

Traces numériques et recherche scientifique au prisme du droit des données personnelles

Article inédit, mis en ligne le 15 novembre 2018.

Julien Rossi

Julien Rossi est doctorant au laboratoire COSTECH de l'Université de technologie de Compiègne. Sa thèse porte sur les politiques publiques de protection des données à caractère personnel, qu'il étudie en adoptant une approche communicationnelle de l'action publique permettant d'analyser le rôle des représentations politiques dans la formulation et la mise en œuvre de ces politiques publiques. julien.rossi@utc.fr

Jean-Edouard Bigot

Jean-Edouard Bigot est doctorant en sciences de l'information et de la communication au laboratoire COSTECH de l'Université de Technologie de Compiègne. Ses recherches portent sur les transformations numériques des méthodes de recherche en sciences sociales. Il développe en particulier une approche techno-sémiotique de l'instrumentation numérique en questionnant les enjeux épistémologiques et politiques associés à l'utilisation de dispositifs d'exploitation de données numériques dans l'étude des phénomènes socioculturels. jean-edouard.bigot@utc.fr

Plan de l'article

Introduction

Promesses et critique des « traces numériques »

Le droit applicable et sa généalogie

La protection des données personnelles sur le terrain d'un projet en SHS

Conclusion

Références bibliographiques

RÉSUMÉ

La disponibilité croissante de masses de traces numériques inspire une nouvelle génération de projets de recherche numériquement instrumentés. Or l'exploitation de ces traces, en tant que le droit les qualifie de données à caractère personnel, entraîne l'application d'un corpus de normes juridiques dont le Règlement général de protection des données récemment adopté. Celui-ci prévoit un cadre dérogatoire pour la recherche scientifique, soutenu par une coalition d'acteurs issus de la recherche médicale, qui s'applique également aux sciences humaines et sociales (SHS). Le présent article présente une synthèse de travaux de recherche sur ce cadre légal et la façon dont il vient interroger le rapport aux traces numériques dans la recherche en SHS.

Mots clés

Traces numériques, droit des données personnelles, sciences humaines et sociales, méthodes numériques.

TITLE

Digital Traces and Scientific Research through the Lens of Data Protection Law

Abstract

The ever-increasing masses of digital traces available online have inspired a new generation of digitally equipped research projects. As digital traces legally qualify as personal data, their use is subject to a set of rules such as the recently adopted General Data Protection Regulation. These regulations provide a special legal framework that is applicable to the use of personal data for scientific research purposes, it is supported by a coalition of actors involved in medical research but applies to humanities and social sciences as well. This article presents a summary of findings from studies on the way this legal framework interacts with digitally equipped research in humanities and social sciences and blurs the lines between digital traces and personal data.

Keywords

Digital traces, personal data protection law, humanities and social sciences, digital methods.

TÍTULO

Huellas digitales e investigación científica a través del prisma de las leyes de protección de datos

Resumen

La disponibilidad creciente de masas de huellas digitales en línea inspira una nueva generación de proyectos de investigación optimizados para trabajo digital. Las leyes de protección de datos personales, entre las cuales se haya el reciente Reglamento general de protección de datos, incluyen esas huellas digitales en la categoría jurídica de los datos personales. Sin embargo, el Reglamento incluye un régimen derogatorio para el empleo de datos personales con fines de investigación científica, que fue apoyada por una coalición procedente de la investigación médica. Este artículo presenta de manera resumida las conclusiones de investigaciones sobre este marco legal, que se aplica también a las ciencias humanas y sociales, y que contribuye a difuminar el límite entre huellas digitales y datos personales.

Palabras clave

Huellas digitales, ley de protección de datos personales, ciencias humanas y sociales, métodos digitales.

INTRODUCTION

Le numérique prend progressivement une place accrue dans le champ des sciences humaines et sociales (SHS) (Bourdaloie, 2014) qui se trouverait ainsi « mis au défi » (Diminescu et Wiewiorka, 2015) au point que certains observateurs proclament l'avènement d'une « troisième génération de SHS » (Boullier, 2015 A et B). Dans ce contexte, « s'équiper numériquement »¹ deviendrait incontournable, même pour des disciplines *a priori*

.....

1. Nous nommons « recherche numériquement équipée » toute recherche s'appuyant sur des dispositifs d'exploitation de données numériques intervenant en position d'instruments dans les processus de connaissance (Bigot, 2018).

éloignées du numérique. Ces projets ont notamment pour matériau les « traces numériques ». Or qu'il s'agisse classiquement d'entretiens, de réponses à des questionnaires, ou de traces numériques, ces données entrent toutes dans la catégorie juridique des données à caractère personnel, entraînant l'application à ces projets du Règlement général de protection des données² (RGPD) adopté en 2016 et entrant en vigueur le 25 mai 2018, venu mettre à jour sans pour autant en remettre en cause les grands principes les lois antérieures telles que la loi française Informatique et Libertés de 1978 (Gellert, 2016).

Comment, si elle a lieu, s'établit la médiation entre ce droit et les projets de recherche évoqués plus haut, consommateurs de « traces numériques »?. Quelle est sa matérialité sur le terrain?

Pour répondre à ces questions nous avons conduit une recherche en quatre principales étapes. En parcourant la littérature académique sur les projets « numériquement équipés » en SHS et en portant un regard réflexif sur nos propres pratiques d'exploitation de données numériques, nous avons cherché à déterminer le régime juridique applicable en théorie à ce type de recherches, ainsi que la généalogie de ce régime et le ou les référentiels³ sous-jacents en ayant légitimé l'adoption. Nous avons étudié les échanges lors de deux journées d'études organisées sur l'application à la recherche en sciences sociales du droit des données personnelles⁴. Ces journées visaient à faire dialoguer des professionnels de la protection des données et des enseignants-chercheurs. Elles ont formé des arènes publiques éphémères permettant l'expression de débats (Badouard, Mabi, et Monnoyer-Smith, 2016) et de controverses (Smadja, 2012) n'ayant pas la place d'être exposés ici en détail, mais cependant intégrés à la présente analyse. Enfin, nous avons mené une campagne d'entretiens avec les chercheurs et chercheuses d'un projet de recherche financé par l'ANR, comprenant notamment des chercheurs en SIC et ayant impliqué le traitement statistique de données numériques personnelles dont des traces collectées en ligne.

Les données que nous avons collectées ont été analysées sous l'angle d'une approche communicationnelle du droit et des politiques publiques. Cette approche nous invite à considérer que les normes juridiques sont des instruments (Lascombes, 2004) dont le contenu est le fruit d'un processus délibératif impliquant des controverses parmi des acteurs impliqués dans un même champ de l'action publique (Dubois, 2014) ou sous-système de politique publique (Sabatier, 1998 ; Bergeron, Surel, et Valluy, 1998). Ces acteurs sont regroupés au sein de coalitions qui visent à imposer leur référentiel politique à l'ensemble du champ pour légitimer les dispositions juridiques concrètes pour lesquelles ils militent. La cartographie de ces coalitions permet de rendre compte de la structuration des controverses politiques en partant de l'hypothèse qu'elles se forment autour de référentiels idéologiques communs. Les textes juridiques rédigés puis adoptés par ces acteurs sont des outils techniques de gouvernementalité dont le fonctionnement repose en effet sur le principe de la performativité. Ceci les soumet à des conditions de succès, parmi

.....

