



HAL
open science

La recherche en SHS au croisement de multiples contraintes - Rapport sur le cadre juridique applicable à l'usage de données issues de réseaux socio-numériques à des fins de recherche scientifique

Julien Rossi

► To cite this version:

Julien Rossi. La recherche en SHS au croisement de multiples contraintes - Rapport sur le cadre juridique applicable à l'usage de données issues de réseaux socio-numériques à des fins de recherche scientifique. CERES - Sorbonne Université. 2023. hal-04113200

HAL Id: hal-04113200

<https://hal.science/hal-04113200v1>

Submitted on 1 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

La recherche en SHS au croisement de multiples contraintes

Rapport sur le cadre juridique applicable à l'usage de données
issues de réseaux socio-numériques à des fins de recherche
scientifique

Julien Rossi – julien.rossi04@univ-paris8.fr

Janvier 2023

Centre d'expérimentation en méthodes numériques pour les recherches en Sciences Humaines et
Sociales, Faculté des Lettres de Sorbonne Université

Avant-propos

En 2010, Sylvain Laurens et Frédéric Neyrat coordonnaient la publication d'un livre intitulé : *Enquêter, de quel droit ? Menaces sur l'enquête en sciences sociales*. Ces menaces évoquées dans le titre, qui les ont amenés à conclure à la nécessité de consacrer un « droit à l'enquête », n'ont pas disparu. De nombreuses contraintes, parfois contradictoires, s'accroissent, avec un formalisme croissant qui crée autant de nouveaux lieux de pouvoir. Entre le droit de la propriété intellectuelle, l'existence de différents secrets protégés par la Loi, dont celui de la vie privée, et la multiplication d'obligations relatives à la gestion et à la sécurité des données, l'espace dans lequel une recherche en sciences humaines et sociales (SHS) peut être légalement effectuée semble parfois de plus en plus contraint. Des données à caractère personnel collectées pour étudier la circulation de messages au sein de certaines communautés sur les réseaux sociaux aux œuvres protégées par le droit d'auteur mobilisées pour l'analyse de la musique contemporaine, en passant par les verbatims d'entretiens au sujet du rapport d'un groupe d'enquêté·e·s à des problèmes santé, toutes ces enquêtes font l'objet d'un ensemble de règles qui peuvent varier d'un domaine à l'autre, d'un cas à l'autre.

Ces contraintes, dont beaucoup trouvent un fort écho dans les préoccupations d'ordre éthique qui donnent lieu depuis longtemps à des pratiques telles que le recueil du consentement, doivent être interprétées en tenant compte d'une part du droit fondamental à la liberté d'expression, et d'autre part du caractère d'intérêt général que revêt, en de nombreuses circonstances, la recherche scientifique. Dans la mesure où les diverses obligations qui pèsent sur l'usage des données de la recherche en SHS découlent du droit de l'Union européenne, il n'est pas inutile de rappeler que la liberté académique est spécifiquement protégée par l'article 13 de la Charte des droits fondamentaux de l'Union européenne. Cet article, lu conjointement avec l'article 11 sur la liberté d'expression de cette même charte, reconnaît un droit à l'enquête scientifique. Tout l'enjeu consiste à garantir l'effectivité de ce dernier sans pour autant ignorer les droits des enquêté·e·s et des tiers.

En partant d'une analyse juridique incluant les textes de droit positif en vigueur, la jurisprudence et la doctrine, des questions et des discussions prises en note lors d'un atelier sur le sujet organisé le 2 décembre 2021 à la Maison de la recherche de Sorbonne Université, et en intégrant des éléments issus de la lecture de plans de gestion des données publiquement disponibles, le présent rapport propose des pistes à la fois théoriques et pratiques pour surmonter l'opposition – souvent crainte à tort, mais parfois confirmée – entre pratique concrète de la recherche en SHS, obligations juridiques et normes éthiques.

Il se termine sur la conclusion de l'existence d'un certain nombre de contradictions qui créent des zones d'insécurité juridique pour la pratique de la recherche en SHS, et par l'esquisse de pistes de recherche juridique de terrain qui pourront utilement compléter un état de l'art jusqu'ici principalement focalisé sur le commentaire – utile, mais insuffisant – du droit positif complété du peu de jurisprudence à notre disposition.

Remerciements

Je remercie le Centre d'expérimentation en méthodes numériques pour les recherches en SHS (CERES) de Sorbonne Université, et en particulier Virginie Julliard et Emmanuelle Papinot, pour le financement de cette étude et pour l'organisation des deux ateliers qui l'ont nourrie.

Merci à Océane Valencia et à Cécile Arènes pour leurs multiples contributions et pour leur participation à ces ateliers. Merci également à Sophie Schaeffert, déléguée à la protection des données de Sorbonne Université, pour avoir répondu à nos questions lors d'un entretien par téléphone.

Merci à toutes les personnes qui ont participé aux ateliers et qui ont posé des questions sur le formulaire en ligne. Merci également aux personnes qui ont paramétré le site web sur lequel est hébergé ce formulaire pour avoir inclus un en-tête P3P dans la requête HTTP de réponse. C'est un protocole de communication dédié à la protection de la vie privée qu'il est rare de rencontrer !

Merci enfin aux personnes sur Twitter qui m'ont aidé à trouver certaines ressources à jour, et notamment à Robin Plique (@juristnumerique), Jonathan Keller (@djokeller), Lionel Maurel (@calimaq), et Dani (@sonderweg45).

Sommaire

Avant-propos.....	2
Remerciements.....	3
Sommaire.....	4
Introduction et méthode.....	5
Données de recherche, éthique et droit.....	8
Panorama du cadre juridique.....	13
Liberté d’expression et libertés académiques.....	13
Les plans de gestion des données, les principes FAIR et l’archivage.....	19
Le droit de la propriété intellectuelle, dont le droit d’auteur.....	22
Droit à la vie privée, à la protection des données à caractère personnel, et recherche scientifique....	28
Le droit d’accès à des données mobilisé dans la recherche scientifique.....	40
Des évolutions à surveiller.....	42
Éléments d’analyse.....	44
Quelques enseignements tirés de la lecture de quinze plans de gestion des données.....	44
Quelle responsabilité des chercheurs·ses dans la détermination de l’équilibre entre <i>open science</i> et protection de la vie privée ?.....	48
Conclusion.....	51
Une sécurité juridique encore insuffisante pour assurer un plein exercice des libertés académiques que le droit cherche pourtant bel et bien à préserver.....	51
Pistes de recherche.....	53
Bibliographie.....	55
Références académiques.....	55
Textes normatifs cités.....	56
Jurisprudence.....	58
Délibérations d’autorités administratives indépendantes.....	59
Articles de presse et de blogs.....	59
Rapports et avis.....	60
Annexe 1 : Table des abréviations.....	61
Annexe 2 : Corpus de plans de gestion des données.....	62
Annexe 3 : Les questions posées en atelier.....	65

Introduction et méthode

Il existe de nombreuses contraintes qui pèsent sur la pratique de la recherche en sciences humaines et sociales (SHS), notamment en matière de traitement et de publication de données de la recherche.

Celles-ci, qui peuvent être définies comme l'ensemble du matériau qui concourt à la production de résultats de recherche, quel que soit sa nature (entretiens, carnets de terrains, données issues de réseaux sociaux, photographies, corpus de presse...), peuvent en effet entrer dans le champ d'application d'un grand nombre de textes qui en encadrent l'usage : droit de la propriété intellectuelle, droit des données à caractère personnel, obligations dans le cadre de dépôts de projets de réaliser un plan de gestion des données, droit d'accès aux documents administratifs et aux données d'intérêt public, droit des archives publiques... Dans le cadre d'un contexte de sous-financement chronique de la recherche¹, la multiplication des contraintes aboutit à la crainte légitime d'une impossibilité matérielle de s'y conformer pleinement voire à une interdiction pure et simple de toute utilisation de donnée en SHS, ce qui reviendrait à signer l'arrêt de mort de ce champ scientifique.

Le Centre d'expérimentation en méthodes numériques pour les recherches en SHS (CERES) de Sorbonne Université a organisé deux ateliers, le premier le 2 décembre 2021 et le second le 17 février 2022, en lien avec les services de la Bibliothèque Universitaire (BU) et des archives, sur les plans de gestion des données, le Règlement général de protection des données (RGPD) et les diverses obligations relatives à l'utilisation des données de recherche. Le premier atelier, organisé par Emmanuelle Papinot et Virginie Julliard, et auquel sont intervenu·es Cécile Arènes, Océane Valencia et l'auteur de ce rapport, a permis à un public d'enseignant·es chercheurs·ses de Sorbonne Université et d'autres universités de discuter de la thématique et de poser des questions. Ces éléments ont été pris en note pour alimenter la réflexion et l'analyse des règles juridiques, décisions de justice, avis d'autorités de contrôle et autres textes de doctrine, présentés dans ce rapport. Des premiers éléments de réponse ont été présentés et discutés lors de l'atelier du 17 février 2022.

Cette analyse s'est voulue, dans une mesure qui n'a que partiellement été atteinte, une analyse juridique ancrée dans le terrain. Comme l'affirmait Jean Carbonnier, « le droit est trop humain pour prétendre à l'absolu de la ligne droite » (Carbonnier, 2001, p. 8). Une règle de droit peut être plus ou moins effective, et lorsqu'elle l'est, avoir une incidence différente de l'intention du législateur, si tant est qu'il soit possible de saisir une intention unique dans des textes qui sont souvent le fruit d'un compromis aboutissant délibérément à une formulation floue, soumise à multiples interprétations concurrentes, pour satisfaire les intentions contradictoires de ses auteur·es (Carbonnier, 2001, p. 150). Ces textes et leurs interprétations circulent dans l'espace public sur des supports médiatiques. Les ressortissant·es, c'est-à-dire le public, d'une règle écrite de droit, l'interprètent chacun·e selon ses propres

1 Malgré une croissance régulière, la part du PIB européen dépensé en recherche et développement n'atteint toujours pas la barre des 3 % que l'UE s'est fixée comme objectif. En 2020, il était de 2,27 % en Europe, soit un niveau sensiblement similaire à celui de la Chine, et environ 1 % de PIB moins élevé qu'aux États-Unis et au Japon. Ce chiffre a diminué en 2021. La France se situe dans la moyenne européenne. Voir le site d'Eurostat : https://ec.europa.eu/eurostat/statistics-explained/index.php?title=R_%26_D_expenditure&oldid=548647 (page consultée le 9 janvier 2023).

représentations, valeurs et référentiels idéologiques (Revillard, 2018). De ce constat dérivent à la fois la nécessité d'étudier de façon sociologique l'application et la compréhension du droit sur un terrain donné, et la liberté d'interpréter un texte dont l'interprétation *vraie*, selon l'adage *res judicata pro veritate habetur*², n'a pas encore été « révélée » par un tribunal de dernière instance. Si, institutionnellement et collectivement, les acteurs·trices de la recherche en SHS n'interprètent pas le droit applicable à leur activité, iels courent le risque de se voir imposer dans les faits une interprétation exogène qui ne tient pas compte des contraintes pratiques du métier, sans pour autant répondre à un besoin éthique ou d'intérêt général particulier.

L'objectif des travaux de recherche présentés dans ce rapport a été de partir des questions et des discussions avec des enseignant·es-chercheurs·ses en SHS pour étudier les règles de droit qui s'appliquent à l'usage des données de la recherche en général et des données à caractère personnel en particulier. L'étude, qui se fonde sur des textes de droit positif, sur la jurisprudence et sur la doctrine, dont les avis d'autorités de contrôle, a également porté sur des plans de gestion de données publiquement disponibles et propose une interprétation des règles tenant compte de la hiérarchie des normes, et cherchant à résoudre, le cas échéant, les conflits de normes. Lorsque ces conflits portent sur des droits fondamentaux concurrents, ce qui est, nous le verrons, assez fréquent, la méthode de résolution s'appuie sur le raisonnement de la Cour européenne des droits de l'Homme (CEDH) et de la Cour de justice de l'Union européenne (CJUE) pour déterminer ce qui est « nécessaire dans une société démocratique » (Sauvé, 2018).

Pour la CEDH, le raisonnement est le suivant : pour qu'une ingérence dans un droit fondamental consacré par la Convention de sauvegarde des droits de l'homme et des libertés fondamentales (Convention EDH) soit autorisée, celle-ci doit être (1) prévue par la loi, (2) poursuivre un objectif légitime (qui peut être la protection d'un droit fondamental concurrent) et être (3) « nécessaire dans une société démocratique », c'est-à-dire à la fois utile à la poursuite de l'objectif légitime identifié en (2) et proportionné à celui-ci. Pour la CJUE, le raisonnement varie dans la forme mais est identique sur le fond. Toute limitation d'un droit fondamental doit selon le juge de l'Union (1) être prévue par la loi, (2) ne pas empêcher le respect du contenu essentiel du droit fondamental visé, (3) poursuivre un objectif d'intérêt général (qui peut être le respect d'un droit fondamental concurrent), et doit être (4) proportionnel, c'est-à-dire (4a) être de nature à atteindre l'objectif visé et (4b) ne pas dépasser ce qui est strictement nécessaire pour atteindre cet objectif.

Lorsque les analyses présentées dans le présent rapport contiennent des propositions d'interprétations visant à concilier la possibilité de mener des recherches en SHS avec des intérêts concurrents, ces interprétations prennent systématiquement appui sur les textes et ne dévient pas de ce qui a déjà été interprété avec force de chose jugée en dernière instance, c'est-à-dire, pour toutes les règles qui ont pour origine le droit de l'Union, soit la CJUE soit, dans la mesure où tous les États membres de l'UE sont aussi parties à la Convention EDH, la CEDH. La jurisprudence française est intégrée à la réflexion tout en tenant compte de ce qu'elle-même doit être conforme au droit européen.

L'étude présentée ici accorde une place centrale à la question de l'usage en recherche de données de réseaux socio-numériques, en la replaçant dans le contexte plus large du droit de la recherche en SHS. Ont toutefois été exclus pour le moment de son champ, par faute de temps et parce qu'il n'y a pas eu de

2 « La chose jugée est tenue pour vérité. »

questions à ce sujet lors du premier atelier organisé dans le cadre du projet, les différents secrets (d'État, industriels, des affaires...) qui peuvent entraver l'accès aux données de la recherche.

Le rapport se conclue par quelques conclusions sur l'état du droit et des suggestions de pistes de recherche.

Données de recherche, éthique et droit

Définir ce qu'est une donnée n'est pas une mince affaire. Le règlement sur la gouvernance européenne des données³ (Data Governance Act, DGA) la définit comme « toute représentation **numérique**⁴ d'actes, de faits ou d'informations et toute compilation de ces actes, faits ou informations, notamment sous la forme d'enregistrements sonores, visuels ou audiovisuels » (art. 2 (1) du DGA). Mais toute donnée est-elle nécessairement numérique ? Si nous suivons la définition de ce qu'est une « donnée à caractère personnel » selon le Règlement général de protection des données (RGPD), il s'agit en effet de « toute information se rapportant à une personne physique identifiée ou identifiable [...] » (art. 4 (1) RGPD), peu importe qu'elle soit sur support numérique ou non, dès lors qu'elle est contenue ou appelée à être contenue dans un fichier (art. 2 (1) RGPD).

Le Plan national pour la science ouverte, en France, définit en reprenant une définition de l'OCDE⁵ les « données de la recherche » comme les « enregistrements factuels (chiffres, textes, images, son, vidéo...) utilisés comme sources primaires pour la recherche et qui sont habituellement acceptés par la communauté scientifique comme étant nécessaires pour valider les résultats de la recherche. » Un enregistrement de données électroniques numériques en provenance des réseaux sociaux, quelle que soit la nature sémiotique de la représentation de cette donnée, entre dans cette définition. Mais ce n'est pas aussi évident pour tous les matériaux produits dans le cadre de recherches en SHS. Ainsi, dans le cadre d'une enquête de terrain ethnographique, par exemple, il n'est pas toujours simple de distinguer le « factuel » de la subjectivité située du chercheur ou de la chercheuse. La question peut se poser de savoir si les carnets de terrain, ou des productions poétiques dans le cadre de méthodes comme le *i-poem* (Koelsch, 2016) pouvant être rédigés par le chercheur lui-même pendant son terrain, sont alors des « données » de la recherche selon cette définition. L'Association des archivistes français (AAF) propose une définition plus large, selon laquelle sont des « données de la recherche [...] l'ensemble des informations et matériaux produits et reçus par des équipes de recherche et des chercheurs. Elles sont collectées et documentées à des fins de recherche scientifique. À ce titre, elles constituent une partie des archives de la recherche » (Pomart, 2014). Cette définition fait équivaloir tout matériau collecté ou produit dans le cadre d'une recherche à une donnée de la recherche, qu'il soit numérique ou non, électronique ou non. Elle inclut ainsi les corpus, dans lesquels peuvent figurer des œuvres protégées au titre du droit d'auteur, des réponses à des questionnaires ou à des entretiens, qui sont souvent des données à caractère personnel, des archives dont certaines peuvent être couvertes par un secret protégé par la loi...

3 Règlement 2022/868/UE du Parlement européen et du Conseil du 30 mai 2022 portant sur la gouvernance européenne des données et modifiant le règlement (UE) 2018/1724 (règlement sur la gouvernance des données) (aussi connu sous nom acronyme anglais de Data Governance Act – DGA).

4 Notre mise en forme.

Nous supposons qu'il s'agit de données « numériques » au sens d'« électroniques » et encodées dans un format binaire, et pas uniquement de nombres, lesquels peuvent être électroniques ou non.

5 Voir les Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics de 2007, page 18.

La directive *Open Data* de 2019⁶, elle, réduit, comme le DGA, le champ aux seules données qui sont sur support numérique, en définissant ces données comme « des documents se présentant sous forme numérique, autres que des publications scientifiques, qui sont recueillis ou produits au cours d'activités de recherche scientifique et utilisés comme éléments probants dans le processus de recherche, ou dont la communauté scientifique admet communément qu'ils sont nécessaires pour valider des conclusions et résultats de la recherche » (art. 2 (9) de la directive *Open Data*).

De nombreuses initiatives de politiques publiques visent à ouvrir le plus largement possible les données de la recherche. Ainsi, le Plan national pour la science ouverte présente l'ouverture des données comme utile à la poursuite d'un intérêt général d'innovation et de croissance économique, notamment dans le domaine du développement de l'intelligence artificielle (IA) (MESR, 2018, p. 6). Cela sert également un intérêt scientifique, puisque l'accès aux données doit contribuer à garantir la réfutabilité des résultats d'une recherche, qui est un pilier de l'épistémologie positiviste des sciences, telle que définie par Karl Popper. Elle doit aussi permettre de garantir un certain niveau de qualité, et de confiance du public, dans la science. La presse relate en effet souvent des scandales liés à des données délibérément truquées (Barthélémy, 2012, Malbœuf, 2017, Bonneau, 2019). Selon un article Dalmeet Singh Chawla publié dans la revue *Nature*, 8 % des chercheurs·ses aux Pays-Bas auraient falsifié ou fabriqué des données (Chawla, 2021).

Les intérêts qui sous-tendent la publication des données de la recherche peuvent néanmoins entrer en conflit avec d'autres intérêts, particuliers ou collectifs. Dans le domaine de la recherche médicale, la publication de données de santé d'êtres humains est une atteinte au droit à la vie privée. La publication de données utilisées dans l'évaluation d'un nouveau prototype d'avion de chasse peut porter atteinte à des intérêts de sécurité nationale et de défense, et celle de données qui portent sur des avions civils peut porter atteinte à des secrets industriels. Les SHS sont loin d'échapper à ces problèmes. En effet, par nature, il s'agit de disciplines scientifiques dont les recherches portent sur des humains, et qui ne peuvent donc se passer – sauf à faire de la théorie pure, déconnectée de tout terrain⁷ – de données à caractère personnel. Lorsque ces recherches se fondent sur des corpus, non seulement ceux-ci peuvent contenir des informations relatives à des personnes physiques identifiées ou identifiables directement ou indirectement, mais elles peuvent aussi porter sur un matériau protégé par le droit de la propriété intellectuelle, limitant ainsi possibilité de le partager.

Ces intérêts entrent en contradiction avec les politiques de science ouverte qui visent notamment à favoriser le plus largement possible l'ouverture des données de la recherche. Ainsi, les responsables de projets de recherche peuvent se retrouver dans l'obligation de publier leurs données, soit dans l'interdiction de le faire, et la marge de manœuvre qui concerne les cas où ce partage est à leur libre appréciation se réduit. Or, la détermination des cas dans lesquels il *faut* soit publier, soit ne pas publier, n'est pas toujours évidente, et soulève de nombreuses questions d'ordre éthique.

La question de l'équilibre entre ouverture et fermeture n'est pas la seule contrainte juridique de nature à générer des difficultés pratiques dans la conduite de projets de recherche en SHS. Les règles relatives à

6 Directive 2019/1024/UE du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public.

7 Et encore. Une bibliographie est, après tout, un fichier de données à caractère personnel, sauf si tous les auteurs cités sont décédés.

la protection de la propriété intellectuelle et le droit à la protection des données à caractère personnel conduisent à de nombreuses interrogations.

Faut-il arrêter de produire des recherches sur les discriminations subies par les minorités sexuelles, au motif qu'elles ne peuvent se faire sans utiliser des données sensibles ? Faut-il interdire toute analyse littéraire sur la saga *Harry Potter* ou sur les romans des dernier·es lauréat·es du prix Goncourt ou motif que leurs œuvres sont protégées par le droit d'auteur ? Est-il interdit d'étudier la circulation d'images et de discours médiatiques sur un réseau social comme Twitter, au motif, qu'associé à un compte, il s'agit de données à caractère personnel ? Que montrer des résultats de ces recherches pour concilier ces différents intérêts contradictoires ?

La Loi ménage fort heureusement une marge de manœuvre assez large qui permet, en principe, à la plupart de ces projets de se réaliser. Nous présenterons ce cadre juridique, les aménagements prévus pour la recherche, ses limites et les contreparties exigées, par exemple, en échange de la possibilité de se passer du consentement pour certaines collectes de données à des fins scientifiques.

Un certain nombre de pratiques pensées sous l'angle de l'éthique, et déjà largement répandues dans la recherche en SHS, contribuent à fournir des réponses aux interrogations ci-dessus. Il en est ainsi de l'anonymisation, qui est un objectif sur lequel insiste le Règlement général de protection des données à son article 89 consacré à la recherche scientifique. Le recueil du consentement des enquêté·es, possible dans un grand nombre d'enquêtes, est lui aussi une pratique développée sur le terrain en même temps qu'il répond à une règle de droit.

Ces pratiques routinières nées de préoccupations éthiques et le droit se complètent. Parfois, il peut exister des contradictions. Si, selon Jean Cassien Billier, il existe une « stricte équivalence conceptuelle » (Billier, 2014, p. 3) entre éthique et morale, Paul Ricœur (1989 ; 1990) propose une distinction entre l'éthique, qui peut être propre à l'individu, et une morale qui renvoie aux mœurs et donc à des faits sociaux. La morale est donc culturelle et historiquement située, et l'éthique peut se concevoir à l'échelle de la démarche individuelle. Elles n'aboutissent pas nécessairement à des conclusions ou à des préceptes rationnels, objectifs et universels. Si ces derniers caractérisent la pensée d'Emmanuel Kant, il est difficile d'en dire de même pour celle de Friedrich Nietzsche, d'un libéralisme utilitariste de Jeremy Bentham ou John Stuart Mill, sans même parler des conséquentialistes égoïstes tels que Bernard Mandeville ou Ayn Rand ou d'une éthique de la vertu privilégiant une démarche éthique subjective. Il n'y a donc pas nécessairement d'équivalence entre Loi, morale et éthique.

Ce rapport a cherché à éclairer des éléments de droit, c'est-à-dire des règles de conduite dont la violation peut entraîner une sanction décidée par l'autorité judiciaire. Ces règles ont, du moins en théorie, et souvent en pratique, une dimension coercitive qui s'exerce nonobstant les mœurs d'un groupe ou les convictions éthiques d'un individu. Quelles que soient les différentes interprétations possibles d'un texte, celles retenues par le ou la juge qui déterminera si une règle a été enfreinte ou non sera celle qui entraînera, ou non, une sanction. Il existe de nombreuses définitions souvent contradictions du Bien et du Mal. Loin d'épuiser le débat, l'affirmation de l'existence d'une règle de droit n'équivaut donc pas à la juger juste, bénéfique ou désirable.

Ainsi, bien que le RGPD oblige à informer toutes les personnes concernées par une recherche des finalités de cette recherche, il peut s'avérer délicat d'informer un·e enquêté·e·s des finalités réelles, soit

parce que cela biaiserait les résultats, soit parce que cela mettrait en danger leurs chercheurs·ses (dans le cas par exemple d'une étude sur des organisations criminelles) soit parce que cela pourrait affecter le bien-être de ces personnes. Selon un point de vue conséquentialiste ou de *care*, il peut être souhaitable de cacher ou d'atténuer certaines informations aux enquêté·e·s, par exemple lorsqu'une étude porte sur la façon dont un trauma est vécu en fonction du niveau d'études et de la classe sociales (cf. Rossi et Bigot, 2018). Une interprétation faisant fi de cette situation pourrait être critiquable selon ces conceptions éthiques⁸.

Le présent rapport de recherche n'est pas une exploration systématique de ce que différentes théories éthiques peuvent proposer comme solution à un problème donné, tel que ceux soulevés par le *web scrapping* ou les dérogations au consentement des enquêté·es. Il est écrit par un chercheur en SHS en partant du principe que la recherche, y compris celle qui porte sur la société et sur l'humain, concourt à la réalisation d'intérêts généraux dans le domaine scientifique et dans le débat public dans une société démocratique. Il cherche à fournir un guide pour aider les collègues chercheurs et chercheuses à mener leurs recherches tout en respectant la Loi, dont l'un des objectifs est la protection de droits humains, fondamentaux, dont la liberté d'expression⁹, la liberté académique¹⁰, le droit à la vie privée¹¹, le droit à la protection des données à caractère personnel¹², le droit à la propriété privée, y compris le droit d'auteur et les autres droits de la propriété intellectuelle¹³. Ces droits, leur contenu, leur désirabilité sur le plan éthique ou moral, font l'objet de nombreux débats. Par exemple, pour ce qui concerne le droit à la vie privée, Amitai Etzioni (1999) et Christian Fuchs (2011) critiquent fortement, pour des raisons différentes, la prépondérance qu'il a dans sa conception libérale vis-à-vis d'intérêts collectifs. Catharine MacKinnon (1989) s'est également prononcée contre le droit à la vie privée. Le droit à la propriété privée fait de longue date l'objet d'une critique dans la tradition philosophique marxiste, et le droit à la propriété intellectuelle fait régulièrement controverse. Quoique ces débats soient d'une grande importance pour déterminer le rôle de la recherche et la place relative de l'intérêt collectif des « progrès » de la science (si l'on se place dans une optique positiviste), ce n'est pas l'objet de ce rapport qui admet aussi, de façon axiomatique, la validité et la désirabilité des droits fondamentaux tels

8 Fort heureusement (selon nous), le considérant 62 du RGPD semble permettre une interprétation conforme, dans cette situation, à ce qu'une éthique soucieuse du vécu des enquêté·e·s – que certain·e·s jugeront peut-être paternaliste – à la volonté d'omettre certains détails pour ne pas exacerber de sentiment de détresse chez les personnes qui se prêtent à l'enquête, ou lorsque la détention d'une telle information les mettrait en danger. Une telle approche serait à rapprocher du droit des patient·e·s à ne pas être informé·e·s de leur diagnostic, à leur demande. Cette demande doit être respectée y compris dans les protocoles de recherche médicale (art. L1111-2 I., quatrième alinéa du Code de la santé publique).

