES) », site web Iot Industriel.
« Anonos nommé par Frost & Sullivan comme le leader de l’innovation technologique dans le secteur britannique des services de désidentification des données des patients », Businesswire, 12 janvier 2023, site web Silicon.
« Chiffrement asymétrique : définition, architecture et utilisation », site web Okta.
« Decoding Data? ECJ’s verdict on Vehicle Identification Numbers as personal data », Lennart Schübler, 9 novembre 2023, site web Bird&bird.
« Fort Knox pour vos données : comment le chiffrement AES sécurise votre entreprise », site web Veritas.
« Introduction aux générateurs de nombres aléatoires », site web Ofcm.
« La chaîne de blocs Ethereum veut faire sa révolution verte », 12 septembre 2022, Agence France Presse, site web Radio Canada.
« La confidentialité différentielle au service de la privacy », 20 décembre 2019, site web In Cyber News.
« La pseudonymisation des données à caractère personnel : une mesure de sécurité à déployer ? », Laure Landes-Gronowski et Marie Miliotis, 6 novembre 2019, site web Agil’it.
« La pseudonymisation des données personnelles dans le cadre du RGPD », 20 décembre 2022, site web Le blog data.
« Les limites de l’anonymisation des données », 22 décembre 2019, site en ligne Malekal.
« Norwegian Confederation of Sport fined for inadequate testing », 2021, site web Datalsynet.
« Pseudonymisation des données : principes, techniques et bonnes pratiques », 7 février 2023, site web Vaadata.

« Quand l’IA garantit la confidentialité des données », 17 octobre 2023, site web French Impact Technologie IOT.
« Sarus lève 2 ME pour accélérer l'innovation digitale », Claude Leguilloux, 22 avril 2020, site web boursier.com.
« Sarus lève deux millions d'euros pour son outil d'analyse de données respectueux de la vie privée », Alice Vitard, 21 avril 2020, site web Usine digitale.
« Sécurité absolue : le chiffrement AES-128 d'Ajax Systems est désormais certifié par le NIST », 9 avril 2024, site web Ajax.
« Tout ce qu'il faut savoir sur le navigateur Tor », J. M. Porup, IDG NS (adapté par Jean Elyan), 20 octobre 2019, site web Le monde informatique.

Notes de cours :

Notes de cours de l’étudiante Jeanne Soubrié « Droit et chaînes de blocs », Charlaine Bouchard, printemps 2023, Québec, Université Laval.
Notes du cours de l’étudiante Jeanne Soubrié « Protection et sécurité des données personnelles à l’ère du numérique », Pierre-Luc Déziel, printemps 2023, Québec, Université Laval.

Autres :

« Will You Be Ready for DORA in January 2025? How Anonos Variant Twins Revolutionize Cybersecurity Under the EU Digital Operational Resilience Act », Gary LaFever, 16 janvier 2024, Linkedin.
« Anonymisation ou pseudonymisation », Commission de contrôle des informations nominatives.
« Vos données sont-elles pseudonymisées ou anonymisées ? », 29 mars 2023, site web Association Pour la Sécurité des Systèmes d’Information de Santé.
« Guide des technologies améliorant la confidentialité (TAC) », Monideepa Mrinal Ro, 25 mai 2022, site web Manage Engine Blog.
« La cryptomonnaie », journaliste Binh An Vu Van, réalisatrice Jeannita Richard, épisode de Radio Canada du 31 mars 2019.
« Privacy by Design : An Overview of Privacy Enhancing Technologies », 26 novembre 2008, Entreprise Privacy Group, site web DSP u Toronto.
« Cloud santé - Anonymisation dynamique », site web Euris.

« Recommendation for Block Cipher Modes of Operation : Methods and Techniques », Morris Dworkin, 2001, National Institute of Standards and Technology.
« Données de santé : anonymat et risque de ré-identification », juillet 2015, n°64, Dossiers Solidarité et Santé.
« Technologies d'amélioration de la confidentialité – Un survol des outils et des techniques », rapport préparé par la Division de l'analyse de la technologie du Commissariat à la protection de la vie privée du Canada, novembre 2017, site web du Commissariat à la Protection de la vie privée du Canada.