2. 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.

3. Nous empruntons la notion de référentiel à Pierre Müller pour qui : « Elaborer une politique publique consiste donc d'abord à construire une représentation, une image de la réalité sur laquelle on veut intervenir. C'est en référence à cette image cognitive que les acteurs organisent leur perception du problème, confrontent leurs solutions et définissent leurs propositions d'action : cette vision du monde est le référentiel d'une politique ». (Müller, 2011, p. 57)

4. Il s'agit de la journée « Données personnelles en milieu universitaire : quelles questions se poser ? » organisée à l'Institut des sciences de la communication le 12 janvier 2017 et de la journée « Données personnelles et sensibles : quels enjeux pour la recherche en SHS ? » organisée par l'Institut des sciences sociales du politique à l'Université Paris Nanterre le 7 novembre 2017.

lesquelles le fait que « les participants aient l'intention d'adopter le comportement impliqué » par l'invocation d'un énoncé performatif qui lui-même suppose « certaines pensées ou sentiments » (Austin, 1970, p. 49) permettant d'en saisir le sens (voir également : Reinach, 2004). Or un texte de loi publié est un support matériel de médiation dont le contenu fait l'objet d'une réinterprétation par son public, et de controverses autour son interprétation, débattue dans des procès. Les rédacteurs d'un texte de loi peuvent d'ailleurs laisser à dessein des ambivalences rendre compatibles des visions incompatibles lorsqu'un compromis est nécessaire (voir sur un sujet similaire : Krieg-Planque, 2010), renvoyant en pratique au pouvoir judiciaire le soin de trancher. La réinterprétation du texte conditionne le sens donné à sa mise en actes, et donc les usages potentiels du texte juridique en tant qu'outil politique. On ne peut supposer d'équivalence entre le sens donné à un texte juridique par son rédacteur et celui que lui donne son lecteur ou son public, ce dernier pouvant ne prendre connaissance de son contenu qu'après une série de médiations successives.

Dans notre cas d'étude, ces textes juridiques sont réinterprétés en fonction de référentiels propres, non pas tant aux convictions individuelles des individus que nous avons interrogés, mais au référentiel sectoriel du champ de la recherche scientifique qui a ses propres normes, discours, valeurs et logiques d'intérêts. Nous verrons d'ailleurs que ce référentiel scientifique et son ethos peuvent parfois entrer en conflit avec les convictions personnelles des chercheurs ayant participé à l'enquête.

PROMESSES ET CRITIQUES DES « TRACES NUMÉRIQUES »

L'avènement d'une recherche numériquement équipée en SHS s'accompagne de discours, plus ou moins utopistes, concernant en particulier un renouvellement « positif », au double sens de « bénéfique » et de « positiviste », des paradigmes (Kuhn, 1962) des disciplines des sciences de la culture grâce à la possibilité offerte par les technologies numériques de se saisir de données sociologiques massivement récoltées sur les réseaux informatiques, envisagées comme des traces de pratiques sociales. Ces discours et les promesses dont ils sont porteurs sont particulièrement présents dans le projet du courant des dites méthodes numériques (Rogers, 2009 et 2013 ; Venturini et Latour, 2009 ; Rieder, 2010 ; Rieder et Röhle, 2012 ; Venturini *et al.*, 2014).

Selon le constat fréquemment émis par les promoteurs des méthodes numériques, le développement des technologies numériques s'accompagne d'une production massive de données censées refléter les échanges sociaux qui se produisent en ligne, et de plus en plus ceux qui ont lieu hors ligne aussi. Ces données numériques massives proviendraient d'abord des données d'utilisation des plateformes de réseautage social comme Facebook ou Twitter mais aussi, plus globalement, de l'usage de n'importe quelle application informatique. Il est commun de reconnaître que l'utilisation des médias informatisés s'accompagne toujours de la production et du stockage de deux types d'informations : d'une part les informations produites par les utilisateurs (création et partage de contenus, informations personnelles délivrées lors de la création de profils, *etc.*) et d'autre part les informations résultant d'un enregistrement automatique des actions effectuées par les utilisateurs (temps passé sur un site, contenus consultés, les liens hypertextes actionnés, *etc.*). C'est ce dernier type de données, distinctes par le caractère non-intentionnel de leur production, que les praticiens des méthodes numériques désignent généralement par le terme de « traces » (voir : Venturini, 2012).

Toute interaction sociale médiée par une application numérique tendrait donc aujourd'hui à être tracée et cette nouvelle traçabilité sociale serait rendue sous la forme d'un vaste ensemble de données numériques largement accessibles sur les réseaux infor-

matiques publics, *via* les API des sites web et *via* des services, publics ou privés, dédiés à la diffusion d'informations numérisées. Les technologies numériques offriraient ainsi un réservoir potentiellement illimité de données sociales dont il serait aisé de se saisir dans un contexte de recherche, celles-ci représentant une opportunité pour les SHS de mieux comprendre les transformations contemporaines de la société.

La disponibilité de ces traces numériques permettrait un accès facilité au terrain pour les chercheurs grâce à la fois à une réduction des « coûts » humains, économiques et temporels de la constitution de corpus, et une ouverture croissante des données qui les rendent plus aisées à obtenir, à agréger et à manipuler. En découlerait l'autorisation d'une observation *in vivo* des pratiques sociales qui se produisent sur le Web :

« L'intérêt des médias électroniques est que toute interaction qui les traverse laisse des traces qui peuvent être facilement enregistrées, massivement stockées et aisément récupérées [...] offrant aux sciences sociales plus de données qu'elles n'en ont jamais rêvées. [...] Nées dans une époque de pénurie, les sciences sociales entrent dans un âge d'abondance ». (Venturini et Latour, 2009)

La possibilité d'une transformation des sciences sociales par l'exploitation appareillée de traces numériques, aux fondements du projet des méthodes numériques, a fait l'objet d'importantes critiques émanant en particulier des SIC. Elle sont relatives à la notion de « trace », à son épistémologie et aux enjeux de pouvoir qui la traversent (Merzeau 2009 et 2013 ; Flon *et al.*, 2009 ; Jeanneret, 2011 et 2013 ; Collomb, 2016). Ces critiques visent fondamentalement à déconstruire l'apparente évidence de la trace et s'opposent à une conception naturalisante des traces numériques. Yves Jeanneret invite ainsi à dépasser une conception indicielle des traces numériques en appelant à réinscrire ces objets dans les logiques communicationnelles complexes qui les sous-tendent. Dans cette perspective communicationnelle, les traces sont des objets sémiotiques qui résultent d'une construction culturelle relevant d'une élaboration documentaire : documentation, médiatisation, archivage, *etc.* Parce qu'elles sont toujours liées à des pratiques d'écriture, les traces numériques doivent être envisagées comme des inscriptions, qu'elles soient intentionnelles ou non. Si ces données sont générées et collectées de manière automatique, elles sont aussi traitées, c'est-à-dire matérialisées dans un espace qui leur donne sens, les organise, leur assigne une place, et les inscrit dans un projet interprétatif. Il y a donc un problème à considérer les données récoltées sur les réseaux comme des traces de pratiques sociales, sans en interroger les conditions de traçabilité et de traitement. Aussi, il convient d'adopter une définition de la trace qui prenne en compte les médiations à la fois logistiques et sémiotiques qui permettent d'interpréter certaines productions communicationnelles comme traces d'identité, traces d'usages ou traces de phénomènes sociaux (Jeanneret, 2011). De la même manière, Cléo Collomb (2016) insiste sur le fait que ce que l'on désigne communément comme des traces doit d'abord être considéré comme un ensemble d'inscriptions réalisées par l'interaction d'un agent humain et d'un programme informatique qui double, selon une logique de supplément et selon des modalités spécifiques à la médiation techno-sémiotique qu'il opère, des activités d'utilisation dans un média informatisé. Nous souscrivons ici à cet ensemble de définitions apporté par les SIC, centré sur la dimension à la fois matérielle, technique, et sémiotique des traces numériques.

