Diplôme national de master

Domaine - sciences humaines et sociales

Mention – sciences de l'information et des bibliothèques

Parcours – archives numériques

Mémoire / août 2017

Exploitation des données personnelles : raison commerciale, raison d'état et opportunités

Mallorie Wozny

Sous la direction de Clément Oury
Head of Data, Network and Standards – ISSN International Centre

Remerciements

Je souhaiterais adresser mes premiers remerciements à Clément Oury, mon directeur de mémoire, ainsi qu'à Dylan, pour son aide précieuse et son soutien.

Je voudrais aussi remercier Donia et Laura, qui m'ont aidée dans mes recherches.

Je remercie également Charlie, Claire, et Krysta, pour leur bonne humeur et leur soutien qu'elles m'ont témoigné.

Résumé :

Les données personnelles sont utilisées par les entreprises du web à des fins commerciales. Ce marché se regule actuellement au cours d'une bataille jurdique entre les Etats et les grandes entreprises du web, les GAFA. Des opportunités apparaissent, notamment pour l'archivage du web.

Descripteurs : Donnée personnelle – donnée sensible – GAFA – archivage du web

Abstract :

Personally Identifiable Informations are used by web industries in a commercial goal. There is a struggle between states and the GAFA over the regulation of the use of PII. Web archiving also takes opportunity to grow.

Keywords : PII – SPI – GAFA – Web archiving

Droits d'auteurs

Sommaire

Sigles et abréviations

AFNIC : Association Française pour le Nommage Internet en Coopération

ARPA : Advanced Research Projects Agency

ARPANET : Advanced Research Projects Agency NETwork

BnF : Bibliothèque Nationale de France

CNIL : Commission Nationale Informatique et Libertés

FAI : Fournisseur d'Accès à Internet

GAFA : Google Amazon Facebook Apple, il s'agit de l'acronyme réunissant les géants du net

ICO : Information Commissioner's Office

INRIA : Institut national de recherche en informatique et en automatique

IRL : In Real Life, littéralement « dans la vraie vie », désigne le temps passé hors des réseaux

JORF : Journal Officiel de la République Française

OAIS : Open Archive Information System

PIB : Produit Intérieur Brut

PII : Personally Identifiable Information, les « données personnelles »

PME : Petites et Moyennes Entreprises

SPI : Sensitive Personal Information, les « données sensibles »

TCP/IP : Transmission Control Protocol/Internet Protocol, il s'agit du protocole standard d'échanges de données

USA : United States of America

INTRODUCTION

Le 28 mars 2017, le démocrate Michael Capuano s'exprimait devant la Chambre des Représentants[1], à l'occasion d'un projet de loi visant à permettre aux fournisseurs d'accès à Internet (FAI), la commercialisation des données des internautes états-uniens. Il a formulé lors de son discours, succinctement et de façon tout à fait frappante, la problématique autour de laquelle s'articule ce mémoire : « *Yesterday, I bought undewears on the Internet. Why should you know which size I take ?* »[2]. La question est posée. Pourquoi devrions-nous savoir quelle taille de sous-vêtements porte monsieur Capuano ? Pourquoi devrions-nous savoir qu'il les achète sur Internet ? Pourquoi devrions-nous même savoir s'il a accès ou non à Internet ? Capuano a la réponse : nous ne devrions pas. « *It's mine. It's mine !* »[3], scande-t-il avec véhémence. Ces informations lui appartiennent pleinement, et personne d'autre ne devrait y avoir accès, ou les utiliser.

L'exemple est parlant ; il exprime efficacement les caractères public, quotidien et très intime des usages du web. Une des particularités du médium qu'est le web réside en la multiplicité des niveaux de confidentialité – réels ou perçus – par les utilisateurs selon leurs pratiques. La consommation passive de l'information se vit comme une pratique somme toute anonyme, et se rapproche de l'expérience spectatrice d'un lecteur. L'expression personnelle revêt sur le net des formes multiples. C'est l'action volontaire de l'individu, qui communique un message à un autre, aux autres. Le niveau de confidentialité d'un message sur le net est perçu à travers deux paramètres : l'importance du nombre des personnes qui le reçoivent, ainsi que le degré d'anonymat de l'émetteur.

Le degré de confidentialité d'un point de vue exprimé au sein d'un forum à la communauté réduite et sélective, possiblement en se dissimulant derrière un pseudonyme et/ou un avatar, peut être considéré par son émetteur comme équivalent à l'anonymat, dans la mesure où il s'adresse à des personnes qui ne sont pas désignées par leur nom, dont il n'a pas à connaître des informations suffisantes à les localiser ou les identifier, et qui en retour n'ont pas connaissance à son propos de ce type d'informations. L'émetteur d'un email ou d'un message *via* une messagerie instantanée considère ces messages comme privés, sans qu'ils ne soient anonymisés, puisqu'ils ne sont destinés qu'à un nombre réduit de personnes. Cela peut revenir à une conversation de vive voix : seuls les participants ont connaissance de ce qui est dit. A l'opposé, l'utilisateur actif des réseaux sociaux comme Facebook ou Youtube exhibe et partage opinions, lieux de vie, contenus multimédia, afin qu'ils soient accessibles au maximum de personnes. Il n'attend aucune confidentialité, il cherche au contraire à ce que ses messages deviennent viraux, c'est-à-dire qu'ils soient relayés à la plus grande échelle possible, afin d'impacter – si possible durablement et positivement – sa réputation.

[1] Michael Capuano devant la chambre des représentants, vidéo vue le 5/8/2017

[2] « Hier, j'ai acheté des sous-vêtements sur Internet. Pourquoi devriez-vous savoir quelle taille je prends ? »

[3] « C'est à moi ! C'est à moi ! »

Les exemples listés ci-dessus établissent une première palette des niveaux de confidentialité attendus et/ou perçus par l'usager selon ses pratiques sur le web.

Cette confidentialité vécue est aux antipodes de la réalité des principales entreprises propriétaires des services web utilisés. Au contraire, ces entreprises collectent les informations de leurs utilisateurs souvent au plus intime. Le système d'exploitation Android de Google, par exemple, mémorise les messages tapés par l'utilisateur. Là se situe une des nuances : tapés, pas envoyés. L'expérience est amusante : en tapant un mot qui n'existe pas dans une tournure de phrase fréquemment utilisée, le mot inventé finira par apparaître dans les propositions automatiques générées par l'application. Cette pratique de Google permet certes d'assister à certaines curiosités, comme de voir un sms « Bonjour maman » transformé en « Bonjour Satan », mais elle est surtout symptomatique de l'économie de la donnée à laquelle s'adonnent sans réserve presque toutes les entreprises du web, qu'elles constituent les GAFA ou qu'elles se partagent les miettes laissées par l'oligopole constitué par Google, Amazon, Facebook et Apple.

C'est tout l'enjeu du travail que constitue ce mémoire que s'intéresser aux pratiques de collecte, de traitement et d'exploitation des données personnelles ainsi que les questionnements et contradictions qu'elles soulèvent, afin d'en proposer un panorama. Cette étude se fera à travers le prisme des GAFA, que le grand nombre de données traitées par ces structures et leur influence sur les individus, sur les sociétés entières, voire sur les Etats, rendent intéressantes.

Le marché sur lequel se situent ces entreprises, après des années de libertés quasi-totale dans la recherche de profit, commence à être la cible de régulations qui prennent leur source à la fois dans le comportement plus conscient des utilisateurs et dans les législations successives sanctionnant les pratiques du web, et qui s'incarnent à la fois dans l'adaptation parfois difficile et souvent contrainte des GAFA mécontents[4]. Ceux-ci ont réussi à ériger leur fonctionnement en un modèle économique courant, voire standard, sur le net. A contre-courant, apparaissent pour des services basés sur la protection de la vie privée par non-intrusion.

L'archivage du web est une pratique essentielle à la mise en mémoire du XXIème siècle[5]. Les données personnelles sont collectées par les entreprises privées dans un but commercial, mais sont aussi susceptibles d'être conservées à des fins archivistiques. Le dépôt légal élargi en 2006 donne à la BnF la mission d'archiver le web. Les campagnes ont pour but de fixer sur des archives un web toujours changeant, et d'ainsi conserver une représentation fidèle du web – et aussi du web social – à des points temporels donnés. Les pratiques quotidiennes sont révélatrices d'une époque et donc utile à la recherche. A ce titre, des articles personnels, des blogs ou des statuts Facebook sont collectés, afin de faire l'instantané d'un moment du web.

[4] "L'influence tentaculaire des géants américains», Alexandre Léchenet et BiG, multinationales.org, vue le 5/8/2017, http://multinationales.org/Lobbying-l-artillerie-lourde-des-geants-du-net-en-France-et-en-Europe

[5] Mussou Claude, « Et le Web devint archive : enjeux et défis », Le Temps des médias, 2012/2 (n° 19), p. 259-266. DOI : 10.3917/tdm.019.0259. URL : http://www.cairn.info/revue-le-temps-des-medias-2012-2-page-259.htm

DONNÉES PERSONNELLES

1. DÉFINITIONS

1.1 La donnée

Le terme de « donnée », largement utilisé depuis l'avènement du numérique, est polysémique[6]. Quelle que soit la définition, la donnée est associée à la réflexion, à la recherche : elle est à la fois pilier d'un raisonnement, son point de départ, et son résultat. Informatiquement, la donnée est la « représentation conventionnelle d'une information en vue de son traitement informatique ». Cela lui donne dans le domaine numérique une position d'autant plus essentielle qu'elle est au centre de toutes les actions entreprises. La donnée n'est pas l'information en tant que telle, mais la mise en forme qui la rend accessible au traitement informatique.

Plus simplement, les données dans le domaine numérique correspondent à des informations numériquement codées de façon de plus ou moins complexe. Elles sont produites en permanence, par chacun de nous, dans un acte parfois volontaire, souvent inconscient. Chaque clic, chaque connexion et chaque identification d'un utilisateur laissent des traces, sous forme de données, automatiquement enregistrées, par le site, le navigateur, l'opérateur, etc. Ces données issues de parcours connectés propre à chaque utilisateurs sont appelées «données personnelles».

Le « numérique » est questionné dans des optiques sociologique, économique, et éthique, mais encore peu en étant qu'objet de pensée à part entière.
En conséquence, les domaines qu'il recouvre restent dans leurs définitions parfois vagues, et souvent diverses. Afin de mener ce travail du mieux possible et à travers les différentes définitions attribuées, il sera tenté de circonscrire ici ce qu'est précisément une « donnée personnelle ».

1.2 La donnée personnelle selon la CNIL

1.2.1 L'organisme

La Commission Nationale de l'Informatique et des Libertés, autorité administrative indépendante, créée en 1978 par la loi du même nom, se présente

[6] "Ce qui est connu ou admis comme tel, sur lequel on peut fonder un raisonnement, qui sert de point de départ pour une recherche (surtout pluriel) : Les données actuelles de la biologie. Idée fondamentale qui sert de point de départ, élément essentiel sur lequel est construit un ouvrage : Les données d'une comédie. Renseignement qui sert de point d'appui (surtout pluriel) : Manquer de données pour faire une analyse approfondie.» Dictionnaire de français en ligne Larousse, consulté le 6/8/2017

"Représentation conventionnelle d'une information en vue de son traitement informatique.»

"Dans un problème de mathématiques, hypothèse figurant dans l'énoncé.»

"Résultats d'observations ou d'expériences faites délibérément ou à l'occasion d'autres tâches et soumis aux méthodes statistiques.

comme l'organisme de régulation des données personnelles[7]. Elle agit pour la préservation des droits de l'individu numérique, à travers l'information, l'accompagnement, le conseil, jusqu'au contrôle et à la sanction des organismes non-conformes.

L'action de la CNIL ne se limite néanmoins pas à la seule mise en œuvre de la conformité du traitement des données à la loi : elle s'est donnée une mission réflexive autour des problématiques éthiques liés aux mutations conséquences du développement des (pas si) nouvelles technologies.

La réflexion de l'organisme autour de la protection des données des individus a scindé ces dernières en deux catégories : les données personnelles et les données sensibles.

1.2.2 La donnée personnelle

La CNIL définit les données personnelles comme « toute information identifiant directement ou indirectement une personne physique »[8].

Autrement dit, la donnée personnelle correspond à une information objective sur les personnes – elle cite en exemple le numéro de téléphone ou l'adresse – et relève typiquement du domaine civil – nom, adresses mail et postale, numéro de téléphone – en ce qu'elles servent à identifier et localiser un individu, ainsi que de référencer des moyens de le contacter. Les données qualifiées de « personnelles » par la CNIL sont donc constitutives de l'identité civile d'un individu, à la fois dans la sphère plus traditionnelle de l'existence « IRL » et dans la sphère moins tangible du web.

1.2.3 Données sensibles

La CNIL différencie la donnée personnelle de la donnée sensible[9], qui est une « information concernant l'origine raciale ou ethnique, les opinions politiques, philosophiques ou religieuses, l'appartenance syndicale, la santé ou la vie sexuelle. ». Elle reste relative à la personne, mais se caractérise en ce qu'elle relève de la seule vie privée. Les exemples cités par la CNIL sont parmis les vingt-cinq critères de discrimination listés par la loi[10].

Elle est sensible en ce qu'elle enregistre ce qu'il y a d'intime chez l'utilisateur : ses origines ethniques ou géographiques, ses préférences sexuelles, ses opinions politiques, ses pratiques et habitudes de consommation. Ces connaissances « sensibles » ne devraient donc être possédées que par la seule autorité légitime à en disposer : l'individu à qui elles appartiennent, ainsi que le cercle auquel il a fait le choix de les partager.

[7] « La CNIL en France » https://www.cnil.fr/fr/la-cnil-en-france, vue le 30/08/2017

[8] « La CNIL en France » https://www.cnil.fr/fr/definition/donnee-personnelle vue le 30/08/2017

[9] « La CNIL en France » https://www.cnil.fr/fr/definition/donnee-sensible vue le 30/08/2017

[10] « Le défenseur des droits » https://www.defenseurdesdroits.fr/fr/institution/competences/lutte-contre-discriminations vue le 30/08/2017

1.3 Wikipédia

1.3.1 L'encyclopédie la plus lue du monde

Sixième site le plus consulté au monde[11], Wikipédia est une encyclopédie en ligne possédant l'intéressante particularité de fonctionner uniquement grâce au travail collaboratifs d'une grande communauté d'internautes, spécialistes ou amateurs éclairés, qui alimentent, rectifient et valident les publications de la plateforme. Une communauté suffisamment vaste pour que soient publiés – sur et selon le site lui-même – des articles en deux cent quatre-vingt-onze langues[12].

Le nombre de contributeurs est très important - chaque mois environ soixante-dix mille contributeurs uniques[13] depuis le début de l'année en cours -, mais plus encore celui des visiteurs : en août 2017, la consultation du site montait à près de quatorze millions de visites journalières.

Utilisé par tous, Wikipédia s'est transformé en qu'on pourrait rapprocher d'un habitus tant sa présence en tête des pages de résultats est fréquente. La définition que donne Wikipédia des données personnelles en devient intéressante, puisqu'il s'agit sans doute de celle qui sera la plus lue et la mieux diffusée, et qui surtout, validée par la communauté Internet, montre ce que comprennent de cette notion les internautes eux-mêmes.

1.3.2 L'article francophone

Ecrit en français, l'article de la plateforme de connaissances reprend les informations diffusées par la CNIL. Il s'intéresse essentiellement aux données personnelles dans le cadre juridique français et européen, en citant l'article 2 de la loi «Informatique et libertés» et en faisant le lien vers la directive 2006/24/CE sur la conservation des données[14]. Cette directive n'est plus en vigueur depuis 2014[15]. On peut imaginer que le sujet complexe des données personnelles et de leur législation est peu connu par les contributeurs francophones de Wikipedia.

Ceux-ci mettent en revanche en lumière la polémique devenue psychose qu'est la possibilité d'espionnage de chaque citoyen à travers le traitement de leurs données personnelles, notamment à travers de grands scandales, comme la révélation au grand public du réseau ECHELON ou la déclaration de séropositivité des personnes malades du SIDA.

Les informations présentes sur l'article ont le mérite de rediriger vers des sources sérieuses (la CNIL, l'*International Working Group on Data Protection in Telecommunication*), néanmoins l'article en lui-même cite plus qu'il n'explique, montre plus qu'il ne pense son sujet. Les exemples et le contenu indiquent que l'actualité de l'article est à remettre en cause.

[11] « Les sites web les plus visités » http://www.trackalytics.com/the-most-visited-website/page/1/, vue le 30/08/2017

[12] https://fr.wikipedia.org/wiki/Wikip%C3%A9dia , vue le 30/08/2017

[13] https://stats.wikimedia.org/EN/TablesWikipediaZZ.htm , vue le 30/08/2017

[14] « législation européenne » http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=celex%3A32006L0024 , vue le 30/08/2017

[15]« La CNIL en France » https://lc.cx/chXd , vue le 30/08/2017

1.3.3 L'article anglophone

La langue la plus consultée sur Wikipédia[16] – et la plus utilisée sur le web[17] – étant l'anglais, il semble également intéressant de retranscrire ici ce qui est sans doute le texte le plus lu sur le sujet des données personnelles, ou dans la langue anglaise les *Personally Identifiable Informations* (PII). L'anglais remplace le terme de « donnée » par celui d' « information », plus précis en ce qu'il a été largement plus théorisé, conceptualisé, et débattu. L'article s'accorde avec la CNIL en ce qu'il existe une différenciation entre les données personnelles, les PII, et les données sensibles, les *Sensitive Personal Informations*.