9 Art. 10 de la Convention EDH, art. 11 de la Charte des droits fondamentaux de l'UE, art. 11 de la Déclaration des droits de l'Homme et du citoyen (DDHC).

10 Art. 13 de la Charte des droits fondamentaux de l'UE, C. Const. décision 83-165 DC du 20 janvier 1984 (reconnaissance de l'indépendance des enseignant·e·s chercheurs·ses comme principe fondamental reconnu par les lois de la République). La CEDH rattache la liberté académique à la liberté d'expression protégée à l'article 10 de la Convention EDH (cf. CEDH 23 juin 2009 *Sorguç c. Turquie*, req. n° 17089/03).

11 Art. 8 Convention EDH, art. 7 de la Charte des droits fondamentaux de l'UE, et rattachement aux « droits imprescriptibles de l'Homme » et à la « liberté » de l'art. 2 DDHC (C. Const., décision 99-416 DC du 23 juillet 1999, cons. 45.)

12 Art. 8 de la Charte des droits fondamentaux de l'UE. La CEDH rattache ce droit à l'art. 8 de la Convention EDH (CEDH 4 décembre 2008 « *S. et Marper contre Royaume-Uni* » Req. 30562/04 et 30566/04, pt. 103). Le C. Const. rattache également ce droit au droit à la vie privée, et ainsi à l'art. 2 DDHC (C. Const., décision 2012-652 DC du 22 mars 2012, cons. 8).

13 Art. 1 du Protocole additionnel à la Convention EDH (cf. CEDH 11 janvier 2007, *Anheuser-Busch Inc. c. Portugal*, req. n° 73049/01, pt. 72). Art. 17 de la Charte des droits fondamentaux de l'UE, en particulier son paragraphe 2, art. 2 et 17 de la DDHC (cf. C. Const. 8 janvier 1991, décision 90-283 DC, C. Const., 27 juill. 2006, décision 2006-540 DC)

qu'ils sont énoncés dans la Convention EDH, dans la Charte des droits fondamentaux de l'UE, et dans le bloc de constitutionnalité français, puis interprétés par la CEDH, la CJUE et le Conseil constitutionnel.

Lorsqu'il y en a, les recommandations *de lex ferenda* sont présentées dans le but de consolider la sécurité juridique des projets de recherche en SHS en tenant compte des droits des tiers, en particulier ceux des enquêtés.

Panorama du cadre juridique

Liberté d'expression et libertés académiques

L'article 10 de la Déclaration des droits de l'Homme et du citoyen (DDHC) proclame que « nul ne doit être inquiété pour ses opinions, même religieuses, pourvu que leur manifestation ne trouble pas l'ordre public établi par la Loi. » L'article suivant, que « la libre communication des pensées et des opinions est un des droits les plus précieux de l'Homme : tout Citoyen peut donc parler, écrire, imprimer librement, sauf à répondre de l'abus de cette liberté dans les cas déterminés par la Loi. » Ces libertés de conscience et d'expression sont énoncées aux articles 9 et 10 de la Convention EDH. Dans la Charte des droits fondamentaux de l'Union européenne, il s'agit des articles 10 et 11.

En Europe, le droit à la liberté d'expression n'est pas un droit absolu. Il existe une différence fondamentale entre la conception absolue du droit à la liberté d'expression en vigueur aux États-Unis d'Amérique, et celle qui a cours sur le Vieux continent (Weiss, 2014). Cette différence saute aux yeux à la lecture comparée du Premier amendement à la constitution américaine et de l'article 10 de la Convention EDH. Là où le premier proclame une interdiction absolue de limiter la libre-expression de la parole, l'article de la Convention EDH consacre un paragraphe entier aux possibilités de restrictions :

Premier amendement à la Constitution des États-Unis d'Amérique	Article 10 de la Convention EDH
« Le Congrès ne fera aucune loi qui touche l'établissement ou interdise le libre exercice d'une religion, ni qui restreigne la liberté de la parole ou de la presse, ou le droit qu'a le peuple de s'assembler paisiblement et d'adresser des pétitions au gouvernement pour la réparation des torts dont il a à se plaindre ¹⁴ . »	« 1. Toute personne a droit à la liberté d'expression. Ce droit comprend la liberté d'opinion et la liberté de recevoir ou de communiquer des informations ou des idées sans qu'il puisse y avoir ingérence d'autorités publiques et sans considération de frontière. Le présent article n'empêche pas les États de soumettre les entreprises de radiodiffusion, de cinéma ou de télévision à un régime d'autorisations. 2. L'exercice de ces libertés comportant des devoirs et des responsabilités peut être soumis à certaines formalités, conditions, restrictions ou sanctions prévues par la loi, qui constituent des mesures nécessaires, dans une société

14 Traduit de l'anglais par la Digithèque de matériaux juridiques et politiques de l'Université de Perpignan, à l'URL : <https://mjp.univ-perp.fr/constit/us1787a.htm> (page consultée le 10 décembre 2021). Texte original : « Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances. »

	démocratique, à la sécurité nationale, à l'intégrité territoriale ou à la sûreté publique, à la défense de l'ordre et à la prévention du crime, à la protection de la santé ou de la morale, à la protection de la réputation ou des droits d'autrui, pour empêcher la divulgation d'informations confidentielles ou pour garantir l'autorité et l'impartialité du pouvoir judiciaire. »
--	--

La CEDH et la CJUE ont donc été amenées à trancher des litiges dans lesquels s'opposaient droit à la vie privée et droit à la liberté d'expression, ces deux droits pouvant se limiter mutuellement.

Dans l'affaire qui a amené à la reconnaissance d'un droit au déréférencement, la CJUE devait trouver un équilibre entre le droit à la vie privée du requérant qui voulait effacer du Web la présence d'informations sur une vente aux enchères immobilière liée à une saisie pratiquée en recouvrement de dettes de sécurité sociale le concernant. Cette information remontait à plus de dix ans avant la plainte déposée suite au refus de la société Google de retirer les liens vers l'information incriminée, et celui de *La Vanguardia* de retirer ladite information. Le compromis trouvé par la Cour, en l'espèce, a été d'obliger le moteur de recherche à retirer toute référence pointant du nom du plaignant vers l'article relatant la vente aux enchères, tout en refusant de faire droit à la demande de retrait du contenu sur le site du journal. Sur le principe, l'arrêt du 13 mai 2014 précise toutefois que « s'il apparaissait, pour des raisons particulières, telles que le rôle joué par [une] personne dans la vie publique, que l'ingérence dans ses droits fondamentaux [du fait de l'inclusion de son nom dans des résultats de moteur de recherche] est justifiée par l'intérêt prépondérant dudit public à avoir, du fait de cette inclusion, accès à l'information en question¹⁵ », alors un moteur de recherche recevant une demande de déréférencement pourrait s'y opposer.

La CEDH s'est aussi exprimée plusieurs fois sur l'équilibre entre vie privée et liberté d'expression. Dans l'affaire *Fressoz et Roire contre France*, qui impliquait la publication dans le *Canard Enchaîné* d'informations sur la rémunération du président directeur-général de l'entreprise Peugeot dans un contexte de conflit social portant sur des revendications salariales, la Cour a estimé que la condamnation des journalistes auteurs de l'article incriminé au motif de l'atteinte à la vie privée constituait une atteinte disproportionnée à la liberté d'expression. En effet, les juges de Strasbourg ont estimé que bien que la mesure soit prévue par la loi et poursuivait le motif légitime de protéger la vie privée elle-même protégée à l'article 8 de la Convention EDH, cette condamnation n'était pas « nécessaire dans une société démocratique » car l'« article [dans le *Canard Enchaîné*] s'inscrivait dans le cadre d'un débat public d'intérêt général¹⁶ », portait sur des informations relatives à la gestion d'une entreprise, et ne visait pas une mise en cause de « la réputation ou [des] droits¹⁷ » de son dirigeant.

Une autre affaire, *Ungváry et Irodalom Kft*, est d'autant plus intéressante pour le présent rapport qu'elle porte sur la publication de travaux de recherche en sciences sociales. Krisztián Ungváry, un historien

15 CJUE 13 mai 2014, Google Espagne contre AEPD, aff. C-131/12.

16 CEDH 21 janvier 1999, *Fressoz et Roire contre France*, req. n° 29183/85, pt. 46.

17 *Idem*

hongrois, avait publié dans *Élet és Irodalom*, une revue hebdomadaire traitant de littérature et de politique, un article dans lequel il relatait la répression par le régime communiste d'un mouvement étudiant à Pécs dans les années 1980. L'article mentionnait le rôle joué par László Kiss, alors juge à la Cour constitutionnelle hongroise. Celui-ci a poursuivi Krisztián Ungváry, qui a été condamné pour diffamation. La CEDH a jugé que cette condamnation constituait une atteinte disproportionnée à la liberté d'expression, dans la mesure où l'article incriminé se contentait d'une description appuyée par une argumentation, sur un ton factuel, de faits historiques, que l'article contribuait à un « débat d'intérêt général¹⁸ », et qu'il fallait tenir compte du rôle public joué par le plaignant, qui devait alors accepter une attention du public dans le cadre d'une société démocratique¹⁹.

Les critères précis dégagés par la CEDH pour établir si la publication d'information au sujet d'une personne est une infraction proportionnée et acceptable au droit à la vie privée sont : (a) la contribution à un débat d'intérêt général, (b) le degré de notoriété de la personne concernée, en lien avec le sujet de la publication, (c) la conduite passée de la personne concernée, (d) la méthode d'obtention de l'information et sa véracité et (e) le contenu, la forme et les conséquences de la publication²⁰. La Cour distingue la divulgation de faits qui, même controversés, contribuent à un débat d'intérêt général, des allégations calomnieuses en rapport avec la vie privée d'une personne²¹.

L'article 85 du RGPD vise lui aussi, précisément, à assurer cet équilibre entre liberté d'expression, droit à la vie privée et droit à la protection des données à caractère personnel :

« 1. Les États membres concilient, par la loi, le droit à la protection des données à caractère personnel au titre du présent règlement et le droit à la liberté d'expression et d'information, y compris le traitement à des fins journalistiques et à des fins d'expression universitaire, artistique ou littéraire.

2. Dans le cadre du traitement réalisé à des fins journalistiques ou à des fins d'expression universitaire, artistique ou littéraire, les États membres prévoient des exemptions ou des dérogations au chapitre II (principes), au chapitre III (droits de la personne concernée), au chapitre IV (responsable du traitement et sous-traitant), au chapitre V (transfert de données à caractère personnel vers des pays tiers ou à des organisations internationales), au chapitre VI (autorités de contrôle indépendantes), au chapitre VII (coopération et cohérence) et au chapitre IX (situations particulières de traitement) si celles-ci sont nécessaires pour concilier le droit à la protection des données à caractère personnel et la liberté d'expression et d'information.

3. Chaque État membre notifie à la Commission les dispositions légales qu'il a adoptées en vertu du paragraphe 2 et, sans tarder, toute disposition légale modificative ultérieure ou toute modification ultérieure les concernant. »

18 CEDH 3 décembre 2013, Ungváry et Irodalom Kft contre Hongrie, req n° 64520/10, pts. 62 et 63.

19 CEDH 3 décembre 2013, Ungváry et Irodalom Kft contre Hongrie, req n° 64520/10, pt. 64.

20 CEDH 3 décembre 2013, Ungváry et Irodalom Kft contre Hongrie, req n° 64520/10, pt. 45.

21 CEDH 3 décembre 2013, Ungváry et Irodalom Kft contre Hongrie, req n° 64520/10, pt. 47.

Le dernier paragraphe de l'article ci-dessus charge chaque État membre de l'Union de préciser les mesures visant à assurer cet équilibre, dans le respect de la Charte des droits fondamentaux. La France a notifié à la Commission l'article 5 de la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles, et l'article 67 de la loi informatique et libertés²². Cette dernière a été modifiée depuis une l'ordonnance du 12 décembre 2018. Ainsi, dans sa rédaction actuelle, c'est son article 80 qui prévoit que « lorsqu'une telle dérogation est nécessaire pour concilier le droit à la protection des données à caractère personnel et la liberté d'expression et d'information » dans le cadre de traitements de données à caractère personnel mis en œuvre à des fins, notamment, « d'expression universitaire », journalistique ou artistique, il est possible de déroger :

- à l'obligation de conserver les données « sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire au regard des finalités pour lesquelles elles sont traitées » (art. 4 (5) de la loi informatique et libertés),
- à l'interdiction de traiter des catégories particulièrement sensibles de données à caractère personnelles, comme les données sur les opinions politiques ou la santé²³ (art. 6 de la loi informatique et libertés),
- à l'interdiction de traiter des données relatives « aux condamnations pénales, aux infractions ou aux mesures de sûreté connexes » (art. 46 de la loi informatique et libertés),
- au droit d'information des personnes concernées (art. 48 et 119 de la loi informatique et libertés),
- au droit d'accès des personnes concernées (art. 49, 118 et 119 de la loi informatique et libertés),
- au droit de rectification des personnes concernées (art. 50 et 119 de la loi informatique et libertés),
- au droit à la limitation du traitement des personnes concernées (art. 53 de la loi informatique et libertés),
- à l'ensemble des règles relative au transfert de données à caractère personnel en-dehors des frontières de l'Espace économique européen (chapitre V du RGPD).

Le bénéfice de cette dérogation pourrait-il permettre, « à des fins d'expression universitaire », de ne pas avoir à recourir au 6° de l'article 44 de la loi informatique et libertés, qui oblige à demander l'avis de la CNIL pour les traitements « nécessaires à la recherche » ? L'articulation entre ces deux textes est peu claire.

Il semble d'abord évident que la publication académique doive être l'objectif du traitement de données effectué. Une telle publication implique forcément une phase de recherche préalable, mais l'inverse

22 Voir :

https://commission.europa.eu/system/files/2018-10/fr_notification_gdpr_articles_49_51_84_85_88_90_publish.pdf

23 Pour être précis, il s'agit « des données à caractère personnel qui révèlent la prétendue origine raciale ou l'origine ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale d'une personne physique ou de traiter des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique. »

n'est pas vrai : nous pourrions imaginer une recherche ne débouchant pas sur une publication. Mais quand est-ce qu'un traitement a une finalité « d'expression universitaire » et non « de recherche » ?

Un élément d'éclaircissement pourrait être trouvé en s'inspirant du raisonnement du juge de la CJUE dans les arrêts *Buivids*²⁴ et *Satamedia*²⁵. Le juge de l'Union a en effet entendu défendre une acception large de la notion de « traitement à des fins journalistiques », dans le but de garantir le droit à la liberté d'expression :

« Afin de tenir compte de l'importance que détient la liberté d'expression dans toute société démocratique, il convient, d'une part, d'interpréter les notions y afférentes, dont celle de journalisme, de manière large. » (arrêt *Satamedia*, pt. 56)

Nous devrions donc nous attendre à ce que si la question lui était posée, la CJUE se prononce en faveur d'une acception elle aussi large de la notion d'« expression universitaire. » Mais il n'existe pas de certitude sur ce point, et la relation entre cette finalité et la recherche scientifique n'est pas évidente, et nous aurons l'occasion d'y revenir.

La réponse à la question de l'équilibre entre le droit à la liberté d'expression d'une part, et le droit de la propriété intellectuelle contenu dans le droit à la propriété privée d'autre, est encore plus incertaine. La CEDH et la CJUE s'opposent en effet sur ce point (Varet, 2020), et il n'existe pas de mécanisme permettant de trancher des différences d'appréciation entre ces deux juridictions.

Selon la CEDH, il faut dégager des exceptions au droit d'auteur qui ne sont pas prévues par le législateur d'un État, si jamais ces exceptions sont nécessaires à l'exercice du droit à la liberté protégé consacré à l'article 10 de la Convention EDH²⁶. Ainsi, « indépendamment et hors des exceptions aux droits d'auteur prévues par les lois nationales, on peut [selon la CEDH] faire prévaloir la liberté d'expression sur le droit d'auteur, et donc, dans une circonstance donnée, admettre qu'une utilisation d'une œuvre soit faite sans autorisation, alors qu'il n'y a pas d'exception qui le justifie » (Varet, 2020, p. 106). Ce n'est pas l'avis de la CJUE, qui, dans trois arrêts rendus en grande chambre le 29 juillet 2019²⁷, s'est limitée à vérifier si la loi nationale prévoyait une conciliation entre droit à la propriété privée et liberté d'expression, et admettait sans véritable examen la validité de l'équilibre choisi par le législateur national. Concrètement, cela signifie qu'il n'est pas possible, selon la CJUE, de se prévaloir de la liberté d'expression pour utiliser une œuvre protégée par le droit d'auteur sans autorisation, dès lors que l'on n'entre pas dans l'une des exceptions prévues par la loi du pays, à savoir, en France, à l'article L122-5 du Code de la propriété intellectuelle.

La Cour de cassation française a admis, à la suite de la CEDH, qu'il était nécessaire d'opérer un équilibre entre le droit à la propriété intellectuelle lorsque cela est nécessaire pour l'exercice de la liberté d'expression protégée à l'article 10 de la Convention EDH²⁸. Mais lorsque la Cour d'appel de

24 CJUE 14 février 2019, *Sergej Buivids*, affaire C-345/17.

25 CJUE 16 décembre 2018, *Tietosuojaalvaututettu contre Satakunnan Markkinapörssi Oy et Satamedia Oy*, affaire C-73/07.

26 Voir : CEDH 10 janvier 2013, *Ashby Donald c. France*, req. n° 36769/08, et CEDH 19 février 2013, *Nejj et Kolmisoppi c. Suède*, req. n° 40397/12.

27 CJUE 29 juillet 2019, *Funke Medien NRW GmbH contre Allemagne*, Aff. C-469/17 ; CJUE 29 juillet 2019, *Pelham GmbH e.a. contre Ralf Hütter et Florian Schneider-Esleben*, Aff. C-476/17 ; CJUE 29 juillet 2019, *Spiegel Online GmbH v Volker Beck*, Aff. C-516/17.

28 Cass. civ. 1ère, 15 mai 2015, pourvoi n°13-27.391.

Versailles a jugé²⁹ l'affaire qui lui avait été renvoyée par la Cour de cassation suite à cet arrêt, elle s'est ouvertement opposée au principe énoncé dans ce dernier, et affirme au contraire une position proche de celle de la CJUE :

« [...] une interprétation large de l'arrêt de la Cour de cassation introduirait la possibilité de paralyser purement et simplement le droit d'auteur ; [...] elle serait source d'une confusion qui fragiliserait considérablement le droit d'auteur ; [...] elle viderait de sa substance l'article L 122-5 du code de la propriété intellectuelle contenant une liste fermée des exceptions ; [...] la notion de « juste équilibre » doit être appréciée au sens le plus strict sauf à engendrer une insécurité juridique ; [...] en outre se pose la question de la méthodologie à adopter pour l'appréciation *in concreto*, la Cour de cassation n'ayant donné aucune directive ; [...] il ne peut s'agir de peser au cas par cas l'importance respective d'une part de l'œuvre litigieuse pour apprécier la liberté créatrice du tiers, et d'autre part l'œuvre revendiquée pour apprécier l'atteinte au droit exclusif et au droit moral de son auteur sauf à ériger le juge en critique d'art ; [et] il y a un risque d'ouvrir une brèche en incitant tous les pseudos artistes [sic] à emprunter à des œuvres préexistantes sous prétexte de s'exprimer [...] »

Ce jugement de la Cour d'appel de Versailles est critiquable à de nombreux égards. Outre un langage qui rappelle une discussion de comptoir, et qui ne l'honore pas, dans laquelle les juges se sont émus du risque, selon eux, que « tous les pseudos artistes » deviennent tout d'un coup autorisés à contrefaire tout ce qu'ils veulent au motif d'une dangereuse liberté d'expression, l'arrêt ne renforce en rien, comme il le prétend, la sécurité juridique en s'opposant à un arrêt de la Cour de cassation et à la jurisprudence de la CEDH.

Par ailleurs, en matière d'enseignement supérieur et de recherche, un point d'importance à rappeler dans la discussion sur l'équilibre entre liberté d'expression et droit à la propriété privée, est de rappeler que les libertés académiques disposent de garanties conventionnelles et constitutionnelles renforcées. En France, l'indépendance des enseignant·e·s chercheurs·ses est reconnue par le Conseil constitutionnel comme un principe fondamental reconnu par les lois de la République, d'ordre constitutionnel, depuis une décision du 20 janvier 1984³⁰.

Au niveau de l'UE, l'article 13 de la Charte des droits fondamentaux de l'UE proclame que « les arts et la recherche scientifique sont libres » et que « la liberté académique est respectée. » Il ne s'agit pas d'une simple déclaration de principe sans densité juridique, puisque le 6 octobre 2020, la CJUE a condamné la Hongrie notamment pour avoir des mesures contre la Central European University ayant eu pour effet « de priver les universitaires concernés de l'infrastructure autonome, nécessaire à la conduite de leurs recherches scientifiques et à l'exercice de leurs activités pédagogiques³¹ », sans que cette privation ne réponde à un besoin d'intérêt général³². Dans leur arrêt, les juges de l'Union ont rappelé que « la liberté académique, dans la recherche comme dans l'enseignement, doit garantir la liberté d'expression et d'action, la liberté de communiquer des informations de même que celle de rechercher et de diffuser sans restriction le savoir et la vérité, étant précisé que cette liberté n'est pas

29 C.A. Versailles 1^{re} chambre 1^{re} section, 16 mars 2018, n° 15/06029

30 C. Const. décision 83-165 DC du 20 janvier 1984

31 CJUE 6 octobre 2020, Commission européenne contre Hongrie, Aff. 66/18, pt. 228

32 CJUE 6 octobre 2020, Commission européenne contre Hongrie, Aff. 66/18, pt. 241

limitée à la recherche académique ou scientifique, mais qu'elle s'étend également à la liberté des universitaires d'exprimer librement leurs points de vue et leurs opinions³³. »

En conclusion de ces propos liminaux sur l'encadrement juridique de l'usage de données dans la recherche en SHS, en Europe, ni la liberté d'expression ni la liberté académique ne sont des libertés absolues. Elles peuvent être limitées par un objectif d'intérêt général, dont un autre droit fondamental, par exemple le droit à la vie privée, à la protection des données à caractère personnel, ou à la propriété privée. Ces limitations doivent toutefois respecter un strict principe de proportionnalité, et il y a donc lieu d'examiner l'ensemble des obligations qui encadrent la recherche scientifique, dont la recherche en SHS, en tenant compte de ces droits fondamentaux et du rôle singulier accordé à la recherche scientifique et aux libertés académiques par des conventions internationales et par la constitution française. Reste à voir si ces garanties théoriques trouveront à s'exprimer, en cas de litige, dans les décisions de justice, la recherche de ces équilibres n'ayant rien d'évident.

Les plans de gestion des données, les principes FAIR et l'archivage

Les données collectées ou produites dans le cadre de recherches scientifiques, qu'elles soient en SHS ou non, peuvent être soumises à des obligations de libreaccès, de publication, et d'archivage.

L'ouverture de l'accès à ces données est prévue non seulement dans un cadre général de droit d'accès aux documents administratifs³⁴. Par exemple, dans un avis de 2015, la Commission d'accès aux documents administratifs (CADA) avait prononcé un avis favorable à une demande d'accès aux valeurs de l'indice de température moyenne en France de 1900 à 2013, à la liste des stations météorologiques et leurs températures annuelles utilisées pour évaluer la valeur de chaque indice de 1900 à 2013 et à la méthode de calcul de chaque indice à partir des relevés des stations³⁵. Il ne s'agit pas de données produites par des universités, et cet avis ne dit rien sur l'usage scientifique fait de ces données, mais il est aisément envisageable qu'une telle demande soit réalisée dans le but de réutiliser les données à des fins de recherche scientifique, ou de vérification de résultats scientifiques.

Le Plan national pour la science ouverte du Ministère de l'enseignement supérieur et de la recherche défend par ailleurs une stratégie de science ouverte, justifiée notamment par des intérêts généraux d'ordre économique, tels que l'objectif de développer l'intelligence artificielle en facilitant l'accès à des jeux de données d'apprentissage (MESR, 2018, p. 6). Le projet de la Commission, qui a donné naissance au Data Governance Act, répondait aussi à cet objectif, en cherchant à inciter les personnes morales de droit public à solliciter l'autorisation des tiers qui détiennent des droits sur les données qu'elles ont en leur possession, de sorte à pouvoir les publier dans une logique de libreaccès (Baloup *et al*, 2021). L'article 5 (6) du DGA prévoit ainsi que « lorsqu'il est impossible d'autoriser la

33 CJUE 6 octobre 2020, Commission européenne contre Hongrie, Aff. C-66/18, pt. 225

34 Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal

35 CADA, Avis 20144561 - Séance du 08/01/2015 – Météo France

réutilisation des données en respectant les obligations [prévues par le texte] et qu’il n’existe pas de base juridique pour la transmission des données au titre du [RGPD], l’organisme du secteur public met tout en œuvre [...] pour aider les réutilisateurs potentiels à demander le consentement des personnes concernées ou l’autorisation des détenteurs de données dont les droits et intérêts peuvent être affectés par cette réutilisation [...]. » Ces obligations ne s’appliquent pas aux établissements d’enseignement, mais il n’est pas clair si les activités de recherche des universités sont exclues. Les établissements de recherche, comme le CNRS, semblent toutefois concernés par ces règles.

D’une façon générale, et au-delà de l’exemple récent du DGA, l’ouverture des données est fortement encouragée dans le cadre de la recherche sur projets, publique ou privée, financée sur des fonds publics. L’article L533-4 du Code de la recherche, introduit par la loi pour une République numérique de 2016, prévoit ainsi l’accès en libre-accès après un délai d’embargo des publications de résultats des recherches financées au moins pour moitié par des fonds publics, et dispose par ailleurs que « dès lors que les données issues d’une activité de recherche financée au moins pour moitié par des dotations de l’État, des collectivités territoriales, des établissements publics, des subventions d’agences de financement nationales ou par des fonds de l’Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu’elles ont été rendues publiques par le chercheur, l’établissement ou l’organisme de recherche, leur réutilisation est libre. » La règle est ainsi devenue le libre-accès et le partage, sauf si une règle, relative par exemple à la vie privée ou bien au droit d’auteur, s’y oppose.