Les réflexions critiques conduites par les SIC au sujet des traces numériques et, plus largement, sur les questions liées aux imaginaires du « *big data* » et leur rapport avec la problématique de la surveillance par les données (Carré et Panico, 2011 ; Rouvroy et Berns, 2013) apportent des éclairages originaux et précieux aux enjeux des données numériques. En outre, il nous apparaît tout aussi important de pouvoir aborder les problématiques associées au cadre spécifique des données à caractère personnel collectées à des fins de recherche, et la façon dont celui-ci est reçu par les chercheurs qui font usage de données numériques.

LE DROIT APPLICABLE ET SA GÉNÉALOGIE

En prenant appui sur les études précédemment menées sur la genèse du droit à la protection des données (Hondius, 1975 ; Newman, 2008 ; Vitalis, 2009 ; Atten, 2013 ; Gonzalez Fuster, 2014) nous avons constitué un corpus de documents d'archives issus principalement des groupes de travail du Conseil de l'Europe, de l'OCDE puis de l'Union européenne, à l'origine des notions juridiques de « donnée à caractère personnel » et de « protection des données ». Ces documents ont été complétés par des entretiens qualitatifs avec 16 décideurs publics, pour certains d'entre eux depuis la fin des années 1960, impliqués dans le champ de l'action publique relative à la protection des données à caractère personnel. Ceux-ci nous ont permis de conclure à l'existence d'un réseau d'acteurs réunis dans une coalition de cause reposant sur le « paradigme de la vie privée » (Bennett et Raab, 2003) inspiré de la philosophie utilitariste de John Stuart Mill (1989) et des travaux de Michel Foucault sur le panoptique (Foucault, 1975 ; Simon, 2002). Pour ces acteurs, l'informatisation de la société aboutit à la mise en place de dispositifs de surveillance panoptiques mettant en péril la liberté et l'autonomie des individus notamment par la mise en place d'une gouvernementalité par le calcul (Rouvroy et Berns, 2013) et par la sophistication de mécanismes de discrimination fondée sur l'exploitation de données personnelles (Lyon, 2015). Cette conception est devenue le référentiel sectoriel dominant dans le sous-système de politique publique de la protection des données. Elle suppose de laisser l'individu maître de décider du sort de ses données personnelles par un mécanisme de « droit à l'autodétermination informationnelle » (Tribunal constitutionnel fédéral allemand, 1983) plutôt que de définir par avance et pour tout le monde ce qui relève de la sphère privée ou de la sphère publique.

Cette logique de droit à l'auto-détermination informationnelle, consacrée par le législateur européen, n'opère pas de choix *a priori* entre ce qui relève du privé et du public. La définition juridique qui s'est stabilisée progressivement dans différents textes juridiques à partir des années 1970 laisse ainsi à l'individu le pouvoir de négocier la frontière entre public et privé :

« 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 du RGPD)

Comme il a été plusieurs fois rappelé par la Cour de justice de l'Union européenne :

« [...] les notions de « données à caractère personnel » [...] et de « données relatives à la vie privée » ne se confondent pas [...] ». (CJUE 2015, pt. 32)

De surcroît, « il n'est pas requis que toutes les informations permettant d'identifier la personne concernée se trouvent entre les mains d'une seule personne » (CJUE 2017, pt. 31) pour qu'une donnée soit qualifiée de personnelle. De nombreux travaux en informatique repris par la doctrine juridique ont en effet montré la difficulté pratique d'anonymiser des données personnelles. Il demeure généralement possible, par croisement de données quasi-identifiantes, de retrouver une personne, même indirectement (Sweeney, 2000 ; G29, 2007 ; Ohm, 2010 ; Mascetti *et al.*, 2013 ; G29, 2014 ; de Montjoie *et al.*, 2015 ; Rossi, 2015), ce qui est suffisant pour entrer dans le cadre de la définition citée ci-dessus. En outre, comme le rappelle le considérant 26 du RGPD au sujet des données pseudonymes :

« [...] Les données à caractère personnel qui ont fait l'objet d'une pseudonymisation et qui pour-

raient être attribuées à une personne physique par le recours à des informations supplémentaires devraient être considérées comme des informations concernant une personne physique identifiable [...] ». (RGPD, cons. 26)

En pratique, toute trace numérique devient alors susceptible d'être qualifiée de donnée à caractère personnel, entraînant l'application d'un grand nombre de contraintes pouvant comprendre le consentement de la personne concernée, la limitation de la durée de conservation des données ou encore une obligation d'information complète et transparente qui peuvent poser des problèmes méthodologiques :

« À mon sens, quand je vous écoute, je me demande si tout un tas d'enquêtes sociologiques, notamment, qui ont été menées je pense, sur les milieux soit très difficiles d'accès, soit ésotériques, auraient été possibles avec ce type d'impératifs légaux ». (ISP, 2017b, 11m 10s secondes à 11m 32s)

« Si on utilisait un dispositif formel d'interrogation d'entretiens, on n'allait pas potentiellement [...] avoir le même type de parole que celle qui nous est dite de manière informelle ». (entretien 8)

De plus, il est parfois impossible de recueillir le consentement des personnes ou de les informer de façon parfaitement transparente. Cela peut même s'avérer dangereux pour les chercheurs notamment lorsqu'ils travaillent sur des groupes sociaux violents, comme en a témoigné Fanny Georges (2018), responsable scientifique du projet MINWEB sur la circulation de messages politiques dans la zone de conflits qu'est le Sahara.

L'article 89 du RGPD fournit à la recherche scientifique un certain nombre de dérogations pour répondre à ces préoccupations, dont la portée exacte doit cependant être précisée par les législateurs nationaux. L'appréciation du caractère d'intérêt public de la recherche menée sera un élément d'appréciation pour la CNIL pour déterminer dans quelle mesure ces dérogations pourront s'appliquer :

« Par exemple, [...] pour des recherches qui seraient considérées comme d'intérêt public, dans le cadre d'une mission de service public, le consentement peut ne pas être recueilli. [...] Même sur le plan de l'information [...] on peut aussi considérer que l'information des personnes représenterait un effort disproportionné par rapport à l'intérêt même de la recherche qui est menée et de la finalité poursuivie ». (Sophie Genvresse, ISP, 2017b, 12m01s à 12m37s)

Or la légitimité de ces dérogations a fait l'objet de remises en cause au cours du processus d'adoption du RGPD.