Comme le précédent, l'article anglophone Wikipedia[18] définit les données personnelles par le cadre légal et juridique. Néanmoins, quand le contenu de l'article francophone cite la loi, et s'intéresse à des scandales relatifs aux données personnelles, celui de l'article anglophone s'attache à expliquer ce que sont les données personnelles à travers le prisme de la législation des Etats-Unis, de l'Union Européenne, de l'Australie. L'article fait également le lien vers les principaux textes et organismes définissant et/ou réglementant les PII au Canada, en Nouvelle-Zélande, et au Royaume-Uni.

Selon l'article, ces derniers définissent les données personnelles dans le droit privé, de la manière suivante : « any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.[19] ». Cette définition se trouve dans le droit américain dans le *United States Code*[20], qui codifie le droit fédéral, applicable donc dans tous les Etats des Etats-Unis.

Sont donc décrites deux sortes de PII : celles qui permettent une identification directe de l'individu, et celles nécessitant un travail de traitement de données pour parvenir à l'identification. L'article liste les informations qui sont ou sont susceptibles d'être considérées comme des *Personally Identifiable Informations*, tout en rappelant l'extrême rapidité de développement des technologies, notamment dans le domaine des algorithmes, qui rend possible l'identification des internautes par le traitement d'informations non présentes dans la liste.

---

[16] https://meta.wikimedia.org/wiki/List_of_Wikipedias , vue le 30/08/2017

[17] « Top 10 des langues les plus parler sur le web », Bruno Texier, http://www.archimag.com/vie-numerique/2014/09/24/top-10-langues-plus-parl%C3%A9es-web-fran%C3%A7ais-9%C3%A8me-place , vue le 30/08/2017

[18] https://en.wikipedia.org/wiki/Personally_identifiable_information , vue le 30/08/2017

[19] "Toute information à propos d'un individu détenu par une structure, incluant (1) toute information pouvant être utilisée afin de distinguer ou tracer l'identité d'un individu, comme son nom, son numéro de sécurité sociale, ses date et lieu de naissance, le nom de jeune fille de sa mère ou ses informations biométriques; et (2) toute information associée ou pouvant être associée à un individu, comme les informations médicales, éducatives, financières, et relatives à sa profession.»

[20] Code des Etats-Unis

L'article semble ne pas faire de réelle distinction entre les *Sensitive Personal Informations* et les *Personally Identifiable Information*. D'ailleurs, le Wikipédia anglophone renvoie automatiquement vers l'article sur les PII lorsque l'on tape Sensitive Personal Information dans la barre de recherche.

La Northeastern University dans sa page «*Safe Computing*»[21] fait le même amalgame en titrant «*What is Sensitive Personal Identifying Informations (PII)* ?» et en donne une définition différente : «Sensitive Personal Identifying Information (PII)[22] is defined as information that if lost, compromised, or disclosed could result in substantial harm, embarrassment, inconvenience, or unfairness to an individual[23]«, puis ajoute «In general terms it is any information that could be used by criminals to conduct identity theft, blackmail, stalking, or other crimes against an individual. Federal and State laws, and University regulations dictate how this information must be stored, transmitted, and processed[24]«. Ces définitions permettent à l'université états-unienne de sensibiliser ses étudiants à la sécurité de leurs données, mais, à l'image du Congrès, ne fait pas de distinction claire entre une donnée personnelle et une donnée sensible.

1.4 La loi fédérale américaine

Selon la définition du Code des Etats-Unis[25], un code fédéral applicable à tous les Etats membres : « The term «sensitive personal information», with respect to an individual, means any information about the individual maintained by an agency, including the following: (A) Education, financial transactions, medical history, and criminal or employment history. (B) Information that can be used to distinguish or trace the individual's identity, including name, social security number, date and place of birth, mother's maiden name, or biometric records.[26]«

La place de la définition des informations personnelles sensibles (SPI) dans le Code des Etats-Unis est intéressante : au sein du titre Veteran's Benefits, partie General Administrative Provisions, au chapitre Records and Investigations, sous-

[21] "Navigation sécurisée» http://www.northeastern.edu/securenu/sensitive-information-2/sensitive-information/ , vue le 30/08/2017

[22] "Que sont les données personnelles sensibles ?»

[23] "Une donnée personnelle sensible est définie comme une information qui, si elle venait à être perdue, compromise, ou divulguée, pourrait conduire à un dommage substantiel, à un embarras, un inconfort ou une injustice pour un individu».

[24] "De façon générale, il s'agit de toute information qui pourrait être utilisée par des criminels pour perpétrer usurpation d'identité, chantage, harcèlement ou d'autres délits portant atteinte à la personne. Les lois fédérales et celles de l'Etat, ainsi que le règlement de l'université dicte la manière dont cette information peut être stockée, transmise et traitée.»"Une donnée personnelle sensible est définie comme une information qui, si elle venait à être perdue, compromise, ou divulguée, pourrait conduire à un dommage substantiel, à un embarras, un inconfort ou une injustice pour un individu».

[25] « Code des USA » https://www.gpo.gov/fdsys/pkg/USCODE-2011-title38/html/USCODE-2011-title38-partIV-chap57-subchapIII-sec5727.htm , vue le 30/08/2017

[26] Le terme "information personnelle sensible», en relation avec un individu, se rapporte à toute information à propos de cet individu détenu par une structure, incluant les informations relatives à : (A) l'éducation, les transactions financières, l'historique médical et les historiques criminel ou professionnel. (B) information qui peut être utilisée afin de distinguer ou de tracer l'identité d'un individu, dont le nom, le numéro de sécurité sociale, les date et lieu de naissance, nom de jeune fille de la mère, ou les informations biométriques.

chapitre Information Security. Cette position particulière permet de déduire que cette définition ne vise pas la population civile américaine, mais bien l'armée. Cela permet aussi de proposer une nouvelle traduction du terme «agency» dans le contexte de cette définition. Au départ traduit par «structure», il apparaît que le terme «agence fédérale» serait plus approprié.

La loi fédérale américaine reconnaît les données personnelles comme une propriété des agences fédérales et l'individu comme objet de ces données. Il ne s'agit pas de protection de la vie privée, mais de celle du secret des agences fédérales. Le même article[27] définit les termes suivants : «*security incident*[28]«, «*national security system*[29]«, «*fraud resolution system*[30]« ou «*identity theft insurance*[31]« ; aucune définition n'est donnée concernant la vie privée ou la protection de la vie privée et la donnée sensible se définit avant tout comme honteuse et pouvant porter préjudice à la personne à laquelle elle se rattache.

1.5 Le Royaume-Uni

1.5.1 Information Commissioner's Office

L'amalgame fait aux Etats-Unis entre SPI et PII n'est cependant pas applicable au monde anglophone dans son entièreté. L'*Information Commissioner's Office*[32] (ICO), un organisme public de régulation britannique, chargé de faire appliquer le droit à l'information dans l'intérêt général. Le ICO peut être vu comme un équivalent britannique à la CNIL française. Sept textes voient leur application appliquée par le Bureau du Commissaire à l'Information : , dont le *Data Protection Act*[33], daté de 1998, est le plus ancien, qui pose les bases de la protection et du contrôle des données personnelles. Il statue comme principe premier que «*Personal data shall be processed fairly and lawfully*[34]«.

1.5.2 Personally Identifiable Informations

Dans cet acte, le terme de «*Personally Identifiable Information*[35]« n'est pas utilisé, au profit de «personal data» : « «personal data» means data which relate to a living individual who can be identified — (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any

[27] Ibid.

[28] Incident de sécurité

[29] Système de Sécurité Nationale

[30] Système de Résolution des Fraudes

[31] Assurance contre l'Usurpation d'Identité

[32] Bureau du Commissaire à l'Information

[33] Acte de Protection des Données

[34] "Les données personnelles doivent être traitées loyalement et licitement»

[35] Donnée personnelle

other person in respect of the individual[36]«. Les données personnelles sont relatives à un individu et permettent de l'identifier par elles-mêmes ou grâce au traitement qui en est fait. La définition qui est faite se rapproche beaucoup de celle de la CNIL : « toute information identifiant directement ou indirectement une personne physique ». Le fait qu'une donnée puisse identifier une personne tient compte non seulement de l'ensemble des données à disposition du contrôleur de données au moment où la donnée est entrée, mais aussi de toutes les données susceptibles d'entrer en résonance avec elle pour identifier une personne. Par exemple, le nom d'une personne et une bibliothèque communale n'ont de rapport que si un troisième élément - un formulaire d'inscription, une attestation de domicile - vient lier les deux premières informations.

L'ICO inclut en plus dans sa définition des données personnelles le rôle de contrôleur de données, qui est «*a person who (either alone or jointly or in common with other persons) determines the purposes for which and the manner in which any personal data are, or are to be, processed*[37]«. L'intention ayant présidé à la collecte et au traitement des données entre dans le cadre de la loi britannique pour déterminer le caractère personnel d'une donnée. Tous les éléments pouvant et ayant pu influer sur les décisions du contrôleur de données sont donc constitutifs du caractère personnel - ou non personnel - de la donnée.

Enfin, existe une distinction qui n'est pas - encore ? - dans la définition légale française : une donnée personnelle est relative à une personne vivante. Les débats relatifs à la mort numérique dans le débat en France posent la question de la gestion des données personnelles des personnes décédées[38]. Le droit britannique permet d'éviter une partie du problème : une donnée n'est pas personnelle si elle n'est pas relative à une une personne vivante. Leur sort dépend donc des décisions prises par les héritiers au cours de la succession.

1.5.3 Sensitive personal data

Les données personnelles dans le *Data Protection Act*, sont des «*sensitive personal data*». Une première apparition du terme introduit une liste des différentes données sensibles possibles.

« *In this Act «sensitive personal data» means personal data consisting of information as to— (a)the racial or ethnic origin of the data subject, (b)his political opinions, (c)his religious beliefs or other beliefs of a similar nature, (d)whether he is a member of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992), (e)his physical or mental health or condition, (f)his sexual life, (g)the commission or alleged commission by*

[36] "donnée personnelle» : signifie une donnée relative à un individu vivant qui peut être identifié - (a) grâce à cette donnée, ou (b) grâce à cette donnée et d'autres informations en possession, ou susceptibles d'entrer en possession du contrôleur de données, et inclut toute expression d'une opinion à propos de cet individu et toute indication concernant les intentions du contrôleur de données ou de toute autre personne en rapport avec cet individu.

http://www.legislation.gov.uk/ukpga/1998/29#commentary-c18184261 , vue le 30/08/2017

[37] "Une personne qui (seule, conjointement ou de conserve avec plusieurs autres personnes), détermine les objectifs pour lesquels, et la manière selon laquelle, chaque donnée personnelle est, ou sera, traitée.»

[38] « Comment organiser la mort numérique », Ariane Vennin https://www.village-justice.com/articles/Comment-organiser-mort-numerique-advient-nos-donnees-apres-notre-deces,24553.html , vue le 30/08/2017

him of any offence, or (h)any proceedings for any offence committed or alleged to have been committed by him, the disposal of such proceedings or the sentence of any court in such proceedings.[39]«

Les modalités de traitement des données sensibles sont décrites dans un chapitre dédié, mais la notion n'est pas plus longuement définie dans le *Data Protection Act*. Une donnée sensible est une donnée personnelle relative à l'intimité de la personne, et doit être traitée plus strictement encore que les autres données personnelles. La position de l'ICO se rapproche de celle de la CNIL ; la liste des types de données sensibles est cependant moins importante que dans le droit français, même si elle en comporte les principaux.

1.6 Différencier la donnée et donnée personnelle

Il semble, à l'étude des définitions proposées, que chaque donnée émise lors du parcours d'un individu sur le web puisse être considérée comme une donnée personnelle, puisque chacune de ces données semble permettre l'identification : adresse IP, par exemple, permet de remonter directement jusqu'au propriétaire de la ligne internet. Ce n'est pourtant absolument pas le cas de beaucoup des autres données émises.

En réalité, la plupart d'entre elles ne permettent pas de retracer l'utilisateur. La localisation sur une application GPS ne permet pas de savoir qui l'utilise. La comptabilisation des visites sur les vidéos d'un Youtubeur ne permet pas de savoir qui les a vraiment consultées. On pourrait alors se dire que l'émission de ces données utilisateurs n'a pas de réel impact sur la navigation utilisateur, ou même sur les services utilisés.

Pourtant, ce sont bien ces données qui sont utilisées pour financer les services les plus utilisés du web. Si c'est le cas, c'est parce qu'effectivement, prises une a une, ces données ne signifie rien. En revanche, une fois passées par un système de traitement de données, il est possible de retracer l'historique des actions et d'en déduire un profil utilisateur.

C'est pourquoi la CNIL et les autres organismes de régulation insistent sur le fait qu'une donnée est personnelle lorsqu'elle permet l'identification[40], de façon directe ou indirecte, de l'individu, sans établir pour de liste exhaustive afin de ne pas limiter l'application du règlement des usages de ces données. En effet, si une liste arrêtée

[39] "Dans cet Acte, les "données personnelles sensibles» sont les "données personnelles» consistant en une information telle que - (a)l'origine raciale ou ethnique du sujet de la données, (b)ses opinions politiques, (c)ses croyances religieuses ou autres croyances d'une nature comparable, (d) s'il est membre d'un trade union (au sens du Syndicat et le Dialogue Social (Consolidation) Acte 1992), (e)sa santé ou sa condition physique ou mentale, (f)sa vie sexuelle, (g)les infractions commises ou supposément commises par lui, or (h)toutes les procédures judiciaires pour toutes les infractions commises ou supposément commise par lui, le traitement de ces procédures judiciaires ou la sentence de tout tribunal dans le cadre de ces procédures.

http://www.legislation.gov.uk/ukpga/1998/29#commentary-c18184261 , vue le 30/08/2017

[40]« La définition des données personnelle selon la CNIL » https://www.cnil.fr/fr/definition/donnee-personnelle , vue le 30/08/2017

permettrait de poser un cadre clarifié et plus simple à envisager pour les gestionnaires de données personnelles moins ou pas sensibilisés au sujet, elle diminuerait le champ d'application de la loi et le pouvoir de contrôle et de sanction d'organismes de régulation en ce que les toujours plus puissants algorithmes de traitements de données, utilisés notamment par Facebook et Google, sont susceptibles d'identifier et/ou localiser un individu en passant par des types d'informations non répertoriés.

2. CONTEXTUALISATION

2.1 Naissance de l'Internet et du web

L'histoire du web et de l'Internet, intrinsèquement liés, a déjà été souvent décrite, aussi le but n'est pas de dresser une chronologie exhaustive dans ce travail. Néanmoins, un retour sur la genèse et l'évolution de ces deux entités semble nécessaire au propos, afin de montrer les changements que ces technologies ont représentés pour la gestion des données personnelles.

La première mise en réseau à avoir largement impacté la transmission de l'information à avoir eu lieu est le développement du télégraphe électrique. Cela a été l'occasion de la création de câbles sous-marins, permettant pour la première fois un acheminement rapide de l'information à travers un pays, puis d'un pays à l'autre ne nécessitant pas un support physique de l'information lors de son transport. La télégraphie a permis de transformer, par l'apport de la communication à longue et très longue distance, le monde des affaires et le commerce. On peut en cela la rapprocher cette invention de l'Internet[41].

Les premiers ordinateurs sont antérieurs à la création du web et de l'Internet. Les «*computer*?» date des années quarante et ne servaient qu'à des fins de calcul mathématique. Licklider, un informaticien, s'intéresse vingt ans plus à utiliser les ordinateurs non plus comme des calculateurs, mais comme des terminaux de communication, dans le cadre d'un projet de l'*Advanced Research Projects Agency* (ARPA). Larry Roberts, un chercheur de la même agence, introduit dans ce projet les travaux de Donald Davis, physicien, et Paul Baran, mathématicien, sur une méthode de communication nouvelle, la commutation de paquets[42].

> «Cette technique découpe le message en blocs, appelés paquets, chacun d'entre eux étant placé dans une « enveloppe » électronique, qui indique son adresse de destination, ainsi que d'autres informations. Les conventions pour définir les informations fournies par ces « enveloppes électroniques » sont baptisées protocoles. Les paquets pouvaient être aiguillés séparément sur différentes routes et à différents moments, et ré-assemblés à leur destination finale. Bien que le coût du routage des paquets semblât à priori déraisonnable, une analyse mathématique méticuleuse (réalisée par le

[41] Tom Standage, The Victorian Internet : The Remarkable Story of the Telegraph and the nineteenth Century's On-Line Pioneers, New York, Walker & Company, 1998

[42] Ceruzzi Paul E, « Aux origines américaines de l'Internet : projets militaires, intérêts commerciaux, désirs de communauté », Le Temps des médias, 2012/1 (n° 18), p. 15-28. DOI : 10.3917/tdm.018.0015. URL : http://www.cairn.info/revue-le-temps-des-medias-2012-1-page-15.htm , vue le 30/08/2017

> Professeur Leonard Kleinrock de UCLA – University of California at Los Angeles) montra que la commutation par paquets s'avérait la meilleure méthode de transmission des données informatiques.[43]«.

