Les données codées dans le contexte de la recherche : personnelles ou anonymes ?

Frédéric Erard*

Il existe aujourd'hui une controverse sur le statut juridique des données pseudonymisées ou codées, en particulier pour déterminer si de telles données conservent ou non leur caractère personnel du point de vue de ceux qui ne disposent pas des moyens de réidentifier la personne concernée. Après avoir examiné les tendances existantes sous l'angle de la LPD et du RGPD, l'auteur parvient à la conclusion que les données codées traitées dans le cadre spécifique de la recherche sur l'être humain conservent leur caractère personnel lorsqu'elles se trouvent en main de ceux qui ne disposent pas de clé de décodage.

Der rechtliche Status von pseudonymisierten oder verschlüsselten Daten ist kontrovers, insbesondere hinsichtlich der Frage, ob Personendaten ihren persönlichen Charakter aus der Sicht derer behalten, die nicht über die Mittel verfügen, die betroffene Person zu identifizieren. Nach Prüfung der bestehenden Tendenzen aus Sicht des DSG und der DSGVO kommt der Autor zum Schluss, dass im spezifischen Kontext der Forschung am Menschen verarbeitete verschlüsselte Daten ihren persönlichen Charakter auch gegenüber denjenigen behalten, die sie nicht entschlüsseln können.

Plan

I. Introduction

Dans le secteur de la recherche biomédicale, les nouveaux outils de collecte et d'analyse de données en masse offrent désormais de nouvelles perspectives, dont celle de développer une médecine dite « personnalisée ». Dans les faits, la grande majorité des données traitées dans le contexte de la recherche le sont sous une forme codée. Or, le statut de telles données n'est pas tout à fait clair. Parmi les questions soulevées en figure une qui a fait l'objet de relativement peu d'attention à ce jour : faut-il considérer les données codées comme anonymes du point de vue de ceux qui ne sont pas en mesure de réidentifier les personnes concernées ? Si cette interrogation peut éventuellement surprendre dans le contexte de la recherche (nombreux sont ceux qui considèrent qu'une donnée personnelle codée est personnelle par défaut), elle est toutefois controversée sous l'angle du droit général de la protection des données. Cette question est pourtant loin d'être anodine puisque les traitements de données anonymisées échappent au champ d'application des législations sur la recherche ou sur la protection des données personnelles.

Après une brève présentation des règles relatives à la réutilisation des données personnelles pour la recherche sur l'être humain, la présente contribution aborde le statut des données pseudonymisées sous l'angle du droit général de la protection des données personnelles, puis analyse enfin l'approche qui doit être adoptée dans le cadre de la recherche. L'objet d'étude se limite aux « données » et ne s'étend pas au statut du matériel biologique.

II. Cadre légal de la recherche sur l'être humain

A. Données personnelles dans le cadre de la recherche

La loi fédérale relative à la recherche sur l'être humain[1] encadre les activités de recherche sur les maladies humaines et sur la structure et le fonctionnement du corps humain. Elle trouve non seulement application lorsque de telles activités sont pratiquées directement sur des personnes, mais aussi lorsqu'elles ont, entre autres, pour objet des « données personnelles liées à la santé » (art. 2 al. 1

* Frédéric Erard, Dr. iur., avocat, juriste senior auprès du SIB Institut Suisse de Bioinformatique. Les réflexions menées dans la présente contribution relèvent de l'opinion personnelle de leur auteur et n'engagent en rien son employeur. L'auteur remercie Me Alexandre Jotterand, Me Célian Hirsch et M. Marc Filliettaz pour leur précieuse relecture et leurs commentaires d'une grande pertinence.

[1] Loi fédérale du 30 septembre 2011 relative à la recherche sur l'être humain (Loi relative à la recherche sur l'être humain, LRH ; RS 810.30).

let. e LRH). À l'inverse, la LRH ne s'applique pas à la recherche pratiquée sur les données qui ont été collectées anonymement ou qui ont été anonymisées (art. 2 al. 2 let. c LRH). Le principe est donc en apparence simple : la LRH régit les traitements de données personnelles liées à la santé dans le cadre de la recherche sur l'être humain, mais les traitements de données anonymes sont exclus de son champ d'application.

La LRH établit un niveau de distinction supplémentaire pour la réutilisation des données personnelles liées à la santé en vue de la recherche (recherche rétrospective ; art. 32 ss LRH). Dans ce contexte, les exigences formelles du consentement requis dépendent notamment de la forme des données dont l'utilisation est envisagée pour la recherche selon le triptyque suivant : « données non codées », « données codées » ou « données anonymisées ». Les règles de consentement établies dépendent de surcroît du caractère « génétique » ou « non génétique » des données. En substance, le consentement éclairé de la personne concernée est exigé pour la réutilisation de données non codées (génétiques ou non) et des données génétiques codées (système d'*opt-in*). Sur la base de ce consentement, la réutilisation de données génétiques non codées est cependant uniquement autorisée pour un projet de recherche spécifique alors que la réutilisation des données génétiques codées ou des données non génétiques et non codées peut se faire « à des fins de recherche » (art. 32 al. 1 et 2 et 33 al. 1 LRH ; consentement général). Les exigences en matière de consentement décroissent pour la réutilisation de données non génétiques codées, qui peuvent être réutilisées si la personne concernée ne s'y est pas opposée après avoir été informée (art. 33 al. 2 LRH ; système d'*opt-out*). L'art. 32 al. 3 LRH prévoit un système similaire pour l'anonymisation des données génétiques à des fins de recherche, processus qui peut être mené si la personne concernée ne s'y est pas opposée après avoir été informée (art. 32 al. 3 LRH ; système d'*opt-out*). Enfin, l'art. 34 LRH permet encore la réutilisation à titre exceptionnel de données personnelles liées à la santé lorsque l'obtention du consentement ou l'information sur le droit d'opposition est impossible ou pose des difficultés disproportionnées (*escape clause*). Une telle réutilisation est soumise à des conditions restrictives, en particulier celle selon laquelle aucun document n'atteste le refus de la personne concernée (art. 34 let. a–c LRH).

Schéma récapitulatif proposé par Junod/Elger[2] :

Quelles données réutilisées dans la recherche médicale ?	**Quel régime d'accord du sujet ?**
Données anonymes (DA)	LRH inapplicable – liberté du chercheur
Données personnelles codées (DPC) non-génétiques	Droit d'opposition
Données personnelles codées (DPC) génétiques	Consentement général
Données personnelles non-codées (DPnC) non-génétiques	Consentement général
Données personnelles non-codées (DPnC) génétiques	Consentement spécifique au projet

Le sujet de recherche qui a donné son consentement peut en tout temps révoquer celui-ci sans avoir à justifier sa décision (art. 7 al. 2 LRH). Les conséquences de cette révocation sont précisées à l'art. 10 de l'ordonnance relative à la recherche sur l'être humain à l'exception des essais cliniques[3], qui indique que les données personnelles liées à la santé doivent être anonymisées après avoir été analysées, tout en prévoyant certaines exceptions à cette anonymisation[4]. Toutefois, les données anonymisées suite à la révocation du consentement de la personne concernée ne sont pas pour autant librement mises à la disposition de la recherche future en général au motif qu'elles tomberaient en dehors du champ d'application de la LRH. Comme l'explique à juste titre Rudin, il faut distinguer deux situations[5]. Si les données en question sont conservées pour des recherches futures, la révocation du consentement doit entraîner leur effacement, à moins qu'une prolongation du stockage ne soit valablement justifiée. Quant aux règles prévues par l'art. 10 ORH (anonymisation ou exceptions à celles-ci), elles se limitent aux projets en cours, de telle manière à limiter l'impact disproportionné d'une révocation sur le bon déroulement d'un projet de recherche spécifique.