L'article 89 n'a fait l'objet que de peu de lobbying au cours du processus d'adoption du RGPD. Ainsi, sur 227 *position papers* en anglais recensés dans notre corpus, seuls 21, soit à peine 10 %, portaient au moins en partie sur le statut de la recherche scientifique. Ces 21 documents ont été rédigés par 15 organismes répartis comme décrit par les graphiques ci-dessous, par secteur d'activité et par type :

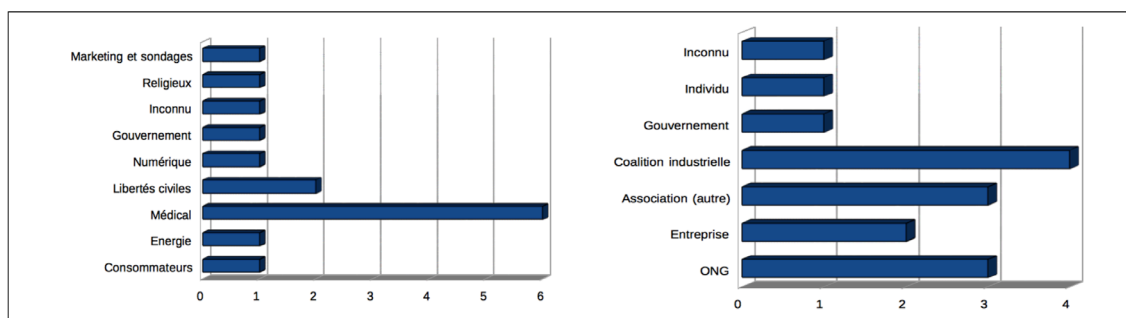


Figure 1. Répartition des organismes par domaine d'activité (à gauche) et par type (à droite).

Ces organismes se regroupent autour de deux coalitions de cause.

La coalition médico-pharmaceutique, avec notamment la Fédération européenne des académies de médecine (FEAM) et Comité de coordination européen de l'industrie radiologique, électromédicale et d'informatique médicale (COCIR), portait une position communautariste au sens d'Amitai Etzioni (1999). En ce sens, ils privilégient l'intérêt de la communauté et les objectifs de santé publique aux intérêts individuels de vie privée, ainsi qu'une confiance dans l'utilité du traitement massif de données de santé passant par la réutilisation des dossiers de patients pour la recherche médicale. Cette position justifiait selon eux des dérogations importantes pour la recherche, comme l'abaissement des exigences d'anonymisation au profit d'une simple pseudonymisation et une limitation considérable des droits des enquêtés quant à leurs données personnelles dès lors qu'elles étaient exploitées, avec ou sans leur consentement, dans une finalité de recherche.

L'autre coalition, portée par les associations Bits of Freedom et European Digital Rights, visait à exclure certaines données sensibles et celles portant sur des mineurs, sauf consentement des enquêtés, selon le cadre dérogatoire de l'article 89. Son périmètre aurait aussi été restreint à des recherches perçues comme plus légitimes que d'autres, par exemple celles bénéficiant de la protection de l'article 13 de la Charte des droits fondamentaux de l'Union européenne, qui protège la liberté académique. Cette coalition adhère au référentiel du paradigme libéral de la vie privée (Bennett et Raab, 2003) précédemment évoqué.

Ainsi pour l'association European Digital Rights :

« Processing of sensitive data for historical, statistical and scientific research purposes is not as urgent or compelling as public health or social protection. Consequently, there is no need to introduce an exception, based on national law, which would put them on the same level as the other listed justifications, which risks undermining fundamental rights, legal certainty and the single market⁵ ». (EDRi, 2012p. 19)

Le copier-coller de ce texte se retrouve dans le projet de rapport de Jan-Philip Albrecht, rapporteur du projet de règlement qui a donné le RGPD (Albrecht, 2013, p. 24), qui voit en les promesses du *Big Data* un avenir dystopique :

« If nothing else, would it not make sense to ignore or even oppose the Big Data efficiency enhancement logic simply because it makes us into machines being driven to the limits of their efficiency. For that logic will make society into a sterile place full of clean clones [...]. The mass analysis of our personal data makes the neoliberal logic of the exploitation of human life into the all-encompassing mantra of the post-modern world⁶ ». (Albrecht, 2015, p. 93)

A défaut de participation au débat par des chercheurs en SHS sur ce qu'est devenu l'article 89 du RGPD dans les arènes institutionnelles pertinentes pour viser à en fonder la légitimité sur d'autres bases que les attentes du *Big Data* pour la recherche médicale et pharmaceutique, cette reprise n'a rien de surprenant.

.....

5. « Le traitement de données sensibles à des fins historiques, statistiques ou scientifiques n'est ni aussi urgent ni aussi impératif que la santé publique ou la protection sociale. Par conséquent, il n'y a aucun besoin d'introduire une exception, fondée sur le droit national, qui mettrait ces justifications au même niveau, car cela porterait atteinte aux libertés fondamentales, au principe de sécurité juridique et au marché unique ». (Traduction des auteurs)

6. « Cela n'aurait-il pas de sens d'ignorer voire de s'opposer à la logique d'accroissement de l'efficacité propre au Big Data, pour la simple raison que cela fait de nouveau des machines poussées à la limite de leur efficacité ? Car cette logique transformera la société en un lieu stérile plein de clones propres [...]. L'analyse de masse de nos données personnelles transforme la logique néolibérale d'exploitation de la vie humaine en un mantra qui englobe l'ensemble de notre monde post-moderne ». (Traduction des auteurs)

Comme nous l'avons vu, le cadre juridique du RGPD, qui n'est qu'une évolution de règles en vigueur en France depuis la loi Informatique et Libertés de 1978, n'est pourtant pas sans poser des difficultés, notamment méthodologiques, aux SHS. Nous avons donc mené une enquête sur un projet de recherche concret pour étudier la réception de ces contraintes juridiques faite sur le terrain par des chercheurs en SHS, dont une forte proportion est située en SIC.

LA PROTECTION DES DONNÉES PERSONNELLES SUR LE TERRAIN D'UN PROJET EN SHS

Parmi les chercheurs que nous avons interrogés dans le cadre d'entretiens semi-directifs, huit ont participé ensemble à un projet de recherche en sociologie et sciences de l'information et de la communication financé par l'ANR⁷. Les questions posées visaient à comprendre leur réception du droit des données personnelles, en lien avec les questionnements éthiques devant lesquels leurs collectes de traces et données numériques les plaçaient.

Le projet de recherche dans lequel les répondants étaient investis collectait des données personnelles de plusieurs façons. D'une part, il s'agissait d'effectuer une collecte de données en ligne numériquement instrumentée, notamment par le biais des réseaux socio-numériques. D'autre part, des données étaient générées par des questionnaires et des entretiens retranscrits avec des enquêtés dont la caractéristique commune était d'avoir vécu le type de traumatisme personnel sur lequel portait le projet. L'ensemble de ces données fit l'objet d'un traitement statistique.