Avec la commutation de paquets, les chercheurs de l'ARPA pouvaient transmettre des données informatiques en connectant un ordinateur à un autre. En 1969, une période supposément éloignée de la technologie actuelle - en 1969, seuls 15,2% des ménages sont équipés d'un téléphone fixe[44] - pourtant nous y retrouvons un vocabulaire familier de notre usage des télécommunications : la transmission de paquets de données décrite dans l'OAIS, le routage (c'est-à-dire l'acheminement des données), les protocoles de transmission de données, la terminologie nous rapproche du web contemporain. L'étape suivante fut de «rechercher une manière de relier non seulement des ordinateurs mais aussi des réseaux possédant des architectures diverses et présentant des degrés variés de fiabilité[45]«, ce qui a donné lieu à l'écriture et mise en place d'un protocole nouveau en 1978 devenu presque familier puisqu'il est aujourd'hui le standard pour la transmission de données : le protocole TCP/IP. Il s'agit d'une combinaison du protocole *Transmission Control Protocol* (TCP), un protocole de transmission destiné à assurer la fiabilité de la transmission des données, et de l'*Internet Protocol*», chargé de transmettre des données. C'est l'ensemble des réseaux, reliés entre eux par le protocole TCP/IP, qui forment Internet[46]. Par extension, il est parfois confondu avec son application principale, le web, une technologie d'hypertexte reliant entre elles des entités numériques, et permettant de passer de l'une à l'autre grâce à un navigateur[47].

2.2 De la recherche à l'utilisation personnelle

Créé grâce à un projet financé par l'armée américaine et qui a pris le nom d'ARPANET en 1967[48], le réseau informatique est d'abord utilisé dans un cadre professionnel, d'autant que l'achat d'une machine individuelle n'est pas permis à toutes les bourses. L'Apple I, à sa sortie en 1976, coûtait 666,66$; la même année, le revenu annuel brut par habitant s'élevait à 8980$[49], soit environ 750$ mensuel. Néanmoins, l'ordinateur pénètre progressivement dans les foyers et son utilisation comme moyen de communication s'étend malgré des coûts dissuasifs. Au fur et à mesure de l'équipement des ménages, de multiples services se développent proposant de faire communiquer entre eux les internautes (forum, par exemple). C'est, selon Paul Ceruzzi, le changement des usages des utilisateurs connectés qui

[43] Ibid, paragraphe 3.

[44] Bodin Jean-Louis, « L'équipement des ménages en téléphone en 1968 », Economie et statistiques 1970/15 (n°1) p. 57-60. DOI : http://www.persee.fr/doc/estat_0336-1454_1970_num_15_1_1984 , vue le 30/08/2017

[45] Ceruzzi Paul E, « Aux origines américaines de l'Internet : projets militaires, intérêts commerciaux, désirs de communauté », Le Temps des médias, 2012/1 (n° 18), p. 15-28. DOI : 10.3917/tdm.018.0015. URL : http://www.cairn.info/revue-le-temps-des-medias-2012-1-page-15.htm paragraphe 3, vue le 30/08/2017

[46] http://www.commentcamarche.net/contents/539-tcp-ip , vue le 30/08/2017

[47] https://www.astuces-aide-informatique.info/70/qu-est-ce-que-le-web , vue le 30/08/2017

[48] Ceruzzi Paul E, « Aux origines américaines de l'Internet : projets militaires, intérêts commerciaux, désirs de communauté », Le Temps des médias, 2012/1 (n° 18), p. 15-28. DOI : 10.3917/tdm.018.0015. URL : http://www.cairn.info/revue-le-temps-des-medias-2012-1-page-15.htm , vue le 30/08/2017

[49]« Le RNB aux USA » http://donnees.banquemondiale.org/indicateur/NY.GNP.PCAP.CD?locations=US , vue le 30/08/2017

a fait passer l'ARPANET - conçu par et pour des chercheurs - à l'Internet actuel[50]. Cette transformation se serait effectuée par l'utilisation de l'ordinateur non plus comme un outil de communication professionnelle, mais comme outil de socialisation.

Internet s'est développé par ARPANET mais aussi grâce à des réseaux parallèles aux architectures plurielles qui se sont reliées plus tard au réseau principal, comme Usenet. Ces réseaux ont été le lieu de l'expression de la culture populaire, notamment des passionnés de la science-fiction. Ces primo-arrivants sur ces réseaux ont également participé à leur amélioration : ils étaient des utilisateurs récursifs[51].

2.3 Le web comme renouveau de la discussion

Ce faisant, et accentuée par l'accroissement des utilisateurs d'Internet, a germé l'idée suivante : quoi qu'un internaute puisse penser, quelles que puissent être ses pratiques, il trouvera sur la toile une ou des personnes avec lesquelles partager ses idées. Cela s'est effectivement vu grâce à la formation de communautés de personnes plutôt marginales[52], qui ont trouvé dans l'Internet et le développement en général de l'informatique une issue à leur solitude, en accédant à des espaces les rassemblant autour de sujets, de passions, de références communes. Cela a entraîné la création de nombreux modèles communautaires, comme les communautés de fans, qui se réunissent autour d'un même sujet de discussion, ou les *forum* de toutes mouvances culturelle, idéologique, politique ou dogmatique. Il s'agit sans doute du mythe le moins mensonger du net, mais comme ailleurs, les communautés ne se contentent pas d'inclure des individus en accord avec elles, il leur arrivent également d'exclure un de leurs membres, ou d'organiser des *vendette* plus ou moins violentes contre un individu appelées cyber-harcèlement[53].

2.4 Culture et technique de l'Internet

2.4.1 Une construction technique

Un algorithme est une suite d'instructions, hiérarchisant des actions pour arriver à un résultat. Un guide d'utilisation est un algorithme, par exemple. Les algorithmes sont au cœur des pratiques informatiques et sous-tendent tous les services connus du net. L'écriture d'un algorithme résulte de l'apprentissage d'un ou plusieurs langages informatiques et de la maîtrise de la grammaire associée. A

[50] Ceruzzi Paul E, « Aux origines américaines de l'Internet : projets militaires, intérêts commerciaux, désirs de communauté », Le Temps des médias, 2012/1 (n° 18), p. 15-28. DOI : 10.3917/tdm.018.0015. URL : http://www.cairn.info/revue-le-temps-des-medias-2012-1-page-15.htm , vue le 30/08/2017

[51] Paloque-Berges Camille, « La mémoire culturelle d'Internet : le folklore de Usenet », Le Temps des médias, 2012/1 (n° 18), p. 111-123. DOI : 10.3917/tdm.018.0111. URL : http://www.cairn.info/revue-le-temps-des-medias-2012-1-page-111.htm , vue le 30/08/2017

[52] Dominique Cardon,Antonio A. Casilli, Qu'est-ce que le Digital labor ?, Paris, Ina Éditions, 2015, 104 p. ISBN 978-2-86938-2299

[53] Le gouvernement français a dédié une page au cyber-harcèlement sur le site : http://www.nonauharcelement.education.gouv.fr , vue le 30/08/2017

ce titre, il s'agit d'une technique et elle bénéficie de l'illusion d'objectivité accordée à toutes les techniques. Bruno Latour a démontré que les techniques ne pouvaient être extraites ou soustraites à leurs dimensions morale et sociale.

2.4.2 De la technique à la culture

Internet est en premier lieu une construction technique, sous-tendue par des protocoles Le terme de «culture Internet» est souvent évoqué sans être clairement délimité. S'agit-il de la culture de la programmation informatique, des jeux en ligne, des vidéos sur Youtube, des blogs personnels, des forums ou des wikis ?

La terminologie «culture de de l'internet» associe la culture à la technique, tant la culture, complexe et multiple, construite par l'être humain semble éloignée de l'objet technique, dépourvu de moral. Ce postulat semble s'imposer très facilement ; c'est pourquoi Bruno Latour le remet en doute dans son article «La fin des moyens[54]«. Il commence par rejeter l'idée selon laquelle les avancées humaines seraient toujours à l'origine du développement technique et jamais sa conséquence, en faisant référence aux travaux sur la préhistoire, qui démontre au moins des inter-relations entre «l'émergence de la technique et celle de l'humain», et aussi qu'il «semble de plus en plus que les humains se soient développés dans un nid ou dans une niche peuplés déjà d'habiletés, de savoir-faire et d'objets techniques[55]«. Il estime qu'un objet ne peut être dépossédé de son aspect moral, mais étudié par à travers son aspect technique. Afin de développer ce point de vue, il utilise la notion de pli. Un objet, lorsqu'on en déplie tous ses aspects, contient à la fois les matériaux qui le composent, les techniques nécessaires à sa création, le temps et l'espace de ses différents usages et usagers.

> «Rien de moins local, de moins contemporain, de moins brutal qu'un marteau, dès que l'on se met à déplier ce qu'il agence ; rien de plus local, brutal et durable que ce même marteau, dès que l'on replie tout ce qu'il a impliqué». Par la misère même où je me trouve quand je suis privé de marteau (qu'on se souvienne du bonheur de Crusoé découvrant l'outillage des caisses rejetées par le naufrage), je mesure les êtres dont ce marteau prend la place. Il remplace d'abord la longue série paradigmatique que les technologues ont eu à cœur de recréer, et qui définirait à travers l'histoire tous les substituts possibles de ce marteau-ci. Aux lieux et aux temps invisibles qu'il faudrait déployer afin de rendre justice à ce marteau, nous devrions donc ajouter, si les historiens, les préhistoriens, les paléontologues et les primatologues nous y autorisaient, la stupéfiante variété des formes dont a hérité mon banal marteau. Mais il prend place encore dans une autre série, syntagmatique cette fois, puisqu'il offre à mon poing une force, une direction, une tenue que le bras maladroit ne se savait pas posséder.

[54] Latour Bruno. La fin des moyens. In: Réseaux, volume 18, n°100, 2000. Communiquer à l'ère des réseaux. pp. 39-58. DOI : 10.3406/reso.2000.2211 URL : www.persee.fr/doc/reso_0751-7971_2000_num_18_100_2211 , vue le 30/08/2017

[55] Ibid, p.44

> Impossible ici de faire comme si le marteau « remplissait une fonction » car il déborde de toute part ce récipient dans les strictes limites duquel il ne saurait se cantonner. C'est de tous les outils (et surtout du marteau) qu'il faut dire que «l'organe crée la fonction». Avec lui en main, les possibles se multiplient, offrant à celui qui le tient des schemes d'action qui ne précédaient pas la saisie.[56]«

L'objet technique n'est pas cantonné à sa fonction, sa fabrication ou à son histoire ; il va au-delà en ce que son existence permet d'envisager à la fois des possibles, c'est-à-dire de tendre vers la création d'une nouvelle réalité, mais aussi d'envisager des réalités alternatives que sa seule existence rend obsolètes. Le danger serait de considérer l'objet technique comme un «ustensile», alors qu'il existe et agit au-delà de sa fonction avec un véritable pouvoir de suggestion, voire de contrainte. L'auteur prend l'exemple d'un bureau dont on ne peut ouvrir les tiroirs qu'un à un ; l'espace numérique, vécu à ses débuts comme un espace de liberté est également plein de ces contraintes. Impossible de lancer deux recherches simultanément dans un moteur de recherche, d'accéder à Facebook sans s'identifier, d'acheter sur Amazon sans communiquer ses coordonnées bancaires et son historique de navigation.

De même, considérer que la technique peut être sujette à maîtrise est une erreur selon Latour, qui prétend que les spécialistes passés par un long apprentissage ne «savent» plus rien de la technique dont ils sont experts : «une fois que le débutant devient expert en montant un à un les apprentissages, une fois que l'invention est devenue innovation grâce à la lente concrétisation exigée par l'industrie et le marché, on finit par pouvoir compter sur une unité d'action tellement fiable qu'elle n'apparaît plus au regard.[57]« Par l'utilisation répétée et habituelle, la technique deviendrait une sorte de fonction mathématique dont la formule resterait un mystère et dont on ne connaîtrait que les entrées et les sorties.

Cela est particulièrement applicable au numérique, dont le nombre d'utilisateurs ne cesse de croître alors que les algorithmes qui le sous-tendent recherche Google était devenu tellement complexe que les programmeurs ne pouvaient en avoir une vision d'ensemble et n'agissaient que sur des parties restreintes[58] ; en 2016, deux intelligences artificielles ont réussi à communiquer dans un langage incompris des humains[59]. L'intelligence artificielle, loin de la «maîtrise» que Latour considère comme une illusion, semble avoir réussi à prendre l'indépendance de fonctionnement que l'automatisation de tous les processus semblaient encourager - plus question désormais d'allumer son ordinateur en tapant des commandes sur le clavier ou d'écrire entièrement les lignes de codes

[56] Ibid, p.44

[57] Latour Bruno. La fin des moyens. In: Réseaux, volume 18, n°100, 2000. Communiquer à l'ère des réseaux. pp. 39-58. DOI : 10.3406/reso.2000.2211 URL : www.persee.fr/doc/reso_0751-7971_2000_num_18_100_2211 p.47, vue le 30/08/2017

[58] Le code source de Google Chrome serait douze fois plus long que celui qui pilotait la navette spatiale américaine http://www.slate.fr/life/79683/longueur-ligne-codes-comparee , vue le 30/08/2017

[59] http://hitek.fr/actualite/google-brain-intelligence-artificielle-langage_11178 , vue le 30/08/2017

d'un jeu vidéo avant de le lancer sur sa console - mais de plus avoir pris une indépendance intellectuelle, qui pose la question de la pensée des machines.

La technique semble dépasser l'humain, en lui faisant apparaître non seulement de nouvelles possibilités ou en créant de nouveaux standards, mais en altérant également le quotidien dans lequel il vit, que ce soit d'un point de vue pratique, économique ou moral.

2.5 La transformation de la gestion des données personnelles

> «Les archives sont, fondamentalement, des documents écrits destinés à conserver la preuve des décisions ou des droits de ceux qui les produisent. Si leur apparition est donc étroitement liée à l'invention de l'écriture, leur organisation est surtout consubstantielle à l'affirmation de l'autorité publique[60]«

Les archives sont produites dans le cadre d'une activité, quelle qu'elle soit, et apportent la preuve de cette activité. L'exercice et l'application du droit implique la création de nombreux documents destinés à devenir des archives. A ce titre, des données personnelles sont depuis longtemps collectées par les acteurs publics. Livret de naissance, acte de mariage et de décès, minutes de procès, casier judiciaire, tous ces documents comportent de nombreuses données personnelles - et sensibles - qui sont archivées par les pouvoirs publics. Les organismes privés collectent également depuis toujours des données personnelles dans le cadre de leurs activités : nom, adresse, informations bancaires, contrat de travail, et plus récemment des informations de localisation avec l'accroissement du nombre de caméras de surveillance. La collecte des données était nécessaire aux activités des organismes responsables de traitement - les administrations comme les entreprises privées - et se faisait à l'initiative de l'utilisateur, qui renseignait lui-même et en plein conscience des informations sur lui-même, au cours de procédures normalisées et parfois obligatoires, comme une déclaration de naissance.

Les réseaux ont apportés un changement majeur dans la perception des données personnelles. Les données personnelles et les droits vis à vis de celles-ci seraient issus d'un «présupposé que la donnée personnelle est autonome, ne renseigne que sur un seul individu, bref ils considèrent la donnée personnelle comme granulaire, indépendante et formant une entité en soi, soumise au droit d'un seul».[61]

Cette conception aurait perdu en pertinence au fur et à mesure que le développement des réseaux reliait les données personnelles. Il ne serait plus

[60] Galland Bruno, « Chapitre premier. Histoire des archives », dans Les archives. Paris, Presses Universitaires de France, « Que sais-je ? », 2016, p. 7-45. URL : http://www.cairn.info , vue le 30/08/2017

[61] Bellanger Pierre, « Les données personnelles : une question de souveraineté », Le Débat, 2015/1 (n° 183), p. 14-25. DOI : 10.3917/deba.183.0014. URL : http://www.cairn.info/revue-le-debat-2015-1-page-14.htm , vue le 30/08/2017

possible d'isoler une donnée personnelle, d'abord parce qu'une donnée personnelle peut renseigner sur d'autres individus et ensuite parce que les algorithmes connectent ces données personnelles entre elles, ce qui renseigne indirectement sur d'autres personnes que le sujet de la donnée. Les données personnelles sont devenues «réticulaires, c'est-à-dire organisées en réseau.

> « Cette intrication forme le réseau des données personnelles qui se substitue, en fait, aux données personnelles isolées du passé ».

Chaque nouvelle information se voyant renseignée, complétée par d'autres données personnelles, le réseau s'étend exponentiellement et «reproduit le réel [car] chaque donnée renseigne sur l'ensemble.[62]« Ce réseau de données personnelles est qualifié par Bellanger[63] de «bien commun [...] qui appartient à tous mais qui ne peut appartenir à personne en particulier». Il montre ensuite la difficulté de faire entrer ce bien commun dans un cadre juridique tenant compte à la fois des droits particuliers des individus sur leurs propres données personnelles et des droits collectifs.

3. LE DÉVELOPPEMENT DU MARCHÉ INTERNET

Si Internet est vécu comme un lieu de socialisation par des passionnés, il est surtout un marché nouveau à conquérir. Plusieurs entreprises commerciales proposent des (micro)ordinateurs de plus en plus performants, des logiciels, des services en ligne. Le réseau, financé par la NSF, est progressivement remplacé par une infrastructure plus performante, fruit d'entreprises commerciales qui en avaient la pleine exploitation dès 1995. A partir de 1991, le *word wide web* de Tim Berners-Lee le navigateur qu'il a conçu facilite la consultation des contenus. Les intérêts scientifiques qui avaient initié la naissance d'Internet s'étaient vus supplantés par des intérêts commerciaux.