Le système de consentements établi par la LRH est complexe et critiqué[6], et son examen détaillé excéderait les limites de la présente contribution, dont l'objet porte sur le statut des données codées. Ces quelques lignes introductives permettent toutefois de faire un premier

[2] Valérie Junod/Bernice Elger, Données codées, non codées ou anonymes : des choix compliqués dans la recherche médicale rétrospective, Jusletter 10 décembre 2018, N 8.

[3] Ordonnance du 20 septembre 2013 relative à la recherche sur l'être humain à l'exception des essais cliniques (Ordonnance relative à la recherche sur l'être humain, ORH ; RS 810.301).

[4] L'art. 10 al. 2 ORH précise ainsi que l'anonymisation des données personnelles n'est pas nécessaire lorsque la personne concernée y a expressément renoncé au moment de la révocation ou lorsqu'il est évident depuis le début du projet de recherche qu'une anonymisation n'est pas possible et que la personne concernée a consenti à participer au projet après en avoir été suffisamment informée.

[5] HFG-Rudin, art. 32 N 11–13, in : Bernhard Rütsche (éd.), Humanforschungsgesetz (HFG), Berne 2015 (cit. SHK HFG-Auteur).

[6] Junod/Elger (n. 2), N 16 ss.

constat : si la LRH ne s'applique pas à la recherche pratiquée sur les données qui ont été collectées anonymement ou qui ont été anonymisées (art. 2 al. 2 let. c LRH), elle définit cependant à tout le moins les conditions de réutilisation des données codées (art. 32 et 33 LRH) ainsi que les conditions auxquelles des données génétiques peuvent être anonymisées (art. 32 al. 3 LRH).

B. Définitions

Comme bien souvent dans les sciences juridiques, la définition des termes employés joue un rôle central. Dans le contexte qui nous occupe, l'examen du statut des « données codées » implique non seulement de définir ce concept, mais aussi et d'abord de définir ce qu'il faut entendre par « données personnelles liées à la santé » et par « données anonymisées ».

La LRH définit les « données personnelles » liées à la santé comme « les informations concernant une personne déterminée ou déterminable qui ont un lien avec son état de santé ou sa maladie, données génétiques comprises » (art. 3 let. f LRH). À l'exception du fait que la LRH se limite aux traitements des données « liées à la santé », la notion de « données personnelles » se recoupe avec celle qui figure dans la LPD, définie à son art. 3 let. b comme « toutes les informations qui se rapportent à une personne identifiée ou identifiable »[7]. La notion de personne « identifiée » pose rarement problème en pratique et vise les situations où l'identité d'une personne (et d'elle seule) ressort directement des informations détenues[8]. Une personne est considérée comme « identifiable » lorsqu'elle peut être identifiée par corrélation d'informations tirées des circonstances ou du contexte[9]. Dans son message relatif à la révision totale de la LPD, le Conseil fédéral a précisé que l'identification pouvait résulter d'un seul élément, à l'exemple du numéro de téléphone, d'immeuble, du numéro AVS, ou d'empreintes digitales, mais également par le recoupement de plusieurs informations telles que l'adresse, la date de naissance et l'état civil[10].

Les données « anonymisées » sont quant à elles définies comme celles « qui ne peuvent être [mises] en relation avec une personne déterminée ou ne peuvent l'être sans engager des efforts démesurés » (art. 3 let. i LRH)[11]. Pour assurer une adaptation plus facile aux évolutions normatives ou techniques, le législateur fédéral a délégué au Conseil fédéral la tâche de préciser les exigences posées pour l'anonymisation des données (art. 35 LRH)[12]. L'art. 25 ORH expose ainsi que l'anonymisation nécessite de rendre définitivement méconnaissable ou de détruire toutes les informations qui, combinées, permettent de rétablir l'identité de la personne sans efforts disproportionnés. Parmi les informations qui doivent être rendues méconnaissables ou être détruites figurent à tout le moins le nom, l'adresse, la date de naissance et les numéros d'identification caractéristiques (art. 25 al. 2 ORH). Le droit suisse n'offre toutefois aucune autre information sur le type d'indicateurs indirects qui permettraient la réidentification d'une personne[13]. Il ressort des dispositions citées que l'anonymisation ne doit pas nécessairement être absolue, puisque la loi l'admet déjà si la réidentification nécessite des « efforts disproportionnés ». Certains auteurs parlent à cet égard d'anonymisation *de facto*[14]. Les efforts requis sont considérés comme disproportionnés si, selon le cours ordinaire des choses, on ne peut pas s'attendre à ce qu'une personne intéressée les mette en œuvre[15]. Pour mener cette appréciation, il faut non seulement prendre en considération l'importance des moyens techniques à déployer, mais aussi l'intérêt propre de celui qui souhaiterait réidentifier les données[16], voire la durée de conserva-

---

7 Message du 21 octobre 2009 sur la loi fédérale relative à la recherche sur l'être humain, FF 2009 7259 ss (cit. Message LRH), 7311 ; SHK HFG-van Spyk/Rudin/Sprecher/Poledna (n. 2), art. 3 N 43. La définition adoptée dans la nouvelle LPD, qui entrera probablement en vigueur au début 2022, est très similaire au droit actuel et se rattache également aux concepts de personne identifiée ou identifiable (art. 5 let. a nLPD).

8 Philippe Meier, Protection des données. Fondements, principes généraux et droit privé, Berne 2011, N 431 ; David Rosenthal, in : David Rosenthal/Yvonne Jöhri, Handkommentar zum Datenschutzgesetz, Zurich 2008 (cit. Auteur, Handkommentar DSG), art. 3 let. a N 20.

9 Message du 15 septembre 2017 concernant la loi fédérale sur la révision totale de la loi fédérale sur la protection des données et sur la modification d'autres lois fédérales, FF 2017 6565 ss (cit. Message nLPD), 6639. ATF 138 II 346 c. 6.1, in : JdT 2013 I 71 ; 136 II 508 c. 3.2, in : JdT 2011 II 446. Célian Hirsch/Emilie Jacot-Guillarmod, Les données bancaires pseudonymisées – Du secret bancaire à la protection des données, RSDA 2020, 151 ss, 158. Plus généralement sur cette question, voir : Thomas Probst, Die unbestimmte « Bestimmbarkeit » der von Daten betroffenen Person im Datenschutzrecht, PJA 2013, 1423 ss ; Rolf H. Weber/Orsolya Fercsik Schnyder, « Was für ‹ne Sorte von Geschöpf ist euer Krokodil? » – Zur datenschutzrechtlichen Qualifikation von IP-Adressen, sic! 2009, 577 ss, 583.