L'absence presque complète de médiation entre les chercheurs interrogés et les acteurs de la protection des données à caractère personnel, qu'il s'agisse d'associations ou de juristes externes à l'université tels que ceux de la CNIL, ou qu'il s'agisse de services juridiques internes et notamment des correspondants informatique et libertés (CIL), est un constat qui s'est très vite imposé dans tous les entretiens.

Les CIL sont des personnes chargées de superviser la mise en conformité aux règles de la protection des données personnelles de leur employeur (Rossi, 2015). Appelées « délégués à la protection des données » dans le RGPD, leur désignation était encouragée depuis 2004 par la loi Informatique et Libertés. En juin 2017, grâce aux jeux de données rendus disponibles sur data.gouv.fr, nous avons ainsi répertorié 137 CIL désignés par des établissements publics d'enseignement supérieur et de recherche, tous internes à l'organisation. Parmi ceux-ci, seuls 24 % disposaient d'un site web permettant d'identifier les coordonnées du CIL, et 20 % n'affichaient même pas les mentions légales obligatoires. Cela est symptomatique d'une faiblesse institutionnelle des CIL, qui n'ont pas forcément les moyens d'exercer leurs missions.

Ainsi même, les enquêtés qui avaient connaissance de l'existence du CIL nous ont indiqué n'avoir jamais reçu de réponse à leurs sollicitations de la part de celui ou celle de leur université :

« Au démarrage, [...] on a pris contact avec le CIL de l'Université XXX. Pour le dire très simplement, on n'a jamais eu de retour de l'Université XXX ». (entretien 2)

.....

7. Nous tenons à remercier chaleureusement chacune de ces personnes pour leurs contributions. Pour des raisons de confidentialité, le nom du projet et, sur leur demande, les noms de certains participants à l'enquête ont été anonymisés.

L'absence de communication entre le champ de la protection des données et celui de la recherche en SHS à l'université contribue à expliquer des approximations juridiques et des confusions exprimées par les membres du projet interrogés, et ce nonobstant l'expression de leurs préoccupations éthiques vis-à-vis des enquêtés et leurs données collectées.

Par exemple, ainsi que l'exprimait un CIL présent lors de la journée d'étude du 12 janvier 2017 :

« Il y a souvent la confusion entre un projet de recherche totalement anonyme [...] et le rendu anonyme, en fait. Au départ on traite quand même bien des données personnelles qui n'ont rien d'anonyme. C'est juste le résultat qui le devient ».

Son constat d'une confusion entre l'anonymat des données et celui du rendu est recoupé par les entretiens que nous avons réalisés :

« J'estimais [que] le côté problématique, ce n'était pas tellement les bases dans les serveurs. [...] Ce qui me posait problème, c'était plutôt au niveau de la publication. C'est-à-dire ce que j'allais pouvoir raconter sur des gens qui n'avaient rien demandé ». (entretien 6)

« Je ne me suis pas posé la question des données informatiques : qu'est-ce qui restait des fichiers par exemple ? Je ne sais même pas où est-ce qu'ils sont ces fichiers aujourd'hui ». (entretien 8)

Il existe par ailleurs une confusion fréquente entre la notion juridique de « donnée à caractère personnel » et celle de donnée relative à l'intimité des personnes, accompagnée du sentiment que travailler sur des données accessibles publiquement permet de contourner l'application de la loi alors que cet élément n'est qu'un élément à prendre en compte parmi d'autres dans l'analyse juridique :

« Les personnes ont mis leurs données sur internet. Donc elles sont censées savoir qu'elles les ont rendues publiques [...]. Et c'est là aussi où je pense que la question juridique n'est pas suffisante pour moi. [...] Oui, d'un point de vue juridique on peut considérer qu'elles sont publiques » (entretien 3).

Si la réponse ne leur semblait pouvoir venir d'un usage du droit des données à caractère personnel, les chercheurs interrogés ont tous fait part de réflexions éthiques. Ces approches, parfois divergentes et sujettes à controverse au sein de l'équipe, ont pour point commun cette importance accordée à la préservation de l'anonymat des enquêtés dans les résultats publiés tant pour des raisons méthodologiques visant à permettre aux enquêtés une expression plus libre et donc de meilleure qualité, que pour des raisons éthiques.

Confirmant les travaux de Latanya Sweeney (2000) et Paul Ohm (2010) sur la difficulté d'anonymiser de façon robuste, c'est-à-dire sans risque aucun de ré-identification des données de recherche par qui que ce soit, les chercheurs interrogés ont fait part de réflexions élaborées sur la façon d'aboutir à l'anonymat dans la publication. L'usage d'un corpus documentaire issu d'extractions de données du Web pose un problème spécifique et nouveau renforçant la difficulté de l'anonymisation :

« Même si j'ai anonymisé tous les noms de personnes [...] j'ai recopié sur mon [document public] des messages tels quels, tels qu'ils sont publiés sur le site internet. C'est-à-dire qu'une personne qui veut retrouver l'identité de la personne qui a écrit, il suffit de taper sur Google le message [...] Donc en fait je n'ai pas du tout anonymisé ». (entretien 3)

La spécificité de l'enquête à partir de données publiées sur le Web est en effet la simplicité de la ré-identification par croisement avec des données publiquement accessibles :

« Si ça avait été fait dans le cadre d'un entretien privé, personne n'aurait pu remonter à l'information. Là [sur internet] tout le monde peut remonter à l'information ». (entretien 3)

Cette question se pose également dans le cas de la collecte par questionnaires. Le questionnaire diffusé par la Société française des sciences de l'information et de la communi-

cation (SFSIC) sur l'éthique de la recherche en ligne est symptomatique de la difficulté rencontrée. Outre divers manquements à la législation tel que l'absence des mentions légales obligatoires, alors qu'il prétendait être anonyme, et qu'il le sera sans doute dans le rendu qui en sera effectué, il contenait des injonctions à la fourniture, sous peine d'être exclu de l'enquête, de renseignements individuels précis. Ces derniers pouvaient aboutir à la possibilité d'identifier une bonne partie des répondants par croisement avec des informations accessibles sur le Web comme la combinaison des langues parlées par la personne (en regardant les langues des publications de la personne, ou son CV en ligne), ou encore leur statut et leur date de naissance notamment lorsqu'il s'agit de dirigeants d'institutions académiques. Ceci était d'autant plus problématique que le questionnaire incitait les répondants à formuler des critiques vis-à-vis d'institutions et de lieux de pouvoir dont la SFSIC elle-même.

The figure displays two screenshots of a questionnaire titled "Ethique de la recherche à l'ère du numérique".