Les projets européens comme les projets de l’Agence nationale de la recherche (ANR) exigent que soient rédigés des plans de gestion des données (PGD). Ainsi, le contrat conclu en avril 2021 entre l’État et cette agence en vertu de l’article L329-2 du Code de la recherche prévoit l’obligation de produire des plans de gestion des données dans le cadre des projets financés, et l’article 39 (4) du Règlement européen Horizon Europe³⁶ dispose que « les bénéficiaires gèrent toutes les données de la recherche générées dans le cadre d’une action au titre du programme, dans le respect des principes FAIR et conformément à la convention de subvention, et établissent un plan de gestion des données. »

Les principes FAIR désignent les objectifs suivants (Wilkison et al, 2016) :

<p>Findable (trouvable)</p>	<p>F1. Chaque donnée et métadonnée se voit assigner un identifié pérenne (de type DOI ou HAL, par exemple) ; F2. Chaque jeu de données de recherche doit être accompagné de métadonnées qui le décrivent ; F3. Les métadonnées incluent clairement et explicitement l’identifiant des données décrites ; F4. Ces métadonnées doivent être indexées et requêtables.</p>
<p>Accessible</p>	<p>A1. Les métadonnées peuvent être requêtées par leur identifiant en utilisant un protocole de communication standardisé, 1. Ce protocole est libre, gratuit, et universellement implémentable, 2. il permet, là où cela est nécessaire, de procéder à une</p>

36 Règlement 2021/695/UE du Parlement européen et du Conseil du 28 avril 2021 portant établissement du programme-cadre pour la recherche et l’innovation «Horizon Europe» et définissant ses règles de participation et de diffusion, et abrogeant les règlements (UE) no 1290/2013 et (UE) no 1291/2013

	<p>authentification et de gérer des droits d'accès ;</p> <p>A2. Les métadonnées sont accessibles même lorsque les données ne sont plus disponibles.</p>
Interoperable	<p>I1. Les données et les métadonnées utilisent un langage formel, accessible, partagé et largement accepté pour la représentation des connaissances ;</p> <p>I2. Les données et les métadonnées utilisent un vocabulaire qui obéit lui-même aux principes FAIR ;</p> <p>I3. Les données et les métadonnées utilisent des références qualifiées vers d'autres données et métadonnées.</p>
Reusable (réutilisable)	<p>R1. Les données et les métadonnées sont décrites de façon riche avec une pluralité d'attributs précis et pertinents ;</p> <p>R2. Les données et les métadonnées sont publiées avec une licence d'utilisation claire et accessible ;</p> <p>R3. Les données et les métadonnées sont associées à une provenance détaillée ;</p> <p>R4. Les données et les métadonnées sont conformes aux standards propres à leur domaine.</p>

Le modèle de plan de gestion de données (PGD) proposé par la Commission européenne pour les projets Horizon Europe³⁷, disponible uniquement en anglais, accorde une place centrale à ces principes dans sa structure :

1. Sommaire des données ;
2. Principes FAIR ;
3. Allocation des ressources (dédiées à rendre les données conformes aux principes FAIR) ;
4. Sécurité des données ;
5. Aspects éthiques ;
6. Autres sujets, notamment si ceux-ci sont indiqués dans des procédures nationales, locales, sectorielles ou liées à des co-financeurs ;
7. Une dernière section contient des liens vers des ressources pour accompagner la rédaction et la mise en œuvre des plans de gestion des données.

L'objectif de ces PGD est de prévoir à l'avance une feuille de route pour le partage sécurisé des données, dans le respect des droits des enquêtés et des tiers, pour aboutir à un partage qui soit, selon une formule souvent employée, « aussi ouvert que possible, aussi fermé que nécessaire » (Galonnier, 2021). La logique est de publier par défaut, sauf si une exception, rendue par exemple nécessaire pour respecter le droit à la vie privée ou le secret des affaires, s'applique, et d'indexer les données de la façon la plus efficace possible. Enfin, à la fin d'un projet de recherche, les données peuvent être archivées, et souvent d'ailleurs souvent soumises à des obligations d'archivage pérenne.

³⁷ Voir sur cette page : https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm (page consultée le 23 décembre 2021).

Les archives sont « l'ensemble des documents et des données, quels que soient leur date, leur lieu de conservation, leur forme, leur support matériel, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité » (art. L211-1 du Code du patrimoine). Sont des archives publiques les documents qui procèdent de l'activité de l'administration dans le cadre de leurs missions de service public (art. L211-4 du Code du patrimoine).

Les établissements publics « contribuant au service public de la recherche et les fondations reconnues d'utilité publique ayant pour activité principale la recherche publique » doivent offrir « les conditions du respect des exigences de l'intégrité scientifique pour les activités et travaux menés en leur sein » et conserver « les résultats bruts des travaux scientifiques réalisés en leur sein afin de permettre leur vérification » (art. L211-2 du Code de la recherche). Il y a donc une obligation, qui pèse sur l'établissement plus que sur le ou la chercheur·se individuel·le de réaliser l'archivage conformément aux dispositions du Code du patrimoine.

L'archivage de données à caractère personnel est permis au-delà de la durée de conservation fixée pour les finalités initiales d'un traitement de données à caractère personnel par l'article 5 (1) sous e) du RGPD.

La communication des archives se fait conformément aux dispositions du Code du patrimoine.

Dans le cadre du dépôt légal, un·e auteur·e ne peut par exemple faire obstacle à ce qu'un organisme dépositaire, comme la Bibliothèque nationale de France (BNF) pour les livres ou l'Institut national de l'audiovisuel pour les productions audiovisuelles, organise la consultation sur place, par des chercheurs dûment accrédités, sur des postes individuels, d'œuvres protégées par le droit d'auteur (art. L132-4 du Code du patrimoine). Les « signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique » font, en vertu de l'article L131-2 du Code du patrimoine, partie du champ couvert par l'obligation de dépôt légal. C'est ainsi que la BNF organise un archivage du Web, y compris de la partie publique de certains réseaux sociaux.

Le droit de la propriété intellectuelle, dont le droit d'auteur

Le droit de la propriété intellectuelle fait partie des droits qui peuvent d'une part limiter l'accès des chercheur·ses à des sources primaires, et d'autre part, limiter la diffusion des corpus de recherche. Il s'agit d'un sujet d'inquiétude majeur pour plusieurs disciplines des SHS. Un rapport de l'Institut national d'histoire de l'art (INHA) et de la Fondation de France évoque même un « constat alarmant » (Denoyelle et al., 2018, p. 5).

Ce corpus de règles dérive du droit fondamental à la propriété privée qui, comme rappelé en introduction du présent rapport, est protégé entre autres par la Constitution³⁸. Le droit de la propriété intellectuelle recouvre plusieurs catégories. Il convient notamment de distinguer la propriété littéraire et artistique (droit d'auteur et droits voisins) de la propriété industrielle (brevets, marque commerciale...).

L'étude des brevets ne pose pas de problème particulier, dans la mesure où, par nature, ceux-ci sont rendus publics en échange de la protection accordée à leur inventeur. Les obstacles à la recherche ou à

³⁸ Cf. C. Const. 8 janvier 1991, décision 90-283 DC.

la partage de données de recherche concernent souvent la propriété littéraire et artistique, et en particulier le droit d'auteur.

Le droit d'auteur protège une œuvre de l'esprit. Une œuvre de l'esprit est protégée « quels qu'en soient le genre, la forme d'expression, le mérite ou la destination » (art. L112-1 du Code de la propriété intellectuelle). Sont notamment considérées comme de telles œuvres les « livres, brochures et autres écrits littéraires et scientifiques », les conférences, les compositions musicales, les œuvres cinématographiques, les illustrations ou encore les logiciels (art. L112-2 du Code de la propriété intellectuelle).

Une œuvre n'est pas une simple idée. Elle doit s'incarner dans une forme, et elle ne doit pas être banale. Elle doit faire l'objet d'une « création » qui porte la marque de son auteur. Dans la pratique, définir ce qu'est précisément une « œuvre de l'esprit » peut s'avérer difficile. Il a fallu plus d'un siècle, par exemple, pour que des photographies puissent être considérées comme telles. Déterminer si une photographie présente un élément de créativité reste aujourd'hui un élément contentieux (Ader, 2005), d'autant que cette originalité doit être appréciée « œuvre par œuvre³⁹. »

Une fois le caractère créatif d'une œuvre de l'esprit établi, l'auteur·e jouit de droits moraux et patrimoniaux du seul fait de cet acte de création (art. L111-1 du Code de la propriété intellectuelle). Au titre des droits moraux, un auteur dispose du droit de divulguer ou non son œuvre, d'un droit de paternité, d'un droit de repentir ou encore d'un droit au respect de son œuvre. Au titre des droits patrimoniaux, il a le droit d'exiger une rémunération au titre de l'usage fait de son œuvre. L'auteur·e et ses ayants droit bénéficient de ces droits patrimoniaux pour une période de 70 ans à compter du 1^{er} janvier de l'année suivant le décès de l'auteur·e (art. L123-1 du Code de la propriété intellectuelle). Pendant toute cette période, iels peuvent notamment autoriser ou interdire la reproduction de l'œuvre, et, le cas échéant, percevoir une rémunération en contrepartie. Toute reproduction non autorisée, soit par l'auteur·e, ses ayants droit ou par le législateur dans le cadre d'une exception prévue au Code de la propriété intellectuelle, est alors une contrefaçon.

Le contenu de certains corpus en SHS sont concernés par le droit d'auteur sans la moindre ambiguïté. C'est le cas par exemple des corpus de presse. Les articles de journaux étant des œuvres de l'esprit. Il en va de même pour un corpus de films ou de morceaux de musique.

Les contenus publiés sur des réseaux sociaux ne sont quant à eux pas tous protégés par le droit d'auteur, car ce ne sont pas tous des œuvres de l'esprit au sens de la loi. Certains peuvent être considérés comme banals, et dépourvus de tout aspect créatif. La difficulté est d'être en mesure de distinguer les contenus protégés, surtout lorsque l'enquête réalisée vise à étudier un volume important de messages, comme cela est souvent le cas dans des recherches en humanités numériques (voir par exemple : Bottini et Julliard, 2017). Même si nous pourrions imaginer entraîner un algorithme d'apprentissage automatique à distinguer les contenus « banals » des contenus « créatifs » au sens du droit de la propriété intellectuelle, et selon le jugement initial des personnes en charge de l'entraînement de cet outil, rien ne garantirait qu'un·e juge en décide autrement. Il apparaît dès lors vain de chercher à expurger un corpus au volume important de toute œuvre de l'esprit. Une telle

39 TGI Paris, 3e Ch. 3e sect., 18 oct. 2006, n° 05/13494.

opération introduirait au demeurant des biais qui, dans de nombreux cas, paraissent peu compatibles avec une démarche scientifique.

Il est à noter que les clauses de cession de droits d'auteur rédigées au bénéfice de réseaux sociaux et couvrant les contenus publiés par leurs utilisateurs ont déjà été jugées comme étant des clauses abusives, réputées non-écrites, dans certaines circonstances⁴⁰. Là encore, il existe une difficulté pratique pour des chercheurs·ses non-spécialistes (et même pour les spécialistes) à distinguer les cas où de telles clauses sont abusives et donc à ignorer, des cas où il convient en effet de solliciter l'autorisation du réseau social en question pour y collecter des données dont certaines peuvent être des œuvres de l'esprit protégées au titre du droit d'auteur. Par souci de sécurité juridique, il n'est pas absurde d'émettre l'hypothèse (à vérifier par des données empiriques) que de nombreux collègues, dans le doute et par souci de sécurité juridique, sollicitent des autorisations même lorsqu'ils ne sont liés par aucune clause juridiquement valable à ce sujet.

La recherche scientifique étant considérée comme relevant d'un intérêt général, le législateur a prévu pour elle quelques assouplissements. Ainsi, malgré la difficulté déjà abordée à voir s'affirmer une exception au droit d'auteur fondée directement sur le droit à la liberté d'expression, voire à la liberté académique et scientifique, il existe dans le droit positif, une fois une œuvre de l'esprit divulguée, une longue liste d'exceptions au droit d'auteur. En dresser la liste n'est pas l'objet de ce rapport, et nous préférons renvoyer les lecteurs et lectrices à l'article L122-5 du Code de la propriété intellectuelle, qui les énumère de façon exhaustive. Nous nous concentrerons ici sur celles de ces exceptions qui sont plus particulièrement susceptibles d'intéresser la recherche en SHS, notamment de celles qui portent sur des données recueillies sur des réseaux sociaux numériques, certaines d'entre elles pouvant être soumises au droit d'auteur.

L'article L122-5 3° sous a) permet tout d'abord les « analyses et courtes citations justifiées par le caractère critique, polémique, pédagogique, scientifique ou d'information de l'œuvre à laquelle elles sont incorporées. » Cela permet d'inclure des *courtes* citations d'œuvres dans des publications et communications scientifiques, mais cette exception est interprétée de façon extrêmement limitative par le juge français. L'ambiguïté autour du terme de *courte* citation génère une grande insécurité juridique. Ainsi, le TGI de Paris avait jugé en mai 2002 que « la longueur des extraits de chansons diffusés par les défenderesses de l'ordre de 30 secondes s'oppose, au regard de la brièveté des œuvres citées d'une durée de l'ordre de 3 minutes chacune, à ce que ces extraits soient considérés comme des courtes citations⁴¹. » Quant aux images, il est très difficile, en France, de se prévaloir du droit de courte citation pour les reproduire, et ce malgré de nombreux appels à l'instauration d'un droit de citation graphique et audiovisuel (cf. Maurel, 2015).

Il existe toutefois des signes d'inflexion dans la jurisprudence récente, notamment sous l'effet du droit européen. Dans une décision de 2009, la Cour de Cassation avait statué que « l'article L. 122-5 du code de la propriété intellectuelle doit être interprété à la lumière de l'article 5-3- c de la directive 2000 / 29 / CE du 22 mai 2000 [sic⁴²] [...] ; que ce texte reconnaissant pour les États membres la faculté de prévoir

40 Au sujet des conditions générales d'utilisation de Facebook, voir : TGI de Paris, 9 avril 2019, 14/07298.

41 TGI de Paris, 3^{ème} chambre, 1^{ère} section. 15 mai 2002, n° 08/01461.

42 Il s'agit en réalité de la Directive 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information.

une exception au monopole du droit d'auteur à des fins d'information, *sans restriction tenant à la nature de l'œuvre, cette exception a vocation à s'appliquer et ne permet pas d'exclure du champ d'application de l'article L. 122-5 3° précité les œuvres photographiques dont la reproduction, fût-elle intégrale, doit recevoir la qualification de courte citation* dès lors qu'elle répond, comme en l'espèce, à un but d'information⁴³. »

Plus récemment, en janvier 2021, l'artiste de rue Combo a été débouté de son action contre une courte évocation d'une fresque murale dans un clip de campagne de la France insoumise par le Tribunal judiciaire de Paris, qui justifia notamment sa décision par la « "liberté de panorama" permet ainsi à toute personne de photographier, filmer, dessiner, etc... les œuvres d'architecture et de sculpture, ainsi que les graffitis dont ils sont éventuellement couverts, dès lors qu'ils sont situés en permanence sur la voie publique, pourvu, prévoit le droit français, que la reproduction soit le fait d'une personne physique, à des fins non commerciales⁴⁴. » Cette liberté de panorama ne doit pour autant pas être confondue avec le droit de courte citation, dont l'usage est soumis à la poursuite d'une finalité « critique, polémique, pédagogique, scientifique ou d'information » et à la présence explicite de la référence à l'œuvre originale⁴⁵.

L'article 5 (3) sous a) de la directive 2001/29/CE sur le droit d'auteur prévoit, dans sa rédaction amendée par la directive sur le copyright de 2019⁴⁶, que les États membres ont la faculté de prévoir des exceptions au droit d'auteur « lorsqu'il s'agit d'une utilisation à des fins exclusives d'illustration dans le cadre de l'enseignement ou de recherche scientifique, sous réserve d'indiquer, à moins que cela ne s'avère impossible, la source, y compris le nom de l'auteur, dans la mesure justifiée par le but non commercial poursuivi [...]. » Il s'agit cependant toujours, y compris après la réforme européenne du droit d'auteur de 2019, d'une simple *faculté* accordée aux États membres d'adopter des exceptions au droit d'auteur. Il n'y a pas d'obligation qui soit issue de cette directive. En France, le Code de la propriété intellectuelle prévoit à son article L122-5 3° sous e) l'existence d'une exception pour :

« La représentation ou la reproduction d'extraits d'œuvres, sous réserve des œuvres conçues à des fins pédagogiques et des partitions de musique, à des fins exclusives d'illustration **dans le cadre de la recherche**, dès lors que cette représentation ou cette reproduction est destinée, notamment au moyen d'un espace numérique de travail, à un public composé majoritairement de chercheurs directement concernés par l'activité de recherche nécessitant cette représentation ou cette reproduction, **qu'elle ne fait l'objet d'aucune publication ou diffusion à un tiers au public ainsi constitué**, que l'utilisation de cette représentation ou cette reproduction ne donne lieu à aucune exploitation commerciale **et qu'elle est compensée par une rémunération négociée sur une base forfaitaire** sans préjudice de la cession du droit de reproduction par reprographie mentionnée à l'article L122-10 ; »

43 C. Cass., Première chambre civile, 22 janvier 2009, pourvoi n° 07-21.063.

44 TJ de Paris, 21 janvier 2021, n° 20/08482.

45 Voir à ce sujet : TJ de Paris, 4 mars 2022, n° 22/34 (affaire des clips de campagne d'Éric Zemmour).

46 Directive 2019/790/UE du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE.

Il s'agit donc d'une possibilité très restreinte, susceptible de poser un obstacle majeur aux projets de recherche, nombreux dans le domaine des SHS, des *Code Studies* à l'Histoire de l'art en passant par la musicologie et la littérature comparée, ayant pour objet des œuvres de l'esprit. Cela est particulièrement vrai pour la recherche privée, en l'absence d'un accord de rémunération versée en compensation de cette exception. Dans le cadre de la recherche publique, plusieurs accords⁴⁷ ont été passés par le Ministère de l'enseignement supérieur et de la recherche. Ceux-ci font l'objet d'une renégociation périodique. Les conditions posées dans ces accords étant assez complexes, elles méritent un traitement séparé qui sort de l'objet d'examen global du droit applicable à l'utilisation des données, notamment issues de réseaux sociaux numériques, dans des projets en SHS. Nous renvoyons donc les lecteurs-trices vers une explication détaillée rédigée par Véronique Ginouvès (2017) sur le carnet Hypothèses « Questions éthique & droit en SHS. »

Une importante source d'insécurité juridique à noter au sujet de cette exception pédagogique et de recherche est que les sites ministériels⁴⁸ renvoient toujours, au sujet de l'usage à des fins pédagogiques d'extraits d'œuvres de la catégorie « livres, presse, images et musique imprimée », au Protocole d'accord sur l'utilisation et la reproduction des livres, des œuvres musicales éditées, des publications périodiques et des œuvres des arts visuels à des fins d'illustration des activités d'enseignement et de recherche du 22 juillet 2016. Or, celui-ci étant prévu pour une durée « qui ne peut excéder 48 mois » (cf. son article 10), il est donc parvenu à expiration le 22 juillet 2020.

C'est ainsi que les modalités de transposition par le législateur français de la directive de 2001 sur le droit d'auteur renvoient donc à un accord entre l'État et des personnes morales de droit privé. Ces dernières ont alors un pouvoir important dans la détermination de l'équilibre entre le droit à la propriété privée – dont le droit à la propriété intellectuelle fait partie – et la liberté d'expression et de recherche scientifique. Or, en vertu de l'article 34 de la Constitution, ce serait plutôt au législateur, et non à un accord entre l'exécutif et des personnes privées, de fixer « les droits civiques et les garanties fondamentales accordées aux citoyens pour l'exercice des libertés publiques ; la liberté, le pluralisme et l'indépendance des médias ; les sujétions imposées par la Défense nationale aux citoyens en leur personne et en leurs biens. » L'ordonnance du 24 novembre 2021⁴⁹, qui reste attaquant devant le

47 Il s'agit du Protocole d'accord sur l'utilisation et la reproduction des livres, des œuvres musicales éditées, des publications périodiques et des œuvres des arts visuels à des fins d'illustration des activités d'enseignement et de recherche du 22 juillet 2016, de l'Accord sur l'utilisation des œuvres cinématographiques et audiovisuelles à des fins d'illustration des activités d'enseignement et de recherche du 4 décembre 2009 et de l'Accord sur l'interprétation vivante d'œuvres musicales, l'utilisation d'enregistrements sonores d'œuvres musicales et l'utilisation de vidéo-musiques à des fins d'illustration des activités d'enseignement et de recherche du 4 décembre 2009.

48 Par exemple : <https://eduscol.education.fr/420/comment-utiliser-des-oeuvres-dans-un-cadre-pedagogique> (page consultée le 23 décembre 2021).

49 Ordonnance n° 2021-1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE.

Conseil d'État jusqu'à l'adoption par le Parlement d'une loi de ratification⁵⁰, transposant la directive 2019, ne fournit pas de remède à cette situation.

Quoiqu'elle n'ait pas résolu les difficultés soulevées précédemment, cette ordonnance a élargi, conformément aux dispositions de la réforme européenne du droit d'auteur, les possibilités de procéder à de la fouille de données à des fins scientifiques. L'article 3 de la directive de 2019 sur le droit d'auteur accorde en effet aux « organismes de recherche » une exception spécifique. Ces organismes de recherche sont définis comme étant :

« [...] une université, y compris ses bibliothèques, un institut de recherche ou toute autre entité, ayant pour objectif premier de mener des recherches scientifiques, ou d'exercer des activités éducatives comprenant également des travaux de recherche scientifique:

a) à titre non lucratif ou en réinvestissant tous les bénéfices dans ses recherches scientifiques; ou

b) dans le cadre d'une mission d'intérêt public reconnue par un État membre;

de telle manière qu'il ne soit pas possible pour une entreprise exerçant une influence déterminante sur cet organisme de bénéficier d'un accès privilégié aux résultats produits par ces recherches scientifiques; » (art. 2 (1) de la directive 2019/790).

Cela signifie que des universités ou organismes de recherche privés reconnus d'utilité publique, comme les établissements d'enseignement supérieur privé d'intérêt général, peuvent bénéficier de cette nouvelle exception, qui pourra s'avérer particulièrement utile pour la recherche sur les réseaux sociaux.

Cette « fouille de textes et de données » est définie à l'article 2 (2) de cette même directive comme « toute technique d'analyse automatisée visant à analyser des textes et des données sous une forme numérique afin d'en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations. » La faculté de procéder à de telles fouilles est accordée en échange d'une obligation de stockage « avec un niveau de sécurité approprié et peuvent être conservées à des fins de recherche scientifique, y compris pour la vérification des résultats de la recherche » (art. 3 de la directive directive 2019/790).

Ces règles ont été transposées à l'article L122-5-3 du Code de la propriété intellectuelle, mettant ainsi fin à l'attente d'adoption par le gouvernement d'un décret d'application pris en application de dispositions allant dans le même sens, qui avaient été introduites par la loi pour une République numérique en 2016. Un décret du 23 juin 2022⁵¹ est venu compléter le dispositif en intégrant aux articles R122-23-I à R122-32 des précisions sur les modalités pratiques de mise en œuvre du droit à la

50 Délai fixé à douze mois par l'article 28 de la loi n° 2020-1674 du 24 décembre 2020 de programmation de la recherche pour les années 2021 à 2030 et portant diverses dispositions relatives à la recherche et à l'enseignement supérieur. Un texte a été déposé au Sénat le 23 février 2022, mais n'a toujours pas fait l'objet d'une première lecture. Voir le dossier législatif du projet de loi ratifiant l'ordonnance n° 2021-1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, sur le site du Sénat : <https://www.senat.fr/dossier-legislatif/pjl21-542.html> (page consultée le 9 janvier 2023).

51 Décret n° 2022-928 du 23 juin 2022 portant modification du code de la propriété intellectuelle et complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE

fouille de textes et de données. Des accords conclus avec les titulaires des droits sur des « bonnes pratiques » en matière de sécurisation des données protégées par droit d'auteur, sont encouragés par le droit de l'Union (art. 3 (4) de la directive 2019/790) et prévus en droit français (art. L122-5-3-II du Code de la propriété intellectuelle).

Malgré cette avancée, le panorama des exceptions au droit d'auteur que nous venons d'effectuer montre qu'en l'état actuel du droit, l'équilibre entre respect du droit à la propriété privée d'une part, et liberté d'expression et liberté académique d'autre part, ne semble pas pleinement atteint. Sous l'effet du droit de l'UE, la loi et la jurisprudence françaises semblent toutefois évoluer dans un sens favorable à la recherche, notamment en ce qui concerne la possibilité de procéder à de la fouille de données sans autorisation.

Droit à la vie privée, à la protection des données à caractère personnel, et recherche scientifique

Présentation du cadre général

Certaines recherches peuvent poser problème du point de vue du droit à la vie privée et à la protection des données à caractère personnel des enquêté·e·s ou de tiers concerné·e·s par l'enquête. L'avis préliminaire du Contrôleur européen à la protection des données (EDPS, de l'anglais European Data Protection Supervisor) sur le sujet évoque ainsi à titre de repoussoir l'exemple d'expériences menées en 2014 par Facebook sur les émotions de ses utilisateurs·trices, ou des expériences menées la même année par OkCupid fondées sur la proposition à des utilisateurs·trices de profils faiblement compatibles, mais présentés comme exceptionnellement compatibles (EDPS 2020, p. 8). L'entreprise Cambridge Analytica se présentait initialement comme une entreprise de recherche, et il par ailleurs est aisé d'imaginer bien d'autres protocoles de recherche intrusifs dans la vie privée des individus, ou qui interfèrent avec leur libre-arbitre. La crainte de ce type d'abus motive l'existence de règles de droit qui encadrent et interdisent un certain nombre de pratiques de recherche au nom du droit à la vie privée. La majorité de ces règles découlent des règles générales qui protègent la vie privée et encadrent la protection des données à caractère personnel. Après une brève présentation du cadre général, nous verrons quelles dérogations ont été rédigées par le législateur européen pour opérer un équilibre entre droit à la vie privée et à la protection des données à caractère personnel d'une part, et, d'autre part droit à la liberté d'expression, libertés académiques et intérêt général poursuivi par la recherche scientifique.

L'article 9 du Code civil français dispose depuis 1970 que « chacun a droit au respect de sa vie privée. » Comme indiqué en introduction, le droit à la vie privée est également un droit fondamental reconnu par la Convention EDH, la Charte des droits fondamentaux de l'UE, et déduit par le Conseil constitutionnel de l'article 2 DDHC.

À ce droit, que Louis Brandeis et Samuel Warren (1890) sont connus pour avoir défini comme étant un « droit à être laissé tranquille » (*right to be let alone*), s'ajoute une nouvelle catégorie de droit : le droit à la protection des données à caractère personnel. Distingués par la Charte des droits fondamentaux de

l'UE et confondus par la Convention EDH, la relation théorique entre ces deux droits est complexe (González Fuster, 2014 et Clément-Fontaine, 2017). En pratique, l'encadrement du traitement de données à caractère personnel pour la recherche en SHS est assuré par le RGPD.