10 Message nLPD (n. 10), 6639. Dans un sens similaire : Meier (n. 8), N 433.

11 Pour une présentation technique du processus de dé-identification des données : Simson L. Garfinkel, De-Identification of Personal Information, Gaithersburg (USA) 2015.

12 Message LRH (n. 7), 7340.

13 Junod/Elger (n. 2), N 16.

14 Meier (n. 8), N 440 ss.

15 SHK HFG-Rudin (n. 5), art. 35 N 7 ; Meier (n. 8), N 440.

16 ATF 136 II 508 c. 3.2, in : JdT 2011 II 446. SHK HFG-Rudin (n. 5), art. 35 N 7.

tion prévue des données[17]. À l'inverse, l'engagement du destinataire des données de ne pas tenter de réidentifier les données n'a pas d'influence sur le caractère anonyme ou non d'une donnée[18]. Eu égard aux évolutions technologiques (*big data*, intelligence artificielle), du caractère toujours moins onéreux de l'accès à ces technologies et à l'augmentation du nombre de bases de données disponibles, la nécessité de déployer des efforts disproportionnés pour rattacher des données à une personne particulière ne saurait toutefois être reconnue facilement[19].

Enfin, les données personnelles « codées » liées à la santé sont définies par le législateur comme « les données qui ne peuvent être mises en relation avec une personne déterminée qu'au moyen d'une clé » (art. 3 let. h LRH). Le codage (aussi appelé « pseudonymisation », surtout dans le contexte du droit de la protection des données) implique donc le remplacement de certaines caractéristiques identifiantes, de telle manière à ce que seuls ceux qui possèdent la clé (en principe au moyen d'une liste de correspondance) puissent remonter aux données personnelles[20]. Le codage est donc par définition réversible. Dans le domaine de la recherche, les données personnelles sont réputées correctement codées lorsqu'elles peuvent être qualifiées d'anonymisées dans l'optique d'une personne qui n'a pas d'accès au code (art. 26 al. 1 ORH). L'ordonnance d'exécution de la LRH précise par ailleurs que le code doit être conservé par une personne qui n'est pas impliquée dans le projet de recherche, séparément des données personnelles (art. 26 al. 2 ORH).

III. Une controverse issue du droit de la protection des données ?

Dans le contexte helvétique de la recherche sur l'être humain, les données codées sont généralement considérées comme des données personnelles, sans autre forme de distinction. Sous l'angle de la LPD, il existe pourtant aujourd'hui une controverse sur les règles applicables en cas de transfert de données pseudonymisées à un tiers ne disposant pas de la clé de réidentification, notamment pour savoir si les données pseudonymisées doivent être considérées comme des données anonymes du point de vue de tels destinataires. Si cette question n'a pas fait l'objet de nombreuses discussions en doctrine, son importance est toutefois centrale. L'adoption d'une telle approche a en effet pour résultat de libérer le destinataire des données considérées comme anonymes de toute obligation en matière de protection des données personnelles, le traitement des données anonymes étant exclu du champ d'application de la LPD. Le résultat serait le même dans le contexte de la recherche puisque, comme on l'a vu, la LRH exclut également de son champ d'application le traitement des données anonymes (art. 2 al. 2 let. c LRH). Avant de prendre position sur cette question dans le contexte de la recherche, il convient d'examiner les développements menés sur cette question sous l'angle du droit de la protection des données.

Deux grandes approches théoriques peuvent être envisagées. Selon une première approche, dite « absolue », il suffit qu'un seul des acteurs de la communication d'une donnée personnelle pseudonymisée (expéditeur ou destinataire) soit en mesure de réidentifier la personne concernée pour que la donnée pseudonymisée soit considérée comme personnelle *« erga omnes »*[21]. À l'inverse, selon l'approche dite « relative », une donnée pseudonymisée est considérée comme personnelle seulement pour la personne qui est en mesure de réidentifier la personne concernée. En d'autres termes, une donnée pseudonymisée serait anonyme lorsqu'elle se trouve en main d'un individu incapable de la rattacher à une personne (en particulier lorsqu'elle n'a pas accès au code de réidentification), mais cette même donnée pseudonymisée constituerait une donnée personnelle à l'égard de la personne qui est en mesure de réidentifier la personne concernée sans déployer d'efforts disproportionnés, généralement si elle est en possession du code de réidentification[22].

L'approche relative est régulièrement préconisée par la doctrine suisse en droit général de la protection des données[23]. Elle reflète aussi le parti pris par le Conseil fédéral dans son Message sur la révision totale de la LPD, qui

---

17 Meier (n. 8), N 443.

18 Junod/Elger (n. 2), N 11.

19 Handelsgericht ZH, HG150170, 30.5.2017, c. 5.3.5.2. BSK DSG-Blechta, art. 3 N 13, in : Urs Maurer-Lambrou/Gabor-Paul Blechta (éd.), Datenschutzgesetz. Öffentlichkeitsgesetz, 3e éd., Bâle 2014 (cit. BSK DSG-auteur) ; SHK HFG-Rudin (n. 5), art. 35 N 7 ; Junod/Elger (n. 2), N 16.

20 Meier (n. 8), N 446. Pour un rapport récent sur les techniques de pseudonymisation et les cas d'application : European Union Agency for Cybersecurity, Data pseudonymisation : advanced techniques & use cases, janvier 2021.

21 Hirsch/Jacot-Guillarmod (n. 9), 160–161. Le Préposé fédéral à la protection des données et à la transparence s'est prononcé en faveur d'une telle approche : PFPDT, 22e Rapport d'activités 2014/2015, Berne 2015, 68.

22 Hirsch/Jacot-Guillarmod (n. 9), 161. Sur cette question, voir aussi le Rapport du Préposé fédéral à la protection des données et à la transparence 22-2014/2015 « Externalisation à l'étranger de données bancaires pseudonymisées ».