The left screenshot shows the "III) Profil" section. It includes a question "Vous êtes :" with radio buttons for "Un homme" and "Une femme". Below this is a "Statut" section with a grid of checkboxes for various roles: Etudiant en Licence ou Master, Doctorant, Docteur, Maître de conférences, Maître de conférences HDR, Professeur des universités, Ingénieur d'étude, Ingénieur de recherche, Salarié du privé, Chef d'entreprise, Chômeur, Retraité, and Autre. A red box indicates "Veuillez sélectionner au moins 1 choix". At the bottom, there is a question "Quelle est votre discipline ?" with a text input field and a red box stating "La réponse à cette question est obligatoire".

The right screenshot shows the "Avez-vous des responsabilités de direction scientifique ?" section with checkboxes for "Direction de laboratoire", "Direction d'axe", "Direction d'équipe", "Direction de projets de recherche collectifs", "Direction d'école doctorale", "Aucune", and "Autre". A red box indicates "Veuillez sélectionner au moins 1 choix". Below this is a "Votre âge" section with a dropdown menu and a red box stating "La réponse à cette question est obligatoire". At the bottom, there is a question "Avez-vous par le passé exercé des activités de recherche à l'étranger ?" with radio buttons for "Oui" and "Non", and a red box stating "La réponse à cette question est obligatoire". Navigation buttons "Précédent" and "Suivant" are visible at the bottom of both screenshots.

Figure 2. Captures d'écran du questionnaire de la SFSIC (2017)

Pour ne pas simplifier les choses, il peut arriver que l'anonymat, y compris lors de la publication, ne soit impossible ou qu'il entre en contradiction avec les finalités d'une recherche perçue comme ayant un intérêt public. Cela est typiquement le cas lorsqu'un chercheur travaille sur une personnalité publique.

Il peut arriver que des pressions s'exercent par ailleurs sur les chercheurs pour qu'ils partagent des données de recherche pourtant confidentielles. Ces pressions peuvent venir des services de police ou de renseignement dans certains cas, de l'industrie, mais également d'institutions de recherche incitant au partage et à l'ouverture des données pour favoriser la transparence de la recherche et le partage des connaissances :

« J'étais forcée pour mon DESS de publier [...] Et j'avais fait une erreur. Parce qu'en fait, mon DESS me forçait de partager un [rapport contenant des données confidentielles] ». (entretien 7)

Cette pression avait également été évoquée par une des conférencières de la journée d'étude du 12 janvier 2017 qui y présenta la façon dont elle avait mobilisé des arguments tirés du droit à la protection des données pour résister à des pressions visant à ce qu'elle abaisse le niveau d'anonymisation de ses résultats d'enquête sur des données au caractère politique sensible :

« Il s'en est suivi une négociation, au cours de laquelle j'ai à nouveau été confrontée à ma question initiale : pour qui ? Notamment, en en discutant au sein de l'institut polonais, j'ai entendu qu'on ne fait pas [...] de recherche pour les [enquêtés]. Il s'agissait donc de respecter à la fois la loi et les normes du milieu, mais aussi de faire un choix de nature plus morale, de se prononcer sur la question de responsabilité et de peser les bénéfices d'un large accès au savoir contre le risque de nuire aux participants et de trahir leur confiance ». (Skowronska, 2017)

Certains chercheurs débutants ont exprimé ressentir un décalage entre leurs convictions personnelles et un certain ethos du partage et de la transparence de la recherche scientifique, qui ne les amène cependant pas, dans le cas du projet ANR que nous avons étudié, à se tourner vers le droit. De façon générale l'analyse des entretiens a montré que le lien entre les convictions individuelles exprimées par les enquêtés et l'usage des outils juridiques de protection des données était généralement faible, même lorsqu'ils avaient conscience par ailleurs de l'existence de ce droit. Les cas où la notion de confidentialité des « données personnelles » est évoqué sous un angle juridique par les enquêtés relevaient en réalité souvent d'une confidentialité au titre du droit de la propriété intellectuelle et visait à se protéger d'une dépossession de ses résultats de recherche face à l'institution universitaire ou à des collègues en qui la confiance n'est pas pleinement assurée.

Cette absence relative de questionnement en sciences humaines et sociales sur le cadre juridique de la collecte de « traces numériques », qui n'empêche en rien le déroulement de réflexions éthiques individuelles, provenait dans certains cas d'un sentiment initial de dépersonnalisation de ces traces et d'impression initiale d'anonymat erronée ou mal appréciée. L'apparition de l'aspect personnel des traces, dès lors qu'elles sont structurées sous une forme en permettant l'analyse, peut cependant prendre le chercheur par surprise :

« J'essayais de faire une variable qui pouvait coder [une caractéristique du comportement des personnes sur lesquelles les données étaient recueillies]. [...] En gros j'attribuais un score à chaque modalité. Plus on allait sur des plateformes d'un certain type, plus on avait un score élevé, en gros. Et c'est vrai que moi j'ai sorti ça, juste en statistique. Et après en fait ça nous donnait [une variable de score comportemental]. Enfin c'est hyper bizarre. Enfin y a des choses comme ça où on le fait statistiquement mais avec notre objet ça colle pas. Et là-dessus, même si c'était pertinent, on a complètement laissé tomber parce que c'était très bizarre en fait. [...] Là ce n'était vraiment pas une question statistique, c'était vraiment une question éthique je pense ». (entretien 5)

Pour finir, même lorsque des données sont générées hors ligne par exemple par la sollicitation d'entretiens, leur analyse suppose aujourd'hui une inscription sur un support numérique, souvent en ligne, qui peut aboutir à une perte de contrôle informationnel sur les données. L'usage de services d'hébergement en ligne des données que les universités, dans le cas que nous avons étudié, refusaient de fournir elles-mêmes, renforçait chez les chercheurs ayant répondu à notre enquête un sentiment d'impuissance quant à leur capacité à sécuriser l'accès aux données qu'ils recueillent et à véritablement garantir le contrôle de leur circulation.

CONCLUSION

Les entretiens que nous avons réalisés montrent une absence de médiation entre les chercheurs impliqués dans le projet étudié et le champ de la protection des données à caractère personnel, en dehors de quelques rares occasions comme les deux journées d'étude évoquées et répertoriées en bibliographie. Dès lors, une multitude de pratiques, reposant essentiellement sur l'anonymisation des publications, sont développées pour préserver la confidentialité des enquêtés. Le passage au numérique affaiblit cependant la frontière entre les données anonymes et les données personnelles. Ainsi toute trace numérique est potentiellement une donnée personnelle au sens de la loi. Ce lien avec l'origine personnelle des traces peut d'ailleurs apparaître lors de la structuration et de la fouille des traces numériques collectées une fois celles-ci structurées de façon exploitable et ce même lorsqu'une anonymisation des résultats publiés demeure possible.

Le cadre légal de la protection des données à caractère personnel ne s'applique pas uniquement à la publication des résultats. Il fixe des exigences en matière de sécurité infor-

matique, en matière de délais de conservation, et des normes encadrant l'information des enquêtés ou encore le recueil de leur consentement sauf motif d'intérêt légitime dans le cadre d'une mission de service public qui doit être justifiée. Cela peut soulever des problèmes méthodologiques de nature à affecter la qualité de la recherche.