Celui-ci définit un traitement de données comme « toute opération ou tout ensemble d'opérations effectuées ou non à l'aide de procédés automatisés et appliquées à des données ou des ensembles de données à caractère personnel, telles que la collecte, l'enregistrement, l'organisation, la structuration, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, la diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, la limitation, l'effacement ou la destruction » (art. 4 (2) RGPD). Les données à caractère personnel, elles, sont constituées de « toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée « personne concernée »); est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale » (art. 4 (1) RGPD). Une donnée est donc à caractère personnel dès qu'il est possible de faire un lien, même indirectement, avec une personne physique, peu importe le contenu (privé ou non, intime ou non, sensible ou non) de cette information.

Comme le RGPD s'applique, en vertu de son article 2 (1), « au traitement de données à caractère personnel, automatisé en tout ou en partie, ainsi qu'au traitement non automatisé de données à caractère personnel contenues ou appelées à figurer dans un fichier ». Les définitions des termes de « traitement » et de « donnée à caractère personnel » précédemment citées confèrent un champ d'application matériel très vaste à ce règlement. La CJUE a ainsi eu l'occasion de rappeler qu'il ne fallait pas ramener la notion de « donnée à caractère personnel » aux seules données relevant de la vie privée⁵², et que des cahiers contenant des notes sur une opération de porte-à-porte⁵³ et des copies d'examens papier⁵⁴ entrent dans le champ d'application du RGPD dès lors que, contenant des données personnelles organisées de sorte à être facilement retrouvables, il fallait considérer qu'il s'agissait de « fichiers » au sens de ce règlement.

Tout traitement de données à caractère personnel doit obéir à des principes énumérés à l'article 5 du RGPD :

- Selon le principe de licéité, loyauté et de transparence, il faut que le traitement soit légal, loyal à l'égard des personnes concernées, qui doivent notamment disposer d'information au sujet du traitement opéré ;
- Selon le principe de limitation des finalités, il déterminer les finalités pour lesquelles des données personnelles sont collectées avant le début du traitement ;

52 CJUE 16 juillet 2015, ClientEarth contre EFSA, Aff. C-615/13P.

53 CJUE 10 juillet 2018, Jehovan Todistajat, Aff. C-25/17.

54 CJUE 20 décembre 2017, Peter Nowak contre Data Protection Commissioner, Aff. C-434/16.

- Selon le principe de minimisation des données, il ne faut collecter et traiter que le strict minimum nécessaire à la réalisation des finalités qui ont été déterminées ;
- Selon le principe d'exactitude, il faut s'assurer de la qualité et de l'exactitude des données ;
- Selon le principe de limitation de conservation, il faut s'assurer de ne pas conserver de données personnelles au-delà de la durée pour laquelle cette conservation est nécessaire pour réaliser les finalités recherchées ;
- Selon le principe d'intégrité et de confidentialité, il faut s'assurer de la sécurité informatique de l'accès à ces données ;
- Selon le principe de responsabilité, le responsable du traitement, c'est-à-dire la personne physique ou morale qui détermine les modalités et les finalités du traitement doit en assumer la responsabilité et doit être capable à tout moment de démontrer sa conformité aux règles du RGPD.

Chaque traitement de données à caractère personnel doit en outre reposer sur un des critères de licéité de l'article 6 (1) du RGPD. Ces critères sont :

- soit le consentement libre, éclairé, spécifique et univoque de la personne concernée, tenant compte de conditions fixées à l'article 7 du RGPD, notamment le fait que ce consentement doit pouvoir être retiré à tout moment ;
- soit la nécessité de collecter des données personnelles pour la réaliser d'obligations contractuelles auxquelles la personne concernée est partie ;
- soit la sauvegarde de la vie humaine ;
- soit la nécessité pour « l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement » (art. 6 (1) (e) RGPD) ;
- soit l'intérêt légitime du responsable du traitement ou d'un tiers, tel que le fait de préparer un contentieux ou, pour un moteur de recherche, d'indexer des données personnelles publiées sur le web dans le cadre de son activité⁵⁵, sous certaines réserves (art. 6 (1) (f) du RGPD).

L'article 9 du RGPD énumère des catégories de données sensibles pour lesquelles il existe des règles plus contraignantes. Il s'agit des « données à caractère personnel [dont le traitement] révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique. » Le traitement de ces données est *interdit*, sauf exceptions, sur lesquelles nous reviendrons.

Les responsables du traitement doivent fournir des informations détaillées sur les traitements de données à caractère personnel qu'ils effectuent (art. 12 à 14 du RGPD), et leur offrir un droit d'accès aux données (art. 15), un droit de faire rectifier des données (art. 16), un droit (dans certains cas) de

⁵⁵ Voir : CJUE 13 mai 2014, Google contre Espagne, aff. C-131/12

demander l'effacement des données (art. 17), un droit à demander la limitation du traitement des données (art. 18) et un droit à la portabilité de leurs données (art. 19). Ils doivent également tenir un registre des traitements de données effectués sous leur contrôle (art. 30), intégrer la vie privée dans la conception de leurs produits et services (art. 25), respecter des obligations de sécurité (art. 32 à 34), et, dans certains cas, désigner un délégué à la protection des données (arts. 37 à 39). Les données à caractère personnel doivent rester dans l'Espace économique européen ou bien être transférées sur le fondement de garanties appropriées (chapitre V). Enfin, « lorsqu'un type de traitement, en particulier par le recours à de nouvelles technologies, et compte tenu de la nature, de la portée, du contexte et des finalités du traitement, est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques, le responsable du traitement effectue, avant le traitement, une analyse de l'impact des opérations de traitement envisagées sur la protection des données à caractère personnel » (art. 35).

Proposer un exposé détaillé du RGPD n'est pas le propos du présent rapport. Nous préférons renvoyer le lecteur ou la lectrice vers d'autres lectures, comme le guide rédigé par l'Agence de l'UE pour les droits fondamentaux et le Conseil de l'Europe (2019) ou le manuel de Fabrice Mattatia (2021). Cette présentation générale des principales obligations qui pèsent sur les responsables de traitements de données à caractère personnel permet toutefois de se convaincre de la difficulté qu'elles peuvent représenter pour la recherche scientifique, dont la recherche en SHS, et celle sur des données issues de réseaux sociaux numériques en particulier.

Le cadre général applicable aux traitements de données à caractère personnel impose de recueillir le consentement éclairé, spécifique, libre et retirable à tout moment de toutes les personnes sur lesquelles portent des informations dans un corpus, ou de trouver une autre base légale. Le recours à l'intérêt légitime du responsable du traitement est une alternative au consentement risquée, surtout dans le cas où celui-ci impliquerait de travailler avec des données sensibles. Il faudrait également informer les personnes de la finalité d'une enquête avant d'entamer n'importe quel terrain, y compris lorsqu'une telle information peut biaiser les observations, voire, dans certains cas, mettre en danger les personnes qui mettent cette recherche. Les données qui portent par exemple sur les opinions politiques, indispensables entre autres aux terrains en science politique, ne pourraient être utilisées que dans des cas bien spécifiques, avec le consentement exprès des personnes concernées. Selon le cadre général, même des opinions politiques partagées publiquement sur un réseau social comme Twitter ou dans la presse ne pourraient pas être utilisées aisément. En effet, même si l'article 9 (2) (e) permet de passer outre l'interdiction du traitement de données particulièrement sensibles si elles ont été « manifestement rendues publiques par la personne concernée », cela ne dispense pas de disposer d'un critère de licéité au titre de l'article 6, ni du respect du principe de limitation des finalités⁵⁶. En d'autres termes, ce n'est pas parce que des données ont été rendues publiques par quelqu'un dans le but de s'exprimer politiquement, que cette personne accepte que ses données soient également traitées à des fins de recherche.

56 Dans son guide sur la collecte de données sur les réseaux sociaux à des fins de communication politique en période électorale, la CNIL propose des critères pour déterminer si des données ont été « manifestement rendues publiques » par la personne concernée : <https://www.cnil.fr/fr/communication-politique-quelles-regles-pour-la-collecte-de-donnees-sur-les-reseaux-sociaux> (page consultée le 9 janvier 2023).

Le cadre spécifique pour les activités de recherche scientifique et d'expression universitaire

Comme le souligne un avis du Contrôleur européen de la protection des données, les règles du RGPD « contiennent un régime spécial qui offre un degré de flexibilité aux vrais projets de recherche qui opèrent dans un cadre éthique et qui ont pour but de faire croître le savoir collectif de la société et son bien-être⁵⁷ » (EDPS, 2020, p. 2). Il existe toutefois un flou important entre le régime applicable aux traitements à des fins de recherche scientifique, et celui applicable à des fins d'expression universitaire.

Nous présenterons d'abord le régime applicable aux finalités de recherche scientifique, puis celui, plus permissif, applicable à l'expression universitaire, laquelle s'apparente aux finalités journalistiques au sens du droit de l'Union.

Plusieurs dispositions du RGPD visent à adapter le règlement aux besoins de la recherche scientifique. Pour commencer, l'article 5 (1) (b) indique que « les données à caractère personnel doivent être: [...] collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités; le traitement ultérieur à des fins [...] de recherche scientifique ou historique [...] n'est pas considéré, conformément à l'article 89, paragraphe 1, comme incompatible avec les finalités initiales (limitation des finalités). » Il est donc possible de réutiliser des données, lorsque c'est pour des finalités de recherche, moyennant le respect de règles énoncées à l'article 89 (1), qui exigent, en échange de ce droit de réutilisation, à ce que les données de recherche soient rendues aussi anonymes que possible eu égard aux objectifs et à la méthodologie de la recherche :

« Le traitement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique, ou à des fins statistiques est soumis, conformément au présent règlement, à des garanties appropriées pour les droits et libertés de la personne concernée. Ces garanties garantissent la mise en place de mesures techniques et organisationnelles, en particulier pour assurer le respect du principe de minimisation des données. Ces mesures peuvent comprendre la pseudonymisation, dans la mesure où ces finalités peuvent être atteintes de cette manière. Chaque fois que ces finalités peuvent être atteintes par un traitement ultérieur ne permettant pas ou plus l'identification des personnes concernées, il convient de procéder de cette manière. » (art. 89 (1) du RGPD)

Cette réutilisation doit trouver une base légale dans l'un des critères de licéité de l'article 6 (1) du RGPD. Contrairement à une idée très répandue, le RGPD n'exige pas dans tous les cas le consentement de la personne concernée. Il est ainsi possible de mener des recherches scientifiques lorsque cela entre dans une mission de service public (art. 6 (1) (e) du RGPD) ou lorsque cela répond à un intérêt légitime du responsable du traitement ou de tiers (art. 6 (1) (f) du RGPD), le RGPD semblant reconnaître dans ses considérants que la recherche scientifique est un intérêt légitime⁵⁸. Dans les circonstances où une recherche ne se fonde pas sur le consentement des personnes concernées au sens du RGPD, il est tout

57 Traduit de l'anglais : « contain a special regime affording a degree of flexibility for genuine research projects that operate within an ethical framework and aim to grow society's collective knowledge and wellbeing. »

58 Ainsi, le considérant 113 évoque « les attentes légitimes de la société en matière de progrès des connaissances. » Le considérant 157 souligne également l'importance de la recherche, dans différents domaines, dont les SHS.

de même possible de demander un accord qui vaut comme une garantie éthique supplémentaire (EDPS, 2020). Le consentement au titre du RGPD est en effet « libre, éclairé, spécifique, univoque » et retirable à tout moment. Le « consentement » tacite, incarnant l'adage « qui ne dit mot consent », n'est pas admis. Il n'est pas non plus valable lorsqu'il existe un déséquilibre dans la relation de pouvoir, et il existe des règles spécifiques relatives au consentement des personnes mineures. Il n'est donc pas certain, par exemple, qu'un accord d'un·e étudiant·e à participer à une recherche menée par l'un·e de ses enseignant·e·s soit valable au titre du RGPD, mais une telle recherche peut malgré tout être envisagée sur un autre fondement (mission d'intérêt public ou bien intérêt légitime) tout en mettant en œuvre une procédure pour vérifier, en complément, l'accord des enquêté·e·s.

Un effort maximal de pseudonymisation et d'anonymisation est demandé en échange de la souplesse accordée aux traitements de données personnelles à des fins de recherche scientifique.

L'anonymisation ne doit pas être confondue avec la pseudonymisation. L'obtention d'un anonymat robuste, c'est-à-dire de l'assurance absolue qu'une donnée ne puisse plus être reliée à un individu, même indirectement, par croisement, par exemple, avec d'autres jeux de données à laquelle un·e adversaire potentiel (comme les services de renseignement d'un État répressif) pourrait accéder, est très difficile à prouver et, de façon générale, très rare (Ohm, 2010). Il est plus juste de parler de degré d'anonymat, qui peut (en partie) se mesurer de façon mathématique (Mascetti *et al*, 2013). C'est cette raison qui amène le RGPD à définir le « 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) du RGPD). Le Groupe de travail de l'article 29, qui a précédé au Comité européen de protection des données, a publié un guide en 2014 sur les techniques d'anonymisation (Groupe de travail de l'Article 29, 2014).

Le considérant 159 donne une forme de définition de ce qu'est la recherche scientifique qui aide à déterminer les cas dans lesquels il est possible de se prévaloir des exceptions prévues à ces fins par le RGPD :

« Aux fins du présent règlement, le traitement de données à caractère personnel à des fins de recherche scientifique devrait être interprété au sens large et couvrir, par exemple, le développement et la démonstration de technologies, la recherche fondamentale, la recherche appliquée et la recherche financée par le secteur privé »

L'avis de l'EDPS (2020) mentionne aussi la définition comprise dans la directive de 2019 sur le copyright. Cette définition est large, et peut inclure des organismes privés, mais précise qu'il ne devrait pas être « possible pour une entreprise exerçant une influence déterminante sur cet organisme de bénéficier d'un accès privilégié aux résultats produits par ces recherches scientifiques » (art. 2 (1) de la directive 2019/790). Mais cette précision de la directive sur le droit d'auteur ne s'applique pas à la question des traitements de données à caractère personnel, et le considérant 159 du RGPD prévoit bien la possibilité d'appliquer les assouplissements accordés aux traitements de données personnelles à des fins de recherche aux organismes privés ou bien aux organismes publics qui effectuent une activité de

recherche sous le contrôle de fonds privés. Notons que la distinction entre la notion de « traitements à des fins statistiques » et « traitements à des fins scientifiques » comporte une part de flou, discutée notamment dans les travaux de Rossana Ducato (2020), qui insiste sur le fait, qu'en pratique, le régime juridique applicable à la recherche scientifique s'applique pour tout traitement dont la finalité est la production de nouvelles connaissances, à condition qu'il n'y ait pas d'impact spécifique sur un individu en particulier.

Au total, un peu moins d'une vingtaine de considérants et d'articles du RGPD portent sur la recherche scientifique. Il s'agit des considérants 33, 50, 52, 53, 62, 65, 113, 156, 157, 159, 161 et 161 et des articles 5, 9, 14, 17, 21 et 89. Lus ensemble, ils offrent la possibilité de déroger, lorsque cela est *strictement nécessaire* à des fins de recherche dans l'intérêt public : au principe de limitation des finalités, à la limite de durée de conservation, à l'obligation d'informer les personnes concernées lorsque les données ne sont pas collectées directement auprès de la personne concernée (art. 14) voire y compris lorsqu'elles le sont directement (considérant 62), et, lorsque la recherche menée est nécessaire à l'exécution d'une mission d'intérêt public, à l'obligation de donner suite aux demandes d'effacement ou au droit d'opposition (art. 21 (6)).

Le considérant 113 facilite par ailleurs la coopération internationale en affirmant que : « Les transferts qui peuvent être qualifiés de non répétitifs et qui ne touchent qu'un nombre limité de personnes concernées pourraient également être autorisés aux fins des intérêts légitimes impérieux poursuivis par le responsable du traitement, lorsque ces intérêts prévalent sur les intérêts ou les libertés et droits fondamentaux de la personne concernée et lorsque le responsable du traitement a évalué toutes les circonstances entourant le transfert de données. [...] À des fins de recherche scientifique ou historique ou à des fins statistiques, il y a lieu de prendre en considération les attentes légitimes de la société en matière de progrès des connaissances. Le responsable du traitement devrait informer l'autorité de contrôle et la personne concernée du transfert. » Cette disposition ne donne toutefois pas de carte blanche à tout transfert de données personnelles en dehors de l'UE, et à ce titre, l'obligation de « consentir » au transfert de données personnelles vers la Chine lors d'un dépôt de projet ANR est une pratique questionnable :

Les données collectées sont indispensables à ce traitement et sont destinées aux services concernés de l'ANR, ainsi que, le cas échéant, à ses sous-traitants, prestataires, agences partenaires, experts et membres de comité d'évaluation. Ainsi cela peut impliquer un transfert de vos données à caractère personnel vers les pays suivants : Turquie, Israël, Egypte, Maroc, Afrique du Sud, Brésil, Inde, Canada, Chili, Chine, Corée du Sud, Etats-Unis, Japon, Mexique, Moldova, Norvège, Singapour, Suisse, Taïwan, Tunisie.

Capture d'écran réalisée le 30 novembre 2020 sur le site de l'ANR.

L'article 9 (1) (j) du RGPD permet de traiter des catégories sensibles de données à caractère personnel lorsque « le traitement est nécessaire à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques ». Le règlement précise que cette exception est valable « sur la base du droit de l'Union ou du droit d'un État membre qui doit être proportionné à l'objectif poursuivi, respecter l'essence du droit à la protection des données et prévoir des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée. » Comme le note l'EDPS (2020, p. 23), en l'absence de règle adoptée par l'UE ou

par un État membre, cette disposition ne semble pas opérationnelle pour se passer du consentement des personnes concernées.

Concrètement, dans ces situations, il n'est possible de traiter des données à caractère personnel à des fins de recherche scientifique seulement avec l'accord exprès de la personne concernée, lorsque la recherche a lieu au sein même d'« une fondation, une association ou tout autre organisme à but non lucratif et poursuivant une finalité politique, philosophique, religieuse ou syndicale » et porte sur des données de membres, en rapport avec la nature de l'organisme (art. 9 (2) (d) RGPD), lorsque les données, comme nous l'avons vu, ont été rendues publiques par la personne concernée elle-même (art. 9 (2) (e) RGPD), ou, lorsque le droit de l'Union ou d'un État membre l'autorise, « pour des motifs d'intérêt public dans le domaine de la santé publique, tels que la protection contre les menaces transfrontalières graves pesant sur la santé, ou aux fins de garantir des normes élevées de qualité et de sécurité des soins de santé et des médicaments ou des dispositifs médicaux. »

Se pose alors la question de savoir que faire des cas où les données de recherche contiennent des données dont le traitement fait apparaître des informations sensibles au titre de l'article 9 du RGPD (comme des opinions religieuses ou un état de santé), mais qu'il n'est pas possible de demander le consentement exprès de la personne. Par exemple, que faire s'il n'est pas possible d'informer pleinement sur les finalités de la recherche avant qu'elles n'y participent les personnes qui acceptent de participer à un questionnaire qui porterait par exemple sur des pratiques religieuses, car cela risquerait de biaiser les réponses ? Comment enquêter sur des groupes Facebook politiques fermés, dans la mesure où l'autorisation accordée par les modérateurs·trices du groupe ne vaut pas consentement exprès de tou·te·s les participant·e·s ?

La réponse à ces questions, en France, se trouve dans différents articles de la rédaction actuelle la loi Informatique et Libertés, notamment en ses articles 78 et 79, qui concernent également le champ des archives publiques.

En matière de traitement de données sensibles sans le consentement exprès de la personne concernée, la réponse a été apportée sous la forme d'obligation de demande d'autorisation à transmettre à la CNIL sur le fondement du sixièmement de l'article 44, sauf lorsque la recherche inclut des données sur la santé des enquêté·e·s et/ou lorsque la recherche a lieu « dans le domaine de la santé. » Dans ce cas, ce sont d'autres articles de la loi informatique et libertés qui s'appliquent. Les articles 72 à 77 de la loi Informatique et Libertés portent sur ce cas de figure. Les obligations varient alors selon les cas de figure (recherche impliquant la personne ou non, interventionnelle ou non, etc⁵⁹.) Il est utile de se renseigner au préalable, car certaines recherches en SHS impliquant des données de santé doivent faire l'objet de déclarations spécifiques à la CNIL et/ou au Health Data Hub et/ou au comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé, et/ou au comité de protection des personnes compétent.

Aucune de ces autorisations n'est cependant nécessaire pour traiter des données personnelles, y compris sensibles, y compris de santé, dès lors que la finalité n'est pas la recherche scientifique... mais « l'expression universitaire. » En effet, par l'inclusion de l'article 85 du RGPD, le législateur européen a tenu à ménager un équilibre entre d'une part le droit à la vie privée et à la protection des données à

59 L'article L1121-1 du Code la santé s'applique alors.

caractère personnel, et d'autre part le droit à la liberté d'expression, notamment académique. Ces dispositions sont précisés à l'article 80 de la loi informatique et libertés, le législateur européen ayant renvoyé au législateur le soin de donner corps à ces dérogations. En tout état de cause, il paraît peu conforme à l'esprit de l'article 85 que ce qui serait autorisé dans le cadre d'une publication journalistique soit interdit dans le cadre d'une publication universitaire. C'est d'ailleurs l'avis de l'Autorité de protection des données de Belgique⁶⁰.

Mais qu'est-ce qui distingue les finalités de recherche scientifique des finalités d'expression universitaire ? Cette question n'a que très peu été traitée. Même la décision de l'APD belge que nous venons de citer ne propose pas de critère strict de distinction entre les deux.

L'expression à l'oral, pendant un cours ou lors d'une conférence, semble évidemment couverte. Une publication dans une revue scientifique semble aussi remplir des finalités d'expression universitaire, en plus de pouvoir remplir des finalités journalistiques dans l'acception large qui découle du droit de l'Union⁶¹. Mais construire un cours, préparer une communication à un colloque ou encore publier dans une revue scientifique suppose d'avoir conduit une recherche scientifique au préalable.

Dans la mesure où le considérant 159 du RGPD définit la recherche scientifique de façon très large⁶², sans faire référence à la finalité de publication, nous pourrions considérer qu'une recherche universitaire publiée entre dans le champ de l'article 85 du règlement, là où une recherche n'aboutissant pas à une publication, mais cantonnée par exemple à de la recherche et développement interne à une entreprise, correspond au régime prévu à l'article 89. Cela signifierait qu'une recherche universitaire, à des fins d'expression universitaire, n'aurait pas besoin de recourir aux autorisations requises par exemple au 6° de l'article 44 de la loi informatique et libertés. Seuls demeurerait les contrôles *ex post*, mis en œuvre par des mécanismes de contrôle interne (sous la supervision du/de la délégué-e à la protection des données), l'autorité de contrôle (la CNIL) ou, en cas de litige, le juge. C'est d'ailleurs à ce dernier qu'il reviendra *in fine* de trancher sur la distinction entre les champs d'applications respectifs des articles 85 et 89 du RGPD, l'interprétation que nous proposons ici étant susceptibles de ne pas faire l'unanimité.

Malgré ces interrogations qui demeurent sur le mécanisme permettant de traiter des données à caractère personnel sensibles dans un projet de recherche en SHS, les nombreuses dérogations qui peuvent être accordées à la recherche scientifique, à condition que la nécessité de leur emploi soit démontrée (et documentée), complétées par l'article 85 de ce règlement, offrent une grande flexibilité. Cette flexibilité ne doit pas être confondue avec une absence de règles. La tenue d'un registre des traitements, ou encore celles concernant la sécurité des données, doivent être respectées en toute circonstance,

60 Voir la décision du 27 janvier 2022 de la chambre contentieuse de l'Autorité de protection des données de Belgique, décision quant au fond, n° 13/2022, DOS-2018-04433, dans l'affaire EU DisinfoLab.

61 L'Autorité de protection des données de Belgique a considéré qu'« exiger la démonstration d'une appartenance professionnelle au journalisme (ex. souscription d'un code de déontologie) ne permettrait pas de donner effet utile au droit à la liberté d'expression prévu à l'article 11 de la Charte dans le contexte où la législation belge applicable au moment des faits ne prévoyait pas encore la possibilité explicite d'invoquer une exception couvrant la liberté d'expression en dehors d'un contexte de journalisme, contrairement à l'article 85 du RGPD. » Ce même raisonnement s'applique en France. Voir notamment les arrêts Buivids et Satamedia de la CJUE (référéncés en bibliographie).

62 Selon celui-ci, la « recherche scientifique devrait être interprété au sens large et couvrir, par exemple, le développement et la démonstration de technologies, la recherche fondamentale, la recherche appliquée et la recherche financée par le secteur privé. »

même dans le cadre de la recherche scientifique et de l’expression universitaire. Il en va de même de l’obligation une évaluation de l’impact d’un traitement sur les droits et libertés des personnes concernées avant de le mettre en œuvre, lorsque celui-ci comporte un « risque élevé » (art. 35 du RGPD), et de consulter l’autorité de protection des données si cette évaluation ne permet pas de dégager de solution satisfaisante (art. 36 du RGPD).

Cette évaluation d’impact est déclenchée lorsque deux critères de risque parmi ceux présentés dans le tableau ci-dessous sont présents⁶³ :

Critères de risque du RGPD		
Évaluation systématique et approfondie d’aspects personnels (profilage)	Traitements à grande échelle de catégories particulières de données	La surveillance systématique à grande échelle d’une zone accessible au public
Critères de risque du Comité européen de protection des données		
Évaluation ou notation	Prise de décision automatisée provoquant un effet juridique ou similaire significatif	Surveillance systématique
Données de l’art. 9 RGPD	Données traitées « à grande échelle »	Combinaison d’ensembles de données
Données concernant des personnes vulnérables (cons. 75 RGPD)	Utilisation de nouvelles « solutions technologiques ou organisationnelles »	Traitements qui peuvent aboutir à exclusion des personnes d’un droit ou du bénéfice d’un contrat ou d’un service

L’utilisation de données transmises par des capteurs collecteurs disséminés dans une ville dans le cadre d’un projet de *smart city*, collectant des données à grand échelle sur la géolocalisation des passant·e·s dans une zone accessible au public, déclencherait l’obligation de mener une évaluation d’impact. Il en va de même lorsque des données sont par exemple collectées pour une étude sur les discours politiques d’une minorité religieuse faisant l’objet de menaces, c’est-à-dire dont les membres sont vulnérables. Beaucoup d’autres projets sont concernées en dehors de ces deux exemples proposés à titre d’exemples.

Ces évaluations d’impact se font préalablement à la mise en œuvre du traitement (donc de la phase de collecte de données) et peuvent donc utilement se faire au même moment que la rédaction des PGD. La Commission a publié un guide interactif utile pour commencer une réflexion sur la protection des données à caractère personnel dans le contexte de la rédaction d’un PGD⁶⁴.

Enfin, les dérogations rendues possibles par le RGPD à des fins de recherche ou d’expression universitaire ne peuvent être utilisées qu’en cas de nécessité, qu’il faut être en mesure de démontrer ou, *a minima*, d’argumenter en cas de contrôle ou de litige. Documenter la réflexion à ce sujet, dans un PGD et/ou une évaluation d’impacts, renforcera la sécurité juridique du projet de recherche.