23 P. ex. Rosenthal, Handkommentar DSG (n. 8), art. 3 let. a, N 36 ; Beat Rudin, in: Bruno Baeriswyl/Kurt Pärli (éd.), Datenschutzge-

affirmait que : « La loi ne s'applique pas aux données qui ont été anonymisées si une ré-identification par un tiers est impossible (les données ont été anonymisées complètement et définitivement) ou ne paraît possible qu'au prix d'efforts tels qu'aucun intéressé ne s'y attèlera. Cette dernière règle vaut aussi pour les données pseudonymisées »[24]. On notera toutefois que la nouvelle LPD, dans sa version finale, prévoit que le traitement de données personnelles à des fins ne se rapportant pas à des personnes (par exemple, recherche, planification ou statistique) peut constituer un intérêt prépondérant susceptible de justifier un traitement de données. Si des données sensibles sont traitées dans ce cadre, l'art. 31 al. 2 let. e ch. 2 nLPD imposera toutefois au responsable du traitement de ne les communiquer à des tiers « que sous une forme ne permettant pas d'identifier la personne concernée ; si cela n'est pas possible, des mesures [devront] être prises qui [garantiront] que les tiers ne traitent les données qu'à des fins ne se rapportant pas à des personnes ». Cette formulation laisse ainsi suggérer que la future LPD contient des règles susceptibles de s'appliquer à des données qui seraient pseudonymisées.

L'approche relative s'oppose à celle qui avait été défendue par le Préposé fédéral à la protection des données et à la transparence (PFPDT) dans son 22e Rapport d'activités 2014/2015. Sur la question de l'externalisation à l'étranger de données bancaires pseudonymisées, il avait préconisé l'approche absolue en expliquant que le législateur avait délibérément choisi une loi techniquement neutre, qui devait permettre au droit de suivre l'évolution de la technologie moderne[25]. Il appuyait par ailleurs son raisonnement sur un arrêt du Tribunal fédéral que je présente brièvement.

Dans l'arrêt dit « Logistep » rendu en 2010[26], le Tribunal fédéral a été amené à examiner l'activité d'une entreprise qui avait développé un logiciel capable de surveiller les réseaux *peer-to-peer* et de collecter des données en lien avec les téléchargements d'œuvres protégées, en particulier l'adresse IP de la connexion par laquelle le téléchargement était opéré ainsi que la date, l'heure et la durée de connexion. Les données étaient ensuite communiquées au détenteur des droits sur les œuvres, qui pouvait déposer une plainte pénale contre inconnu, puis consulter le nom de celui ou celle qui avait téléchargé l'œuvre en faisant valoir son droit de consulter le dossier pénal. Les adresses IP collectées étaient dites dynamiques, c'est-à-dire qu'elles changeaient lors de chaque connexion. Dans cette configuration, les détenteurs de Logistep faisaient valoir que la LPD ne leur était pas applicable puisqu'ils ne traitaient pas de données personnelles, n'étant eux-mêmes en possession d'aucun moyen permettant de réidentifier les personnes derrière les connexions. Tout en affirmant qu'il fallait se placer du point de vue du détenteur de l'information pour déterminer si celle-ci pouvait être associée à un individu déterminable et donc constituer une donnée personnelle, le Tribunal fédéral a toutefois estimé qu'il suffisait qu'une partie des destinataires soit en mesure de réidentifier la personne pour que l'information constitue une donnée personnelle. Il a ainsi jugé que les détenteurs de Logistep, même s'ils n'avaient pas eux-mêmes la possibilité de réidentifier les personnes concernées, traitaient donc des données personnelles dans la mesure où les personnes à qui ils remettaient ces données (détenteurs des droits d'auteurs) disposaient quant à eux de moyens pour réidentifier les personnes concernées, même si ces moyens nécessitaient des efforts passablement importants (plainte pénale, rattachement de l'adresse IP dynamique à une connexion, consultation du dossier, vérification de la personne qui se trouvait bien derrière l'ordinateur, etc.)[27].

Il ressort ainsi de l'arrêt Logistep qu'une donnée peut être considérée comme personnelle à partir du moment où l'une des deux parties à l'échange de données est en mesure de réidentifier la personne concernée. Une telle approche se rattache à la conception absolue, qui permet un parallèle avec qualification des données pseudonymisées, comme l'a fait le PFPDT dans son 22e Rapport 2014/2015. Cet arrêt a toutefois suscité une certaine critique chez une partie de la doctrine, qui estimait cette approche excessive et préconisait de limiter ce raisonnement aux faits particuliers de la cause[28].

Sept ans après l'arrêt Logistep, en 2017, le *Handelsgericht* zurichois a au contraire jugé que des données bancaires pseudonymisées – c'est-à-dire des données pour lesquelles la banque gardait une table de concordance – ne constituaient plus des données personnelles à l'égard de leur destinataire (en l'occurrence, le *Department of Justice* américain) à partir de l'instant où les mesures de pseudonymisation entreprises empêchaient effectivement

setz (DSG), Berne 2015 (cit. Auteur, Stämpflis Handkommentar zum DSG), art. 3 N 14 ; Hirsch/Jacot-Guillarmod (n. 9), 162.

24 Message nLPD (n. 9), 6640.

25 PFPDT, 22e Rapport d'activités 2014/2015, Berne 2015, 68.

26 ATF 136 II 508, in : JdT 2011 II 446.

27 ATF 136 II 508 c. 3.4 et 3.5, in : JdT 2011 II 446, c. 3.4 et 3.5.

28 Pour l'approche jugée excessive, par ex. Philippe Meier, Préposé fédéral à la protection des données et à la transparence c. Logistep AG (recours en matière civile), 8 septembre 2010, JdT 2011 II 446, 462 ss. Pour une limitation de cet arrêt aux faits de la cause, voir en particulier : Hirsch/Jacot-Guillarmod (n. 9), 161 et plus précisément les références citées en note 89.

l'identification de la personne concernée par le destinataire (approche relative)[29]. Il a cependant jugé qu'il appartenait à l'expéditeur des données de prouver qu'il avait adopté les mesures suffisantes pour empêcher la réidentification[30]. Or, précise-t-il, les technologies disponibles aujourd'hui, notamment celles liées au *big data*, laissent peu de place pour l'anonymisation irréversible (ou la pseudonymisation équivalente)[31]. En l'occurrence, la banque qui entendait envoyer des données aux Etats-Unis n'avait pas prouvé qu'elle avait pris les mesures suffisantes pour empêcher la réidentification. Le Tribunal fédéral a rejeté le recours de la banque sans développer d'argumentation particulière sur ce point, se limitant essentiellement à affirmer que la banque qui procédait à l'anonymisation ou à la pseudonymisation procédait bien à un traitement de données personnelles au sens de la LPD, même si le résultat de ce traitement n'était plus une donnée personnelle, respectivement une donnée protégée[32]. Il cite à cet effet Rosenthal/Jöhri qui, à la référence concernée, expliquent que l'anonymisation constitue un « traitement de données » au sens de l'art. 3 let. e LPD même si les données qui en résultent ne sont plus personnelles, sans toutefois aborder dans ce cadre la question de la pseudonymisation[33].