3.1 La gratuité sur le net

3.1.1 La gratuité comme valeur pionnière

Les pionniers ont beaucoup donné pour construire ce qui constitue aujourd'hui l'environnement quotidien de plusieurs centaines de millions de personnes, même s'ils ne prévoyaient sans doute pas l'Internet actuel. En remontant plus loin, ce sont des initiatives publiques qui ont conduit jusqu'à l'Internet. C'est gratuitement que se fait le partage d'opinions, d'idées, de valeurs, et c'est également gratuitement que se consomme les biens culturels présents sur le web, parfois et surtout au début très indépendamment de la volonté de leur créateur/propriétaire, ce qui a été à l'origine du crash de l'industrie du disque dans les années 2000, au fur et à mesure que le *peer to peer* et des plateformes de vidéos

[62] Ibid.

[63] Fondateur de Skyrock et auteur de "La Souveraineté Numérique».

comme Youtube et Dailymotion, créées toutes deux en 2005, se développaient.

Cette gratuité a été une des valeurs fondatrices du net, liée à l'idéologie universaliste qui a présidé à sa création comme à ses développement et popularisation. L'information est vu comme un bien public pur[64]. Elle est à la base d'un transfert de données gratuites susceptibles d'être utilisées par des entreprises exploitantes à des fins marchandes[65]. Les FAI et les entreprises proposant des services en ligne ont souvent recours à la gratuité. Les premières peuvent proposer une gratuité partielle, en offrant un «bonus» à leurs abonnés[66]. Certains producteurs de contenus, comme les créateurs de jeux vidéos ou des journaux en ligne, offre un accès *freemium* à une partie de leur contenu afin de motiver les utilisateurs à payer pour le jeu ou l'article dans sa totalité. La gratuité a alors valeur d'essai pour l'utilisateur.

3.1.2 La gratuité comme condition de succès

La gratuité précédemment évoquée ne doit pas faire oublier une chose qui semble évidente, mais qui mérite d'être clairement posée ici : les services du web sont produits par des entreprises. Avec la rapide et très forte démocratisation d'Internet[67] se sont développés plusieurs services orientés utilisateurs : moteurs de recherche, plateformes multimédia, réseaux sociaux, etc. La nature même de la pratique Internet abstrait cet aspect de l'expérience de l'utilisateur, mais l'accroissement démographique de la population Internet, rend l'investissement en ressources physiques – notamment en support de mémoire – indispensable pour les producteurs, que ces derniers ne peuvent répercuter sur les utilisateurs au risque de les voir tourner le dos. Il a donc fallu pour tous les producteurs, des GAFA jusqu'au bloggeur photographe semi-professionnel, trouver une autre mâne que les consommateurs.

Le fait que les coûts des transactions numériques aient été élevés lors de l'installation de services marchands sur le web n'est pas étranger à la gratuité[68]. Cependant, les données personnelles constituent surtout la monnaie du système de financement du numérique. Pour faire une analogie, les données personnelles sont aux entreprises du web ce que le pétrole était aux industries du carburant du XXème : une matière première indispensable, que l'on prélève et commercialise massivement sans envisager les conséquences d'une pénurie. Ce sont les données personnelles communiquées aux entreprises du net qui permettent la gratuité, en permettant d'une part à ces firmes de proposer plus facilement des services innovants et attractifs et d'autre part en servant de monnaie d'échange.

[64] Dang Nguyen Godefroy, Pénard Thierry, « La gratuité à la croisée des nouveaux modèles d'affaires sur l'Internet », Réseaux, 2004/2 (no 124), p. 81-109. DOI : 10.3917/res.124.0081. URL : http://www.cairn.info/revue-reseaux1-2004-2-page-81.htm , vue le 30/08/2017

[65] Ibid.

[66] Fin 2015, pour un abonnement Internet, Bouygues offrait deux mois d'accès à un bouquet de chaînes privées.

[67] " Les utilisateurs d'internet sont passés de 150 000 en 1995 à 26 millions en 2005 en France.» https://www.insee.fr/fr/statistiques/1280834 , vue le 30/08/2017

[68] Dang Nguyen Godefroy, Pénard Thierry, « La gratuité à la croisée des nouveaux modèles d'affaires sur l'Internet », Réseaux, 2004/2 (no 124), p. 81-109. DOI : 10.3917/res.124.0081. URL : http://www.cairn.info/revue-reseaux1-2004-2-page-81.htm , vue le 30/08/2017

3.1.3 Les coûts de la gratuité

Derrière la gratuité des plateformes en ligne se cachent d'importants coûts, assumés par d'autres acteurs. L'accès à Internet implique la production d'électricité[69], et représente 4% à 8% de la production mondiale d'électricité. Cela induit de forts coûts écologiques : « l'envoi des emails d'une entreprise de 100 personnes générait chaque année pas moins de 13,6 tonnes équivalent CO2, soit environ 13 allers-retours Paris-New York ».

L'Etat est un acteur majeur du développement d'Internet. Aux Etats-Unis, les premières dorsales qui distribuaient Internet à travers le territoire ont été financées par de l'argent public.

En France, des initiatives étatiques, visant à réduire la fracture numérique règlementent et financent directement ou indirectement plusieurs d'action. Une loi oblige les constructeurs depuis 2011, à prévoir l'équipement d'une infrastructure internet pour les nouveaux immeubles collectifs dans les zones au trafic Internet denses et les lotissements. Une autre permet aux foyers modestes de bénéficier d'une aide de l'Etat pour maintenir leur connexion à Internet[70].

Les régions financent en partie l'accès à Internet par satellite, lorsque les utilisateurs se trouvent dans des zones qui ne peuvent pas être desservis par les FAI classiques[71].

3.2 Le plus grand marché du monde financé par la publicité

3.2.1 Un marché inédit

Le marché Internet est né de plusieurs constats. Le premier est celui des producteurs de services, et notamment des moteurs de recherche, comme Google, et répond à une logique d'entreprise : pour perdurer, s'étendre, et faire du profit, une entreprise doit pouvoir commercialiser quelque chose, et ce quelque chose ne peut pas être leur production en tant que telle.

Le second est celui de l'évolution technologique : au fur et à mesure que les machines donnant accès à Internet se démocratisent, se multiplient, et se rendent mobiles, le temps moyen passé sur celles-ci ne cesse de s'accroître et celui passé sur les média traditionnels de se réduire.

Le troisième appartient aux publicitaires et poursuit le point précédent : l'accroissement démographique toujours vivace de la population Internet, couplé au temps dédié à sa consommation, en fait un lieu de trafic exceptionnellement fréquenté, en conséquence un lieu idéal de publicité. La publicité a donc pris sa place sur Internet : ce sont les annonceurs qui financent les producteurs en achetant de la place sur leurs pages en fonction de la visibilité permise et de la

[69] Quel est le coût écologique de votre surf sur Internet ?, http://www.leparisien.fr/environnement/energies/quel-est-le-cout-ecologique-de-votre-surf-sur-internet-17-11-2015-5285997.php , vue le 30/08/2017

[70] https://www.legifrance.gouv.fr/eli/loi/2016/10/7/2016-1321/jo/article_108 , vue le 30/08/2017

[71] http://www.ariase.com/fr/observatoire/subventions-satellite.html , vue le 30/08/2017

fréquentation du site. A titre d'exemple, Google a perçu plus de 79 milliards de la publicité[72].

3.2.2 Une nouvelle approche publicitaire

Le système de financement par la publicité est le même en ligne et IRL - pour les journaux par exemple -, à ceci près que les possibilités de traitement des données par des machines toujours plus performantes et des algorithmes toujours plus sophistiqués ont fait basculer la façon dont la publicité est envisagée. Depuis la consommation de masse, la logique des publicitaires étaient de faire parvenir le message au plus grand nombre afin que le plus de consommateurs possible soient touchés. Il existait déjà des vélléités de ciblage marketing, mais elles reposaient plus sur le comportement imaginé de leur cible par les publicitaires et les responsables marketing que sur ses appétences réelles. Par exemple, les publicités pour les produits ménagers s'adressaient et s'adressent encore aujourd'hui quasi-exclusivement à des femmes, à des "ménagères" alors que les hommes se retrouvent également devant la nécessité de nettoyer les vitres ou de faire la vaisselle.

Le traitement des données des utilisateurs recueillies par les producteurs permet de créer un profil de chacun, ou des profils types plus élaborés que ceux de la publicité traditionnelles, dans la mesure où il est possible d'avoir accès à l'individualité de chacun : les statistiques s'approfondissent, elles deviennent plus représentatives dans la mesure où elles ne s'établissent plus à partir d'un panel mais du public tout entier. Il est possible pour les publicitaires, grâce à l'action des producteurs, de cibler plus ou moins largement les consommateurs potentiels, voire d'orienter leurs achats grâce à des systèmes plus ou moins agressifs allant de la mention indicative "Les autres clients ont aussi consulté" à la diffusion massives de publicité pour des produits similaires à ceux dont l'utilisateur vient de consulter la page. Cela rend possible pour les annonceurs de payer au *pro-rata* du public ciblé, et d'envisager leur campagne publicitaire comme une frappe chirurgicale.

3.3 Redéfinition d'un modèle économique

C'est une des grandes spécificités de l'Internet : le consommateur n'est pas le principal client. Bien sûr, les utilisateurs sont aussi sollicités en tant que consommateurs par les entreprises de l'Internet ; l'action des entreprises concernées reste cependant traditionnelle, ne s'inscrit pas dans les changements induits par le médium digital. C'est le cas par exemple des sites de vente en ligne, *pure players* ou nouvellement engagés sur le marché du net. C'est bien de cela qu'il s'agit : d'un lieu d'expression-communication, a émergé le plus grand marché du monde.

[72] "Google et Facebook contrôlent 20% du total des investissements publicitaires», Les Echos, Véronique Richebois, https://www.lesechos.fr/tech-medias/medias/030372465160-google-et-facebook-controlent-20-du-total-des-investissements-publicitaires-2092549.php , vue le 30/08/2017

Mêlant intimement dans un lieu intangible transactions financières et production culturelle, consommation de masse et gratuité, capitalisme et contre-culture, Internet redistribue les rôles pour un nouveau modèle d'économie de marché qui dissocie le client du consommateur. Le client est le publicitaire, qui va dépenser de grosses sommes afin de pouvoir atteindre les consommateurs. Ces derniers accèdent gratuitement aux services du net, à la condition de permettre aux producteurs de ces services un accès plus ou moins ouvert aux informations à leurs données personnelles. Les producteurs de service trouvent leur profit en jouant sur un principe de vases communicants : la nécessité d'amener toujours plus de consommateurs vers ses services afin de valoriser ses statistiques de fréquentation et les informations de ses utilisateurs auprès des annonceurs, qui en retour achètent de la place publicitaire, et financent ainsi le développement des services proposés par les producteurs qui pourront attirer un public plus grand.

Chacune des parties financent l'autre dans un cercle économiquement vertueux. Néanmoins, il semble que dans ce cercle, seules deux des trois parties soient réellement actives, et seules deux des trois parties en tirent bénéfice - et quels bénéfices, les seuls français ayant dépensé soixante-douze milliards dans le e-commerce sur l'année 2016. Dans ces soixante-douze milliards ne sont bien sûrs pas comptés les millions engendrés par les transactions entre les producteurs de services du net et leurs clients annonceurs.

EXPLOITATION COMMERCIALE DES DONNÉES PERSONNELLES

1. LA CAPITALISATION DES DONNÉES PERSONNELLES

1.1 Acteurs et environnement juridique

1.1.1 Les grands acteurs du web

Le web n'a pas été créé dans un but commercial, mais de recherche. Pour autant, le phénomène de démocratisation du web a pris de l'ampleur en même temps que les origines de recherche scientifique cédaient à des intérêts commerciaux. De 1995 à aujourd'hui, quatre entreprises ont pris une telle importance sur le marché de l'Internet qu'elles ont reçu la qualification de «géants du web». Il s'agit de Google, Apple, Facebook et Amazon, aussi désignées toutes les quatre par l'acronyme «GAFA». A l'exception de Facebook, il s'agit d'entreprises pionnières dans leur domaine. Avant toute chose, il convient de dresser un bref horizon de l'importance de ces entreprises, en ligne et dans le monde. Le but de cette partie est avant tout de présenter les GAFA ; une description plus longue sera faite pour les entreprises Facebook et Google, parce qu'elles sont celles qui retirent le plus de bénéfices de la monétisation des données personnelles.

1.1.1.1 Amazon : la confiance par l'interinfluence

Jeff Bezos a lancé le site de librairie en ligne Amazon en 1995. L'entreprise s'est vite diversifiée, étendant la boutique à tous les objets du quotidien - de la gamelle pour chat à la machine à laver - et innovant aussi dans d'autres domaines, bataillant ferme dans le marché du *hardware*[73] avec la liseuse numérique Kindle, les tablettes et investissant le service de cloud avec AWS[74].

La stratégie d'Amazon repose sur une stratégie de bénéfices sur le long terme imprimé par son président-directeur-général et fondateur Jeff Bezos[75]. Celle-ci consiste à réinvestir la totalité des résultats opérationnels dans le but d'élargir son offre. Une fois le nouveau secteur d'activité investi, Amazon casse les prix et les reins de la concurrence. Si celle-ci possède une technologie intéressante, elle est ensuite rachetée et intégrée. Une fois en situation de monopole, Amazon commence alors à faire des bénéfices, grâce à des économies d'échelle, réutilisés ensuite pour dominer un autre secteur. A titre d'exemple, Amazon a en 2005

[73] Dahan Michel, « Une guerre économique d'une violence inédite », Le journal de l'école de Paris du management, 2014/3 (N° 107), p. 36-42. DOI : 10.3917/jepam.107.0036. URL : http://www.cairn.info/revue-le-journal-de-l-ecole-de-paris-du-management-2014-3-page-36.htm , vue le 30/08/2017

[74]« En vingt ans Amazon a conquis Wall Street », Jérôme Marin , http://www.lemonde.fr/economie/article/2017/05/18/en-vingt-ans-amazon-a-conquis-wall-street_5129744_3234.html , vue le 30/08/2017

[75]«La stratégie d'Amazon à l'épreuve des marchés », La stratégie d'Amazon à l'épreuve des marchés, http://www.lemonde.fr/economie/article/2014/01/31/la-strategie-d-amazon-a-l-epreuve-des-marches_4357711_3234.html , vue le 30/08/2017

racheté Mobipocket, qui vendait des livres numériques, et c'est le format mobi, à peine retouché, qui est utilisé dans la Kindle d'Amazon depuis sa sortie en 2007.

Amazon a aussi contribué à l'appropriation du web par les utilisateurs, en leur donnant le pouvoir de s'influencer les uns les autres, insérant dans l'acte d'achat, voire de simple consultation, une dimension sociale[76]. Chaque action de l'internaute sur chaque produit est comptabilisée pour qu'Amazon puisse déterminer les produits les plus vus, les plus populaires, mais également ceux que les utilisateurs ont préféré (par un système d'évaluation de la satisfaction) et pour un produit donné, indiquer ce que les autres acheteurs ont également consommé. Cela se fonde sur l'idée que des personnes qui achètent des produits identiques auront sans doute d'autres points communs. Cela tient parfois à la simple logique : si un utilisateur recherche une médaille pour chat, les produits apparaissant dans la catégorie «les acheteurs de ce produit ont également acheté ça» sont logiquement des colliers, des laisses et des produits de soin aux animaux. Amazon transforme, grâce à la recommandation généralisée, une plateforme de vente de biens en ligne en une communauté d'acheteurs, valorisant ainsi l'expérience de ses utilisateurs en un service qui encourage l'utilisateur a acheté, puisque guidé par l'expérience des précédents acheteurs, devenus, sans doute sans que ceux-ci n'en aient eu l'intention, des gatekeepers[77].

Aujourd'hui, Amazon représente 2,4 millions de tablettes dans le monde, soit 6,4% de part de marché en 2017 ; il a aussi absorbé 43% des ventes de e-commerce aux Etats-Unis, un secteur qui représentait à la même période 395 milliards de dollars[78].

1.1.1.2 Facebook : l'exposition de soi

Le réseau social à succès Facebook, créé par Mark Zuckerberg en 2004 à destination des étudiants de Harvard[79], est accessible à tous deux ans plus tard. Le nombre d'utilisateurs ne cesse de grimper.

C'est un service en ligne qui propose à ses utilisateurs une structure d'expression combinée associée à la mise en réseau des utilisateurs choisis entre eux, et qualifié d'»amis». Le financement de Facebook se fait exclusivement grâce à la publicité, en commercialisant les données des utilisateurs.

«La particularité de Facebook est de proposer aux internautes un espace d'expression à la fois public et privé, en « clair-obscur[80] »«.

Les utilisateurs, s'ils le souhaitent peuvent paramétrer le champ de diffusion de leur profil, constitué des renseignements qu'ils ont eux-mêmes consentis à inscrire sur Facebook (posts, like, profil), mais aussi ce que les autres ont diffusé à

[76] Merzeau Louise, « Présence numérique : les médiations de l'identité », Les Enjeux de l'information et de la communication, 2009/1 (Volume 2009), p. 79-91. URL : http://www.cairn.info/revue-les-enjeux-de-l-information-et-de-la-communication-2009-1-page-79.htm , vue le 30/08/2017

[77] Ibid.