63 Art. 35 (3) du RGPD et Lignes directrices sur les évaluations d’impact du Groupe de travail de l’Article 29, adoptées le 4 avril 2017.

64 Voir : <https://ec.europa.eu/assets/rtd/ethics-data-protection-decision-tree/index.html> . Voir également : https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_fr_en.pdf .

L'application de ce cadre spécifique aux recherches mobilisant des données issues de RSN

Les dérogations prévues à des fins de recherche scientifique et à des fins d'expression universitaire s'appliquent aux projets de recherche portant sur des données à caractère personnel issues de réseaux socio-numériques.

Les principales difficultés pour ce type de projet sont :

- lorsque le volume de données est massif, il peut être difficile de recueillir le consentement de chaque personne, d'autant qu'il n'est pas toujours possible de prendre contact avec l'auteur·e d'un message sur un réseau social (voir à ce sujet : Boyadjian, 2016) ;
- l'information préalable des personnes concernées peut, dans certains cas, porter atteinte à la viabilité de la méthodologie (en modifiant les pratiques des acteurs du terrain) voire mettre en danger les personnes qui mènent l'enquête (par exemple s'il s'agit de documenter la propagande de groupes terroristes en ligne) ;
- il est difficile, voire impossible, de garantir l'absence de collecte de données sensibles relevant de l'article 9 du RGPD (ex : données de santé, opinions politiques, orientation sexuelle...) et, si dans certains cas, les données ont été manifestement rendues publiques, y compris en tenant compte de l'ensemble des critères dégagés pour parvenir à cette conclusion par la CNIL en suivant l'avis de l'EDPB⁶⁵, il existe aussi des cas où :
 - les données sont recueillies depuis des groupes fermés (sur Facebook, Telegram, WhatsApp...) ou sur des réseaux dont les messages ne sont accessibles qu'aux utilisateurs et utilisatrices inscrit·es (c'est le cas de la plupart des applications de rencontre, par exemple) ;
 - il est dans la nature même du projet de recherche de générer des informations couvertes par l'article 9 du RGPD en raison de son objectif, par exemple dans le cas d'une étude sur la réception de messages politiques qui nécessiterait de catégoriser des messages de réseaux sociaux en fonction d'une opinion politique déduite du contenu des messages.

65 Voir : <https://www.cnil.fr/fr/communication-politique-queles-regles-pour-la-collecte-de-donnees-sur-les-reseaux-sociaux> (page consultée le 9 janvier 2023). Les critères à prendre en compte sont les suivants :

« - le paramétrage par défaut de la plateforme de médias sociaux (c'est-à-dire si la personne concernée a fait une action pour changer ses paramètres privés par défaut en paramètres publics) ; ou
- la nature de la plateforme de médias sociaux (c'est-à-dire si cette plateforme a vocation à connecter des connaissances proches avec la personne concernée ou à nouer des relations intimes, telles que les plateformes de rencontres en ligne), ou s'il est destiné à fournir un plus large éventail de relations interpersonnelles, telles que les relations professionnelles, ou le « microblogging », le partage de médias, les réseaux sociaux plateformes de partage d'avis en ligne, etc. ; ou
- l'accessibilité de la page où les données sensibles sont publiées (c'est-à-dire si les informations sont accessibles au public ou si, par exemple, la création d'un compte est nécessaire avant d'accéder aux informations) ; ou
- la visibilité de l'information sur le caractère public des contenus publiés par les personnes concernées (c'est-à-dire s'il y a par exemple une bannière continue sur la page, ou si le bouton de publication informe la personne concernée que les informations seront rendues publiques...) ; ou
- si la personne concernée a elle-même publié les données sensibles, ou si, à l'inverse, les données ont été publiées par un tiers (par exemple une photo publiée par un ami qui révèle des données sensibles) ou déduites.»

Les recherches portant sur des données à caractère personnel issues de réseaux sociaux ne sont pas pour autant interdites. Nous pouvons voir dans le texte du RGPD et dans son interprétation par les autorités de contrôle qu'il existe des solutions aux problèmes ci-dessus.

À défaut du consentement de la personne concernée, nous avons vu qu'un projet de recherche mené par une équipe de recherche publique pouvait se fonder sur sa mission de service public, ou, à défaut, s'il s'agit d'une équipe de droit privé, sur l'intérêt légitime de la recherche scientifique, évoqué dans plusieurs considérants du RGPD. Ainsi, comme l'a admis la chambre contentieuse de l'autorité belge de protection des données (l'APD) dans une décision du 27 janvier 2022 sur une recherche de l'association EU DisinfoLab sur l'analyse de la circulation de tweets en rapport avec l'affaire Benalla en France :

« Une démarche scientifique pourra quant à elle trouver justification dans l'intérêt légitime du responsable de traitement, ou trouvera son fondement dans un intérêt public si elle relève de l'exercice de l'autorité publique dont est investi le responsable de traitement de droit privé ou public, ou encore si elle s'opère dans l'exécution d'une mission d'intérêt public légalement reconnue » (APD belge, 2022, p. 24)

Ensuite, la CNIL a déjà admis dans une délibération de 2018 qu'un projet de recherche portant sur 20 000 profils Facebook pouvait réutiliser, à des fins de recherche scientifique, des données publiées par les détenteurs·trices de ces profils⁶⁶. Cela est rendu possible par les articles 5 (1) (b), 89 et le considérant 50 du RGPD, comme l'a admis l'APD belge (2022, p. 25).

La CNIL notait dans la suite de sa délibération de 2018 que :

« [...] il ne peut être exclu que les profils d'utilisateurs de réseaux sociaux concernés par la collecte du LORIA contiennent des informations concernant leurs opinions politiques ou religieuses, par exemple. Ces données sont nécessaires pour l'analyse devant être menée par le responsable de traitement puisqu'il s'agit de mettre en évidence ce que peuvent révéler sur la vie privée d'une personne, ses publications en ligne. » (CNIL, délibération du 3 mai 2018)

La CNIL a toutefois estimé sur le fondement du droit alors en vigueur, c'est-à-dire sur le fondement de la directive de 1995⁶⁷ et de la loi informatique et libertés, que le traitement pouvait être autorisé « compte tenu de l'intérêt public poursuivi par l'enquête. » Elle exigea des conditions renforcées de conservation sécurisée.

Elle autorisa également les porteurs·ses du projet à déroger à l'obligation d'information, en tenant compte notamment du fait « qu'une telle information risquerait de conduire les personnes concernées à modifier leur comportement en ligne biaisant de ce fait la recherche » et que « le responsable de traitement justifie de cet effort disproportionné » notamment par le nombre de personnes à contacter.

Elle exigea en revanche un effort d'information générale, en insérant sur le site du laboratoire les mentions obligatoires. Aujourd'hui, avec l'entrée en application du RGPD, une évaluation d'impact serait nécessaire en raison du nombre important (plusieurs milliers) de personnes concernées, et de la

66 Voir : CNIL, délibération n°2018-151 du 3 mai 2018.

67 Directive n° 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.

nature sensible des données collectées. L'inscription du traitement dans un registre des traitements tenu par l'établissement de recherche aurait aussi été obligatoire, de même que la conclusion de contrats avec les sous-traitants qui respectent les dispositions de l'article 28 du règlement. Ces trois obligations ont été rappelées en 2022 par la chambre contentieuse de l'APD belge.

Cette dernière a établi, dans la même décision, et à l'instar de ce que la CNIL avait admis en 2018, qu'un traitement à des fins scientifiques, dès lors que cette finalité est bien établie, peut permettre de déroger à l'obligation d'information des personnes concernées, si cela est « susceptible de rendre impossible ou de compromettre gravement la réalisation des objectifs scientifiques du traitement » (APD belge, 2022, p. 28). Elle a toutefois précisé que les interviews par voie de presse ne suffisent pas à permettre une information adéquate. L'APD a relevé qu'il aurait été nécessaire, au minimum, de mettre à la disposition du public des mentions d'information sur le site de l'association EU DisinfoLab (APD belge, 2022, p. 29).

Ces deux décisions illustrent ainsi les deux principales conclusions que nous pouvons tirer de l'analyse du droit applicable au traitement de données personnelles issues de réseaux socio-numériques à des fins de recherche en SHS : l'existence de nombreuses dérogations mobilisables à condition de justifier et de documenter leur nécessité aux fins de la recherche envisagée, et le maintien d'un certain nombre de règles, en particulier en matière de sécurité, de confidentialité, d'obligation de moyens de viser une anonymisation maximale compte tenu des finalités et de la méthodologie requise, d'obligation de conduire des évaluations d'impact, et de réfléchir à l'information à fournir aux personnes concernées *a minima* sur le site de l'institution pour le compte de laquelle l'étude est menée.

Le droit d'accès à des données mobilisé dans la recherche scientifique

Il existe des règles qui encadrent la façon dont les chercheurs·ses peuvent utiliser les données de la recherche, et, le cas échéant, les partager, mais il existe aussi des règles qui contribuent à faciliter l'accès de ceux-ci à des données. C'est le cas des règles qui régissent l'accès aux fonds d'archives, mais pas uniquement.

Ainsi, depuis la loi du 17 juillet 1978⁶⁸, aujourd'hui codifiée au Code des relations entre le public et l'administration, les documents administratifs sont en principe librement communicables. La Commission d'accès aux documents administratifs est en charge de veiller au respect de cette obligation par l'administration.

La loi pour une République numérique du 7 octobre 2016 a introduit, par ses articles 17 à 24, une nouvelle catégorie de données en droit français : la donnée d'intérêt général. Cette catégorie est cependant très hétérogène. Toutes ces données d'intérêt général ne sont pas nécessairement accessibles au public. Le législateur français a en effet adopté une approche sectorielle, chaque type de données

68 Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal.

d'intérêt général faisant l'objet de ses propres règles. Par exemple, l'article 17 de cette loi a introduit des dispositions à l'ordonnance n° 2016-65 du 29 janvier 2016 relative aux contrats de concession, qui ont pour effet d'obliger les concessionnaires auquel la gestion d'un service public est déléguée de fournir certaines données à l'administration (mais pas nécessairement au public). L'article 20 organise, lui, la publication des décisions de justice et sa conciliation avec le droit à la vie privée des justiciables. Cette dernière revêt une plus grande utilité pour la recherche.

ETALAB, qui est un département de la direction interministérielle au numérique, publie un grand nombre de jeux de données qui peuvent être utiles à la recherche. Ses missions sont définies dans un décret de 2019⁶⁹. De nombreuses collectivités territoriales ont aussi développé leurs propres services de mise à disposition de données. C'est le cas par exemple de la ville de Paris au moyen du site opendata.paris.fr.

Plus récemment, le législateur européen a entendu permettre l'accès de chercheurs·ses à des données détenues par les très grandes plateformes numériques et les très grands moteurs de recherche. Cette catégorie concerne les plateformes en ligne et les moteurs de recherche qui disposent de plus de 45 millions d'utilisateurs mensuels sur le territoire de l'Union, en plus des plateformes que la Commission européenne aura spécifiquement désigné au terme d'une procédure spécifique.

Le considérant 96 du Règlement sur les services numériques⁷⁰ (*Digital Services Act* en anglais, ou DSA) affirme que « les enquêtes menées par des chercheurs sur l'évolution et la gravité des risques systémiques en ligne sont particulièrement importantes. » Dès lors, les chercheurs·ses peuvent demander à l'autorité compétente (en France, il s'agit de l'ARCOM) un agrément qui leur permet d'exiger l'accès à des données. Cette nouvelle possibilité vise précisément à faciliter les enquêtes scientifiques sur le fonctionnement des réseaux socio-numériques.

Aux termes de l'article 40 du DSA, cette demande d'agrément doit être « dûment motivée. » Le chercheur ou la chercheuse qui fait la demande doit démontrer qu'elle répond aux critères suivants :

- Être affilié·e à un organisme de recherche au sens de la directive 2019/790 sur le droit d'auteur, c'est-à-dire à tout organisme ayant « pour objectif premier de mener des recherches scientifiques », « à titre non lucratif ou en réinvestissant tous les bénéfices dans ses recherches scientifiques » ou « dans le cadre d'une mission d'intérêt public reconnue par un État membre », « de telle manière qu'il ne soit pas possible pour une entreprise exerçant une influence déterminante sur cet organisme de bénéficier d'un accès privilégié aux résultats produits par ces recherches scientifiques » ;
- Être indépendant·e de tout intérêt commercial ;
- Indiquer la source du financement des recherches ;
- Démontrer leur capacité à respecter les exigences de sécurité et confidentialité, en particulier lorsque les données auxquelles un accès est demandé sont à caractère personnel ;

69 Décret n° 2019-1088 du 25 octobre 2019 relatif au système d'information et de communication de l'Etat et à la direction interministérielle du numérique.

70 Règlement 2022/2065/UE du Parlement européen et du Conseil relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques).

- Démontrer que les accès demandés sont « nécessaires et proportionnés » (art. 40 (8) (e) du DSA) au regard des objectifs de leur recherche ;
- Démontrer que cette recherche a pour objectif de contribuer « à la détection, au recensement et à la compréhension des risques systémiques dans l'Union⁷¹ » (art. 40 (4) du DSA) et à « l'évaluation du caractère adéquat, de l'efficacité et des effets des mesures d'atténuation des risques » (art. 40 (4) ;
- S'engager à mettre le résultat de leurs travaux de recherches gratuitement à la disposition du public, sous réserve du respect des règles de protection des données à caractère personnel.

Nous le verrons un peu plus loin dans ce rapport, les obligations du type de celles contenues dans cette dernière condition revêtent bien souvent un caractère antithétique, et placent les chercheurs-ses dans une situation d'injonction contradictoire qui, parfois, engage leur responsabilité à titre individuel. Notons pour l'instant qu'une fois l'agrément accordé, le fournisseur de service de très grande plateforme en ligne ou de très grand moteur de recherche a l'obligation de fournir les données demandées. Il sera utile de suivre la façon dont ces dispositions vont être mises en œuvre en pratique, les conditions à remplir pouvant être interprétées de façon très restrictives.

Notons enfin l'existence d'initiatives qui visent à explorer l'usage du droit d'accès qu'un individu détient à ses données à caractère personnel dans le cadre de projets de recherche. Cette démarche se fonde sur l'article 15 du RGPD (Pop Stefanija et Pierson, 2020). Elle suppose pour être utile que les responsables du traitement auxquels les demandes sont adressées coopèrent et fournissent effectivement l'ensemble des données qu'ils ont collectées et générées. Comme ce n'est pas toujours le cas – dans des proportions qui mériteraient d'être documentées – cette méthode supposera certainement une action de la part de la autorités de contrôle pour inciter les responsables du traitement à se mettre en conformité avec leur obligation de fournir un droit d'accès effectif. Enfin, cette méthode suppose généralement, pour être efficace, la coopération entre plusieurs personnes concernées par un même traitement de données à caractère personnel. Cela est notamment nécessaire pour étudier le fonctionnement de mécanismes de profilage et de personnalisation des interfaces.

Des évolutions à surveiller

D'autres évolutions mériteront l'attention de futures recherches sur le droit applicable aux données de recherche en sciences humaines et sociales.

La réforme du droit de l'Union sur le droit d'auteur, intervenue en 2019⁷², a ainsi, comme nous l'avons vu, introduit dans le droit positif une autorisation de procéder à la fouille de données. Avant même cette

71 Ces risques sont ceux définis à l'article 34 du DSA. Il s'agit de la diffusion de contenus illicites, des effets négatifs sur l'exercice des droits fondamentaux, de « tout effet négatif réel ou prévisible sur le discours civique, les processus électoraux et la sécurité publique », et « tout effet négatif réel ou prévisible lié aux violences sexistes et à la protection de la santé publique et des mineurs et les conséquences négatives graves sur le bien-être physique et mental des personnes. »

72 Directive 2019/790/UE du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE.

réforme, la directive sur le droit d'auteur de 2001⁷³ avait permis à la Cour de cassation d'assouplir un peu les conditions dans lesquelles il était possible de citer des œuvres graphiques.

D'autres textes ont pour but d'imposer ou d'encourager l'ouverture des données, qu'il s'agisse ou non de données de la recherche. C'est le cas du règlement Horizon Europe de 2021, et de la directive de 2019 sur l'*open data*⁷⁴, qui affirme à son considérant 27 qu'« afin d'être en mesure de faire face efficacement et globalement à des défis de société qui vont croissant, il est devenu crucial et urgent de pouvoir accéder à des données de différentes sources, secteurs et disciplines, les combiner et les réutiliser. [...] » Cette directive impose par son article 10 aux États membres de prévoir des règles sur le libre-accès des données de la recherche selon les principes FAIR. Le Data Governance Act va plus loin, en incitant fortement les établissements publics à obtenir l'autorisation des personnes qui disposent de droits sur des données qu'elles détiennent (par exemple en vertu du droit à la vie privée ou d'un droit de la propriété intellectuelle) pour que celles-ci soient publiées en libre-accès (Baloup et al., 2021).

Une autre tendance est celle qui voit l'émergence de nouvelles institutions chargées de veiller au respect par la recherche scientifique de principes éthiques. Le Contrôleur européen de la protection des données encourage ainsi le dialogue entre autorités nationales de protection des données et comités d'éthique.

La dernière tendance, qui consiste à reconnaître sur le plan européen l'existence de libertés académiques inscrites dans la Charte des droits fondamentaux de l'UE, apporte (du moins en théorie) des garanties à la recherche scientifique en imposant aux États membres de prendre les mesures nécessaires pour garantir l'indépendance de la recherche. Dans son arrêt du 6 octobre 2020, la CJUE a rappelé aux États membres leur obligation de fournir aux « universitaires [...] l'infrastructure autonome, nécessaire à la conduite de leurs recherches scientifiques⁷⁵. » En l'espèce, l'affaire portait sur l'éviction d'une université par le gouvernement hongrois. Ceci dit, semble se dégager des principes dégagés dans cette décision d'espèce la possibilité de plaider que la superposition par le législateur d'un nombre croissant d'obligations procédurales qui s'imposent aux chercheurs·ses, sans que cette croissance ne soit accompagnée des moyens nécessaires pour permettre à la recherche de se faire de façon légale, constitue de fait une entrave aux libertés académiques garanties à l'article 13 de la Charte des droits fondamentaux de l'UE.

73 Directive 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information.

74 Directive 2019/1024/UE du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public.

75 CJUE 6 octobre 2020, Commission européenne contre Hongrie, Aff. C-66/18, pt. 228.

Éléments d'analyse

Quelques enseignements tirés de la lecture de quinze plans de gestion des données

Dans le cadre de cette étude, nous avons constitué un corpus de quinze plans de gestion de données (PGD) récupérés à partir de trois sites web :

- celui du portail OPIDoR du CNRS, à l'adresse : https://dmp.opidor.fr/public_plans,
- celui du Digital Curation Centre, à l'adresse : <https://dmponline.dcc.ac.uk/>,
- et de celui du logiciel DMPTool de l'Université de Californie : https://dmptool.org/public_plans.

Ce modeste corpus, dont la valeur est purement exploratoire, a eu pour objectif de fournir une matière empirique à l'étude de la façon dont l'opposition entre la volonté d'ouvrir les données de la recherche et la nécessité de protéger les données personnelles sont traitées dans les plans de gestion des données.

Les cinq PGD les plus récents ayant un lien avec au moins une discipline des SHS, rédigés soit en français, soit en anglais, soit en castillan, ont été sélectionnés. La liste détaillée est disponible en annexe 1. Le choix d'un corpus de taille limité a été motivé par des contraintes de temps, mais il serait pertinent de poursuivre l'étude sur un plus grand nombre de documents, tout en affinant l'analyse et les questions étudiées à partir de ceux-ci, par exemple en étudiant la variation du contenu en fonction des disciplines concernées.

Un tiers des documents recueillis provient de France. Il s'agit des plans de gestion des données mis en ligne sur OPIDoR. Deux PGD proviennent des États-Unis d'Amérique, deux autres du Royaume-Uni, et il y a également un PGD espagnol, un sud-africain, un brésilien, un néerlandais, un irlandais et un suédois. Tous ces documents ont été rédigés en 2021, l'année de la réalisation de l'étude.

Au moins 13 de ces 15 PGD portaient notamment sur des données à caractère personnel au sens du RGPD. Un PGD portait uniquement sur le code source d'un logiciel, et un autre sur un corpus de documents du Moyen-Âge. Le considérant 27 du RGPD précise que ce règlement ne s'applique pas aux données relatives à des personnes décédées, ce qui est le cas pour toutes les personnes qui pourraient être identifiées dans ce corpus. Enfin, dans un troisième cas, il n'a pas été possible de déterminer la présence ou non de données à caractère personnel, car le jeu de données n'était pas décrit et que, d'une façon générale, en dehors de quelques métadonnées identifiant le projet et d'un bref sommaire, le PGD était vide.

Sans grande surprise, nous avons pu noter une grande diversité dans les types de données à caractère personnel traitées dans le cadre de projets de recherche en SHS. Il peut s'agir en effet de notes de terrain, de photographies, d'enregistrements d'entretiens avec ou sans retranscriptions associées, d'archives épistolaires, d'archives sonores, de dossiers complétés par des travailleur·e·s sociaux·les, de réponses à des questionnaires en ligne, de messages et images échangés sur des réseaux sociaux et de

captures d'écran associés, de scores de sentiment associés à de tels messages et générés par des logiciels d'analyse de sentiment, de réponses à des questionnaires, de formulaires de consentement complétés (qui sont aussi des données à caractère personnel), d'analyses associées à des enquêtés, des dossiers de patient·e·s ou de personnes suivi·e·s par une administration publique, ou encore de données quantitatives associées à une personne, par exemple liées à son niveau de revenu. Ces données contiennent souvent des informations sensibles au sens de l'article 9 du RGPD, sur l'état de santé, sur les opinions politiques, ou encore sur la religion. Elles peuvent également porter, par exemple dans le cas d'une étude (PGD #10) sur un état de vulnérabilité socio-économique qui, sans entrer dans le cadre des dispositions de l'article 9 du RGPD, peuvent néanmoins revêtir une certaine sensibilité pour les personnes concernées.

Notons que les modèles de PGD semblent souvent peu adaptés pour traiter du sujet de la vie privée et de la protection des données à caractère personnel. Certains projets traitent des données à caractère personnel, y compris des données qui entrent dans l'une des catégories soumises à un degré renforcé de protection au titre de l'article 9 du RGPD, sans pour autant qu'il y ait de section spécifique sur la protection des données. C'est le cas par exemple du PGD #6. Cela n'entraîne pas nécessairement une absence totale de considération pour les questions relatives à la protection des données, mais celles-ci se retrouvent alors traitées de façon éparse et peu détaillée. D'autres modèles entretiennent une confusion entre l'éthique de la recherche et la protection des données. Cela aboutit à une situation dans laquelle, d'une façon générale, les PGD que nous avons lus ne contiennent pas le degré d'information requis pour fournir une information conforme aux articles 12 à 14 du RGPD (sur l'information à fournir aux personnes concernées) ou à entrer en vertu de l'article 30 de ce même règlement dans le registre des traitements de l'institution à laquelle les porteurs·ses du projet sont affilié·e·s. Un seul projet mentionne la réalisation d'une étude d'impact sur la vie privée (PGD #8), sans en détailler le contenu.

Certains projets remettent la question de la protection des données à plus tard. Ainsi, il est indiqué dans le PGD #2, rédigé par Juan Carlos Conde, que « si des problèmes sérieux de données potentiellement sensibles ou compromettantes, ou qui soulèvent des enjeux de type éthique, nous chercherons des conseils juridiques qualifiés⁷⁶. » Cette attitude peut s'expliquer par la nature du projet, dans lequel il est question de recueillir des archives épistolaires et de les structurer sous forme d'archives utilisables pour la recherche. Nous pouvons supposer que plupart des auteurs des lettres concernées sont décédés, et il n'est pas toujours possible de deviner à l'avance la nature (plus ou moins sensible) de données contenues dans des archives qui n'ont pas encore été ouvertes.

De nombreux PGD de projets impliquant le traitement des données à caractère personnel évoquent le recueil du consentement des enquêtés. Ce n'est cependant pas le cas de tous les projets. Les deux projets qui traitent des données issues de réseaux sociaux numériques (PGD #5 et #7) ne semblent pas recueillir de consentement des personnes dont ils traitent les messages. Qu'ils utilisent le consentement – le cas échéant comme garantie supplémentaire – ou non comme base légale en vertu de l'article 6 du RGPD ou d'autres lois applicables pour les PGD hors-UE, la base légale n'est pas précisée de façon explicite.

76 Traduit du castillan : « En caso de que se planteen serios problemas de datos potencialmente sensibles o comprometedores, o que planteen interrogantes de tipo ético, se buscará consejo legal cualificado. »

Un des projets de recherche étudiés propose de recueillir le consentement des « communautés » avec lesquelles il est en lien. Cela peut répondre à des principes éthiques portés par l'équipe de recherche, mais cela ne correspond pas à l'esprit du RGPD, qui ne prévoit qu'un consentement *individuel* de chaque personne concernée, sauf à ce que le traitement se fonde sur un autre critère de licéité de l'article 6.

L'anonymisation et/ou la pseudonymisation sont aussi très souvent évoquées comme garanties de sécurité. Certains projets donnent un niveau fourni de détails autour du degré d'anonymat auquel ils espèrent aboutir, tout en précisant les limites et en modulant l'accès futur aux données et à leur partage pour tenir compte des limites démontrées de l'anonymat (Ohm, 2010). C'est le cas du PGD #10, dont les auteurs, Per Liv et Maria Lindström, précisent :

« Même si tous les identifiants sont supprimés, nous jugeons qu'il y a un risque de menace potentielle à l'intégrité, et il n'est pas possible de protéger les identités des sujets et le type de données que nous allons collecter, en raison la zone relativement restreinte d'où nous sommes en train de recueillir des données. En conséquent, nous ne prévoyons pas de partager les données avec d'autres personnes que celles impliquées dans le groupe de recherche et les co-auteur·e·s associé·e·s⁷⁷. »

D'autres PGD proposent la mise en place d'une procédure de demande d'autorisation pour avoir accès aux données, même lorsqu'elles sont pseudonymisées. C'est le cas du PGD #6, dont les auteures, Elisabeth de Moor et Susan Branje écrivent :

« Si un·e chercheur·se (doctorant·e ou plus haut) souhaite accéder aux données [du projet] NIPPA pour une analyse scientifique, il ou elle doit soumettre un Formulaire de demande de données, incluant un bref projet (de recherche) et une liste de données requises, au responsable scientifique, qui prend alors une décision. Un modèle défini est utilisé à cet effet. Un principe important est que les données sont toujours émises pour une unique question de recherche ou pour un ensemble cohérent de questions de recherche qui sous soumises au consentement éclairé des participant·e·s. Après approbation par la responsable scientifique du projet, elle préparera le fichier de données⁷⁸. »

Dans plusieurs autres PGD, la confusion entretenue entre pseudonymat et anonymat amène leurs auteur·e·s à conclure que les données seront librement diffusées sur des plateformes de partage de données scientifiques car « anonymes », alors même que dans certains cas, la robustesse de cet anonymat n'est non seulement pas démontrée mais également peu probable étant donné la nature des données qu'il est prévu de partager. Cela peut entraîner des risques importants de ré-identification (cf. Barbaro et Zeller, 2006 et , Ohm, 2010) qui peuvent porter préjudice aux enquêté·e·s.