Certains auteurs estiment que l'arrêt du Tribunal fédéral rejetant le recours contre l'arrêt du *Handelsgericht* zurichois apporte une clarification bienvenue et valide l'approche dite relative, mettant même fin aux incertitudes liées à l'arrêt Logistep[34]. En l'absence d'argumentation du Tribunal fédéral sur ce point, une telle affirmation semble toutefois quelque peu prématurée. Parallèlement, il semble surtout important de souligner un autre point soulevé par le Tribunal fédéral, soit celui selon lequel l'opération de pseudonymisation constitue elle-même un traitement de données personnelles, et qu'elle ne peut donc être réalisée que si elle est compatible avec la LPD. Cela s'applique certainement à plus forte raison à tous les traitements effectués par le maître du fichier avec des données pseudonymisées, y compris leur communication à des tiers. Cela dit, l'examen de la doctrine et de la jurisprudence mené dans la présente section rend aujourd'hui certainement compte d'une tendance en Suisse à favoriser l'approche dite relative pour la qualification des données pseudonymisées sous l'angle de la LPD. Avant d'analyser plus en détail la position à adopter pour les traitements de données tombant dans le champ d'application de la LRH, il convient encore de présenter brièvement l'état de situation en droit européen.

IV. Aperçu du droit européen

Le RGPD s'applique aux traitements des données à caractère personnel (art. 2 ch. 1 RGPD). Les données à caractère personnel sont définies par celui-ci comme toute information se rapportant à une personne physique identifiée ou identifiable, étant entendu qu'une personne physique identifiable est celle qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant. Ces identifiants peuvent prendre la forme d'un nom, d'un numéro d'identification, mais aussi d'un ou plusieurs éléments propres à l'identité physique, physiologique, génétique ou psychique. Le RGPD définit par ailleurs explicitement 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 » (art. 4 5) RGPD).

Le consid. 26 du RGPD, qui sert d'outil d'interprétation, expose que « il y a lieu d'appliquer les principes relatifs à la protection des données à toute information concernant une personne physique identifiée ou identifiable. Les données à caractère personnel qui ont fait l'objet d'une pseudonymisation et qui pourraient être attribuées à une personne physique par le recours à des informations supplémentaires devraient être considérées comme des informations concernant une personne physique identifiable ». Le même considérant ajoute 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, tels que le ciblage ». Enfin, il n'y a « pas lieu d'appliquer les principes relatifs à la protection des données aux informations anonymes, à savoir les informations ne concernant pas une personne physique identifiée ou identifiable, ni aux données à caractère personnel rendues anonymes de telle manière que la personne concernée ne soit pas ou plus identifiable. Le présent règlement ne s'applique, par

[29] HGer ZH, HG150170, 30.5.2017, c. 5.3.5.2.
[30] HGer ZH, HG150170, 30.5.2017, c. 5.3.5.8.
[31] HGer ZH, HG150170, 30.5.2017, c. 5.3.5.2.
[32] TF, 4A_365/2017, 26.2.2018, c. 5.2.2.
[33] David Rosenthal, Handkommentar DSG, art. 3 let. e, N 63.
[34] Hirsch/Jacot-Guillarmod (n. 9), 162.

conséquent, pas au traitement de telles informations anonymes, y compris à des fins statistiques ou de recherche ».

La lettre du RGPD, et plus particulièrement celle de son consid. 26, tend ainsi à laisser penser que les données pseudonymisées doivent par principe être considérées comme des données personnelles (ou plus précisément un sous-groupe de données personnelles)[35]. À la lumière du RGPD, la pseudonymisation est d'ailleurs essentiellement vue comme une mesure supplémentaire de sécurité (consid. 28 RGPD). Plus encore, la lecture du consid. 26 indique que la détermination du caractère identifiable d'une donnée doit prendre en compte les moyens raisonnablement susceptibles d'être utilisés, à l'exemple du « ciblage » *(singling out)*. En d'autres termes, le fait qu'une personne puisse être « différenciée » sans qu'elle ne soit forcément identifiée doit être pris en compte pour déterminer si une donnée ou un jeu de données permet ou non la réidentification de la personne concernée[36].

L'acception large donnée aux données personnelles par le RGPD reflète également dans une certaine mesure l'avis de l'*Art. 29 Data Protection Working Party*, aujourd'hui remplacé par le Comité Européen de la Protection des Données. Dans une opinion du 10 avril 2014 sur les techniques d'anonymisation[37], l'*Art. 29 Data Protection Working Party* avait ainsi précisé que la pseudonymisation réduisait la possibilité d'établir un lien entre des données et l'identité d'une personne, mais qu'il ne s'agissait pas d'une mesure d'anonymisation[38]. Plus encore, il a estimé qu'il était « crucial de comprendre que, dans le cas où un responsable du traitement des données n'efface pas les données originales (identifiables) au niveau des événements individuels et transmet une partie de cet ensemble de données (par exemple, après avoir supprimé ou masqué les données identifiables), l'ensemble de données résultant constitue encore des données à caractère personnel. Ce n'est que si les données sont agrégées par le responsable de leur traitement à un niveau où les événements individuels ne sont plus identifiables que l'ensemble de données résultant peut être qualifié d'anonyme »[39].

Même si l'opinion de l'*Art. 29 Data Protection Working Party* ne déploie pas directement d'effets juridiques, les éléments qui précèdent, notamment le texte du RGPD, tendent à montrer que ce dernier préconise une approche absolue, selon laquelle les données pseudonymisées qui se trouvent en main d'une personne incapable de les réidentifier, faute de posséder la clé de réidentification, doivent être considérées comme des données personnelles dont le traitement est soumis aux règles du RGPD[40]. Cette opinion est aujourd'hui celle qui est généralement défendue par les autorités nationales de protection des données en Europe[41], mais aussi par le Comité européen de la protection des données[42], même si ce dernier n'a pas publié de rapport spécifique sur cette question à ce jour. Cette approche large s'inscrit au demeurant dans la ligne de la jurisprudence de la CJUE, qui avait jugé sous l'empire de l'ancienne directive européenne 95/46 qu'une personne pouvait être considérée comme responsable conjointe d'un traitement de données personnelles même si elle n'était pas en mesure d'accéder aux données identifiantes (en l'occurrence, son traitement se limitait à la collecte des données ; seul l'autre co-responsable du traitement étant en mesure d'accéder aux données à caractère personnel collectées)[43].

L'approche absolue susmentionnée est toutefois critiquée par une partie de la doctrine, qui soutient notamment que la pseudonymisation peut s'apparenter dans les faits à une anonymisation[44]. Les critiques s'appuient parfois

---

35 David Peloquin/Michael DiMaio/Barbara Bierer/Mark Barnes, Disruptive and avoidable: GDPR challenges to secondary research uses of data, European Journal of Human Genetics 2020, 697 ss, 699. Voir aussi : Julien Rossi, Protection des données personnelles et droit à la vie privée : enquête sur la notion controversée de « donnée à caractère personnel », Paris 2020, 558, qui procède à une analyse détaillée des discussions préalables à l'adoption du RGPD qui voyaient différents groupes s'affronter, notamment les défenseurs de la vie privée qui souhaitaient inscrire la notion de *singling out* directement dans un article du RGPD et non dans un considérant contre des groupes industriels qui ont à l'origine proposé l'insertion du concept de pseudonymisation pour faciliter les traitements de données.