[78] «Chiffre d'affaires de l'e-commerce aux Etats-Unis », www.journaldunet.com/ebusiness/commerce/1087478-chiffre-d-affaires-e-commerce-etats-unis/ , vue le 30/08/2017

[79]« Facebook », http://www.numerama.com/startup/facebook , vue le 30/08/2017

[80] Bastard Irène, Cardon Dominique, Charbey Raphaël et al., « Facebook, pour quoi faire ? Configurations d'activités et structures relationnelles », Sociologie, 2017/1 (Vol. 8), p. 57-82. DOI : 10.3917/socio.081.0057. URL : http://www.cairn.info/revue-sociologie-2017-1-page-57.htm , vue le 30/08/2017

leur propos (commentaires, partages). Facebook permet à un individu de se mettre en scène massivement, et de s'exposer aux autres[81] et à leur jugement.

Fin 2016, on décomptait sur le site 1, 86 milliards d'utilisateurs actifs[82] par mois, soit plus d'un quart de l'humanité. Environ 93% de ces utilisateurs utilise un support mobile[83].

1.1.1.3 Apple : enfermer l'utilisateur dans un environnement propriétaire

Apple est une entreprise californienne fondée par Ronald Wayne, Steve Wozniak et Steve Jobs, en 1976, c'est-à-dire au cours de la démocratisation des ordinateurs personnels. La première activité de la firme est la construction et la vente d'ordinateurs. Les produits de la firme sont reconnus autant pour la qualité de leur matériel informatique que pour celle des logiciels et applications qui l'accompagnent[84]. Ils sont particulièrement réputés pour leur navigation instinctive. Apple ne répond pas aux besoins des utilisateurs, mais les crées.

Apple propose son propre système d'exploitation, ses propres logiciels : la stratégie de la firme consiste à amener ses utilisateurs à intégrer un environnement logiciel certes performant, mais surtout cloisonné. Cela oblige les utilisateurs à toujours acheter Apple, puisqu'ils fonctionnent ensemble de façon optimale[85]. Une fois utilisateur d'un Mac, par exemple, le mobile le plus logique à acquérir pour pouvoir profiter de toutes ses possibilités - application, cloud - est l'Iphone.

Dans le même temps, la sortie de chaque produit est soigneusement marketée par les équipes d'Apple, selon un modèle similaire[86]. Aucune information ne circule, jusqu'à ce qu'une poignée de journalistes choisis par la marque soient invités à tester le prototype dans une ambiance détendue, ce qui les amène au moins à l'indulgence. Enfin, lors de la sortie du produit, la firme donne l'illusion que son produit est rare et surtout très recherché en contrôlant les stocks disponibles de son nouveau produit.

A travers le monde, le système d'exploitation appelé les systèmes d'exploitation iOS et MacOSX équipaient à travers le monde en 2016 presque 18% des smartphones[87], puis en 2017 6,72% des tablettes et MacOSX 10, 49% des ordinateurs[88].

---

[81] Brodin Oliviane, Magnier Lise, « Le développement d'un index d'exposition de soi dans les médias sociaux : phase exploratoire d'identification des indicateurs constitutifs », Management & Avenir, 2012/8 (N° 58), p. 144-168. DOI : 10.3917/mav.058.0144. URL : http://www.cairn.info/revue-management-et-avenir-2012-8-page-144.htm , vue le 30/08/2017

[82] «Nombre d'utilisateurs de Facebook dans le monde » www.journaldunet.com/ebusiness/le-net/1125265-nombre-d-utilisateurs-de-facebook-dans-le-monde/ , vue le 30/08/2017

[83] Ibid.

[84]« Apple » http://www.numerama.com/startup/apple , vue le 30/08/2017

[85]« Apple nous oblige gentiment à acheter toujours plus Apple » Will Oremus, traduit par Antoine Bourguilleau http://www.slate.fr/story/123279/apple-oblige-acheter-apple , vue le 30/08/2017

[86] « Apple récit d'un succès » ,http://radio-londres.fr/2016/10/apple-recit-dun-succes/ , vue le 30/08/2017, vue le 30/08/2017

[87] «Chiffres clés : les OS pour smartphones » , Christophe Auffray http://www.zdnet.fr/actualites/chiffres-cles-les-os-pour-smartphones-39790245.htm vue le 30/08/2017

[88] https://fr.statista.com/statistiques/557282/parts-de-marche-mondiales-de-systeme-exploitation-ordinateur-tablette/ , vue le 30/08/2017

1.1.1.4 Google : un moteur de recherche omniprésent

Google est l'entreprise de Sergei Brin et Larry Page, créée en 1998, dont le premier service a été de proposer un moteur de rechercher plus simple et gratuit, plus sobre et surtout plus performant que ses concurrents de l'époque grâce à la technologie du PageRank, un algorithme qui classe les sites web selon les liens hypertextes qui les relie entre eux. La firme s'est par la suite diversifiée en développant entre autre un navigateur et un système d'exploitation pour smartphone. En 2015, 88, 66% des recherches sur le net étaient effectuées via le moteur de recherche de Google ; en mai 2017 Google Chrome représentait 63, 36% du marché des navigateur en mai 2017[89], et son système d'exploitation Android équipait 3,36% des tablettes la même année et 85% des smartphones en 2017[90].

Le sociologue Dominique Cardon s'est intéressé à l'algorithme utilisé par Google. Le PageRank, d'abord, a pour mission d'indexer le web. Alors qu'au départ, les moteurs de recherche ne permettaient d'accéder qu'à peu de pages web, en raison de la difficulté de trouver une méthode d'indexation qui rendrait à la requête de l'utilisateur une réponse pertinente :

> «Avant Google, le web était une vaste loterie. Les réponses aux requêtes des internautes étaient hasardeuses, souvent fantaisistes, massivement truquées et occasionnellement pertinentes. Les premiers moteurs de recherche fonctionnaient à partir de mots clés et mesuraient la densité de la présence du terme recherché sur les différentes pages du web. En 1990, les pionniers, Archie et Veronica, n'indexaient que le titre du document, avant qu'en 1994, le WebCrawler de Brain Pinkerton ne prenne en compte l'ensemble du texte de la page[91]«.

L'algorithme de Larry Page a permis d'indexer de façon plus efficace que les autres moteurs de recherche de la même époque, en s'intéressant à la structure des pages et non à leur contenu.

Le succès du moteur de recherche de Google le rend désormais presque seul décideur de l'ordre d'indexation et donc du degré d'accessibilité de l'information. Google se sert de la reconnaissance des internautes pour classer les sites web, à la manière du Science Citation Index (SCI), un index permettant d'identifier quels document sont les plus souvent cités dans la recherche scientifique, créé à l'initiative d'Eugène Garfield en 1964, dont « l'objectif n'était pas alors de mesurer la réputation des chercheurs, mais de « fournir au scientifique consciencieux un moyen de connaître les critiques dont les articles qu'ils citent ont fait l'objet » en révélant des « associations d'idées [92]»«.

Chaque lien vers une autre page fonctionne comme un vote : plus une page est citée, plus son pouvoir de prescription grandit, et plus ce qu'elle cite prend de l'importance et remonte dans les résultats du moteur de recherche.

---

[89]«Chiffres clés : les OS pour les navigateurs internet » , Christophe Auffray http://www.zdnet.fr/actualites/chiffres-cles-les-navigateurs-internet-39381322.htm , vue le 30/08/2017

[90]«Chiffres clés : les OS pour smartphones » , Christophe Auffray http://www.zdnet.fr/actualites/chiffres-cles-les-os-pour-smartphones-39790245.htm , vue le 30/08/2017

[91] Cardon Dominique, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », Réseaux, 2013/1 (n° 177), p. 63-95. DOI : 10.3917/res.177.0063. URL : http://www.cairn.info/revue-reseaux-2013-1-page-63.htm , vue le 30/08/2017

[92] Ibid

Pour Cardon, le SCI présuppose cinq qualités : extériorité, abstraction, procéduralisme, neutralité, honnêteté[93]. Des qualités que le moteur de recherche cherche à atteindre, sans toujours y parvenir. Google se revendique comme extérieur aux résultats produits par son PageRank. L'algorithme ne s'intéresse qu'à la structure des pages web et aux liens de citation qui les unissent, il ne peut donc pas être considéré comme responsable des résultats. C'est la défense du moteur de recherche pour chaque polémique dont il est victime, particulièrement lorsqu'une personne fait valoir son droit à l'oubli : «*our third philosophy: no manual intervention*[94]«. Le caractère procédural du traitement des pages web élimine toute subjectivité des résultats.

Le PageRank fournit une représentation hiérarchisée de la production des internautes, en partant du principe que ces internautes ne prennent pas en compte dans leur production les méthodes de ranking. Or, les méthodes pour permettre aux personnes d'améliorer leur ranking sur Google sont légion[95]. Pour Cardon, l'abstraction dans le cadre du traitement de citation par des liens hypertextes, a conduit à «monétiser» la citation. Du classement de Google dépend grandement l'audience d'un site web.

L'action des webmestres consiste alors à connaître les méthodes de ranking pour les exploiter à son avantage. Une action compréhensible, mais qui fausse en partie les résultats et force la firme à «s'affranchir de son procéduralisme[96]« en intervenant sur l'algorithme pour créer de nouvelles règles afin de lutter contre ces pratiques, ce qui contribue à fragiliser la neutralité du moteur de recherche.

En 1998, les fondateurs de Google déclaraient «qu'il existe une incompatibilité de principe entre la recherche d'information et la publicité». Pourtant, outre de faire partie de heureux élus à être jugés les plus pertinents pour une recherche donnée, la solution qui consiste, pour des annonceurs à acheter des mots-clés à Google par un système d'enchères, afin d'apparaître en premier dans les recherches des internautes, à condition que l'apparition soit pertinente.

En réalité, Google ne vend pas les données personnelles qu'il collecte à proprement parler. En revanche, il croise des données utilisateurs afin de vendre une audience ciblée et susceptible d'achat. Prenons l'exemple d'une entreprise de bricolage qui paie pour diffuser sa publicité. La diffusion des annonces concernera les utilisateurs qui ont recherché un mot-clé associé comme «étagère» ou «serre-joint» ; également, ceux qui à travers le contenu qu'ils créent dans les services de Google (mail, Sheets, agenda, etc.) auront souvent employé un des mots-clés associés.

Payer pour exister en tête des résultats Google fonctionne, mais il ne s'agit pas d'entrer de façon immédiate et invisible dans le classement. Google, qui refusait à a dressé une barrière visuelle infranchissable - surnommée la «muraille de Chine» - entre les résultats jugés les plus pertinents et les résultats financés par les annonceurs.

---

[93] Ibid

[94]« *Introduction to Google ranking* » Amit Singhal https://googleblog.blogspot.fr/2008/07/introduction-to-google-ranking.html , vue le 30/08/2017

[95] En combinant les mots-clés "amélioration» "Google» et "référencement», le moteur de recherche fait remonter 17 400 résultats

[96] Cardon Dominique, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », Réseaux, 2013/1 (n° 177), p. 63-95. DOI : 10.3917/res.177.0063. URL : http://www.cairn.info/revue-reseaux-2013-1-page-63.htm , vue le 30/08/2017

Cette volonté de séparer les résultats méritocratiques des résultats obtenus par la vente de données personnelles marque une véritable volonté de se conformer aux valeurs des pionniers en conservant aux premiers une vraie autorité. Un consensus qui permet à Google de garder la confiance des utilisateurs tout en engrangeant de très grosses sommes en provenance des annonceurs. Un business qui s'est révélé très lucratif : au dernier trimestre 2017, la publicité permettait à Google - et à sa société mère Alphabet - d'atteindre 22, 4 milliards de recettes[97].

1.1.2 Un environnement juridique permissif

1.1.2.1 Les Etats-Unis et la vie privée

Historiquement, les Etats-Unis sont le premier pays à avoir utilisé Internet et étendu leur économie sur le territoire numérique. La population des Etats-Unis est également la plus connectée au monde, avec un taux de pénétration d'Internet de XX%. Pourtant, la problématique du respect de la vie privée qui préoccupe l'Europe depuis les années 1970 y est beaucoup moins prégnante. A l'exception de la loi sur les données personnelles et sensibles en lien avec l'armée, les Etats-Unis ne disposent pas de texte d'application fédérale destiné à protéger la vie privée des citoyens dans leur usage de l'Internet, notamment face à la commercialisation qui en faite.

La Constitution des Etats-Unis, publiée en 1787, n'évoque pas le droit au respect de la vie privée. Les amendements apportés par la Déclaration des Droits (ou *Bill of Rights*), à l'initiative du président Washington en 1791[98] offre une première reconnaissance de la vie privée des citoyens avec le droit à la libre religion, la libre expression, le droit à la propriété. Cette reconnaissance réside pourtant uniquement dans le rapport qu'entretient l'Etat avec les citoyens. La Déclaration des Droits restreint le pouvoir de l'Etat sur la vie des citoyens, en leur garantissant une non-interférence totale des pouvoirs étatiques avec la vie des citoyens, en dehors du cadre prévu par la loi[99]. Elle ne leur garantit aucun moyen de protéger contre une violation de la vie privée dont l'Etat ne serait pas à l'origine.

Internet, longtemps resté sous l'égide du ministère américain du commerce, et fidèle à la tradition de non-intervention de l'Etat dans les marchés[100], est peu réglementé aux Etats-Unis, comparativement à l'Europe. La loi fédérale sur la vie privée en ligne s'intéresse essentiellement à la protection de la vie privée des enfants, n'interdit pas la collecte des données personnelles des mineurs de moins de treize ans, mais exige seulement des collecteurs d'obtenir le consentement des enfants : «operators of commercial web sites and online services to provide notice and get verifiable parental consent before collecting personal information from children under the age of 13».

---

[97] « Alphabet (Google) : un bénéfice net de 19,5 milliards de dollars en 2016, et pourtant... » http://www.latribune.fr/technos-medias/alphabet-google-un-benefice-net-de-19-5-milliards-de-dollars-en-2016-et-pourtant-633769.html , vue le 30/08/2017

[98] https://www.herodote.net/17_septembre_1787-evenement-17870917.php , vue le 30/08/2017

[99] Par exemple, le troisième amendement protège les citoyens contre le réquisitionnement abusif : "Aucun soldat ne sera, en temps de paix, logé dans une maison sans le consentement du propriétaire, ni en temps de guerre, si ce n'est de la manière prescrite par la loi.» URL : http://mjp.univ-perp.fr/constit/us1787a.htm , vue le 30/08/2017

[100] « Le marché et l'État », Revue de l'OFCE, 2012/2 (n° 121), p. 73-82. DOI : 10.3917/reof.121.0073. URL : http://www.cairn.info/revue-de-l-ofce-2012-2-page-73.htm , vue le 30/08/2017

Les droits de la personne concernant le respect de sa vie privée sur en ligne n'existent pas dans la loi fédérale pour les adultes. En revanche, elle protège les agences fédérales contre les intrusions informatiques.

1.1.2.2 La Californie : une législation très permissive sur le traitement de données personnelles

Les principaux services en ligne et tous les GAFA sont originaires de Californie et exercent leur activité selon le droit californien. Ce dernier étant très permissif, il était difficile et coûteux pour les utilisateurs non-californiens qui s'estimaient lésés par l'aspiration et la revente de leurs données personnelles de porter leurs revendications devant un tribunal.

La principale régulation de la loi californienne visant la collecte de données personnelles par des organismes privés est d'ordre contractuel et date de 2004. Elle impose aux exploitants des données personnelles de mettre à la disposition de leurs utilisateurs une politique de confidentialité, laquelle doit lister les catégories de données personnelles qui sont collectées, ainsi que les personnes ou organismes tiers qui sont susceptibles d'accès aux informations collectées[101]. Les organismes privés sont susceptibles d'être poursuivis s'ils refusent de mettre en place une politique de confidentialité et s'ils ne respectent pas leur propre politique de confidentialité. Aucune restriction n'est mentionnée concernant le contenu de la politique de confidentialité.

Les autres lois ont des champs d'action plus précis : par exemple, le piratage d'un ordinateur situé en Californie par un particulier est interdit ; les agences fédérales doivent obtenir le consentement d'un individu avant de partager entre elles les données personnelles qui lui sont relatives ; des données personnelles comme les coordonnées physiques (adresses, numéro de téléphone) des élus, des personnes ayant eu recours à des services génésiques ainsi que les médecins associés, des femmes victimes de violences participant à un programme de protection[102]. D'autres lois concernent plus les atteintes à la vie privée que les données personnelles. Par exemple, les coupables de cyberbullying[103] sont punis de suspension ou d'exclusion de leur école[104]. L'essentiel des lois de protection des données personnelles et de la vie privée concerne les mineurs. Une loi récente autorise même les mineurs à demander la suppression de toutes les traces qu'ils ont laissée sur Internet, notamment sur les réseaux sociaux[105].

La loi aux Etats-Unis se nourrit aussi de jurisprudence. Lorsqu'une personne s'estime lésée, et qu'un précédent établit que le comportement qui en est à l'origine a déjà été condamné par un tribunal - s'il a par exemple été jugé déloyal - une autre décision de justice peut interdire le comportement incriminé, quand bien même la loi ne le décrit pas comme sujet à sanction. Inversement, une jurisprudence en faveur d'un comportement controversé peut motiver l'abandon des poursuites. Les services en ligne ont connu de nombreux procès, qui ont aussi

[101] « *Privacy law* »https://oag.ca.gov/privacy/privacy-laws# , vue le 30/08/2017

[102] Ibid.

[103] "Cyberbizutage» : concrètement, il s'agit de harcèlement en ligne, perpétré sur un élève par un ou des autres élève(s).