77 Traduit de l'anglais : « Even if all IDs are removed, we judge that there is a risk of potential threat to integrity, and not possible to protect the identities of subjects and type of data we will collect, due to the relatively restricted area from which we are collecting data. Therefore, we are not planning to share the data with anyone other than those involved in the research group and associated co-authors. »

78 Traduit de l'anglais : « If a researcher (PhD student or higher) wants access to NIPPA data for scientific analysis, they must submit a Data Request form, including a brief (research) proposal and a list of required data, to the principal investigator who then takes a decision. A fixed model is used for this. An important principle is that data is always issued for a single research question or set of related research questions that falls within the informed consent of participants. After approval of the principal investigator, she will prepare the data file. »

Environ la moitié des PGD du corpus (8 sur 15) comportent des paragraphes dédiés à la sécurité et aux garanties techniques de la confidentialité des données. Il s'agit généralement d'une protection par mot de passe, avec parfois une journalisation des accès voire, dans le cas du PGD #7, une alerte reçue par la coordinatrice. Le PGD #10 prévoit d'interdire les envois par voie électronique de données personnelles jugées sensibles, préférant passer par le courrier postal. La disponibilité étant souvent considérée comme un des critères de sécurité des systèmes d'information, ce même projet prévoit de garder des copies des données dans un lieu ignifugé. Souvent, la description des mesures de sécurité renvoie aux mesures proposées par le logiciel en Software-as-a-Service (service de cloud) utilisé dans le cadre du projet, tel que OneDrive, iCloud, Google Drive, ou dans un certain nombre de projets, des services propres à l'institution d'affiliation ou des infrastructures de recherche comme Huma-Num en France.

Enfin, en ce qui concerne le droit de la propriété intellectuelle, celui-ci est relativement peu abordé, même dans les projets dans lesquels il est prévu de capter des données issues de réseaux sociaux. Quelques PGD comportent des déclarations de principe, comme le PGD #2, qui promet de respecter « la législation sur le copyright et le droit de la propriété intellectuelle en vigueur tant en Espagne qu'au niveau international au moment de la réalisation du projet⁷⁹. » D'autres projets évoquent la propriété intellectuelle non pas pour parler de la façon dont sera obtenu – si nécessaire – la permission d'utiliser un bien protégé à des fins de recherche, mais pour décrire les conditions dans lesquelles les données de recherche seront partagées et pour donner des précisions sur la propriété intellectuelle des données générées dans le cadre du projet.

L'exploration rapide de ce petit corpus de quinze PGD ne permet pas de tirer de grandes conclusions universelles. Une première impression générale qui se dégage toutefois à leur lecture est que, dans leur ensemble, les modèles utilisés ne semblent pas favoriser un examen détaillé de la protection des données à caractère personnel, de la vie privée et, le cas échéant, des autres droits fondamentaux des personnes concernées par une enquête et par le partage des données de l'enquête. Les questions d'ordre éthique permettent de compléter des angles morts de la Loi. Par exemple, il peut arriver qu'une enquête sur un groupe social donné ait des répercussions sur l'ensemble du groupe bien que des individus qui en font partie donnent leur consentement à leur participation à titre individuel. Cela n'est pas pris en compte en tant que tel par un texte comme le RGPD. Mais le format actuel fait que les questions des droits des enquêtés ou des tiers tels que garantis par la Loi ne sont pas traités suffisamment, l'accent étant mis sur le format des données et sur la description de comment elles seront diffusées. Une suggestion d'évolution des pratiques, à droit constant, en ce qui concerne la protection des données, serait de combiner la réalisation des évaluations d'impact vie privée prévus à l'article 35 du RGPD (lorsqu'ils sont nécessaires) avec la réalisation des plans de gestion des données. L'essentiel des conclusions de ces évaluations d'impact, et l'ensemble des informations nécessaires à la fourniture des renseignements exigés aux articles 12 à 14 (transparence et information des personnes concernées) et 30 (registre des traitements) devrait figurer dans le PGD.

79 Traduit du castillan : « [...] la legislación sobre copyright y derechos de propiedad intelectual en vigor tanto en España como internacionalmente en el momento de realización del proyecto. »

Quelle responsabilité des chercheurs·ses dans la détermination de l'équilibre entre *open science* et protection de la vie privée ?

Nous venons de le voir : il peut exister une contradiction, en pratique, entre les politiques d'ouverture des données et le souci de garantir la confidentialité des données à caractère personnelle confiées par les enquêté·e·s. Or, le choix de partager ou non des données est un choix binaire. Il n'existe pas d'entre-deux prudent qui offrirait une forme de sécurité juridique aux chercheurs·ses confronté·e·s à cette injonction contradictoire. Il arrive ainsi que certain·e·s fassent l'objet de sanctions à titre personnel pour ne pas avoir trouvé le bon équilibre. Deux exemples illustrent cette situation inconfortable : une affaire qui concernait un professeur des universités suédois, Lars Christopher Gillberg, sanctionné pour avoir détruit les données d'enquêté·e·s à qui il avait promis la confidentialité, et celle d'un chercheur belge de l'Université catholique de Louvain, Nicolas Vanderbiest, sanctionné pour avoir obéi à une injonction de publication de données.

Lars Christopher Gillberg est un psychiatre suédois qui a mené, entre 1977 et 1992, des recherches à l'Université de Göteborg, sur le trouble déficitaire de l'attention avec hyperactivité. À cette occasion, le comité d'éthique de cette université aurait posé comme condition préalable au projet que les données sensibles des participant·es ne soient accessibles qu'à l'équipe du projet. En conséquence, il avait promis une confidentialité absolue. En 2002, Eva Kärfve, une sociologue d'une autre université et un pédiatre, Leif Elinder, critiques des travaux de Lars Christopher Gillberg, demandèrent d'avoir accès à ces données, et ont obtenu d'un tribunal administratif suédois un jugement obligeant l'université de Göteborg à leur accorder cet accès. Cet accès fut accordé sur le fondement du droit constitutionnel consacré en droit suédois d'accès aux documents administratifs. Le professeur Gillberg refusa toutefois d'obtempérer, et, avec des collègues, détruisit les données. Il fut condamné pour abus de fonction à une amende équivalente à un montant d'environ 4000 euros, malgré ses arguments en défense fondés sur l'affirmation d'un principe éthique de confidentialité des données collectées, sur sa parole donnée aux enquêtés, et sur des principes d'éthique médicale⁸⁰. La CEDH n'a pas jugé qu'il y ait eu violation des droits à la liberté d'expression ou à la vie privée contenus aux articles 10 et 8 de la Convention EDH. Concernant ce dernier droit, la Cour a considéré qu'il n'y avait pas eu d'atteinte au droit à la vie privée du professeur Gillbert⁸¹. Le jugement aurait peut-être été différent si c'était une personne sur laquelle portent les données visées qui avait fait un recours au motif de l'atteinte à sa vie privée. Quoi qu'il en soit, cette affaire illustre le risque pris personnellement par des chercheurs·ses face à l'injonction contradictoire entre l'ouverture des données et la protection de leur confidentialité, et l'opposition qu'il peut y avoir entre, d'une part, l'éthique (du chercheur sanctionné) et la loi (obligeant à accorder un accès), et deux éthiques opposées, celles du professeur Gillbert (brandissant l'étendard de la

80 La description des faits dans cette affaire a été reconstituée à partir des éléments disponibles dans la décision de la CEDH, et dans la biographie de Lars Christopher Gillbert sur le site Wikipedia : https://en.wikipedia.org/wiki/Christopher_Gillberg#Legal_case (page consultée le 10 janvier 2023). Voir : CEDH 3 avril 2012, Gillberg contre Suède, requête 41723/06.

81 CEDH 3 avril 2012, Gillberg contre Suède, requête 41723/06.

confidentialité médicale), et celles de ses critiques (favorables à une politique de transparence des données au nom de la rigueur scientifique).

Nicolas Vanderbiest est un chercheur en sciences de l'information et de la communication, qui, à l'époque où il était chercheur doctorant à l'Université Catholique de Louvain, a mené pour l'association EU DisinfoLab une étude sur la circulation de tweets au sujet de l'affaire Benalla. En 2022, la Chambre contentieuse de l'Autorité de protection des données (APD) de Belgique a condamné cette association à 2700 € et Nicolas Vanderbiest à 1200 € d'amende pour plusieurs infractions, parmi lesquels la publication des jeux de données collectées. Ces dernières contenaient notamment le nom d'utilisateur des comptes étudiés, et une indication sur les opinions politiques associés à chaque compte du corpus⁸².

L'APD belge met notamment en doute l'existence d'un intérêt légitime de *publier* les données brutes de la recherche, au lieu d'accorder un accès limité à quelques personnes agréées au cas-par-cas aux seules fins nécessaires de vérification du sérieux de la démarche de recherche (APD belge, 2022, p. 39 et 40). Elle critique également l'absence d'effort d'anonymisation ou *a minima* de pseudonymisation des données de l'étude mise en ligne.

Deux éléments sont à retenir de cette décision.

Le premier est que l'APD a retenu une part de responsabilité individuelle à Nicolas Vanderbiest, comme responsable du traitement *conjointement* à l'association pour le compte de laquelle il a conduit son étude. L'APD distingue deux traitements, avec la répartition des responsabilités suivante :

- la réalisation de l'étude, pour laquelle EU DisinfoLab et M. Vanderbiest étaient responsables conjointement, dans la mesure où ce dernier « a volontairement rédigé l'Étude et a décidé de recourir au logiciel Visibrain pour traiter les données personnelles » (APD belge, 2022, p. 16)
- la publication des fichiers de données utilisées pour l'étude :
 - certains fichiers ont été publiés par l'association, et elle seule est responsable de cette publication ;
 - certains autres fichiers ont été mis en ligne par M. Vanderbiest (à qui la réalisation de l'étude a été confiée par l'association), de son propre chef, et il est responsable de ce traitement.

Fidèle à l'adage *dura lex, sed lex*, la Chambre contentieuse de l'APD belge retient la responsabilité de Nicolas Vanderbiest malgré la bonne volonté qu'elle lui attribue :

« Il apparaît que M. Vanderbiest n'était pas animé d'une mauvaise intention eu égard à la finalité qu'il poursuivait, notamment, contribuer à un débat public ; il a agi dans un contexte difficile (contestation de l'Étude qu'il a réalisée et de son intégrité scientifique), et il a rapidement mis fin à sa propre mise à disposition des fichiers de données litigieux via un lien Dropbox de son entreprise SaperVedere. Les faits précités n'en portent pas moins atteinte au RGPD » (APD belge, 2022, page 48)

82 Voir : Chambre contentieuse de l'Autorité de protection des données de Belgique, décision quant au fond 13/2022 du 27 janvier 2022, DOS-2018-04433, « EU DisinfoLab ».

Enfin, elle prend note de la circonstance atténuante selon laquelle « la suggestion de publier les données brutes en ligne lui ait été faite par une personne liée à l'institution scientifique à laquelle il était alors attaché, ce qui peut tendre à démontrer qu'il a procédé à la diffusion de données personnelles concernées en lien avec des catégorisations politiques alléguées, sans conscience d'enfreindre des principes du droit fondamental à la protection des données personnelles » (APD belge, 2022, p. 51).

Le fait que cette « personne liée à l'institution scientifique » ait donné un conseil en faveur d'une ouverture de données qui n'auraient pas dû faire l'objet d'une publication montre comment une mauvaise prise en compte du droit des données personnelles dans les politiques d'ouverture des données de la recherche peut conduire à des erreurs engageant la responsabilité personnelle des chercheurs·ses.

À l'inverse, dans l'affaire impliquant le professeur Gillbert, ce dernier et plusieurs de ses collègues ont fait l'objet de sanctions personnelles pour ne pas avoir suffisamment ouvert l'accès à des données.

Rappelons qu'à ces sanctions administratives peuvent s'ajouter des sanctions de nature pénale, prévues aux articles 226-16 et suivants du Code pénal.

Dans les deux affaires, le juge suédois comme l'autorité belge de protection des données sont d'accord pour estimer que la loi peut, dans le premier cas, imposer, dans le second, autoriser le partage limité au strict nécessaire des données à des personnes agréées qui en font la demande et démontrent un intérêt légitime à obtenir cet accès. Ce qui ouvre la question de la nature de cet intérêt légitime, et des conditions dans lesquelles il faut informer les personnes concernées qu'un tel accès a lieu, et des modalités dans lesquelles celles-ci peuvent alors s'opposer sur le fondement de l'article 21 (6) du RGPD.

Enfin, ces réflexions sur la responsabilité individuelle des chercheurs·ses amènent deux questions de nature plus politique. Dans la mesure où il existe un principe d'autonomie des (enseignant·es-)chercheurs·ses qui peuvent librement déterminer les recherches qu'ils et elles mènent, dans quelle mesure devons-nous accepter une part de responsabilité individuelle en tant que responsables (conjoint) du traitement, et dans quelle mesure cette responsabilité doit-elle être assumée par l'employeur ? Dans l'hypothèse où l'institution devient seule responsable du traitement – ce qui est plus protecteur – dans quelle mesure cela conduit-il à instaurer une forme de contrôle de l'autorité hiérarchique *a priori* sur les projets de recherche ? Le droit n'offre pas de réponse à ces questions. En revanche, quel que soit le choix retenu, il paraît difficilement acceptable que des collègues soient sanctionné·e·s individuellement pour s'être soumis·es à des injonctions (surtout si elles sont contradictoires) venant de cette même autorité hiérarchique.

Conclusion

Une sécurité juridique encore insuffisante pour assurer un plein exercice des libertés académiques que le droit cherche pourtant bel et bien à préserver

En l'état actuel du droit, plusieurs constats semblent s'imposer.

Le premier est que la recherche scientifique est considérée par le droit de l'Union comme étant d'intérêt public, et que la liberté d'expression tout comme l'indépendance de la recherche scientifique sont des droits fondamentaux. Ces éléments doivent être pris en compte dans l'équilibre à trouver avec les droits tout aussi fondamentaux des personnes qui font l'objet d'une étude scientifique et avec les droits des tiers. Plusieurs dispositions du RGPD illustrent dans cette recherche d'équilibre la volonté de préservation un fort espace de liberté pour la recherche scientifique et l'expression universitaire. Des évolutions introduites dans la directive de 2019 sur le droit d'auteur, et dans le règlement sur les services numériques de 2022, créant une nouvelle exception au droit d'auteur pour la fouille de données et ouvrant la possibilité d'un accès aux données des très grandes plateformes et moteurs de recherche en ligne, facilitent le travail d'enquête en sciences sociales sur les réseaux socio-numériques en particulier. Le droit préserve ainsi un large espace de liberté pour la recherche scientifique, y compris en SHS, et celui-ci a tendance à être consolider. Ce premier constat est nuancé par les cinq autres qui suivent.

Le second constat est en effet qu'il existe un nombre croissants d'obligations formalisées dans des règles de droit qui encadrent la recherche, auxquelles une quantité croissante de temps doit être consacrée alors même que les moyens matériels et humains sont très fortement contraints. Les chercheurs·ses doivent ainsi assurer un niveau adéquat de sécurité informatique aux données à caractère personnel traitées, et rédiger des plans de gestion des données et des évaluations d'impacts relatives à la protection des données. Les objectifs poursuivis par ces obligations sont parfois contradictoires. L'incitation à l'ouverture des données se heurte ainsi à la difficulté pratique de garantir l'anonymat des données et donc la confidentialité de ce que les enquêté·e·s confient à la recherche. Elle se heurte également à un droit de la propriété intellectuelle qui, au moins en France, restreint fortement la possibilité non seulement de diffuser, mais même d'utiliser des œuvres protégées à des fins de recherche scientifique. Si ces différentes règles visent à atteindre un certain nombre d'objectifs éthiques et à garantir des droits fondamentaux comme celui à la propriété privée ou celui à la vie privée, la détermination d'un équilibre clair n'est pas encore acquis. La rédaction simultanée des plans de gestion des données et des évaluations d'impact sur la protection des données personnelles, lorsque ces documents sont exigés, est une piste pour faciliter l'application cohérente de ces règles et la résolution des contradictions grâce à une analyse juridique fine réalisée en amont et *in concreto*. Le développement de cette habitude ne réduira toutefois pas totalement la situation d'insécurité juridique

dans laquelle les chercheurs·ses se trouvent, et qui ont été illustrées par les affaires Gillberg contre Suède et EU DisinfoLab discutées dans ce rapport.

L'équilibre entre le droit de la propriété intellectuelle et la liberté d'expression n'est toujours pas satisfaisant, malgré quelques avancées récentes. Le droit de courte citation est insatisfaisant pour ce qui concerne la citation d'œuvres graphiques ou audiovisuelles. L'exception pédagogique et de recherche ne devrait ainsi pas être laissée par le législateur à la discrétion d'un accord conclu avec des personnes privées, comme c'est le cas actuellement, mais devrait être prévue et encadrée par la loi. L'article 5 (3) sous a) de la directive 2001/29/CE pourrait ainsi acquérir son plein effet en droit national.

En matière de protection de la vie privée et des données à caractère personnel, il existe un flou important sur la distinction opérée par le RGPD entre les traitements réalisés à des fins de « recherche scientifique » ou « historique », et ceux qui sont réalisés à des fins « d'expression universitaire. » Ces derniers sont rattachés au régime applicable aux activités de journalisme et à la création artistique. Il permet de traiter des données, même sensibles, sans autorisation préalable de la CNIL, alors que cette dernière est requise par le 6° de l'article 44 pour la recherche scientifique. Cette dernière procédure étant peu connue, peu pratiquée et mal documentée par la CNIL, dont la capacité à se prononcer sur la pertinence de méthodologies scientifiques est discutable, ce flou entretient une grande insécurité juridique pour de nombreux projets de recherche en SHS.

Or, au-delà du formalisme imposé par de multiples demandes d'autorisations qui peuvent significativement freiner le développement de projets de recherche, il ne semble pas que les moyens à disposition de la recherche en SHS pour respecter l'ensemble des obligations, notamment en matière d'évaluations d'impacts pour les projets à hauts risques, ou en matière de sécurité, soient suffisants. Nombreuses sont les universités à faire reposer leurs systèmes d'information sur des solutions d'infonuagique (*cloud computing*) proposées par des grandes entreprises américaines, sans chiffrement de bout-en-bout des données. Dans un contexte de précarisation de la recherche, les collègues sont amené·e·s à souvent changer d'employeur, voire à alterner des périodes d'emploi avec des périodes de chômage, pendant lesquelles leurs accès aux systèmes d'information universitaires sont régulièrement coupés, et les données perdues. Ceci conduit à prendre l'habitude d'échanger via des outils à l'accès pérenne, comme les services proposés par les « GAFAM », au lieu de pouvoir se reposer sur des systèmes universitaires sécurisés, conformes à la loi, dont l'offre est parfois lacunaire. Ces contraintes pratiques ne font que renforcer un constat d'injonctions paradoxales, qui génèrent là encore une forme d'insécurité juridique. Celle-ci ne se résoudra pas en abaissant le niveau d'exigence en matière de droit à la vie privée et à la protection des données des enquêté·e·s. Notre propos ne vise surtout pas à plaider pour une remise en cause de ces droits fondamentaux. Au contraire, il nous semble nécessaire de poursuivre la réflexion sur le renforcement du niveau de protection des données à caractère personnel dans les projets de recherche, en approfondissant les connaissances empiriques et en sortant d'une approche réduite à un commentaire du droit positif certes utile, mais insuffisant. Le droit ne peut apporter, à lui seul, de solution pleinement satisfaisante.

Enfin, concernant le secret des sources, il serait pertinent d'entamer une réflexion sur le rapprochement entre le statut des journalistes et ceux des chercheurs·ses, prenant acte du rôle joué par les un·e·s comme par les autres dans le débat public dans une société démocratique. Il paraîtrait alors pertinent de

sortir d'une conception organique du journalisme, préférée en droit national, au bénéfice d'une conception matérielle plus conforme au droit de l'Union.

Pistes de recherche

Outre la conduite d'un travail minutieux de documentation des obstacles concrets à la conduite de projets de recherche scientifique, l'étude présentée dans le présent rapport pourrait être utilement complétée par d'autres travaux de recherche juridique de terrain, à la dimension empirique encore étoffée.

Une première piste concerne les PGD, et l'articulation des obligations de fourniture de données de la recherche en libre-accès avec le droit à la vie privée et à la protection des données à caractère personnel des enquêté·e·s d'une part, et le droit de la propriété intellectuelle d'autre part. Il pourrait être utile d'étudier les jeux de données de recherche mis en ligne sous la promesse de l'anonymat des enquêté·e·s pour étudier le degré de robustesse de cet anonymat, en partant des plans de gestion publiés des jeux de données publics associés. Un tel projet impliquerait une interdisciplinarité entre droit et informatique. Une autre piste consiste à développer une méthodologie commune pour la réalisation des plans de gestion des données et des évaluations d'impacts sur la protection des données, dans le cadre de projets de recherche scientifique. Le suivi de la rédaction de plans de gestion des données et d'évaluations d'impacts, sous forme d'observation participante complétée par des entretiens, aiderait à identifier les difficultés pratiques auxquelles s'atteler de façon institutionnelle.

Il peut également être utile de réaliser une étude comparée sur le droit applicable à la recherche en SHS dans les pays où s'applique le RGPD, voire au-delà, dans d'autres pays qui disposent de lois de protection des données à caractère personnel. Il en va de même pour l'étude comparée de la mise en œuvre de l'article 5 (3) (a) de la directive de 2001 sur le droit d'auteur. Cela pourrait contribuer à l'étude de la conciliation entre droit de la propriété intellectuelle et liberté d'expression et liberté académique et des sciences dans les différents pays de l'Union européenne, en cherchant si des solutions pratiques ont pu être développées pour répondre à la tension entre la jurisprudence de la CJUE et celle de la CEDH, mises en évidence par Vincent Varet (2020).

Ce rapport ne contient que peu de références à des contentieux et décisions de justice ayant opposé des organismes de recherche à des enquêté·e·s ou à des tiers. Il pourrait être intéressant de mener une étude sur un corpus de décisions plus étendu, et d'interroger les services juridiques de plusieurs universités sur la réalité de ces litiges sur le terrain, incluant par-là ceux des contentieux qui trouvent une résolution avant d'être portés devant le tribunal. Nos quelques demandes et discussions informelles avec des interlocuteurs·trices au sein de Sorbonne Université semblent toutefois indiquer que les contentieux dans ce domaine sont rares.

Une autre manière de compléter par des observations empiriques la réflexion sur les règles de droit qui encadrent la recherche en SHS serait de partir d'une documentation fournie des pratiques du droit et de l'éthique observées sur le terrain. Les travaux présentés dans l'ouvrage de Sylvain Laurens et Frédéric Neyrat (2010) montrent bien qu'il existe une longue histoire de pratiques éthiques antérieures ou parallèles au développement des règles de droit qui encadrent la recherche. Elles pourraient utilement

compléter par des exemples concrets des guides et formations comprenant souvent des recommandations certes utiles, mais parfois abstraites, comme peuvent l'être aussi les engagements pris dans les quelques PGD que nous avons parcourus. Une telle enquête permettrait aussi d'identifier de façon plus fine les manques et les besoins, par exemple en matière d'outils permettant une sécurisation des données personnelles des enquêté·e·s.

Enfin, il est à noter que la multiplication des règles applicables, et des comités chargés de veiller à l'éthique et/ou à la conformité des projets de recherche multiplie les lieux de pouvoir. Aux décisions des financeurs s'ajoutent ainsi les décisions de celles et ceux qui sont chargé·e·s de veiller au respect d'un corpus de règles croissant. Les auteurs et titulaires des droits d'auteurs peuvent empêcher certains travaux et publications au nom du droit à la propriété privée. Selon la composition des comités d'éthique, certains travaux de recherche pourront être considérés comme irrecevables sur le fondement de motifs politiques ; car là où il n'existe qu'une seule loi, dont l'interprétation est (idéalement) unifiée par la jurisprudence, il existe potentiellement autant d'éthiques que de convictions individuelles.

L'affaire Cambridge Analytica est emblématique des cas d'infractions délibérées au droit, et notamment au droit à la protection des données à caractère personnel, et elle démontre ainsi la nécessité de disposer de mécanismes qui assurent la conformité des projets de recherche, y compris en SHS. Ces mécanismes peuvent aussi être utiles pour éviter, grâce à une prise en compte précoce des impératifs juridiques et éthiques en amont de la recherche, des incidents tels qu'un défaut de sécurité, qui peut entraîner des conséquences préjudiciables pour les personnes qui ont confié à la recherche des informations, potentiellement sensibles. Des cas d'abus de pouvoir ne peuvent toutefois être exclus, surtout si le nombre d'autorisations à obtenir se multiplie. L'exclure serait de façon absolue et catégorique serait selon nous naïf. Une étude sur un corpus large d'avis de différents comités serait donc utile pour vérifier la nature des critères pouvant donner lieu à l'éventuelle censure de projets. Une telle étude pourra le cas échéant suggérer des mécanismes pour assurer la transparence et l'équité des décisions des instances chargées de donner des avis sur le volet éthique des projets de recherche.

Ces quelques pistes ne sont pas une liste exhaustive. Comme les autres conclusions de ce rapport, elles sont ouvertes à la discussion. Le sujet, qui revêt une grande importance pour la recherche en SHS – que celle-ci porte sur les réseaux socio-numériques ou non – est loin d'être épuisé.

Bibliographie

Références académiques

- ADER Basile. (2005) « L'évolution de la notion d'originalité dans la jurisprudence », *Legicom*, n° 34, p. 43-49.
- BALOUP Julie, BAYAMLIOĞLU Emre, BENMAYOR Alike, DUCUING Charlotte, DUTKIEWICZ Lidia, LALOVA Teodora, MIADZVETSKAYA Yuliya, PEETERS Bert (2021), « White Paper on the Data Governance Act », CiTiP Working Paper 2021, <http://dx.doi.org/10.2139/ssrn.3872703>
- BELLIVIER Florence et NOIVILLE Christine. (2016) « L'adieu à l'article 1128 du Code civil : l'ordre public suffit-il à protéger le corps humain ? », *Revue des contrats*, 2016/3, p. 505 et s.
- BILLIER Jean-Cassien. (2014) *Introduction à l'éthique*, collection Quadrige, Paris : Presses Universitaires de France.
- BOTTINI Thomas et JULLIARD Virginie. (2017) « Entre informatique et sémiotique. Les conditions technométhodologiques d'une analyse de controverse sur Twitter », *Réseaux*, n° 204, p. 35-69.
- BOYADJIAN Julien. (2016) « La science politique face aux enjeux du « big data » et de la protection des données personnelles sur internet », *Revue du droit public*, n° 1, p. 7 et s.
- CARBONNIER Jean. (2001) *Flexible droit*. Dixième édition. Paris : LGDJ.
- CLÉMENT-FONTAINE Mélanie. (2017) « L'union du droit à la protection des données à caractère personnel et du droit à la vie privée », *Legicom*, vol. 59, n° 2, p. 61-68.
- DUCATO Rossana. (2020) « Data protection, scientific research, and the role of information », *Computer Law and Security Review*, vol. 37, 105412
- ETZIONI Amitai. (1999) *The Limits of Privacy*, New York : Basic Books.
- FUCHS Christian. (2011) « Towards an alternative concept of privacy », *Journal of Information, Communication and Ethics in Society*. vol. 9, n° 4, p. 220-237.
- GONZÁLEZ FUSTER Gloria. (2014) « Fighting For Your Right to What Exactly? The Convolved Case Law of the EU Court of Justice on Privacy and/or Personal Data Protection », *Birkbeck Law Review*, vol. 2, n° 2, p. 263-278.
- KOELSCH Lori. (2016) « The Use of I Poems to Better Understand Complex Subjectivities », in : GALVIN Kathleen et PRENDERGAST Monica (dir.), *Poetic Inquiry II – Seeing, Caring, Understanding. Using Poetry as and for Inquiry*, Rotterda :, Sense Publishers, p. 169 - 179
- LAURENS Sylvain et NEYRAT Frédéric. (2010) *Enquêter : de quel droit ? Menaces sur l'enquête en sciences sociales*. Bellecombe-en-Bauges : Éditions du Croquant.
- MACKINNON Catharine. (1989) *Towards a Feminist Theory of the State*. Cambridge, Mass. : Harvard University Press.