36 Rossi (n. 35), 558. Pour un avis selon lequel la « singularisation » d'une donnée n'est pas suffisante pour en faire une donnée personnelle sous le RGPD : David Rosenthal, Das neue Datenschutzgesetz, Jusletter 16 novembre 2020, N 20 ; David Rosenthal, Personendaten ohne Identifizierbarkeit ?, digma 2017, 198, 198 ss.

37 Groupe de travail « Art. 29 » sur la protection des données, Opinion 05/2014 sur les Techniques d'anonymisation, 10.4.2014.

38 Groupe de travail « Art. 29 » sur la protection des données, Opinion 05/2014 sur les Techniques d'anonymisation, 10.4.2014, 22.

39 Groupe de travail « Art. 29 » sur la protection des données, Opinion 05/2014 sur les Techniques d'anonymisation, 10.4.2014, 10.

40 Dans un sens similaire, par ex. Marit Hansen, in : Spiros Simitis/Gerrit Hornung/Indra Spiecker (éd.), Datenschutzrecht. DSGVO mit BDSG. Grosskommentar, Baden-Baden 2019, art. 4 Nr. 5 N 15.

41 Peloquin/DiMaio/Bierer/Barnes (n. 35), 699.

42 Voir p. ex. European Data protection Board, Lignes directrices 4/2020 relatives à l'utilisation de données de localisation et d'outils de recherche de contacts dans le cadre de la pandémie COVID-19, 21.4.2020, N 17.

43 CJUE C-40/17, 29.7.2019, *Fashion ID*, N 69 et 82.

44 Sur cette controverse, voir p. ex. Wolfgang Ziebart, in : Gernot Sydow (éd.), Europäische Datenschutzgrundverordnung. Handkommentar, 2e éd., Baden-Baden Vienne Zurich 2018, art. 4, N 95 ss.

sur l'arrêt de la CJUE *Breyer c. Allemagne*[45] de 2016. Bien que rendu sous l'empire de l'ancienne directive européenne 95/46, l'arrêt apporte des éléments intéressants sur l'interprétation à donner à la notion de donnée personnelle. La CJUE devait déterminer si une adresse IP dynamique ainsi que les données relatives à la date et l'heure de connexion constituaient des données personnelles aux yeux d'un fournisseur de services de médias en ligne, étant entendu que ces données ne permettaient pas une réidentification sans un recoupement avec des données détenues par le fournisseur d'accès à internet. Selon la CJUE, les données sont seulement personnelles s'il existe un moyen qui peut raisonnablement être mis en œuvre pour identifier la personne concernée. Tel n'est toutefois pas le cas si l'identification est interdite par la loi ou irréalisable en pratique. Dans le cas particulier, la CJUE est toutefois parvenue à la conclusion qu'il existait visiblement des moyens de réidentification, par exemple en s'adressant à l'autorité compétente en cas de cyberattaque[46]. Vu les risques particulièrement faibles pris en compte par le CJUE pour reconnaître le caractère personnel des données, certains auteurs en ont conclu qu'il fallait retenir sur cette base une approche absolue. D'autres ont critiqué ce point de vue et appelé à revoir les critères de qualification des données personnelles à la lumière des possibilités de réidentification mises à disposition de la personne qui détient les données[47].

Il est certainement possible de déduire de l'arrêt *Breyer c. Allemagne* que la qualification d'une donnée personnelle dépend essentiellement des moyens de réidentification à disposition de celui ou celle qui détient les données. Toutefois, comme cela a été exposé, le texte même du RGPD tend aujourd'hui à favoriser une approche absolue, consistant à considérer les données pseudonymisées comme des données personnelles, sans égard au fait que celui qui les détient possède la clé pour les réidentifier.

V. Discussion sur le statut des données codées de recherche en Suisse

Il ressort des deux sections précédentes qu'il existe aujourd'hui une controverse sur la manière d'appréhender et de qualifier les données pseudonymisées sous l'angle des législations sur la protection des données. Si l'approche relative semble aujourd'hui l'emporter dans le contexte d'application de la LPD, l'approche absolue semble plus communément adoptée sous l'empire du RGPD, même si elle est parfois combattue. Reste maintenant à déterminer quel régime doit être appliqué aux données dites « codées » au sens de la LRH.

Il existe indéniablement des avantages à affirmer que les données codées collectées dans le cadre de la recherche sur l'être humain seraient anonymes une fois tombées dans les mains de ceux qui ne seraient pas en mesure de réidentifier le sujet de recherche, en particulier ceux qui ne disposeraient pas de clé de réidentification. On pense notamment au bon fonctionnement des biobanques, qui pourraient ainsi facilement mettre à disposition des chercheurs et des chercheuses l'ensemble de leurs échantillons biologiques et leurs données codées à la simple condition que la clé de réidentification ne leur soit pas communiquée. Les projets de recherche multicentriques seraient quant à eux d'autant plus facilités puisque le codage des données écarterait toute réglementation en matière de protection des données, laissant champ libre aux échanges de données codées entre institutions. Les chercheurs et chercheuses ne seraient guère plus limités que par des règles de conduite propres aux institutions ou aux bonnes pratiques scientifiques. Cela étant, même si cette approche trouverait certainement des partisans dans le milieu de la recherche, elle doit être rejetée pour les raisons qui suivent.

À titre liminaire, il convient d'éclaircir les relations entre la LRH et la LPD, voire entre la LRH et le droit cantonal de la protection des données puisque ce dernier s'applique régulièrement aux activités de recherche sur l'être humain (recherche menée au sein des institutions publiques cantonales). La LPD et les lois cantonales sur la protection des données se limitent à poser un cadre général pour les traitements de données personnelles et elles cèdent le pas face aux (nombreuses) législations qui contiennent des règles spéciales en matière de protection des données (*lex specialis*). En l'occurrence, la LRH constitue une réglementation spéciale qui impose des règles particulières pour les traitements de données dans ce contexte[48]. Les règles générales du droit de la

[45] CJUE C-582/14, Patrick Breyer c. Bundesrepublik Deutschland, 19.10.2016.

[46] CJUE C-582/14, Patrick Breyer c. Bundesrepublik Deutschland, 19.10.2016, N 45–48.

[47] Sur cette question, voir en particulier : Daniel Groos/Evert-Ben Van Veen, Anonymised Data and the Rule of Law, European Data Protection Law Review, vol. 4, 2020, 1 ss, 1-11, qui proposent l'établissement d'un test de réidentification spécifique construit sur le raisonnement de la CJUE. Voir aussi : Peloquin/DiMaio/Bierer/Barnes (n. 35), 699.

[48] SHK HFG-Brunner (n. 5), Vorbemerkungen art. 56–61 N 4.

protection des données continuent toutefois à s'appliquer en parallèle lorsque la LRH reste muette, notamment pour l'application des principes généraux du droit de la protection des données[49]. La détermination du statut des données codées doit donc être analysée en premier lieu au regard de la LRH et non des législations générales sur la protection des données personnelles.