[104]*« Privacy law »* https://oag.ca.gov/privacy/privacy-laws# , vue le 30/08/2017

[105] http://www.usine-digitale.fr/article/une-loi-californienne-devrait-permettre-aux-mineurs-d-effacer-leur-passe-sur-internet.N205809 , vue le 30/08/2017

été des occasions pour le public d'améliorer leur connaissance des pratiques des ces entreprises exploitant leurs données. Par exemple, la cour d'appel de Californie a statué en faveur de Facebook, qui avait refusé de fermer la page d'un groupe dont les membres appelaient à la violence et menaçaient de mort le rappeur Mikel Knight. Un autre jugement a sanctionné Amazon pour être entré en possession de plus de données personnelles sur ses utilisateurs que sa politique de confidentialité le déclarait[106].

Il faut noter malgré tout une évolution vers une plus grande protection des données personnelles dans le pays des géants de l'Internet. L'Etat de Californie, avant la loi de 2015 sur le droit à l'oubli des mineurs, proposait en 2011 un projet de loi (avorté) destiné à donner aux utilisateurs la possibilité de refuser entièrement tout traitement de leurs données à des fins commerciales[107].

1.2 Collecte des données personnelles

1.2.1 Un objectif commercial

Internet a vu les principaux services qu'il a permis de proposer se structurer autour d'un système de rémunération par la publicité. Il est intéressant de noter que les plus répandus proposent voire exigent de l'usager une « identification ». Comme lors d'une interpellation, Facebook demande avant même de souhaiter la bienvenue au visiteur, qu'il lui communique des informations sur sa personne : civilité, âge, adresse mail, patronyme et prénom, parfois adresse postale ou numéro de carte bancaire. Toutes ces informations sur l'usager ne sont pas nécessaires en tant que telles au fonctionnement du site. Les mêmes services seront proposés à chaque usager, indépendamment des informations qu'il a fournies. Un utilisateur Facebook, par exemple, peu importe qu'il soit âgé de treize[108] ou soixante-dix ans, pourra accéder aux mêmes fonctionnalités du site parmi lesquelles post, like, et commentaires.

La récolte de ces informations répond, en conséquence, à d'autres objectifs que de procurer des services, contrairement à ce que déclarent les sites qui affirment que le refus de leur procurer une information pourrait entraîner des dysfonctionnements dans l'expérience de l'utilisateur. Ces dysfonctionnements sont moins consécutifs du refus de l'utilisateur de donner une information que de la punition infligée par le site suite à ce refus.

1.2.2 Les informations collectées par les GAFA

1.2.2.1 Les données renseignées par les utilisateurs

Les données collectées par les géants du web sont légion[109] : de l'adresse mail fournie à l'ouverture d'un compte Facebook à la manière dont l'utilisateur a

[106] https://www.cnet.com/news/amazon-unit-settles-privacy-lawsuit/ , vue le 30/08/2017

[107] « Facebook et Google s'opposent a une proposition de loi califonienne sur la vie privé » http://www.lemonde.fr/technologies/article/2011/05/09/facebook-et-google-s-opposent-a-une-proposition-de-loi-californienne-sur-la-vie-privee_1519254_651865.html , vue le 30/08/2017

[108] Il s'agit de l'âge minimum requis pour gérer un compte Facebook selon la réglementation du réseau social.

[109] Voir annexe 1.

scrollé sur la page, la vie de l'individu sur Internet est enregistrée par de multiples traceurs. Certaines de ces données sont renseignées par l'action volontaire des utilisateurs, par exemple lors de l'ouverture d'un compte Gmail, lorsque l'utilisateur répond à un questionnaire de satisfaction, ou renseigne son numéro de carte bancaire pour une transaction. Il s'agit presque toujours de données personnelles.

1.2.2.2 Les données d'activité des plateformes des GAFA

D'autres données sont collectées par des systèmes automatisés des plateformes. Par exemple, lorsqu'un utilisateur partage sur Facebook un contenu avec un groupe d'amis, Facebook (qui dispose déjà des informations personnelles renseignées par l'utilisateur) enregistre le contenu partagé, les heure, date et fuseau horaire du partage, le nom du groupe dans lequel ce contenu a été partagé, la liste des personnes qui en sont membres, ainsi que le nombre de likes, de commentaires, leurs auteurs, la date, heure et fuseau horaire de ces commentaires, le contenu de ces commentaires. Toutes ces informations sont collectées sur la plateforme de Facebook, et permettent au site de retracer ses activités.

1.2.2.3 Les données collectées grâce aux traceurs

D'autres informations qui ne concernent pas l'activité sur les plateformes sont collectées par les GAFA, grâce à des moyens indirects comme l'installation de cookies ou de balises web[110] sur le hardware de l'utilisateur. Les balises web - ou pixels invisibles - correspondent à des fichiers images invisibles pour l'utilisateur, affichées sur une page web et stockées sur un serveur, qui sont utilisées pour mesurer le trafic sur un ou plusieurs sites web utilisant le même serveur. Les balises web peuvent - et sont souvent - utilisées de pair avec les cookies.

Un cookie est un fichier texte stocké par un site ou une application le plus souvent sur le répertoire du navigateur de l'utilisateur. Ces cookies peuvent avoir de multiples fonctions, allant de la personnalisation de l'accès à un site - langue préférée, couleur d'affichage - au pistage de l'internaute. Par exemple, certains cookies ou balises affectent à chaque appareil un nom aléatoire, qui va permettre au site de reconnaître l'appareil et de décompter le nombre de visites d'utilisateurs uniques, ou de connaître le nombre et la durée des visites, les pages visitées, ou le contenu du panier d'achat s'il s'agit d'un site de e-commerce. Il ne s'agit que d'exemples, les utilisations des cookies couvrent un bien plus large panel de possibilités.

Les cookies ne sont pas définitivement implantés sur le terminal de l'utilisateur, ils ont une durée de vie plus ou moins longue selon la volonté du concepteur du site. Ils sont stockés même hors connexion, ce qui facilite les choses pour créer des profils utilisateurs, en utilisant le numéro unique attribué par le cookie pour identifier l'utilisateur. Quand Amazon installe un cookie traceur - par exemple, traçant l'activité d'un utilisateur sur les pages de amazon.com - que l'utilisateur ait ou non choisi de s'authentifier sur le site, les pages qu'il a parcourues sont connues d'Amazon. Dans ce cas, le cookie «crée» de l'information en générant un identifiant machine aléatoire. Cependant, il existe aussi des cookies

[110] http://www.allaboutcookie , vue le 30/08/2017

qui collectent de l'information sur vos appareils, comme l'identifiant unique de votre matériel - PC, tablette, téléphone - ou le système d'exploitation, voire le nom et le type de fichiers qui y sont stockés. Les cookies se répartissent en catégories selon leur niveau d'utilité ou leur fonction pour l'organisme émetteur : préférence de l'utilisateur, sécurité, analyse d'audience, état de session, publicité ou encore processus[111].

Un cookie est une information texte implantée par un site visité par l'internaute. Existent également les cookies tiers[112], qui eux sont des cookies implantés par d'autres entités que le site visité via la présence d'un tag tiers[113] dont l'utilisateur ne connaît pas toujours l'existence avant d'accepter les cookies d'un site. Souvent, un bandeau déclare que le site utilise des cookies pour le bon fonctionnement de ses services, mais beaucoup plus rarement que d'autres entreprises se servent du même site pour collecter des données à fins publicitaires. Les cookies tiers ne diffèrent pas des cookies dans leur fonctionnalité. Ils sont capables de tracer l'activité de l'internaute - et donc ses comportements sur le net - mais aussi de collecter des informations sur son matériel.

1.2.3 Modalités de collecte

1.2.3.1 Obtenir le consentement

La collecte de données des utilisateurs ne peut plus se faire sans le consentement actif de l'utilisateur[114] et sa pleine information préalables à la mise en place des cookies ou des traceurs. Par exemple, un bandeau sur un site installant des cookies avec un lien redirigeant vers une page permettant à l'utilisateur de s'informer. Le consentement doit être actif. Une case pré-cochée ne permet pas, selon la CNIL de s'assurer du consentement réel d'un individu[115]. Ce consentement passe bien entendu par la possibilité pour les internautes d'accéder à un document leur permettant de bien saisir la finalité de la collecte, le destinataire des données collectées, les transferts possibles des données hors de l'Union Européenne. Les utilisateurs doivent aussi, pour éclairer leur choix connaître leur droits sur ces données, ainsi que «le caractère obligatoire ou facultatif des réponses et des conséquences d'un défaut de réponse[116]«.

Cependant, les exceptions à cette obligation de consentement rendent la loi «d'application molle» selon le docteur en droit Thiébaut Devergranne, particulièrement la cinquième, trop large et donc sujette à interprétation :

---

[111] https://www.google.fr/intl/fr/policies/technologies/types/ , vue le 30/08/2017

[112] https://www.definitions-marketing.com/definition/cookie-tiers/ , vue le 30/08/2017

[113] "Un tag tiers ou marqueur tiers est un tag qui est présent sur un site Internet et qui provoque au niveau du navigateur du visiteur un appel de données vers un serveur ou domaine tiers distinct de celui qui est visité.

C'est à partir de ces appels générés par les tags tiers que peuvent être implémentés des cookies tiers sur la machine du visiteur.» https://www.definitions-marketing.com/definition/tag-tiers/ , vue le 30/08/2017

[114] https://www.cnil.fr/fr/respecter-les-droits-des-personnes , vue le 30/08/2017

[115] Ibid.

[116] https://www.cnil.fr/fr/cookies-traceurs-que-dit-la-loi , vue le 30/08/2017

« La réalisation de l'intérêt légitime poursuivi par le responsable du traitement ou par le destinataire, sous réserve de ne pas méconnaître l'intérêt ou les droits et libertés fondamentaux de la personne concernée[117]«.

Alors que la majorité des sites des GAFA sont faciles d'accès - rien de plus simple que de faire une recherche sur Google, ou de naviguer sur Amazon ou Facebook - les pages concernant la collecte de données personnelles sont souvent dissimulées : en page de la page, en police réduite, comme sur Amazon, ou dissimulé dans le menu, sur Facebook. Déterminer la légitimité de la collecte de données, et particulièrement de données personnelles, pose en effet sinon problème, au moins question de la légitimité à laquelle peuvent prétendre Google ou Facebook dans l'action de la collecte des mails ou des messages instantanés

La Commission Nationale Informatique et Libertés a établi de que le consentement donné ponctuellement par un individu ne peut être confondu avec un accord contractuel durable. Autrement dit, ce n'est pas parce qu'un utilisateur n'a pas supprimé les cookies du répertoire de son navigateur que l'émetteur d'un cookie conserve son consentement. Au-delà d'un seuil de treize mois à partir de la mise en place du cookie dans l'ordinateur, la tablette ou le téléphone de l'utilisateur, le cookie doit être supprimé[118]. La CNIL précise aussi que dans le cas de visite à répétition sur un site, les cookies installés par ce dernier doivent toute de même être supprimés treize mois après la première visite, c'est-à-dire après que la personne ait donné son consentement. Il doit de nouveau récolter le consentement de l'utilisateur.

1.2.3.2 Securité des données

La CNIL a également établi, une liste de bonnes pratiques et de comportements à adopter par les responsables de traitement afin de sécuriser les données qu'ils collectent[119]. Pour appréhender la sécurité sur leur site, la CNIL recommande une approche par évaluation des risques, tenant compte de la loi en vigueur :

«Le responsable du fichier doit identifier les risques sur la vie privée des personnes concernées engendrés par son traitement avant de déterminer les moyens adéquats pour les réduire. Pour ce faire, il convient d'adopter une vision globale et d'étudier les conséquences sur les personnes concernées[120]«

Les recommandations de la Commission portent surtout les traitements de données «communs» dans le cadre de l'activité des PME ; dû à la trop grande complexité des grands organismes de traitement de données, il est difficile d'établir des règles générales de sécurité[121]. Ces grands organismes de traitement, dont les GAFA font partie, doivent néanmoins assurer la sécurité des données de

[117] Thiébaut Devergranne, docteur en droit, sur son site https://www.donneespersonnelles.fr , vue le 30/08/2017

[118] https://www.cnil.fr/fr/cookies-traceurs-que-dit-la-loi , vue le 30/08/2017

[119] https://www.cnil.fr/fr/etude-dimpacts-sur-la-vie-privee-suivez-la-methode-de-la-cnil , vue le 30/08/2017

[120] Etude d'impact sur la vie privée, en anglais Privacy Impact Assessment

[121] https://www.cnil.fr/fr/garantir-la-securite-des-donnees , vue le 30/08/2017

leurs utilisateurs et répondre de tout dysfonctionnement[122] devant les organismes de régulation.

1.3 Conservation des données

1.3.1 Les enjeux du stockage des données

Le premier enjeu du stockage des données se révèle avant tout économique, puisqu'il s'agit d'un poste de dépenses particulièrement important chez les GAFA. Apple, par exemple a dépensé environ deux milliards de dollars pour la construction de deux centres de données en Europe.

Au fur et à mesure que l'utilisateur dématérialise toutes ses données, se rematérialise la concurrence entre les géants d'Internet. Le développement numérique annoncé comme l'innovation permettant de virtualiser presque tous les aspects de la vie - achats, conversation, écriture - en réalité reconcrétise les processus de collecte, d'enregistrement, de traitement, sous la forme d'imposants datacenters. Plus les activités en réseaux se développent, moins la performance des machines individuelles est requise : plus besoin de disque dur à grande capacité mémorielle pour stocker tous les documents produits ou reçus par un individu lorsqu'il lui est loisible de tout placer dans le *cloud*. Directement, la construction de telles infrastructures a de fortes retombées sur d'autres secteurs, comme la fabrication de matériel informatique (Intel, par exemple), les sociétés spécialisées dans la production d'énergie, dans la sécurité ou dans la construction.

Il est possible pour les entreprises exploitant les données de leurs utilisateurs de sous-traiter le stockage et le traitement à d'autres entreprises. Cependant, conserver et accéder aux données concernant ses clients sur ses propres infrastructures permet de rester seul le « propriétaire » - au moins le receleur – des données.

1.3.2 La localisation : un bref horizon des data centers

Les recherches effectuées ont conduit à formuler l'hypothèse suivante : les géants de l'Internet gèrent de façon similaire leurs données, notamment en ce qui concerne la sécurité de celles-ci. Ces recherches s'appuient essentiellement sur les déclarations des groupes concernés dans leurs politiques de confidentialité respectives. Dans celles-ci, la problématique de la sécurité des données des utilisateurs est mise en avant, et les mêmes assurances explicitées plus ou moins longuement, sont exposées aux utilisateurs : les données collectées sont automatiquement chiffrées, copiées et stockées dans plusieurs des *datacenters*. La copie constitue une sauvegarde simple car facilement automatisée, destinée à pallier un problème technique, comme un incendie par exemple. Des algorithmes de traitement effectuent ensuite le travail consistant à croiser ces données pour obtenir un profil utilisateur, et lui envoyer des publicités susceptible de répondre à ce que l'algorithme a déterminer être ses attentes.

Les GAFA stockent toutes les données nécessaires à leurs activités - et en conséquence les données personnelles de leurs utilisateurs - dans des *datacenters*, c'est-

[122] "Délibération de la formation restreinte SAN –2017-006 du 27 Avril 2017 prononçant une sanction pécuniaire à l'encontre des sociétés FACEBOOK INC. et FACEBOOK IRELAND» les enjoignant notamment à "prendre toutes mesures nécessaires pour garantir la sécurité des données à caractère personnel des inscrits, notamment en renforçant la robustesse des mots de passe des comptes». https://www.legifrance.gouv.fr/affichCnil.do?id=CNILTEXT000034728338 , vue le 30/08/2017

à-dire d'immenses complexes construits pour accueillir des serveurs toujours plus nombreux et surtout plus puissants. A titre d'exemple, le *datacenter* d'Apple à Prineville (Oregon) représente plus de trente mille mètres carrés de hardware destinés à faire tourner rapidement les applications - comme l'AppStore - et à accueillir les données des utilisateurs.

Les GAFA ne déclarent pas tous le même nombre de *datacenters*, et se montrent plus ou moins transparents : Google montre à ses utilisateurs une carte géographique pointant la localisation de ses centres de données[123], quand Facebook n'y donne pas directement accès par ses services. Ces recherches sur la localisation des *datacenters* des GAFA ont permis de formuler la remarque suivante : les GAFA construisent leurs *datacenters* aux mêmes endroits, majoritairement dans les états du Sud des Etats-Unis : l'Oregon comme la Californie accueillent l'entièreté des GAFA ; la Caroline du Sud les centres de Google, la Caroline du Nord, Apple et Facebook.

En Europe, ce sont les états du Nord qui sont les plus prisés par les GAFA, puisque tous ont établis leur *datacenters* en Irlande. Google s'est également installé en Finlande, en Belgique et aux Pays-Bas, Apple au Danemark, Facebook en Suède, Amazon en Angleterre, et en Allemagne.

Les quatre membres des GAFA se livrent également à une course aux *datacenters* : tous vont faire construire de nouvelles infrastructures d'ici 2018, majoritairement en Europe, dont le marché devient un véritable enjeu. Avec l'avancement du débat sur les données et l'adaptation de la législation européenne à l'environnement numérique, les GAFA ont tout intérêt à pouvoir répondre aux exigences à venir de l'Union Européenne.