Malgré l'absence d'acteurs issus des SHS dans les discussions sur le cadre dérogatoire applicable à la recherche scientifique prévu par le RGPD, il demeure une possibilité de déroger à certaines de ces règles pour des impératifs matériels ou méthodologiques. Mais la possibilité de recourir à ces règles est soumise à une interprétation dont la compétence, au sens juridique, n'appartient plus exclusivement au chercheur.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Albrecht, Jan Philipp (2015), *Hands off our data! Bruxelles* : Jan Philipp Albrecht.
- Atten, Michel (2013), « Ce que les bases de données font à la vie privée, What databases do to privatelife », *Réseaux*, n° 178-179, p. 21-53.
- Austin, John Langshaw (1970), *Quand dire, c'est faire*, Paris : Éditions du Seuil.
- Badouard, Romain ; Mabi, Clément ; Monnoyer-Smith, Laurence (2016), « Le débat et ses arènes », *Questions de communication*, n° 30, p. 7-23.
- Bennett, Colin J., ; Raab Charles D. (2003), *The Governance of Privacy. Policy Instruments in Global Perspective*. Aldershot : Ashgate.
- Bergeron, Henri ; Surel, Yves, et Valluy, Jérôme (1998), « L'Advocacy Coalition Framework. Une contribution au renouvellement des études de politiques publiques? », *Politix* 11 (41), p. 195-223.
- Bigot, Jean-Édouard (2018), *Instruments, pratiques et enjeux d'une recherche numériquement équipée en sciences humaines et sociales*, Thèse en Sciences de l'information et de la communication, Université de Technologie de Compiègne.
- Boullier Dominique (2015A), « Les sciences sociales face aux traces du *Big Data* », *Revue française de science politique*, 5-6, p. 805-828.
- Boullier, Dominique (2015B), « Pour des sciences sociales de troisième génération (SS3G) : des traces numériques aux répliques », in Menger, Pierre-Michel (coord.), *Big Data, entreprises et sciences sociales*, Paris : OpenEdition.
- Bourdeleio, Hélène (2014), « Ce que le numérique fait aux sciences humaines et sociales. épistémologie, méthodes et outils en questions », *Tic & société*, vol. 7, n° 2, p. 7-38
- Carré, Dominique et Panico, Robert (2011), Le contrôle social à l'heure des technologies de mobilité et de connectivité, du fichage ciblé des individus au traçage continu des agissements, *Terminal*, n° 108-109, p. 17-31
- Collomb, Cléo (2016), *Un concept technologique de trace numérique*, Thèse en philosophie, Université de Technologie de Compiègne.
- Dubois, Vincent (2014), « L'Etat, L'action Publique et La Sociologie Des Champs », *Swiss Political Science Review*, 20 (1), p. 25-30.
- Etzioni, Amitai (1999), *The Limits of Privacy*. New York : Basic Books.
- Flon, Émilie ; Davallon, Jean ; Tardy, Cécile et Jeanneret, Yves (2009), « Traces d'écriture, traces de pratiques, traces d'identité », *Rétrospective et Perspective*, Actes du colloque H2PTM, Paris : Hermès-Lavoisier.

- Foucault, Michel (1975), *Surveiller et punir. Naissance de la prison*, Paris : Gallimard.
- Gellert, Raphael (2016), « We Have Always Managed Risks in Data Protection Law : Understanding the Similarities and Differences between the Rights-Based and the Risk-Based Approaches to Data Protection », *European Data Protection Law Review*, vol. 2, n° 4, p. 481-492.
- Gonzalez Fuster, Gloria (2014), *The Emergence of Personal Data Protection as a Fundamental Right of the EU*, Dordrecht : Springer.
- Groupe de travail de l'Article 29 (2007), *Avis 04/2007 sur le concept de donnée à caractère personnel, adopté le 20 juin*, Bruxelles : Commission européenne
- Groupe de travail de l'Article 29 (2014), *Avis 05/2014 sur les techniques d'anonymisation, adopté le 10 avril*, Bruxelles : Commission européenne
- Hondius, Frits Willem (1975), *Emerging data protection in Europe*. Amsterdam : Elsevier.
- Jeanneret, Yves (2011), « Complexité de la notion de trace. De la traque au tracé » (p. 59-86), in Galinon-Méléneq, Béatrice (coord.), *L'homme trace. Perspectives anthropologiques des traces contemporaines*, Paris : CNRS Éditions.
- Jeanneret, Yves (2013), « Les chimères cartographiques sur l'internet, panoplie représentationnelle de la traçabilité sociale » (p. 235-267), in Galinon-Méléneq, Béatrice ; Zlitni, Sami (dirs.), *Traces numériques : de la production à l'interprétation*, Paris : CNRS Éditions.
- Krieg-Planque, Alice (2010), « La formule "développement durable" : un opérateur de neutralisation de la conflictualité », *Langage et société*, n° 134, p. 5-29.
- Kuhn, Thomas (1962), *La structure des révolutions scientifiques*, Paris : Flammarion.
- Lascoumes, Pierre (2004), « La Gouvernamentalité : de la critique de l'État aux technologies du pouvoir », *Le Portique*, n° 13-14, p. 1-15.
- Lyon, David (2015), *Surveillance After Snowden*, Cambridge, Mass. : Polity Press.
- Mascetti, Sergio ; Monreale, Anna ; Ricci, Annarita ; Gerino, Andrea (2013), «Anonymity : A Comparison Between the Legal and Computer Science Perspectives » (p. 85-115), in Gutwirth, Serge ; Leenes, Ronald ; de Hert, Paul ; Pouillet, Yves ; Flinn, Rachel ; Wright, David ; Friedewald, Michael (coord.), *European Data Protection : Coming of Age*, Dordrecht : Springer.
- Merzeau, Louise (2009), « Du signe à la trace : l'information sur mesure », *Hermès*, n° 53, p. 23-29.
- Merzeau, Louise (2013), « L'intelligence des traces », *Intellectica*, n° 59, p. 115-135.
- Mill, John Stuart (1989), *On liberty ; with The subjection of women ; and chapters on socialism*, Cambridge et New York : Cambridge University Press.
- de Montjoie, Yves-Alexandre ; Radaelli, Laura ; Singh, Vivek Kumar ; Pentland, Alex «Sandy» (2015), « Unique in the shopping mall : On the reidentifiability of credit card meta-data », *Science*, vol. 347, n° 6221, pp. 536 - 539
- Müller, Pierre (2011), *Les politiques publiques*, Paris : Presses universitaires de France.
- Newman, Abraham (2008), *Protectors of Privacy. Regulating Personal Data in the Global Economy*, Ithaca : Cornell University Press.
- Ohm, Paul (2010), « Broken Promises of Privacy : Responding to the Surprising Failure of Anonymization », *UCLA Law Review*, n° 57, p. 1701-1777.

Reinach, Adolphe (2004), *Les fondements a priori du Droit Civil*, Paris : Librairie Philosophique Vrin.

Rieder, Bernhard (2010), « Pratiques informationnelles et analyse des traces numériques : de la représentation à l'intervention », *Études de communication*, n° 35, p. 91-104.

Rieder, Bernhard ; Röhle, Theo (2012), « Digital Methods : Five Challenges » (p. 67-85), in Berry, David M. (coord.), *Understanding Digital Humanities*, Londres : Palgrave Macmillan.