La LRH a pour but de protéger la dignité, la personnalité et la santé de l'être humain dans le cadre de la recherche (art. 1 al. 1 LRH). Elle concrétise ainsi l'art. 118*b* Cst. qui donne mandat à la Confédération de légiférer sur la recherche sur l'être humain, « dans la mesure où la protection de la dignité humaine et de la personnalité l'exige ». Cette même disposition consacre le principe du consentement éclairé dans la recherche, précisant que la loi peut prévoir des exceptions et qu'un refus est contraignant dans tous les cas. La Confédération doit cependant aussi veiller à la liberté de la recherche et tenir compte de l'importance de la recherche pour la santé et la société (art. 118*b* al. 1 Cst.). Le principe du consentement du sujet de recherche et plus généralement de son droit à l'autodétermination et à la vie privée est aujourd'hui reconnu comme un pilier fondamental de la recherche et est consacré par les plus grands textes internationaux en la matière[50]. La gestion des bases de données de santé et celle des biobanques font, par exemple, l'objet d'une déclaration spécifique de l'Association Médicale Mondiale, la Déclaration de Taipei[51], qui rappelle notamment que « les droits à l'autonomie, à la vie privée et à la confidentialité habilitent aussi les individus à exercer un contrôle sur l'utilisation de leurs données personnelles et de leur matériel biologique », ou encore que les personnes concernées ont le droit d'obtenir des informations sur leurs données et leur utilisation, ou encore de modifier leur consentement à tout moment.

Pour concrétiser le mandat de protéger l'intégrité des sujets de recherche, la LRH contient des règles particulières pour la réutilisation des données de santé pour la recherche. Comme on l'a vu (cf. *supra* II.A.), elle institue en particulier un système de consentements différenciés dont les exigences de forme varient selon le type de données en jeu, en distinguant notamment les données non codées, codées ou anonymisées (également selon leur nature génétique ou non). Ces règles spécifiques (*lex specialis* par rapport à la LPD notamment) exposent, par exemple, que les données génétiques peuvent être réutilisées à des fins de recherche sous une forme codée lorsque la personne concernée a donné son consentement éclairé (art. 32 al. 2 LRH), ou encore que les données personnelles non génétiques liées à la santé peuvent être réutilisées à des fins de recherche sous une forme codée lorsque la personne concernée ne s'y est pas opposée après avoir été informée (art. 33 al. 2 LRH). Il ressort de ces deux dispositions que la LRH reconnaît au sujet de recherche un droit d'autodétermination sur ses données codées. Celui-ci doit, par exemple, pouvoir limiter l'utilisation de ses données codées à un seul projet en particulier[52]. Or, si l'on considère que les données codées sont anonymes pour les chercheurs qui les reçoivent et qui se trouvent dans l'impossibilité de réidentifier le sujet de recherche (approche relative), la mise en œuvre du consentement du sujet de recherche (voire de son droit d'opposition) garanti par la LRH deviendrait tout simplement sans objet. Une fois parvenues dans les mains des chercheurs, les données codées sortiraient en effet du champ d'application de la LRH et/ou des lois sur la protection des données personnelles. Les chercheurs pourraient alors librement en disposer puisqu'elles auraient perdu leur caractère personnel. Une telle approche donnerait par ailleurs lieu à des situations surprenantes. À défaut d'avoir obtenu le consentement nécessaire pour mener une recherche, un chercheur pourrait, par exemple, décider de coder les données dont il dispose et de les envoyer à des fins d'analyse à une équipe de recherche tierce qui ne disposerait pas de clé. En adoptant l'approche relative, les analyses menées par l'équipe tierce ne tomberaient pas sous le coup de la loi puisque les données seraient considérées comme anonymes pour cette équipe de recherche. Il suffirait alors au chercheur de se faire envoyer le résultat des analyses, éludant ainsi de manière insoutenable les règles posées par la LRH. Le fait que l'art. 26 ORH exige que les données codées soient qualifiées d'anonymes du point de vue de celui n'a pas d'accès au code n'y change rien. On doit même plutôt y voir une confirmation que la loi traite distinctement les données anonymes et les données codées.

Le même raisonnement peut s'appliquer pour la portée du consentement donné « à des fins de recherche », comme c'est le cas du consentement général proposé aujourd'hui dans tous les hôpitaux universitaires suisses. Le consentement général autorise l'utilisation les données

[49] SHK HFG-Brunner (n. 5), Vorbemerkungen art. 56–61 N 6.

[50] Ex. : art. 5, 10 et 16 v) Convention sur les Droits de l'Homme et la biomédecine, RS 0.810.2 ; art. 9 et 32 Déclaration d'Helsinki de l'Association Médicale Mondiale (version 15.2.2017).

[51] Déclaration de l'Association Médicale Mondiale sur les considérations éthiques concernant les bases de données de santé et les biobanques (version 22.3.2017).

[52] Message LRH (n. 7), 7337. Voir aussi : Lea Schläpfer, Clinical Data Sharing: Nutzen, Risiken und regulatorische Herausforderungen, recht 2016, 136 ss, 141.

codées aux seules fins de recherche. Si l'on considère que les données codées sont anonymisées dès qu'elles parviennent en main des chercheurs qui ne disposent pas du code, ceux-ci auraient toutefois champ libre pour communiquer ces données à des tiers qui pourraient ensuite les utiliser à des fins étrangères à la recherche (par exemple, le secteur assurantiel), les données anonymisées étant exclues du champ de la LRH. Une telle approche n'est pas non plus soutenable.

L'examen des conditions auxquelles des données de recherche peuvent être anonymisées débouche également sur un résultat similaire. L'art. 32 al. 3 LRH prévoit, par exemple, que les données génétiques peuvent être anonymisées à des fins de recherche à condition que la personne concernée ne s'y soit pas opposée après avoir été informée. L'idée qui sous-tend cette exigence vise à éviter que la personne concernée soit privée d'informations utiles sur sa santé qui pourraient devenir disponibles dans le futur[53]. En d'autres termes, admettre que la seule communication de données génétiques codées aurait pour effet de rendre anonymes les données aux yeux de leur destinataire éluderait la règle posée par l'art. 32 al. 3 LRH, le sujet de recherche n'étant alors pas nécessairement informé du processus d'anonymisation (*de facto*, par la simple communication) de ses données génétiques codées. Adopter l'approche relative sur la pseudonymisation reviendrait de surcroît à admettre que la réutilisation des données génétiques codées à des fins de recherche (sans communication de la clé) serait possible si le sujet de recherche ne s'y est pas opposé (*opt-out*), alors que l'art. 32 al. 2 LRH exige le consentement éclairé (*opt-in*) pour ce type de réutilisation.