MASCETTI Sergio, MONREALE Anna, RICCI Annarita et GERINO Andrea. (2013) « Anonymity : A Comparison Between the Legal and Computer Science Perspectives », in : GUTWIRTH Serge, LEENES Ronald, DE HERT Paul et POULLET Yves (éds.), *European Data Protection: Coming of Age*, Dordrecht : Springer, p. 85-115.

MATTATIA Fabrice. (2021) *RGPD et droit des données personnelles*. Paris : Eyrolles

OHM Paul. (2010) « Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization », *UCLA Law Review*, vol. 57, p. 1701-1777.

POP STEFANIJA Ana et PIERSON Jo. (2020) « Practical AI Transparency —Revealing Datafication and Algorithmic Identities ». *Journal of Digital Social Research* vol. 2, n° 3, p. 84-125.

REVILLARD Anne. (2018) « Saisir les conséquences d'une politique à partir de ses ressortissants. La réception de l'action publique », *Revue française de science politique*, vol. 68, n° 3, p. 469-491.

RICŒUR Paul. (1989) « L'éthique, la morale et la règle », *Autres Temps. Les cahiers du christianisme social*, n° 24, p. 52 – 59.

RICŒUR Paul. (1990) *Soi-même comme un autre*. Paris : Seuil.

ROSSI Julien et BIGOT Jean-Édouard. (2018), « Traces numériques et recherche scientifique au prisme du droit des données personnelles », *Les Enjeux de l'information et de la communication*, vol. 19, n° 2, p. 161-177.

SAUVÉ Jean-Mars. (2018) « Le principe de proportionnalité, protecteur des libertés ? », *Les Cahiers Portalis*, n° 5, p. 9-21.

SWEENEY Latanya. (2000) « Uniquement of Simple Demographics in the U.S. Population », Laboratory for International Data Privacy, Working Paper LIDAP-WP4.

VARET Vincent. (2020) « Droit d'auteur et liberté d'expression : de la balance des intérêts au conflit de juridictions ? », *Légipresse*, n° 63, p. 105-110.

WARREN Samuel et BRANDEIS Louis. (1890) « The Right to Privacy », *Harvard Law Review*, vol. 4, n° 5, p. 193-220.

WEISS Marie-Andrée. (2014) « Liberté d'expression sur les réseaux sociaux. Regards croisés États-Unis/Europe », *Documentaliste-Sciences de l'information*, vol. 51, n° 3, p. 20-22.

WILKINSON Mark, DUMONTIER Michel, IJsbrand Jan Aalbersberg *et al.* (2016) « The FAIR Guiding Principles for scientific data management and stewardship », *Scientific Data*, 3, article n° 160018.

Textes normatifs cités

Convention de sauvegarde des droits de l'Homme et des libertés fondamentales du 4 novembre 1950.

Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics de 2007 : <https://www.oecd.org/fr/science/inno/38500823.pdf> .

Protocole additionnel à la Convention de sauvegarde des droits de l'homme et des libertés fondamentales du 20 mars 1952.

Directive 2001/29/CE du 22 mai 2001 du Parlement européen et du Conseil sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information.

Règlement 2016/679/UE du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

Directive 2019/790/UE du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE.

Directive 2019/1024/UE du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public.

Règlement 2021/695/UE du Parlement européen et du Conseil du 28 avril 2021 portant établissement du programme-cadre pour la recherche et l'innovation «Horizon Europe» et définissant ses règles de participation et de diffusion, et abrogeant les règlements (UE) n° 1290/2013 et (UE) n° 1291/2013.

Règlement 2022/868/UE du Parlement européen et du Conseil du 30 mai 2022 portant sur la gouvernance européenne des données et modifiant le règlement (UE) 2018/1724 (règlement sur la gouvernance des données) (*Data Governance Act – DGA*).

Règlement 2022/2065/UE du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (*Digital Services Act – DSA*).

Code civil.

Code de la recherche.

Code du patrimoine.

Code pénal.

Code des relations entre le public et l'administration.

Loi du 29 juillet 1881 sur la liberté de la presse.

Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal.

Loi n° 2010-1 du 4 janvier 2010 relative à la protection du secret des sources des journalistes.

Loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles.

Loi n° 2020-1674 du 24 décembre 2020 de programmation de la recherche pour les années 2021 à 2030 et portant diverses dispositions relatives à la recherche et à l'enseignement supérieur.

Contrat pluriannuel État-ANR d'avril 2021.

Jurisprudence

CEDH 21 janvier 1999, Fressoz et Roire contre France, requête 29183/85.

CEDH 23 juin 2009 Soruç c. Turquie, req. requête 17089/03.

CEDH 3 avril 2012, Gillberg contre Suède, requête 41723/06.

CEDH 3 décembre 2013, Ungváry et Irodalom Kft contre Hongrie, requête 64520/10.

CEDH 10 janvier 2013, Ashby Donald c. France, requête 36769/08.

CEDH 19 février 2013, Nejj et Kolmisoppi c. Suède, requête 40397/12.

CJUE 13 mai 2014, Google Espagne contre AEPD, aff. C-131/12.

CJUE 16 juillet 2015 « ClientEarth contre EFSA » Aff. C-615/13P.

CJUE 24 septembre 2019 GC e.a. contre CNIL, aff. C-136/17.

CJUE 20 décembre 2017, Peter Nowak contre Data Protection Commissioner, Aff. C-434/16.

CJUE 10 juillet 2018, Jehovan Todistajat, Aff. C-25/17.

CJUE 16 décembre 2018, Tietosuojavaltuutettu contre Satakunnan Markkinapörssi Oy et Satamedia Oy, affaire C-73/07.

CJUE 14 février 2019, Sergej Buivids, affaire C-345/17.

CJUE 29 juillet 2019, Funke Medien NRW GmbH contre Allemagne, Aff. C-469/17.

CJUE 29 juillet 2019, Pelham GmbH e.a. contre Ralf Hütter et Florian Schneider-Esleben, Aff. C-476/17.

CJUE 29 juillet 2019, Spiegel Online GmbH v Volker Beck, Aff. C-516/17.

CJUE 6 octobre 2020, Commission européenne contre Hongrie, Aff. C-66/18.

C. Const. 20 janvier 1984, décision 83-165 DC.

C. Const. 8 janvier 1991, décision 90-283 DC.

C. Const., 23 juillet 1999, décision 99-416 DC.

C. Const., 22 mars 2012, décision 2012-652 DC.

C. Cass. 2^e ch. civile, 10 mars 2004, n°01-15.322.

C. Cass., 1^{ère} ch. civile, 22 janvier 2009, n° 07-21.063.

C. Cass., 1^{ère} ch. civile, 5 avril 2012, n° 11-15328.

C. Cass., 1^{ère} ch. civile, 31 octobre 2012, n° 11-26.941.

C. Cass., chambre commerciale, 25 juin 2013, n°12-17037.

TGI de Paris, 3^{ème} chambre, 1^{ère} section. 15 mai 2002, n° 08/01461.

TGI de Paris, 3^{ème} chambre, 3^{ème} section, 18 oct. 2006, n° 05/13494.

TGI de Paris, 9 avril 2019, n° 14/07298.

TJ de Paris, 21 janvier 2021, n° 20/08482.

TJ de Paris, 4 mars 2022, 3^{ème} chambre, 2^{ème} section, n° 22/34.

Délibérations d'autorités administratives indépendantes

CADA, Avis 20144561 - Séance du 08/01/2015 – Météo France

Chambre contentieuse de l'Autorité de protection des données de Belgique, décision quant au fond 13/2022 du 27 janvier 2022, DOS-2018-04433, « EU DisinfoLab ».

CNIL, délibération n°2018-151 du 3 mai 2018, projet de recherche du LORIA.

Articles de presse et de blogs

BARBARO Michael et ZELLER Tom Jr. (2006), « A Face Is Exposed for AOL Searcher No. 4417749 », *The New York Times*, 9 août 2006.
<https://www.nytimes.com/2006/08/09/technology/09aol.html>

BARTHÉLEMY Pierre. (2012) « Le scandale Stapel, ou comment un homme seul a dupé le système scientifique », *Le Monde*, 9 décembre.
https://www.lemonde.fr/passeurdsciences/article/2012/12/09/le-scandale-stapel-ou-comment-un-homme-seul-a-dupe-le-systeme-scientifique_5986311_5470970.html

BONNEAU Cécile. (2008) « Quand les scientifiques trichent : enquête dans les coulisses des labos », *Science & Vie*, 31 octobre. <https://www.science-et-vie.com/archives/quand-les-scientifiques-trichent-38760>

CHALWA Dalmit Singh. (2021) « 8% of researchers in Dutch survey have falsified or fabricated data », *Nature Career News*, 22 juillet. <https://www.nature.com/articles/d41586-021-02035-2>

GALONNIER Juliette. (2021) « Penser la protection des chercheurs, de leurs enquêtés et de leurs données à l'heure de la science ouverte », *La vie de la recherche scientifique*, SNCS FSU, pp. 10-12.
<https://hal.archives-ouvertes.fr/hal-03419665/>

GINOUVÈS Véronique. (2017) « L'exception pédagogique et de recherche », *Questions éthique & droit en SHS*. <https://ethiquedroit.hypotheses.org/1485>

MALBŒUF Marie-Claude. (2017) « Les tricheurs de la science », *La Presse*, 12 septembre.
<https://www.lapresse.ca/actualites/enquetes/201709/12/01-5132442-les-tricheurs-de-la-science.php>

MAUREL Lionel. (2015) « Pourquoi nous avons besoin d'un droit de citation audiovisuelle », *Blog S.I. Lex*. <https://scinfolex.com/2015/10/17/pourquoi-nous-avons-besoin-dun-droit-de-citation-audiovisuelle/>

POMART Julien. (2014) « AAF / Section Aurore: Un groupe de travail sur les données de la recherche », *Archives de la Fondation Maison des sciences de l'Homme*.
<https://archivesfmsh.hypotheses.org/1209>

Rapports et avis

Agence de l'UE pour les droits fondamentaux et Conseil de l'Europe. (2019) *Manuel de droit européen en matière de protection des données*, Luxembourg : Office des publications de l'Union européenne.

CNIL (Commission nationale de l'informatique et des libertés). (non daté) *Sécurité des données personnelles*. https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle.pdf

DENOYELLE Martin, DURAND Katie, DANIEL Johanna et DOULKARIDOU-RAMANTANI Elli. (2018) *Droit des images, histoire de l'art et société*. Rapport de l'Institut national de l'histoire de l'art et de la Fondation de France. https://www.inha.fr/attachments/images-usages-ressources/INHA_rapport_images_usages.pdf?download=true

EDPS (Contrôleur européen de la protection des données). (2020) *A Preliminary Opinion on Data Protection and Scientific Research*. 6 janvier. https://edps.europa.eu/sites/default/files/publication/2001-06_opinion_research_en.pdf

Groupe de travail de l'Article 29. (2014) *Avis 05/2014 sur les techniques d'anonymisation*. 10 avril. <https://www.dataprotection.ro/servlet/ViewDocument?id=1288>

Groupe de travail de l'Article 29. (2017) *Lignes directrices concernant l'analyse d'impact relative à la protection des données (AIPD) et la manière de déterminer si le traitement est «susceptible d'engendrer un risque élevé» aux fins du règlement (UE) 2016/679*. 4 avril. <https://ec.europa.eu/newsroom/article29/items/611236/en>

INSHS (Institut des sciences humaines et sociales). (2019) *Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte. Guide pour la recherche*. https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf

MESR (Ministère de l'enseignement supérieur et de la recherche). (2018) *Plan national pour la science ouverte*. <https://www.enseignementsup-recherche.gouv.fr/fr/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-tous-49241>

Annexe 1 : Table des abréviations

AAF	Association des archivistes français
ANR	Agence nationale de la recherche
APD	Autorité de protection des données (Belgique)
BNF	Bibliothèque nationale de France
CADA	Commission d'accès aux documents administratifs
C. Cass.	Cour de cassation
C. Const.	Conseil constitutionnel
CEDH	Cour européenne des droits de l'Homme
CERES	Centre d'expérimentation en méthodes numériques pour les recherches en SHS
CJUE	Cour de justice de l'Union européenne
CNIL	Commission nationale de l'informatique et des libertés
CNRS	Centre national de la recherche scientifique
Convention EDH	Convention de sauvegarde des droits de l'homme et des libertés fondamentales

DDHC	Déclaration des droits de l'Homme et du citoyen
DGA	Data Governance Act (règlement sur la gouvernance des données)
DSA	Digital Services Act (Règlement sur les services numériques)
EDPS	European Data Protection Supervisor (en français : Contrôleur européen de la protection des données)
FAIR	Findable, Accessible, Interoperable, Reusable
IA	Intelligence artificielle
INSHS	Institut des sciences humaines et sociales
PGD	Plan de gestion des données
RGPD	Règlement général de protection des données
SHS	Sciences humaines et sociales
TGI	Tribunal de grande instance
TJ	Tribunal judiciaire
UE	Union européenne

Annexe 2 : Corpus de plans de gestion des données

Critères de sélection : plans de gestion des données disponibles publiquement sur Internet, sur dmptools.org, dmponline.dcc.ac.uk, et https://dmp.opidor.fr/public_plans, en français, anglais ou espagnol, et portant sur des projets impliquant au moins une discipline de SHS. Pour disposer d'un nombre qu'il soit possible de traiter avec le temps et les moyens disponibles, seuls les cinq PGD les plus récents répondant à ces critères ont été choisis pour chacun de ces sites.

#	Titre du projet	Auteur·e·s	Institution	Pays	Lien
1	Westchester Local Food Project	Allison Turcan	D.I.G. Farm Inc.	États-Unis	https://dmptool.org/plans/73513/export.pdf?export%5Bquestion_headings%5D=true
2	Biblioteca de epistolarios de intelectuales hispánicos del siglo XX (BEPHISsXX). Patrimonio textual y configuración histórica de un campo Intelectual	Juan Carlos Conde	Universidad de Salamanca	Espagne	https://dmptool.org/plans/73225/export.pdf?export%5Bquestion_headings%5D=true
3	Factors affecting higher education readiness amongst first-year undergraduate students, who reside at Azalea Residence, at the University of Pretoria	Deidre Huysamen	University of Pretoria	Afrique du Sud	https://dmptool.org/plans/71780/export.pdf?export%5Bquestion_headings%5D=true
4	Patient Understanding of Anesthesia and Consent	Justin Cha	John Hopkins University	États-Unis	https://dmptool.org/plans/70597/export.pdf?export%5Bquestion_headings%5D=true
5	Chat Apps Analytics: the use of Telegram during the pre-campaign to 2022 Brazilian presidential election	Giulia Tucci	Universidade Federal do Rio de Janeiro	Brésil	https://dmptool.org/plans/69645/export.pdf?export

#	Titre du projet	Auteur·e·s	Institution	Pays	Lien
					%5Bquestion_headings%5D=true
6	The Narrative Identity and PsychoPathology in Adolescence (NIPPA) Project	Elisabeth de Moor, Susan Branje	Utrecht University	Pays-Bas	https://dmponline.dcc.ac.uk/plans/75262/export.pdf?export%5Bquestion_headings%5D=true
7	Dublin Castle Christmas Market Social Media Review	Caoimhe Creed	Inconnue	Irlande	https://dmponline.dcc.ac.uk/plans/90375/export.pdf?export%5Bquestion_headings%5D=true
8	Exploratory Factor Analysis of a sexually objectifying media scale (Media-SOS) for cisgender women and men	Zhuozhuo hu	University of Sheffield	Royaume-Uni	https://dmponline.dcc.ac.uk/plans/90697/export.pdf?export%5Bquestion_headings%5D=true
9	Improved understandings of health services for male and LGBTQI+ survivors of conflict related sexual violence in Iraq	Elsie Cullen	London School of Economics and Political Science	Royaume-Uni	https://dmponline.dcc.ac.uk/plans/89031/export.pdf?export%5Bquestion_headings%5D=true
10	Effect, Cost-effectiveness, and Usefulness of the Everyday Life Rehabilitation Intervention	Per Liv ; Maria Lindström	Umeå University	Suède	https://dmponline.dcc.ac.uk/plans/86475/export.pdf?export%5Bquestion_headings%5D=true
11	CLIODIFLE 2 : Des institutions,	Vincent de Lavenne	Université de Paris 3 -	France	https://dmp.opidor.fr/

#	Titre du projet	Auteur·e·s	Institution	Pays	Lien
	enseignants, des étudiants... Une contribution à l'histoire de la francophonie		Sorbonne Nouvelle		plans/11501/ export.pdf
12	SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech	Brigitte Bigi	CNRS	France	https:// dmp.opidor.fr/ plans/4296/ export.pdf
13	E-NdP	Pierre Brochard	CNRS	France	https:// dmp.opidor.fr/ plans/10339/ export.pdf
14	Project Places	William Kelleher	Université Rennes 2	France	https:// dmp.opidor.fr/ plans/10402/ export.pdf
15	Decolonizing Southeast Asian Sound Archives	Joséphine Simonnot	CNRS	France	https:// dmp.opidor.fr/ plans/10610/ export.pdf

Annexe 3 : Les questions posées en atelier

Un questionnaire a été mis en ligne et diffusé en amont de l'atelier organisé le 2 décembre 2021 par le CERES⁸³, pour recueillir des questions, témoignages et avis sur les règles de droit applicables à l'usage de données de la recherche en SHS. Des questions ont également été prises en note pendant cet atelier. Les questions ainsi collectées sont traitées dans cette section du rapport, sans ordre précis.

- « **Remplacer le nom par enquêté 1, 2, 3, 4...** est-ce que c'est problématique au regard du RGPD ? »

Cela n'est pas problématique. Au contraire, cela constitue une pseudonymisation qui est encouragée par le RGPD en échange de la flexibilité accordée à la recherche. Le remplacement d'un nom par un numéro unique n'est toutefois pas toujours une garantie suffisante pour parler d'anonymat, ni pour garantir, le cas échéant, la confidentialité et la sécurité des enquêtés (cf. Barbaro et Zeller, 2006 et Sweeney, 2000).

Des techniques existent (k-anonymat, vie privée différentielle...) pour construire une réflexion plus fine autour du degré d'anonymat souhaitable dans le cadre d'une recherche (cf. Groupe de travail de l'article 29, 2014).

- Les chercheuses et chercheurs s'interrogent sur le traitement à réserver **aux données accessibles sur internet, comme par exemple des forums d'échanges publics**. Lorsqu'ils récupèrent ces données pour constituer un corpus de travail, **ont-ils l'obligation de pseudonymiser ou d'anonymiser** ces données et **quel usage peuvent-ils en faire ?**

Des chercheurs·ses peuvent utiliser des données accessibles publiquement sur Internet, même lorsqu'il s'agit de données appartenant aux catégories sensibles de l'article 9 (données de santé, politiques, sur la vie sexuelle, etc.) dès lors qu'elles ont été manifestement rendues publiques par la personne elle-même (art. 9 (2) (e) du RGPD). En l'état actuel du droit, et étant donné qu'un flou juridique subsiste sur l'exception à des fins de recherche sur l'usage de données sensibles (art. 9 (2) (j) du RGPD), il vaut mieux, pour plus de sécurité juridique, recueillir le consentement des personnes concernées. Cela est particulièrement le cas lorsque les données en question ont été publiées contre la volonté de la personne concernée. En cas d'impossibilité de recueillir ce consentement, il est possible soit de demander l'autorisation à la CNIL (sur le fondement du 6° de l'article 44 de la loi informatique et libertés française) soit de tenter de se prévaloir de la liberté d'expression universitaire prévue à l'article 85 du RGPD, mais dont la définition et surtout la distinction avec le régime applicable à la recherche scientifique n'est pas d'une grande clarté.

83 Voir : <https://enquetes.ent.upmc.fr/lime/index.php/557411?lang=fr> (page consultée le 19 décembre 2021).

La possibilité de collecter ces données à des fins de recherche, donc pour une fin autre que celle qui avait été envisagée par la personne au moment de publier ces informations, est possible en vertu de dispositions qui imposent une obligation d'effort d'anonymisation. Concrètement, cela signifie qu'il faut s'interroger sur le degré maximal d'anonymat qu'il soit possible d'accorder à ces données sans qu'elles perdent tout intérêt pour la recherche.

L'usage qu'il est possible d'en faire est d'une part de les traiter à des fins de recherche, à des fins d'archivage, et enfin à des fins de liberté d'expression universitaire (art. 85 du RGPD). Cela inclut donc la possibilité de publier ces informations, dans la stricte limite de ce qui est nécessaire et utile au débat public (voir la jurisprudence de la CEDH étudiées dans les sections sur la liberté d'expression).

- Quelle serait la responsabilité du chercheur en France concernant **des données achetées auprès d'entreprises à l'étranger**, par exemple les États-Unis, où l'on peut acheter des données issues des réseaux sociaux, données qui sont collectés par ces entreprises ?

L'ancien article 1128 du Code civil disposait qu'« il n'y a que les choses qui sont dans le commerce qui puissent être l'objet des conventions. » La Cour de Cassation française avait jugé⁸⁴ sur la base de cet article qu'un contrat de vente d'un fichier de données à caractère personnel qui n'avait pas été déclaré à la Commission nationale de l'informatique et des libertés (CNIL), comme la Loi l'exigeait avant l'entrée en application du RGPD, était frappé de nullité. Malgré la disparition de cet article du Code civil depuis la réforme de 2016 (Bellivier et Noville, 2016), et l'évolution du droit des données personnelles, il est probable que la jurisprudence de la Cour de cassation reste la même sur le fond, peut-être en se fondant sur l'article 1598 du Code civil, qui dispose que « tout ce qui est dans le commerce, peut être vendu. »

En tout cas, même la jurisprudence venait à évoluer, une personne physique ou morale qui achèterait un jeu de données personnelles obtenues sans base légale, et/ou sans en informer les personnes concernées, serait elle-même en tort.

Il convient donc avant d'acheter un jeu de données de vérifier la licéité de leur provenance, et de réfléchir à l'information à fournir aux personnes dont les données ont été achetées.

Lorsque les données viennent de l'étranger, il faut vérifier qu'elles aient été collectées et exportées de façon légale depuis le pays de provenance, et une fois importées sur le territoire de l'Espace économique européen, le RGPD s'applique. Cela signifie qu'il faut remplir l'un des critères de licéité de l'article 6 (1), s'assurer qu'il n'y ait pas de données sensibles ou s'assurer qu'elles bénéficient d'une des exemptions au traitement de ce type de données, etc.

- Dans le cas de sondages **auprès d'étudiants**, est-il **suffisant de mentionner** que ces données seront utilisées, en début de sondage ? La condition du consentement est-elle remplie ici ?

84 C. Cass., chambre commerciale, 25 juin 2013, pourvoi n°12-17037

Quelles que soient les personnes auprès de qui des données sont collectées, il faut les informer, selon l'article 13 du RGPD, de :

- l'identité et les coordonnées du responsable du traitement et, le cas échéant, du représentant du responsable du traitement
- le cas échéant, lui transmettre les coordonnées du délégué à la protection des données;
- donner des informations sur les finalités du traitement auquel sont destinées les données, ainsi que sur la base juridique du traitement;
- lorsque le traitement est fondé sur l'article 6, paragraphe 1, point f) (c'est-à-dire sur la notion d'intérêt légitime), les intérêts légitimes poursuivis par le responsable du traitement ou par un tiers doivent être clairement indiqués ;
- les destinataires ou les catégories de destinataires des données à caractère personnel, s'ils existent, doivent être précisés ;
- si un transfert de données en-dehors de l'Espace économique européen est prévu, il faut l'indiquer et indiquer ce que sont les garanties en place pour que les données soient traitées dans des conditions adéquates ;
- la durée de conservation des données à caractère personnel ou, lorsque ce n'est pas possible, des critères utilisés pour déterminer cette durée;
- lorsque cela s'applique, l'existence du droit de demander au responsable du traitement l'accès aux données à caractère personnel, la rectification ou l'effacement de celles-ci, ou une limitation du traitement relatif à la personne concernée, ou du droit de s'opposer au traitement et du droit à la portabilité des données;
- lorsque le traitement est fondé sur le consentement, l'existence du droit de retirer son consentement à tout moment, sans porter atteinte à la licéité du traitement fondé sur le consentement effectué avant le retrait de celui-ci;
- du droit d'introduire une réclamation auprès d'une autorité de contrôle;
- des informations sur la question de savoir si l'exigence de fourniture de données à caractère personnel est obligatoire ou pas ;
- de l'existence d'une prise de décision automatisée (y compris le profilage), et, au moins en pareils cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée.

S'il manque une ou plusieurs de ces informations, alors le consentement n'est pas réputé valable au titre du RGPD. Il faut donc se tourner vers une autre base légale, comme la mission de service public. Cela n'exclut pas de demander un accord à l'enquêté-e, mais cet accord seul n'équivaudra pas à un consentement au titre du RGPD.

Enfin, le considérant 43 précise que le consentement « ne constitue pas un fondement juridique valable pour le traitement de données à caractère personnel dans un cas particulier lorsqu'il existe un

déséquilibre manifeste entre la personne concernée et le responsable du traitement, en particulier lorsque le responsable du traitement est une autorité publique et qu'il est improbable que le consentement ait été donné librement au vu de toutes les circonstances de cette situation particulière. » Faire consentir valablement ses propres étudiant·e·s paraît donc délicat.

Si un autre fondement juridique que le consentement des étudiant·e·s, par exemple la mission de service public à laquelle obéit la recherche publique, est mobilisé pour légitimer la collecte de leurs données, il est tout de même possible de leur demander un « accord » qui ne vaudra pas « consentement » au titre du RGPD, mais permettra néanmoins d'apporter des garanties supplémentaires de respect de leurs droits et de leurs intérêts. Il sera aussi possible de limiter l'information fournie sur les modalités et les finalités de l'enquête, dans la stricte mesure où cela est absolument indispensable à la réalisation des objectifs de la recherche, comme précisé au considérant 62 du RGPD.