Rogers, Richard (2013), *Digital Methods*, Cambridge : MIT Press.

Rogers, Richard, (2009), *The End of the Virtual. Digital Methods*, Amsterdam : Vossiuspers UvA.

Rossi, Julien (2015), « Les quatre modèles de Data Protection Officers en Europe » (p. 27-42), in Rasle, Burno (coord.) *Correspondant informatique et libertés : bien plus qu'un métier*, Paris : AFCDP.

Rossi, Julien (2015), « Données de recherche et vie privée : l'anonymat règle-t-il le problème? » (p. 319-332), in Rasle, Burno (coord.) *Correspondant informatique et libertés : bien plus qu'un métier*, Paris : AFCDP.

Rouvroy, Antoinette ; Berns, Thomas (2013), « Gouvernamentalité algorithmique et perspectives d'émancipation : Le disparate comme condition d'individuation par la relation? », *Réseaux* n° 177, p. 163-196.

Sabatier, Paul A. (1998), «The Advocacy Coalition Framework : revisions and relevance for Europe », *Journal of European Public Policy*, n° 5:1, p. 98-130.

Simon, Bart (2002), «The Return of Panopticism : Supervision, Subjection and the New Surveillance », *Surveillance & Society* 3 (1), p. 1-20.

Smadja, David (2012), « La boîte noire de la controverse », *Raisons politiques*, n° 47 (octobre) : 5-11.

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

Venturini, Tommaso (2012), « Great Expectations : Méthodes Quali-quantitative et Analyse des Réseaux Sociaux » (p. 39-51), in Fourmentaux, Jean-Paul (dir.), *L'Ère Post-Media. Humanités digitales et Cultures numériques*, Paris : Hermann.

Venturini, Tommaso ; Latour, Bruno (2009), « *Le tissu social/the social fabric : traces numériques et méthodes quali-quantitatives* », *Proceedings of Future En Seine*.

Venturini, Tommaso ; Cardon, Dominique ; Cointet, Jean-Philippe (2014), « Méthodes digitales. Approches quali/quantitative des données numériques », *Réseaux*, n° 188, p. 9-21.

Vitalis, André (2009), « "Informatique et libertés" : une histoire de trente ans », *Hermès*, n° 53, p. 137-43

Documents téléchargés sur le site web d'ETALAB

Commission nationale de l'informatique et des libertés (CNIL) (2017), «Correspondants Informatique et Libertés (CIL)», Etalab [en ligne], Consulté le 26 juin 2017, <http://www.data.gouv.fr/fr/datasets/correspondants-informatique-et-libertes-cil/>

Documents issus des deux journées d'étude étudiées

COSTECH (2017), « Synthèse la journée d'étude : "données personnelles en milieu universitaire : quelles questions se poser"? », Université de technologie de Compiègne [en

ligne], Consulté le 20 février 2018, <http://www.costech.utc.fr/spip.php?article121>

Institut des sciences sociales du politique (ISP) (2017), « Retour en vidéos de la journée d'études "Données personnelles et sensibles, quels enjeux pour la recherche en SHS?" », Université Paris Nanterre [en ligne] consulté le 20 février 2018, <https://www.parisnanterre.fr/actualite-de-la-recherche/retour-en-vidéos-de-la-journée-d-etudes-donnees-personnelles-et-sensibles-quels-enjeux-pour-la-recherche-en-shs-809751.kjsp>

Pasquier, Florent ; Rossi, Julien (2017), Journée d'étude «Données personnelles en milieu universitaire : quelles questions se poser? », Synthèse de la journée [En ligne], Compiègne - Paris : COSTECH - ISCC, Disponible sur : http://www.costech.utc.fr/IMG/pdf/synthese_donnees_personnelles_en_milieu_universitaire.pdf

Ressources ayant contribué à la constitution du corpus documentaire de documents issus du processus d'adoption du RGPD

Albrecht Jan-Philipp (2013), Draft report on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), Commission des libertés civiles, de la justice et des affaires intérieures, Parlement européen [en ligne], Consulté le 20 février 2018, <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+COMPARL+PE-501.927+04+DOC+PDF+V0//EN&language=EN>

Commission européenne (2009) « Consultation on the legal framework for the fundamental right to protection of personal data », DG HOME, [en ligne], Consulté le 20 février 2018, https://ec.europa.eu/home-affairs/what-is-new/public-consultation/2009/consulting_0003_en

European Digital Rights (EDRi) (2012), EDRi's suggested amendments to the Commission's Proposal for a Regulation on the Protection of individuals with regard to the processing of personal data, and on the free movement of such data (General Data Protection Regulation), Document "EDRi amendments.pdf", Github [en ligne], Consulté le 20 février 2018, <https://github.com/lobbyplag/lobbyplag-data/blob/master/raw/lobby-documents/EDRi%20Amendments.pdf>

Lobbyplag (n.d.), « lobbyplag-data/raw/lobby-documents at master », Github [en ligne], Consulté le 20 février 2018, <https://github.com/lobbyplag/lobbyplag-data/tree/master/raw/lobby-documents>

Parlement européen (n.d.), « Fiche de procédure : 2012/011(COD) », Observatoire Législatif [en ligne], Consulté le 20 février 2018, [http://www.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2012/0011\(COD\)&l=fr](http://www.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2012/0011(COD)&l=fr)

Corpus de jurisprudences citées

Verdict du Tribunal constitutionnel fédéral allemand (BVerfG) du 15 décembre 1983 (BVerfG Urteil vom 15. Dezember 1983 Az. 1 BvR 209/83, 1 BvR 484/83, 1 BvR 420/83, 1 BvR 362/83, 1 BvR 269/83, 1 BvR 440/83) (Décision « *Volkszählungsurteil* »)

Cour de justice de l'Union européenne (CJUE) 16 juillet 2015 « ClientEarth contre EFSA » Aff. C - 615/13P

Cour de justice de l'Union européenne (CJUE) 20 décembre 2017 «Peter Nowak contre Data Protection Commissioner» Aff. C-434/16

Entretiens et enregistrements cités dans l'article

Genvresse, Sophie, dans : Institut des sciences sociales du politique (ISP) (2017b), « Débat 1 : "Données personnelles et sensibles, quels enjeux pour la recherche en SHS?" », Youtube [en ligne], Consulté le 20 février 2018, <https://www.youtube.com/watch?v=NyrbzTesLTg&t=132s>

Georges, Fanny (2018), Entretien réalisé à Paris le 29 janvier 2018

Skowronska, Kaja (2017), Entre responsabilité éthique et contrainte formelle – le traitement des données confiées aux chercheurs lors des enquêtes qualitatives. Notes de lectures non publiées. Intervention à la journée d'étude du 12 janvier 2017 « données personnelles en milieu universitaire : quelles questions se poser? ». Enregistrement audio disponible en ligne sur le site web du COSTECH, et consulté le 20 février 2018, http://www.costech.utc.fr/IMG/mp3/08_kaja_skowronska_entre_responsabilite_ethique_et_contrainte_formelle.mp3