1.3.3 Amazon, un cas particulier

Le leader du commerce en ligne Amazon n'a cessé d'étendre son éventail d'activités et dans celui-ci des services de cloud computing, qui permettent aux entreprises comme aux particuliers de bénéficier de louer les infrastructures d'Amazon. A ce titre, la firme dispose de très nombreux datacenters, au moins quarante-deux à travers le monde, dont plus de la moitié répartis aux USA et en Europe. L'estimation du nombre de datacenters qui est faite ici est la plus basse, basée sur les déclarations peu précises d'Amazon[124], qui dans une carte indique quarante-deux «zones de disponibilité» définies comme «[comprenant] un ou plusieurs centres de données discrets hébergés dans des installations séparées [mises] en réseau et [interconnectées]». Amazon communicant peu sur ses installations, il n'est pas possible d'établir un nombre précis ou seulement réaliste de datacenters.

De plus, l'activité de cloud computing nécessitant ces datacenters, il n'est pas possible de savoir si les informations collectées par Amazon sont stockées dans des centres dédiés ou répartis dans tous.

Au départ, l'activité servait à rentabiliser les coûteuses infrastructures du géant du e-commerce ; selon le journal en ligne *Silicon.fr*, elle représentait sur le

[123] « Localisation des datacenters » https://www.google.fr/about/datacenters/inside/locations/index.html , vue le 30/08/2017

[124] https://aws.amazon.com/fr/about-aws/global-infrastructure/ , vue le 30/08/2017

troisième trimestre 2016 les trois-quarts du bénéfice. C'est ce qui explique que la société de Jeff Bezos possède beaucoup plus de centres de données que ses trois concurrents des GAFA. Concurrents, mais aussi clients, puisque parmi les prestigieux clients du cloud d'Amazon - Spotify, la Société Générale - se trouve Apple, qui possède peu de centres de données, mais fait appel à d'autres partenaires pour le stockage des données. L'ouverture de deux *datacenters* en Europe et l'agrandissement de celui de Mesa permet de supposer néanmoins que l'entreprise a pris conscience des enjeux économique comme politique de la propriété des données et pourrait désormais souhaiter conserver ses données dans ses propres locaux.

2. RÉGULATION DES DONNÉES PERSONNELLES

2.1 Une régulation nécessaire

2.1.1 La vie privée des personnes exposée

Les personnes ne produisent plus seulement des données utiles à l'administration, comme une déclaration de naissance ou d'impôts. Le numérique a bouleversé les comportements en s'invitant dans chaque partie de la vie des personnes : de la voiture connectée qui sait faire des créneaux à l'application pour ouvrir ou clore les volets de chez soi à distance, en passant par les montres connectées, des collecteurs d'informations deviennent souvent des objets du quotidien, à tel point que les individus ne se rendent pas nécessairement compte de toutes les informations qu'ils communiquent sur eux et leur quotidien.

Les assurances proposent à leurs clients des objets connectés pour les équiper. Générali, «[propose]en Allemagne, en Espagne et en Italie un boîtier connecté recueillant des informations sur les conducteurs : distances parcourues, environnement de conduite, comportement au volant.[...] Demain, nous n'aurons plus un boîtier externe : il sera directement intégré aux véhicules[125]«.

Aux Etats-Unis, les assurances ont le droit d'adapter leurs tarifs en fonction du mode de vie des individus : autrement dit, une assurance peut imposer à son client une augmentation des tarifs, si la montre connectée de cet assuré, capable de mesurer le pouls, produit des données susceptibles de prouver une dégradation de son hygiène de vie.

Ce n'est qu'un exemple parmi tant d'autres. Le traitement des données permet à toutes sortes d'acteurs d'en apprendre toujours plus sur les individus, dans le fonctionnement de leurs relations sociales, comme dans leur vie quotidienne. Tout ce savoir sur les personnes, exploité à des fins commerciales, ou non, a longuement été questionné sur son bien-fondé et en particulier sur les nombreuses atteintes à la vie privée dont il est le résultat. Est-il légitime de suivre l'activité sur Internet ? Les données des individus peuvent-elles être collectées puis commercialisées ? Les plus grands collecteurs de données accordent-ils de l'importance à leur sécurité ou sont-elles susceptibles d'être la cible de vols

[125] https://www.generali.fr/actu/objets-connectes-assurance/ , vue le 30/08/2017

réguliers ? Ce sont des questions de ce type, pouvant mener à une certaine paranoïa, relative à la surveillance des citoyens, que les autorités se sont appropriées. Du fait même de la structure du traitement, inédit jusqu'à récemment, le rapport de la personne aux entreprises du web collectrices de données penchait à vers ces dernières. Comment punir le comportement, même à l'évidence répréhensible, d'une structure qui n'est ni décrite, ni même envisagée par la loi ?

La loi française s'est attaquée relativement tôt au problème de l'informatique et des traitements de données, en 1978[126]. Cependant, les données personnelles sont au coeur d'un système économique en construction et en tant que tel, la loi ne peut prévoir ses évolutions, mais seulement évoluer en fonction du marché. Les régulations du marché de la protection des données personnelles sont un enjeu législatif pour plusieurs raisons : en premier lieu importe le droit des citoyens à voir leur vie privée respecter ; vient ensuite la nécessité d'apporter un cadre à un secteur économique dont l'importance ne cesse de grandir ; enfin, vient la préoccupation des Etats pour la conservation de leur souveraineté concernant les informations produites par des citoyens.

2.1.2 Protéger les personnes

La volumétrie des données transitant sur Internet n'a jamais cessé d'augmenter[127], passant de cent gigaoctets par jour - l'équivalent de cinquante minutes de visionnage de vidéos en haute définition sur Youtube - à presque vingt-neuf mille gigaoctets. Selon une infographie publiée en mars 2016 par le cabinet de consulting Markentive, les données produites en une seule journée représente deux quintillons et demi d'octets, soit l'équivalent de dix millions de CDs Blu-ray[128]. La pénétration d'Internet dans la vie des individus ne décroit pas : la population mondiale ayant accès à Internet aurait connu une augmentation de 14,3% entre 2011 et 2013 pour atteindre trois milliards de personnes, soit l'équivalent de la population mondiale en 1960[129].

La régulation de ce volume de données personnelles et la protection des données comme des personnes physiques qui en sont à l'origine représente un défi, dans la mesure où le législateur se retrouve à régenter un territoire nouveau, conquis et exploité par quelques précurseurs, souvent surnommés des «pionniers» - parmi lesquels Larry Page et Sergei Brin[130] ou Jeff Bezos[131].

M. Anciaud et Mme. Farchy dans leur article « Données personnelles et droit de propriété : quatre chantiers et un enterrement » file la métaphore de la conquête de l'Ouest en évoquant le numérique comme «un sous-continent devant être investi, construit et domestiqué par les acteurs, qui peuvent acquérir un droit de propriété

[126] Loi dite "Informatique et Libertés», 1978.

[127] « Combien de temps sur internet avec mon forfait (50 Mo,1 Go,20 Go…) » http://astuto.fr/combien-de-temps-sur-internet-avec-mon-forfait-50-mo-1-go-5-go-20-go-50-go/ , vue le 30/08/2017

[128] https://www.markentive.fr/blog/infographie-chiffres-cles-big-data/ , vue le 30/08/2017

[129] Ibid.

[130] Page et Brin sont les fondateurs du moteur de recherche Google.

[131] Fondateur et actuel PDG du site de commerce en ligne Amazon.

dès lors qu'ils prouvent leur utilisation et la mise en valeur d'une parcelle de territoire» et les données personnelles comme un «or noir du XXIeme siècle».

L'image est éloquente, et au fur et à mesure que les pionniers connaissaient de grandes réussites ou l'oubli, créaient de nouveaux marchés, l'outil législatif s'est attelé à la définition et la régulation de ces activités pionnières. A titre d'exemple, en 1978, la loi 78-17 «Informatique et Libertés» a posé les premiers pas de la protection des personnes et leur vie privée vis-à-vis des traitements de données ; et le *Data Information Act* de 1998, définit le rôle des acteurs centraux comme suit :

> *»data controller» means, subject to subsection (4), a person who (either alone or jointly or in common with other persons) determines the purposes for which and the manner in which any personal data are, or are to be, processed;*
>
> *«data processor», in relation to personal data, means any person (other than an employee of the data controller) who processes the data on behalf of the data controller;*
> *«data subject» means an individual who is the subject of personal data;*[132]*«*

2.1.3 Encadrer le marché de la donnée : une question de souveraineté

La souveraineté se définit comme «Pouvoir suprême reconnu à l'État, qui implique l'exclusivité de sa compétence sur le territoire national (souveraineté interne) et son indépendance absolue dans l'ordre international où il n'est limité que par ses propres engagements (souveraineté externe).[133]«

Les Etats conservent jalousement les informations concernant leur population, en témoignent certains cas de non-coopération internationale[134] ou le scandale des écoutes américaines en 2015. Le président Hollande avait qualifié «d'agressions» les écoutes de la NSA[135]. Si l'écoute de trois citoyens - fussent-ils au coeur du pouvoir - par une puissance étrangère est une agression, qu'en est-il de la collecte des données personnelles, suivi de la commercialisation, de 86% des ménages français par des entreprises étrangères[136] répondant au droit états-unien ?

[132] " le contrôleur de données» est, sous réserve du paragraphe (4), une personne qui (seule, conjointement ou de conserve avec plusieurs autres personnes), détermine les objectifs pour lesquels, et la manière selon laquelle, chaque donnée personnelle est, ou sera, traitée;

" le responsable de traitement de données», en relation avec les données personnelles, est toute personne (autre qu'un employé du contrôleur de données) qui procède à un traitement de données pour le compte du contrôleur de données,

" le sujet de la donnée» est l'individu qui est le sujet de la donnée personnelle;»

[133] URL : http://www.larousse.fr/dictionnaires/francais/souverainet%C3%A9/74000?q=Souverainet%C3%A9#73171 , vue le 30/08/2017

[134]« L'insuffisante coopération des Etats à la justice pénale internationale », Philippe Weckel ,URL : http://www.sentinelle-droit-international.fr/?q=content/linsuffisante-coop%C3%A9ration-des-etats-%C3%A0-la-justice-p%C3%A9nale-internationale , vue le 30/08/2017

[135]« Francois Hollande qualifie les écoutes Amériques d'agression, mais parle d'affaire classée » , http://www.lemonde.fr/pixels/article/2015/06/26/a-washington-le-gouvernement-condamne-les-ecoutes-americaines-mais-dedramatise_4662213_4408996.html , vue le 30/08/2017

[136] https://www.insee.fr/fr/statistiques/2385835#tableau-Donnes , vue le 30/08/2017

L'exclusivité dans la gestion des données personnelles, dont les Etats étaient les principaux collecteurs avant la démocratisation de l'Internet, est bien sûre largement remise en cause, voire tout simplement remisée. L'Etat français connaît d'un citoyen son état civil (date, lieu de naissance, statut marital, date et lieux de décès), son apparence, son adresse, ses revenus déclarés, sa profession et son lieu d'exercice, son lieu de résidence, s'il a été condamné par la justice ; ainsi que les mêmes informations sur les membres français de sa famille, et cela constitue déjà de solides renseignements sur cet individu, d'autant qu'il se voit contraint légalement de donner ces renseignements.

Facebook sait tout cela, puisque l'individu peut renseigner ces données dans son «profil», mais Facebook sait aussi qui sont les amis de cet individu. Il peut même lui proposer de contacter des personnes que le citoyen ne se rappelle plus avoir connues, en fonction de son lieu de résidence ou du lycée qu'il a fréquenté. Facebook a aussi accès à toute l'individualité de cette personne, même dans ce qu'il y a de plus intime : ce qu'elle écrit à ses contacts, les livres et séries qu'elle aime, son orientation sexuelle, sa religion, ses citations préférées, ses photos de vacances. L'individu est même traçable par Facebook physiquement, et jusque dans ses émotions, grâce à la possibilité qui a été ajoutée aux fonctions «like» et «commentaire» : le bouton «réaction». La collecte de données est donc bien plus massive et systématique sur d'importantes plateformes que celle des services de l'Etat.

La souveraineté externe n'est pas beaucoup plus grande dans le rapport GAFA/Etat français. La France - comme la majorité des autres pays accédant à Internet - n'est pas indépendante sur le territoire international dans son accès et sa consommation Internet. Bien sûr, même devenu mondial, Internet est d'abord né aux Etats-Unis, et ce sont logiquement les entreprises américaines qui ont eu le privilège de porter Internet et d'en diversifier les services en premier lieu. Toutes les innovations, même majeures de l'internet ne sont pas l'œuvre de chercheurs américains ; néanmoins, ce sont les Etats-Unis qui ont su attirer les innovateurs[137]. Cela a permis à certaines de s'installer, puis de péricliter, et à d'autres, comme Google, Amazon, Facebook et Apple, de s'installer plus durablement. Cependant, la logique de la bonne initiative au bon moment, si elle porte toutes les innovations, ne suffit pas à expliquer la dominance américaine.

Dans les années 1990, alors qu'Internet s'étendait, les fournisseurs d'accès européens ont au départ décidé de connecter directement l'Europe aux points d'accès des USA[138] ; la majorité des câbles sous-marins par lesquels transite l'information passe aujourd'hui encore par les USA. C'est d'ailleurs ce qui a permis à la NSA de mener à bien les écoutes qui lui sont reprochées actuellement[139]. Ce sont également les Etats-Unis qui sont chargés de contrôler

[137] Tim Berners-Lee, l'inventeur du World Wide Web, est britannique et a créé le web à Genève au début des années 1990. En 1994, il occupe une chaire au MIT.

[138] Ceruzzi Paul E, « Aux origines américaines de l'Internet : projets militaires, intérêts commerciaux, désirs de communauté », Le Temps des médias, 2012/1 (n° 18), p. 15-28. DOI : 10.3917/tdm.018.0015. URL : http://www.cairn.info/revue-le-temps-des-medias-2012-1-page-15.htm , vue le 30/08/2017

[139]« Les câbles sous-marins, ces autoroutes du Web prisées par les espions » , Jules Darmanin , http://www.lefigaro.fr/secteur/high-tech/2015/07/01/32001-20150701ARTFIG00339-les-cables-sous-marins-ces-autoroutes-du-web-prisees-par-les-espions.php , vue le 30/08/2017

l'attribution des noms de domaines, jusqu'en 1986, où l'INRIA prend le relais en France, suivi par l'Association Française pour le Nommage Internet en Coopération (Afnic)[140]. De plus, les médias européens se sont beaucoup plus intéressés dans les années 1990 aux dangers d'Internet qu'aux opportunités créées[141]. En bref, les Etats-Unis, non seulement sont le territoire de naissance et de développement de l'Internet et de ses principaux protocoles, mais dans le monde numérique sont aussi dominants dans la répartition du réseau, dans la gouvernance de l'Internet, et dans les services proposés à l'internaute.

La souveraineté interne et externe pose donc problème dans le cadre du numérique. La souveraineté numérique est définie par Bellanger comme «la maîtrise de notre présent et de notre destin tels qu'ils se manifestent et s'orientent par l'usage des technologies et des réseaux informatiques.». Dans son article «De la Souveraineté Numérique» en 2012, il dresse un portrait peu rassurant, voire parfois alarmiste, de la situation. En particulier, il s'indigne du transfert à très grande échelle des données personnelles des européens :

> «Les carnets d'adresses, les listes d'amis, les messages intimes, les photos, les secrets, les ombres et le reste sont stockés sur des serveurs situés à dix mille kilomètres de nous et répondant de la compétence du tribunal de Sacramento.
> Nous avons accepté des contrats que nous n'avons jamais lus, aux clauses obscures et changeantes par ailleurs. Nous avons cédé les droits et la propriété de souvenirs, d'images, de notre passé, de nos pensées à des sociétés de services informatiques sur un autre continent régi par un autre droit, une autre langue et sur lequel la moindre procédure judiciaire est d'un coût dissuasif. Nous avons fait preuve là d'une ingénuité aussi paradoxale que la phobie du réseau de la fin des années 1990. Comment a-t-on pu laisser faire cela sans s'en rendre compte un seul instant ?[142] «

Il considère le retard de l'Europe sur les Etats-Unis comme une cession de pouvoir, comme un «un transfert de souveraineté, de maîtrise de notre destin numérique, massif et silencieux[143]«.

Il est vrai que les données personnelles des individus aux mains d'entreprises étrangères, au-delà d'enjeux économiques, posent le problèmes du respect de la vie privée - et surtout de la sanction du non-respect de la vie privée - comme celui de la mémoire - celles des échanges comme celles des documents plus rédigés. Comment protéger les données des citoyens quand celles-ci ne trouvent plus dans la même juridiction que les personnes auxquelles elles se rattachent ? La solution proposée par Bellanger consiste en ce que la communauté européenne impose ses règles de collecte et de

[140] https://www.afnic.fr/ , vue le 30/08/2017

[141] Bellanger Pierre, « De la souveraineté numérique », Le Débat, 2012/3 (n° 170), p. 149-159. DOI : 10.3917/deba.170.0149. URL : http://www.cairn.info/revue-le-debat-2012-3-page-149.htm , vue le 30/08/2017

[142] Ibid.

[143] Ibid.

transactions de données[144]. En particulier, il souhaiterait que les données des européens ne puissent être stockées que sur le territoire européen.

2.2 Des régulations multiples

2.2.1 Les rapports de force dans la régulation du marché des données personnelles

L'échange des données personnelles nécessite trois grands acteurs : les utilisateurs, les entreprises exploitantes sur le net, et les états (Bien sûr, d'autres acteurs moins visibles sont essentiels, comme les FAI ou les constructeurs de ligne Internet). Françoise Benhamou définit trois formes de régulation du marché des données[145], selon les acteurs qui la pratiquent et les rapports de force qu'ils entretiennent : coercitive, coopérative, et marchande.