- Est-ce que les chercheurs peuvent bénéficier de **la protection des sources** au même titre que des journalistes ?

Non. La protection des sources ne concerne que les seuls journalistes. Dans sa rédaction actuelle, issue de la loi du 4 janvier 2010 relative à la protection du secret des sources, l'article 2 de la loi du 29 juillet 1881 sur la liberté de la presse indique à son premier paragraphe que « le secret des sources des journalistes est protégé dans l'exercice de leur mission d'information du public. » Il est précisé immédiatement après qu'« est considérée comme journaliste au sens du premier alinéa toute personne qui, exerçant sa profession dans une ou plusieurs entreprises de presse, de communication au public en ligne, de communication audiovisuelle ou une ou plusieurs agences de presse, y pratique, à titre régulier et rétribué, le recueil d'informations et leur diffusion au public. » Cela n'est pas le cas des chercheurs·ses.

- Est-il possible de **réutiliser** des données issues de recherches antérieures ?

Oui. L'article 5 paragraphe 1 sous b) du RGPD dispose que : « Les données à caractère personnel doivent être: [...] collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités; **le traitement ultérieur** [...] à des fins de **recherche scientifique** ou historique ou à des fins statistiques **n'est pas considéré**, conformément à l'article 89, paragraphe 1, **comme incompatible avec les finalités initiales** [...]. »

- Est-ce que nous sommes tenus de **sécuriser** les données sur l'ordinateur ?

Oui, c'est impératif, et il n'existe aucune dérogation à ce sujet. L'établissement qui accueille la recherche est responsable de la sécurité et de la confidentialité des données, et doit fournir les moyens d'un traitement sécurisé.

Cette sécurité doit tenir compte de l'état de la menace, dans le contexte de la recherche menée. L'article 32 du RGPD dispose que « compte tenu de l'état des connaissances, des coûts de mise en œuvre 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 physiques, le responsable du traitement et le sous-traitant mettent en œuvre les mesures techniques et organisationnelles**

appropriées afin de garantir un niveau de sécurité adapté au risque. » Ce même article préconise notamment l'emploi de la pseudonymisation et du chiffrement.

En cas de violation des données, c'est-à-dire de diffusion ou d'accès non-autorisé à ces données, il faut notifier la CNIL sous 72 heures (art. 33 du RGPD) et, dans certains cas, les personnes concernées par cette fuite (art. 34 du RGPD).

Le guide⁸⁵ de l'Institut des sciences humaines et sociales (INSHS) du Centre national de la recherche scientifique (CNRS) sur la protection des données contient quelques conseils (page 24 du guide), et la CNIL publie également un guide sur son site web⁸⁶.

- Les données ont-elles la même valeur **sur ordinateur et sur papier** ?

Le RGPD s'applique pour tout traitement de données à caractère personnel « automatisé en tout ou partie », mais aussi « au traitement non automatisé de données à caractère personnel contenues ou appelées à figurer dans un fichier » (art. 2 (1) du RGPD). La CJUE a déjà jugé que des copies d'examen⁸⁷ et des cahiers dans lesquels des Témoins de Jehovah prenaient des notes sur des personnes prospectées au cours d'opérations de porte-à-porte⁸⁸ entraient dans le champ d'application du RGPD. Dans cette seconde décision, la CJUE avait noté que :

« les données à caractère personnel collectées dans le cadre de l'activité de prédication de porte-à-porte en cause au principal sont **structurées** selon des critères retenus en fonction de l'objectif poursuivi par cette collecte, qui est de préparer des visites ultérieures et de gérer les listes de personnes ne souhaitant plus être démarchées. Ainsi qu'il ressort de la décision de renvoi, ces critères, au nombre desquels figurent notamment le nom et l'adresse des personnes démarchées, leurs convictions ou encore leur souhait de ne plus recevoir de visites, sont choisis de sorte qu'ils puissent permettre de **retrouver aisément les données relatives à des personnes déterminées**⁸⁹. »

Les carnets de terrains et autres notes manuscrites utilisées dans le cadre d'un projet de recherche en SHS sont généralement aussi structurés de telle sorte à être facilement retrouvables. Dans la grande majorité des cas, des données personnelles au format papier (ou sur support analogique) et des données numérisées seront donc soumises de la même façon au RGPD.

- Peut-on **utiliser le pseudo Twitter** d'une personne dans une **publication scientifique** ?

Un pseudo Twitter est une donnée à caractère personnel. Cela ne signifie pas que sa publication soit interdite. Le législateur européen a en effet prévu une conciliation entre le droit à la vie privée et/ou à la protection des données à caractère personnel d'une part, et la liberté d'expression d'autre part. Comme nous l'avons vu dans la section de ce rapport dédié à la liberté d'expression et aux libertés académiques, la jurisprudence de la CEDH défend le droit des chercheurs·ses à publier des informations qui participent à un débat d'intérêt général, et divers éléments de droit, dont l'article 179, paragraphe 1, du traité sur le fonctionnement de l'Union européenne, considèrent la recherche scientifique comme d'intérêt général.

85 INSHS, 2019 : https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf

86 CNIL, non daté : https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle.pdf

87 CJUE 20 décembre 2017, Peter Nowak contre Data Protection Commissioner, Aff. C-434/16,

88 CJUE 10 juillet 2018, Jehovan Todistajat, Aff. C-25/17

89 CJUE 10 juillet 2018, Jehovan Todistajat, Aff. C-25/17, pt. 60.

Cette liberté a des limites. Ainsi, il faut distinguer les cas où une personne occupe une position dans la société qui fait d'elle une personnalité publique, ou pas. Il convient aussi de vérifier si les données et les réflexions associées à ce pseudonyme ne sont pas de nature sensible (au titre de l'article 9 du RGPD). Si c'est le cas, il reste possible de publier ces données en vertu de l'article 85 du RGPD si cela correspond vraiment à un débat d'intérêt public, mais pas s'il le seul but de la publication du pseudonyme associé au contenu sensible sert à appuyer une démonstration scientifique.

Les règles qui interdisent la diffamation continuent de s'appliquer.

- Doit-on obligatoirement effectuer une **déclaration à la CNIL** lorsqu'on mène une **enquête par 1) entretiens et 2) questionnaire**, même lorsqu'on ne collecte pas de **données "sensibles"**?

Le RGPD a supprimé les obligations de notification à la CNIL dans l'écrasante majorité des cas.

Il reste quatre types d'exceptions :

- les cas où une étude d'impact est obligatoire en vertu de l'article 35 du RGPD, mais qu'elle n'a pas permis de résoudre tous les risques, et que le projet présente encore des risques élevés pour les droits et libertés des personnes concernées (voir : article 36 du RGPD) ;
- certains traitements, qui intéressent la sûreté de l'Etat, la défense ou la sécurité publique, ont pour objet la prévention, la recherche, la constatation ou la poursuite des infractions pénales ou l'exécution des condamnations pénales ou des mesures de sûreté, ou ceux mis en œuvre pour le compte de l'État, agissant dans l'exercice de ses prérogatives de puissance publique, qui portent sur des données génétiques ou sur des données biométriques nécessaires à l'authentification ou au contrôle de l'identité des personnes, sont autorisés par décret en Conseil d'État après avis de la CNIL, en vertu de procédures fixées aux articles 31 à 36 de la loi Informatique et Libertés ; ces cas de figure sont peu susceptibles de concerner la recherche en SHS ;
- Les demandes d'autorisation qui concernent la recherche en santé (y compris lorsqu'elle comporte un volet en SHS), régies par les articles 66 et 76 de la loi Informatique et Libertés.
- Le cas de la demande d'avis de la CNIL lorsque nous collectons des données sensibles au titre de l'article 9 du RGPD, sans le consentement explicite de la personne concernée ou sans que le traitement ne porte sur des données manifestement rendues publiques, et que cela entre dans le cadre de ce qui est nécessaire à la recherche publique au sens de l'article L. 112-1 du code de la recherche, sous réserve que des motifs d'intérêt public important les rendent nécessaires (art. 44 6° de la loi Informatique et Libertés), selon une procédure fixée à l'article 34 de la loi Informatique et Libertés ; la CNIL dispose d'un délai de huit semaines, qui peut être prolongé de six semaines, pour répondre, faute de quoi son avis est réputé favorable.

Il existe également une obligation de la part de l'université de s'assurer que tous les traitements de données à caractère personnel effectués soient bien inscrits dans un registre des traitements tenu par le/la délégué-e à la protection des données.

Aucune déclaration n'est à faire à la CNIL lorsque le traitement envisagé ne comporte pas de données pouvant révéler des informations sensibles, sauf lorsqu'il fait malgré tout courir des risques élevés aux droits et libertés de personnes physiques, et que l'évaluation d'impact menée en application de l'article 35 du RGPD ne permet pas de dégager de pistes de diminution des risques substantiels identifiés (voir : article 36 du RGPD).

- Pourrez-vous préciser quelle **différence** fait la loi entre « **données sensibles** » et « **données sociodémographiques** »?

Le droit de la protection des données à caractère personnel ne connaît pas la notion de « donnée sociodémographique. »

L'article 9 définit toutefois une liste d'informations jugées particulièrement sensibles, et interdit (sauf exceptions énumérées au même article) tout traitement de données à caractère personnel révélant ces informations. Il s'agit des traitements qui révèlent « l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique. »

Un point du RGPD régulièrement critiqué est l'omission des données dont le traitement révèle des informations sur la situation financière des personnes de la liste de l'article 9.

Notons en outre l'existence d'un article 10 du RGPD qui encadre le « traitement des données à caractère personnel relatives aux condamnations pénales et aux infractions ou aux mesures de sûreté connexes », en prévoyant que celui-ci « ne peut être effectué que sous le contrôle de l'autorité publique, ou si le traitement est autorisé par le droit de l'Union ou par le droit d'un 'État membre qui prévoit des garanties appropriées pour les droits et libertés des personnes concernées. » Il n'existe pas de dérogation propre à la recherche. L'article précise toutefois qu'il ne concerne que les traitements fondés sur l'article 6 (1) du RGPD, et non l'article 6 (4). Cette dernière disposition vise notamment les cas où « le traitement à une fin autre que celle pour laquelle les données ont été collectées n'est pas fondé sur le consentement de la personne concernée. » Or, en vertu de l'article 5 (1), « le traitement ultérieur à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques n'est pas considéré, conformément à l'article 89, paragraphe 1, comme incompatible avec les finalités initiales. » Il est donc possible, sur cette base, de mener des recherches sur les condamnations pénales à des fins de recherche.

- Existe-t-il des **normes de stockages des données à respecter** pour les entretiens et questionnaires lorsqu'on garantit la confidentialité à l'interviewé/répondant?

Si le traitement de données effectué est véritablement et robustement anonyme, c'est-à-dire que toute identification de la personne, même indirectement, est strictement impossible, alors il ne s'agit pas de données à caractère personnel et le RGPD ne s'applique pas.

En revanche, dès lors que ces données contiennent des quasi-identifiants (noms, codes postaux, âges, professions...) qui pourraient permettre, par croisement avec d'autres données, d'identifier les personnes concernées ou même, simplement, de distinguer (*single out* en anglais), alors il faut prendre les mesures de sécurité appropriées pour garantir leur confidentialité et leur sécurité.

Le RGPD, en particulier son article 89, encourage très fortement la pseudonymisation des données de la recherche, c'est-à-dire le fait de s'assurer de garder le moins d'information possible pouvant mener à l'identification. Le remplacement du nom d'état civil par un nom fictif, l'approximation des âges et du lieu de résidence, l'introduction de bruit (un faible pourcentage d'informations fausses) sont des exemples de techniques qui peuvent être mises en place pour limiter le risque d'identification. Ces techniques sont plus simples à mettre en œuvre dans les méthodes quantitatives et dans des questionnaires fermés qu'avec des méthodes qualitatives.

Les mesures appropriées s'apprécient en fonction du risque et de l'état de l'art en matière de sécurité informatique. Dans un certain nombre de cas, il est nécessaire de conduire une évaluation d'impact avant de commencer la recherche. Ce processus peut être l'occasion de prendre contact avec la DSI de l'université ou de l'institut où se déroule la recherche, pour mettre en place les mesures appropriées. C'est à l'employeur, en tant que responsable du traitement, de fournir les moyens appropriés.

Une bonne pratique, de façon générale, consiste à généraliser le chiffrement des données et des communications entre collègues d'une équipe de recherche partageant des données à caractère personnel. Cela peut se faire par exemple avec des applications de messagerie chiffrée comme Signal, ou par le chiffrement de courriers électroniques avec PGP. Le chiffrement des répertoires de stockage des données voire même leur conservation sur un support hors-ligne peuvent aussi, dans certains cas, être envisagés.

- Doit-on **mentionner un certain article de loi (CNIL?)** dans l'introduction d'un questionnaire en ligne? Y-a-t-il d'autres informations obligatoires à indiquer au répondant pour être en conformité?

Les mentions obligatoires qui doivent être portées à la connaissance d'un·e enquêté·e, surtout dans l'hypothèse où son consentement est recherché, sont énumérées aux articles 13 du RGPD pour les cas où les informations sont recueillies directement auprès de cette personne, et 14 dans les autres cas. Il s'agit notamment des informations suivantes :

- l'identité et les coordonnées du responsable du traitement ;
- les coordonnées du ou de la délégué·e à la protection des données (lorsqu'il y en a un·e) ;
- les finalités, c'est-à-dire l'objectif, auxquelles sont destinées les données collectées ;
- la base juridique du traitement : s'agit-il d'un consentement ? D'un traitement nécessaire pour une mission de service public ? (voir l'article 6 (1) du RGPD) ;
- les catégories de données qui sont concernées par la collecte ;
- les destinataires des données ;
- le cas échéant, l'existence d'un transfert de ces données en dehors de l'Espace économique européen ;
- la durée de conservation des données ;
- lorsque le traitement est fondé par l'objectif de la poursuite d'un intérêt légitime du responsable du traitement ou d'un tiers, il faut indiquer quel est cet objectif, et la possibilité pour l'enquêté·e de faire jour son droit d'opposition ;
- l'existence de droits d'accès, de rectification, d'opposition (pour les traitements fondés sur l'intérêt légitime) ou d'effacement (pour les traitements qui reposent sur le consentement) ;
- le droit, le cas échéant, de retirer son consentement à tout moment ;
- le droit d'introduire une réclamation auprès de la CNIL ;
- la source d'où proviennent les données à caractère personnel et, le cas échéant, une mention indiquant qu'elles sont issues ou non de sources accessibles au public ;
- le cas échéant, l'existence d'une prise de décision automatisée fondée sur les données collectées, et les logiques sous-jacentes au mécanisme de prise de décision.

Cette liste est longue et peut paraître impressionnante de prime abord, mais tous ces éléments ne sont pas nécessairement pertinents dans tous les cas de traitements de données à des fins de recherche. Il est par exemple peu probable qu'il y ait besoin de mentionner l'existence d'une prise de décision automatisée affectant l'enquêté·e dans le cadre d'un projet de recherche.

Notons qu'il est possible d'omettre une partie de ces informations lorsque les données ne sont pas collectées directement auprès de la personne concernée si cela s'avère nécessaire d'un point de vue méthodologique, à condition qu'il soit possible, en cas de contrôle, de démontrer la nécessité d'omettre certaines informations. Il est utile de documenter cette nécessité au préalable, par exemple dans l'étude d'impact fondée sur l'article 35 du RGPD, dans le plan de gestion des données et/ou dans le registre des traitements. Cette possibilité de déroger en tout ou partie aux obligations de transparence est prévue par l'article 14 (5) b) du RGPD.

Le considérant 62 offre cette même possibilité, en cas de stricte nécessité, qu'il faut être mesure de démontrer, dans le cas où les données sont collectées directement auprès de la personne concernée.

- Quel est le **statut d'images photographiques prises pendant une enquête** de terrain (pour le documenter et/ou comme donnée intégrante du protocole scientifique) :
 - a. Quel **droit à l'image** s'applique dans ce cas là, notamment **si des personnes sont présentes** sur les images ? Doivent-elles signer une décharge et/ou être de trois quart / de dos (donc non identifiable) ?
 - b. Quels sont les **droits de publication de mes propres images d'enquête** en tant que chercheur dans une revue scientifique par exemple ? Est-ce que le régime varie en fonction du sujet photographié, et/ou du support de diffusion (en ligne, papier, en ligne et papier) ?

Lorsque les données sont collectées directement auprès de l'enquêté·e, l'article 13 du RGPD ne prévoit pas de dérogation possible à l'obligation d'information. Toutefois, cet article doit être lu en combinaison avec le considérant 62, qui affirme qu'« il n'est pas nécessaire d'imposer l'obligation de fournir des informations [...] lorsque la communication d'informations à la personne concernée se révèle impossible ou exigerait des efforts disproportionnés. Tel pourrait être le cas, notamment, lorsqu'il s'agit d'un traitement à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques. À cet égard, devraient être pris en considération le nombre de personnes concernées, l'ancienneté des données, ainsi que les garanties appropriées éventuelles adoptées. »

Les images de personnes identifiables sont des données à caractère personnel. Les images sont aussi couvertes par le droit à l'image, qui découle du droit au respect de la vie privée prévu à l'article 9 du Code civil.

Le fait qu'il s'agisse de données à caractère personnel fait que toute photographie prise et traitée dans le cadre d'un projet de recherche est un traitement de données à caractère personnel. Il est donc soumis à des obligations, comme par exemple l'obligation d'informer les personnes concernées et, le cas échéant, d'obtenir leur consentement. Cette information peut se faire par la mise en place de panneaux d'information similaires à la façon dont se fait l'information sur la vidéosurveillance.

Le droit à l'image concerne plus particulièrement la publication de l'image d'une personne.

L'article 226-1, 2° du Code pénal punit par ailleurs « d'un an d'emprisonnement et de 45 000 euros d'amende le fait, au moyen d'un procédé quelconque, volontairement de porter atteinte à l'intimité de la vie privée d'autrui [...] en fixant, enregistrant ou transmettant, sans le consentement de celle-ci, l'image d'une personne se trouvant dans un lieu privé. »

Le fait que la photographie soit prise dans un lieu public n'exclut toutefois pas la protection de la vie privée. Il est possible de photographier une foule, tant qu'il n'y a pas de gros plan sur une personne en particulier qui soit isolée de la foule dans l'image, et facilement identifiable⁹⁰.

Le fait que les personnes ne soient plus reconnaissables, par exemple parce que leurs visages sont dissimulés (comme sur Google Street View), fait qu'il n'est plus possible de parler d'atteinte au droit à l'image ou à la vie privée⁹¹.

Il faut aussi noter que « la liberté de communication et d'information autorise la publication d'informations et de photographies qui en constituent l'illustration pertinente, lorsqu'elles sont relatives à un événement d'actualité, même si ces informations et photographies sont susceptibles d'apporter des révélations sur la vie privée des participants à cet événement⁹². »

Les images peuvent par ailleurs relever du droit d'auteur lorsqu'il s'agit d'une œuvre de l'esprit, ce qui est souvent le cas, même si la reconnaissance du droit d'auteur sur les photographies est un sujet complexe, soumis le cas échéant à la libre appréciation du juge, qui peut revêtir une part d'arbitraire. Les droits d'auteur sont détenus par l'auteur de l'image, et non par les sujets qui y figurent. Cela signifie, concrètement, qu'un·e chercheur·se fait ce qu'il ou elle souhaite de ses images, dès lors que cela n'enfreint pas les droits des tiers.

- Est ce que **le type d'ouverture** (partielle ou totale) est **conditionné par la nature des données** ?

Effectivement, l'ouverture des données est conditionnée par leur nature.

Par exemple, l'article L 213-2 du Code du patrimoine dispose entre autres que les archives publiques « dont la communication porte atteinte au secret des délibérations du Gouvernement » ne peuvent être communiqués qu'après une période d'embargo de 25 ans, et de 50 ans pour les documents dont la communication peut porter atteinte à la vie privée.

D'autres secrets, comme les secrets d'affaires, les secrets industriels, le secret bancaire, le secret médical, peuvent s'opposer à la divulgation d'information.

Quant au partage de données qui ne sont pas encore transmises à des archives publiques, celui-ci n'est possible que s'il ne porte pas atteinte à la vie privée des personnes concernées, si jamais celles-ci ont été informées qu'un tel partage aura lieu, et qu'elles y ont consenti ou qu'elles ne se soient pas opposées à l'intérêt légitime de la personne à l'initiative du partage de ces données.

- Travailler sur la modération dans les réseaux sociaux numériques ou les forums implique parfois **d'analyser des publications supprimées**. Dans quelle mesure cela contrevient-il au droit à l'oubli ? Dans quelle mesure pouvons-nous éviter de contrevenir au droit à l'oubli ? L'anonymisation suffit-elle ?

90 C. Cass. 2^e ch. civile, 10 mars 2004, n°01-15.322.

91 C. Cass., 1^{ère} ch. civile, 5 avril 2012, n°11-15328.

92 C. Cass., 1^{ère} ch. civile, 31 octobre 2012, n° 11-26.941.

Au moment de la collecte, les données de ces tweets avaient été rendues publiques par leur auteur, dans le but de s'exprimer dans le contexte de Twitter. La réutilisation de ces données à une autre finalité, celle de la recherche scientifique, est autorisée par l'article 5 (1) (b), sous réserve de respecter les dispositions de l'article 89 du RGPD, qui soumet cette possibilité à des « garanties appropriées pour les droits et libertés de la personne concernée. »

Dès lors qu'il n'est pas possible de savoir si un tweet a été supprimé par la plateforme ou bien par son auteur, une interprétation possible du RGPD est de considérer qu'il demeure possible de traiter cette donnée à des fins de recherche, mais pas de publier le tweet, sauf si sa publication revêt un intérêt de nature journalistique et relève de l'article 85 du RGPD. En effet, le traitement à des fins de recherche, dont la finalité est compatible (nous dit le RGPD) avec la finalité initiale (le fait de publier quelque chose sur Twitter), ne repose pas nécessairement sur le consentement de la personne concernée. Dès lors, le retrait de ce consentement, manifesté par la suppression du contenu, affecte la licéité du traitement initial (la publication sur Twitter) mais pas le traitement subséquent (à des finalités de recherche), qui peut reposer sur la mission de service public de recherche ou l'intérêt légitime. Dans ces deux dernières hypothèses, la personne concernée aurait un droit d'opposition.

L'anonymisation peut être une façon de respecter le droit à l'oubli dans la mesure où une donnée anonyme n'est pas soumise au RGPD. Cet anonymat doit toutefois être robuste, c'est-à-dire qu'il doit être possible de démontrer l'impossibilité de retrouver l'auteur·e du tweet.

Enfin, le droit à l'effacement ou le droit d'opposition, qui constituent ensemble le droit à l'oubli, connaissent des limites⁹³, connaît des exceptions dans le but de ne pas empiéter avec le droit à la liberté d'expression. Ainsi, l'article 17 (3) du RGPD prévoit que le droit à l'effacement ne s'applique pas lorsqu'il s'opposerait à un traitement de données à caractère personnel nécessaire :

« a) à l'exercice du droit à la liberté d'expression et d'information ;

b) pour respecter une obligation légale qui requiert le traitement prévue par le droit de l'Union ou par le droit de l'État membre auquel le responsable du traitement est soumis, ou pour exécuter une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement ;

c) pour des motifs d'intérêt public dans le domaine de la santé publique, conformément à l'article 9, paragraphe 2, points h) et i), ainsi qu'à l'article 9, paragraphe 3 ;

d) à des fins archivistiques dans l'intérêt public, **à des fins de recherche scientifique** ou historique ou à des fins statistiques conformément à l'article 89, paragraphe 1, dans la mesure où le droit visé au paragraphe 1 est susceptible de rendre impossible ou de compromettre gravement la réalisation des objectifs dudit traitement ; ou

e) à la constatation, à l'exercice ou à la défense de droits en justice. »

- Travailler sur des corpus d'images diffusées sur le Web implique de pouvoir les montrer (publications scientifiques). Il semblerait que cela contrevienne au droit d'auteur. **L'accord de l'auteur d'une image (caricature, photo) suffit-elle pour la publication lorsque celle-ci s'est trouvée transformée par ses circulations** (même internet) ?

Si l'image est une œuvre de l'esprit protégée par le droit d'auteur, dans ce cas, en dehors des parodies n'ayant pas l'intention de nuire, qui relèvent d'une exception au droit d'auteur, les autres réutilisations

⁹³ Pour les limites au droit à l'oubli dans le cadre d'une demande de déréférencement, voir : CJUE 13 mai 2014, Google Espagne contre AEPD, aff. C-131/12 et CJUE 24 septembre 2019 GC e.a. contre CNIL, aff. C-136/17.

ne peuvent généralement se faire qu'avec l'accord de l'auteur. Dans le cas contraire, il s'agit de contrefaçons, dont les responsables ne peuvent pas avoir la qualité d'auteurs. Dès lors, l'accord de l'auteur initial suffit.

- Nous travaillons sur un projet commun réunissant des collectes de tweets réalisées en 2013 ; 2014-2017 ; puis 2019-2021. Cela nous amène à analyser des comptes (users) et des images.
L'antériorité à l'adoption RGPD a-t-elle une incidence sur ce qu'il est possible de faire ?

Le RGPD n'a pas modifié en profondeur le droit des données à caractère personnel. Si le traitement de données effectué dans le cadre de cette recherche antérieurement à l'entrée en application du RGPD le 25 mai 2018 était conforme, il y a de fortes chances pour que cela soit encore le cas. Il peut toutefois s'avérer utile de le vérifier, ce qui nécessite de disposer d'informations complémentaires, notamment sur la nature de ces tweets, la façon dont ils ont collectés et la finalité de la recherche.

- Quand nous publions dans une revue américaine, sommes-nous soumis à des règles différentes eu égard au fait que le RGPD ne s'applique pas aux États-Unis ? (question sur le **champ d'application territorial** du RGPD)

Les États-Unis disposent de leurs propres règles en matière de respect de la vie privée. D'une façon générale, le droit états-unien est plus protecteur que le droit européen en matière de liberté d'expression. La revue est toutefois accessible depuis l'Europe, et sa publication implique le traitement de données à caractère personnel de personnes situées sur le territoire de l'Union. Dès lors, le RGPD s'applique en vertu de son article 3.

- **Les règles** en matière de protection des données **sont-elles les mêmes que les recherches soient financées** (Emergence SU ou ville de Paris, ANR) **ou non** (thèse sans contrat) ?

Oui. Dans les deux cas, l'établissement où se déroule la recherche contribue à fixer les modalités et les finalités du traitement. Dans le cadre d'une thèse sans contrat, cela vient du fait que les modalités d'obtention du diplôme (de doctorat) déterminent la façon dont la recherche sera menée. Il est dès lors de la responsabilité de l'université de fournir les moyens y compris à ses étudiant·e·s et doctorant·e·s, financées ou non, de mettre leurs traitements de données à caractère personnel en conformité.