La reconnaissance du droit à l'autodétermination du sujet de recherche sur ses données codées est par ailleurs confirmée par l'art. 27 ORH qui énonce les situations dans lesquelles des données peuvent être décodées. Ainsi, le décodage de données de recherche est notamment permis lorsqu'il est nécessaire pour garantir les droits de la personne concernée, notamment son droit de révocation (art. 27 let. c ORH). Dans ce dernier cas, le processus de décodage doit permettre d'effacer les données codées et s'assurer que ces données ne seront plus mises à disposition à des fins de recherche[54].

Il ressort de ce qui précède que le législateur a établi dans la LRH un système dans lequel les données codées sont soumises à un régime différent de celui applicable aux données anonymisées, sans égard au fait que le destinataire des données codées soit ou non en mesure de réidentifier les personnes concernées.

On pourrait certes rétorquer aux arguments qui précèdent que le consentement du sujet de recherche ne se trouve pas nécessairement à la base de tous les traitements de données effectués sous l'égide de la LRH et que le rôle du consentement pour la réutilisation de données auxquelles on applique un processus équivalent à l'anonymisation (du point de vue d'une personne sans clé) devrait être relativisé. L'*escape clause* de l'art. 34 LRH permet, par exemple, à la commission d'éthique compétente d'autoriser une réutilisation secondaire de données en l'absence de consentement du sujet de recherche. Cela étant, outre le caractère exceptionnel de l'art. 34 LRH, c'est à mon sens précisément dans ce type de situations que les garanties offertes par le droit de la protection des données sont le plus nécessaires, notamment parce qu'elles offrent à la personne concernée des outils d'investigation pour déterminer si des données qui le concernent sont traitées à des fins de recherche (droit d'accès) et, si tel est le cas, de s'assurer qu'elles le sont de manière licite. À ce propos, il est intéressant de noter que sous l'empire du RGPD, il existe une controverse pour déterminer sur quelle base justifier les traitements de données personnelles en vue de la recherche. Ces traitements peuvent certes être justifiés par le consentement, mais aussi éventuellement par l'intérêt public en lui-même ou l'intérêt de la recherche scientifique (art. 9 par. 2 RGPD). Ainsi, à considérer que la réutilisation des données de recherche soit possible sans le consentement du sujet de recherche, l'avantage du RGPD consiste à contrebalancer cette concession par un champ d'application large, qui couvrirait aussi les données pseudonymisées selon une approche absolue.

Enfin, il convient encore de souligner que l'évolution des moyens technologiques combinée avec les bases de données désormais à disposition plaide pour une reconnaissance particulièrement prudente du caractère anonyme d'une donnée. Si cela a été mis en lumière pour le traitement des données bancaires (cf. *supra* III.), cette mise en garde doit être prise avec d'autant plus de sérieux pour les données de santé, qui sont intimement liées aux caractéristiques physiologiques d'une personne déterminée, et par conséquent plus difficiles à masquer, en particulier à l'époque de la médecine de précision[55]. Cet argument est certainement compliqué à manier car, s'il

[53] Philippe Ducor, Protection de la personnalité des sujets de recherche, in : Margareta Baddeley et al. (éd.), Facettes du droit de la personnalité. Journée de droit civil 2013 en l'honneur de la Professeure Dominique Manaï, Genève 2014, 167 ss, 175–176.

[54] SHK HFG-Rudin (n. 5), art. 35 N 25.

[55] Voir p. ex. Ghislaine Issenhuth-Scharly, Autonomie individuelle et biobanques, Genève 2009, 212 ss.

était poussé à l'extrême, toute donnée deviendrait personnelle. Il ne saurait toutefois être sous-estimé. Des études montrent aujourd'hui que les risques de réidentification des données de santé sont bien réels, en particulier au sein des populations restreintes, auxquelles on peut certainement assimiler la Suisse, ou lorsque des données génétiques sont en jeu (via l'établissement progressif de bases de données génétiques sur des sites de génétique récréative par exemple)[56]. Les risques marqués de réidentification des données de santé ouvrent par ailleurs la question de savoir si l'approche relative aujourd'hui souvent prônée en droit helvétique de la protection des données ne devrait pas être nuancée, de manière générale, pour les données de santé. Cette problématique, qui excède le cadre de la présente contribution, mériterait une étude approfondie.

VI. Conclusion

En définitive, il ressort de cette analyse que la LRH établit un système qui consacre une approche absolue, selon laquelle les données personnelles codées qui tombent dans son champ d'application doivent être considérées comme personnelles aussi bien par celui qui détient le code que par celui qui en est privé. Cette approche est celle qui correspond à la volonté du législateur lorsqu'il a opéré un arbitrage entre la protection des sujets de recherche et la liberté de la recherche, cette dernière étant laissée libre de toute contrainte lorsqu'elle porte sur des données réellement anonymes. On peut toutefois légitimement se demander si cette approche absolue n'est pas trop radicale dans certaines situations, notamment lorsque des données codées tombent dans les mains d'un chercheur ou d'une chercheuse qui a toutes les raisons de penser que ces données ont été correctement anonymisées[57]. Dès lors, on peut également se demander si, dans le contexte de la recherche, il faudrait éventuellement envisager une approche relative « renforcée » fondée non seulement sur le calcul des efforts et sur l'intérêt à la réidentification, mais aussi sur le fait que le chercheur ou la chercheuse sait ou peut raisonnablement se douter qu'il travaille sur des données codées. En présence d'un doute raisonnable, qui serait fréquent en pratique, une telle approche imposerait au chercheur ou à la chercheuse de s'en assurer auprès de la source des données et d'entreprendre les mesures nécessaires pour respecter les exigences posées par la loi en vue de leur réutilisation.

Les incertitudes soulevées dans la présente contribution soulignent au demeurant la nécessité d'établir un cadre contractuel clair et complet pour les échanges de données dans le contexte de la recherche. Même si de tels contrats, souvent complexes, sont parfois vus comme un frein au bon déroulement de la recherche en raison de la charge administrative qu'ils impliquent, ils constituent le moyen le plus concret et efficace pour permettre au responsable du traitement de s'assurer qu'il traite ou laisse des tiers traiter les données qu'il collecte en conformité avec la loi ainsi qu'avec ses propres engagements (cf. les formulaires de consentement éclairé, qui définissent ce qui sera fait des données).

Dans un futur plus ou moins proche, les règles de réutilisation des données dans la recherche posées par la LRH devraient certainement être repensées, notamment parce que les catégories de données utilisées ne correspondent plus nécessairement aux réalités actuelles de la recherche (par exemple, qu'est-ce qu'une donnée génétique aujourd'hui ?)[58]. Lorsque cela sera entrepris, il serait bienvenu d'établir clairement le statut des données codées dans la recherche.

---

[56] À ce sujet : Junod/Elger (n. 2), N 15.3 et 15.8 et réf. citées.

[57] Junod/Elger (n. 2), N 16.6 : selon l'expérience personnelle des auteures qui siègent dans des commissions d'éthique de la recherche, l'anonymisation de données est régulièrement reconnue dans le contexte de la recherche, des données sur cette problématique étant toutefois indisponibles.

[58] À ce sujet : Junod/Elger (n. 2).