La régulation coercitive est une réglementation, une norme, un mode de fonctionnement. Elle ne se négocie pas, mais s'impose aux acteurs, qui doivent la respecter peu importe leur opinion à son égard, tant qu'elle est en vigueur. C'est une coercition législative[146], qui permet à l'Etat de poser des conditions préalables et obligatoires à l'exercice de toute activité, afin d'encadrer le marché des données personnelles. Ces conditions déterminent les pratiques abusives et crée des sanctions pour punir ces pratiques. Une fois la loi édictée au bout d'un processus très long, il est très difficile de revenir en arrière. En France, l'organisme coercitif - c'est-à-dire chargé de faire respecter la loi - dans le cadre du marché des données personnelles est la CNIL.

Les acteurs de la régulation coopérative sont surtout les utilisateurs et les entreprises exploitant les données personnelles. La régulation coopérative est envisagée comme «[relevant] des règles formelles ou informelles adoptées par les communautés d'internautes, ou des négociations entre le régulateur, les sites et les fournisseurs d'accès[147]«. Elle relèverait donc de l'accord, c'est-à-dire du domaine du contrat. C'est une régulation courante dans le cadre des usages de l'Internet, retrouvée notamment dans l'aspect formel des conditions d'utilisation qui conditionnent l'accès à un service. Cette régulation peut apparaître aussi sou un aspect plus informel, comme une question sur l'âge de l'utilisateur à son entrée sur un site, ou un bandeau lui signifiant l'utilisation de cookies. Dans tous les cas, la régulation coopérative - ou régulation contractuelle - prend ses racines à la fois dans un rapport d'intérêts convergents - le service a un intérêt commercial dans l'exploitation de données commerciales, et les utilisateurs ont intérêt à accéder à des services gratuits - et de confiance ponctuelle réciproque entre les acteurs - l'utilisateur fait confiance au service en ligne pour une utilisation légale et loyale de ses données personnelles, et le service en ligne donne accès à son contenu car il a confiance en la capacité de l'utilisateur à respecter les conditions d'utilisation propres à ce service.

[144] Ibid.

[145] Benhamou Françoise, « L'État et l'internet. Un cousinage à géométrie variable », Esprit, 2011/7 (Juillet), p. 96-110. DOI : 10.3917/espri.1107.0096. URL : http://www.cairn.info/revue-esprit-2011-7-page-96.htm , vue le 30/08/2017

[146] Ibid, paragraphe 27.

[147] Ibid, paragraphe 28.

La régulation marchande repose sur l'action des acteurs du marché, hors du cadre de la loi. Elle correspond à la régulation du marché de la donnée par la seule confrontation des acteurs sur un marché. Autrement dit, ce sont des intérêts économiques qui conduisent les entreprises exploitantes à adopter un comportement plus prudent vis à vis des données personnelles[148].

2.2.2 Typologies de régulation

En 2011, Alain Rallet et Fabrice Rochelandet s'interrogent sur les possibilités de réguler la collecte et l'exploitation des données personnelles, devenues monnaie d'échange sur les services du web[149]. Pour établir une typologie, ils précisent des niveaux de régulation un premier critère de temporalité avec les régulations ex ante et ex post : la première, la régulation ex ante, est une forme de régulation préventive, qui encourage les acteurs à adopter avant la collecte et le traitement des données un comportement approprié à la protection de celles-ci ; la seconde régulation intervient après la collecte, c'est une régulatrice réparatrice, qui intervient après un dommage. Un second critère est l'orientation que prennent ces régulations : soit vers les personnes, soit vers les exploitants (Amazon, ou Apple, par exemple). MM. Rallet et Rochelandet croisent ces deux critères dans un tableau pour en tirer quatre typologies.

2.2.2.1 Les régulations anticipatrices

La régulation ex ante orientée vers les exploitants «consiste à émettre des règles collectives plus ou moins contraignantes, que leur émetteur soit une institution publique ou privée.[150]«. Il s'agit souvent d'une régulation publique.

Le but est d'anticiper des dangers potentiels et de mettre en place des barrières de protection destinées à se mettre en conformité avec les normes de protection. La privacy by design est sans doute la pratique la plus significative.

Cela consiste en l'obligation pour les exploitants de données personnelles à anticiper dès la création d'un nouveau service ou d'une nouvelle application à tous les dangers potentiels en ce qui concerne la sécurité des individus et le respect de leur vie privée, et de dresser dans la structure de leur service des barrières adaptées en prévision des dangers potentiels détectés, ainsi qu'une procédure de réaction à une éventuelle faille dans le système (désigner le service chargé de récupérer des données, par exemple, ou de sécuriser un serveur).

La régulation ex ante orientée vers les individus est «[l'incitation des] individus à adopter des comportements prudentiels vis-à-vis des activités exploitant des données personnelles susceptibles de leur causer des préjudices[151]«.

---

[148] Rallet Alain, Rochelandet Fabrice, « La régulation des données personnelles face au web relationnel : une voie sans issue ? », Réseaux, 2011/3 (n° 167), p. 17-47. DOI : 10.3917/res.167.0017. URL : http://www.cairn.info/revue-reseaux-2011-3-page-17.htm , vue le 30/08/2017

[149] Rallet Alain, Rochelandet Fabrice, « La régulation des données personnelles face au web relationnel : une voie sans issue ? », Réseaux, 2011/3 (n° 167), p. 17-47. DOI : 10.3917/res.167.0017. URL : http://www.cairn.info/revue-reseaux-2011-3-page-17.htm , vue le 30/08/2017

[150] Ibid, paragraphe 24.

[151] Ibid, paragraphe 29.

Elle est le plus souvent prise en charge par les autorités publiques. Cela passe concrètement, d'abord par la création juridique de mesure de protection des utilisateurs, puis par des campagnes de prévention et de sensibilisation, la mise à disposition d'informations par les autorités aux consommateurs relatives à leurs droits[152], et aussi par des subventions, qui permettent d'accéder à des technologies plus tournées vers la protection des données personnelles.

> « Tout le problème est de réduire l'asymétrie informationnelle et la méconnaissance des conséquences des comportements de divulgation, sans faire porter une charge cognitive trop élevée sur les individus, d'où la nécessité d'un processus d'apprentissage adaptatif fondé sur l'utilisateur, ni paralyser leur action[153] »

La *privacy by using*, dépend en effet de la compréhension qu'ont les individus utilisateurs des modalités et des conséquences de la collecte parfois sauvage qui est faite de leurs données personnelles. Les utilisateurs désormais pleinement conscients peuvent modifier leur comportement en mesure de ce qu'ils acceptent ou non. Des pratiques individuelles, qui une fois stabilisées, pourront permettre d'aller vers un consensus. C'est ce en quoi consiste la *privacy by using* :

> «De ces comportements éclairés naîtront de nouvelles normes de privacy. Ces normes fonctionneront alors comme de nouvelles conventions, codifiables par le droit, destinées à économiser le savoir des individus[154]«.

2.2.2.2 Les régulations réparatrices

La régulation ex post orientée vers les individus permet à ces derniers de faire valoir leur droit au respect de leur vie privée[155]. De nombreux cas de litiges ont opposé des individus à des services en ligne qui exploitent leurs données personnelles. La revendication des droits des individus à l'effacement de leurs données ou à leur déréférencement constitue une forme de régulation ex post par les individus. L'autorégulation par les réseaux sociaux en constituerait une autre : une dénonciation d'un service abusif ou frauduleux sur un réseau social pouvant vite devenir virale, et causer de grands dommages à la réputation de ce service, et aussi à sa santé financière, si cela des utilisateurs à abandonner le service. MM. Rallet et Rochelandet utilisent le terme d' «équilibre de la terreur». Les exploitants peuvent refuser de fournir un service dont l'utilisateur a besoin s'il refuse de communiquer ses données, ce qui crée un déséquilibre dans la relation individu/exploitant. En se massant contre un service abusif, les utilisateurs rééquilibrent la relation individu/exploitant en menaçant de cesser d'être clients. Ils peuvent alors influer sur la politique de l'entreprise.

---

[152] Des informations relatives aux droits des individus quant à leurs fichiers et informations personnelles sont mises dispositions par l'Etat français et accessibles à l'URL suivante : https://www.service-public.fr/particuliers/vosdroits/F2024 , vue le 30/08/2017

[153] Rallet Alain, Rochelandet Fabrice, Zolynski Célia, « De la Privacy by Design à la Privacy by Using. Regards croisés droit/économie », Réseaux, 2015/1 (n° 189), p. 15-46. DOI : 10.3917/res.189.0015. URL : http://www.cairn.info/revue-reseaux-2015-1-page-15.htm , vue le 30/08/2017

[154] Ibid, paragraphe 37.

[155] Rallet Alain, Rochelandet Fabrice, « La régulation des données personnelles face au web relationnel : une voie sans issue ? », Réseaux, 2011/3 (n° 167), p. 17-47. DOI : 10.3917/res.167.0017. URL : http://www.cairn.info/revue-reseaux-2011-3-page-17.htm , vue le 30/08/2017

Le rôle de la régulation ex post orientée vers les exploitants «est de faire payer les exploitants jugés fautifs et non pas de contraindre tous les exploitants dont l'activité présente des risques en matière de vie privée[156]«.

C'est un acte de sanction, qui punit des pratiques jugées abusives. L'article différencie deux applications de cette régulation : d'une part, celle que peuvent exercer les consommateurs, en sanctionnant par leur non-consommation d'un service l'entreprise fautive[157]. Cette stratégie inclut dans son action les médias, chargés d'alerter les consommateurs, ainsi que les entreprises elles-mêmes, qui souhaiteront redorer leur image après un scandale ou se différencier de la concurrence en investissant dans la sécurité des données. La seconde forme de cette régulation réside en la sanction par des autorités judiciaires des entreprises ne s'étant pas conformées aux normes en vigueur - si ces entreprises sont sises sur le territoire national et entrent dans le champ d'application de la loi - ou si elles ont failli au respect des termes du contrat qu'elles ont passé avec les utilisateurs.

2.3 Une régulation faillible

Comme chaque choix est contestable, chacune des typologies de régulation précédemment décrites a ses failles. La législation accuse toujours un retard sur l'innovation, et le législateur n'est pas spécialistes du numérique : cela peut mener à une « inadéquation [qui] peut de ce fait aboutir à l'institution de règles tantôt trop laxistes et insuffisamment dissuasives, [...] tantôt trop restrictives, empêchant certaines utilisations innovantes des données personnelles, pourtant bénéfiques aux individus ».

Les régulations sont bien sûr grandement dépendantes de la législation en vigueur, et cela peut se transformer en faille. Les exploitants sont des multinationales qui évitent de s'établir dans les pays les plus contraignants, exerçant donc leur activité au-delà et échappant au cadre juridique des territoires dans lesquels évoluent pourtant les individus auprès desquels sont collectées les données personnelles.

La réglementation ex-ante orientée sur les exploitants répartit les coûts entre les exploitants, chargés de se conformer aux normes - ce qui peut impliquer un long processus d'audit - et les institutions publiques, qui s'occupent de vérifier la bonne conformité des pratiques des premiers. Elle requiert des coûts importants[158] des deux parties, ce qui freine l'action sur l'objectif à atteindre : la conformité aux normes réglementaires en vigueur sur le traitement approprié des données personnelles et le respect de la vie privée des individus. Les contraintes budgétaires limitent l'action de la CNIL et celles des entreprises[159].

[156] Ibid, paragraphe 27.

[157] Ibid, paragraphe 28.

[158] En 2017, le budget voté pour le financement de la CNIL dépasse 17 millions d'euros. URL : https://www.data.gouv.fr/fr/datasets/budget-de-la-cnil-1/ , vue le 30/08/2017

[159] En 2010, 82 % des entreprises et administrations ne respecteraient pas les obligations que leur impose la loi Informatique et Libertés en matière de droit d'accès des individus à leurs données personnelles. URL : https://www.cairn.info/revue-reseaux-2011-3-page-17.htm#re7no7 , vue le 30/08/2017

C'est parce que les individus font valoir leurs droits qu'il est possible pour les autorités judiciaires de punir les comportements non conformes des exploitants. Autrement dit, même si cela n'est pas systématique, la régulation ex-post orientée vers les individus conduit à la régulation ex-post orientée vers les exploitants. Cependant, la connaissance des personnes sur leurs droits est loin d'être complète[160]. Cette méconnaissance des utilisateurs freine évidemment la valorisation de leurs droits auprès des autorités compétentes : d'une part, il est difficile d'avoir conscience d'une violation de ses droits lorsque ces derniers ne sont pas bien appréhendés, et d'autre part les procédures de plainte ne sont pas toujours connues et nécessitent un apprentissage de leur part, qui leur demandera du temps, de la volonté et un effort cognitif. De fait, « arbitrant entre ces coûts et l'utilité immédiate des services qui leur sont proposés, les individus ont une forte propension à ne pas contrôler le respect de leurs droits » : la régulation ex-ante orientée vers les individus ne peut s'exercer dans ce cadre.

La régulation ex-ante orientée vers les individus peut être biaisée par le comportement des entreprises exploitantes, qui modifient sans en prévenir les utilisateurs leurs conditions d'utilisation[161] et créent une « illusion de contrôle » à leur attention, en leur proposant de paramétrer les services en fonction de ce qu'ils souhaitent partager à un public large ou restreint. Ces paramètres n'influent pourtant ni sur le nombre de données collectées ni sur l'utilisation qui en sera faite par la suite. La plus grande faille de la régulation ex-ante orientée vers les individus ne vient pourtant pas des entreprises exploitantes, pas même du comportement des individus qui renseignent des données sur eux-mêmes. Les individus ne peuvent maîtriser dans cette approche que ce qu'ils publient et non ce qui est publié sur eux[162]. Aucun contrôle n'est possible de la part de l'individu, aussi averti soit-il et aussi prudentes ses pratiques numériques soient-elles, sur des personnes tierces, qui peuvent à loisir révéler des informations personnelles. Cela a donné lieu à des pratiques malveillantes, dont l'objectif est la destruction de la réputation d'une personne, comme le *revenge porn*[163] (désormais sanctionné par la loi).

Les régulations ex-post orientées vers les exploitants, lorsqu'elles sont judiciaires, sont impactées par les faibles possibilités de recours des individus. Lors d'une régulation ex-post par le marché, c'est-à-dire par la concurrence entre les entreprises et par « l'équilibre de la terreur » des consommateurs, c'est sa réputation - et donc ses utilisateurs - qu'il souhaite préserver. Cela implique le bon fonctionnement des médias qui traqueraient et relaieraient les manquements des

[160] 67% des citoyens connaissent la CNIL, et sur ce chiffre, seulement 53% déclarent voir précisément de quoi il s'agit, selon le rapport annuel de la CNIL de 2016. URL : https://www.cnil.fr/sites/default/files/atoms/files/cnil-37e_rapport_annuel_2016.pdf , vue le 30/08/2017

[161] "Apple se réserve le droit, à sa seule discrétion et à tout moment, de changer, modifier, compléter ou supprimer des parties de ces Conditions d'utilisation. Il est de votre responsabilité de consulter périodiquement ces Conditions d'utilisation pour voir si des modifications y ont été apportées». Informations légales du site web d'Apple. URL : https://www.apple.com/fr/legal/terms/site.html , vue le 30/08/2017

[162] Facebook explicite dans ses Conditions qu'un utilisateur ne peut supprimer que ce qu'il a publié, ce qui exclut les données personnelles d'un individu A publiées par un individu B.

[163] "Le revenge porn (ou "vengeances pornographiques») est le fait de diffuser sur internet, les réseaux sociaux, ou d'envoyer par des moyens de télécommunication des photos intimes et/ou à caractère sexuel obtenues dans le cadre de relations intimes. La diffusion de ces photos par un partenaire est le plus souvent liée à une volonté de chantage ou de nuire à la suite d'une rupture.» URL : https://fondationdesfemmes.org/le-delit-de-revenge-porn-adopte-dans-le-code-penal/ , vue le 30/08/2017

exploitants, ainsi qu'une certaine réactivité de la part des consommateurs, dont on a vu plus tôt qu'ils étaient loin d'avoir une connaissance exhaustive de leurs droits. S'ajoutent à ces bases fragiles que les asymétries informationnelles entre les individus et les exploitants sont largement en faveur des exploitants, qui possèdent la compétence technique et juridique. Les entreprises exploitant des données personnelles, qui dans cette configuration de régulation marchande n'ont pas à crainte de contrôle de la part des autorités publiques, et bénéficient de plus d'une « probabilité très faible d'être sanctionnées par les consommateurs » sont alors incitées par le contexte qui est le leur à jouer sur deux tableaux en faisant de la défense des données personnelles un argument marketing pour ensuite « tricher en revendant les [données personnelles] ou en sous-investissant dans la sécurisation de leurs bases de données pour augmenter leurs revenus ». Rien n'incite donc, dans cette configuration, les entreprises à investir sérieusement dans la sécurité des données Françoise Benhamou notait même en 2011 « la quasi-absence d'un marché de la protection de la vie privée, alors qu'existe un énorme marché des données personnelles[164] ». Il s'agit alors d'un marché régulé par la seule concurrence entre des entreprises privées, où la course à l'innovation ne se fait pas au profit, mais bien « au détriment des droits individuels qui sont ici protégés par les seules règles du droit commun ». C'est une régulation marchande de ce type qui est pratiquée aux Etats-Unis, le pays duquel sont originaires les quatre GAFA et la majorité des services les plus connus sur le net.

3. LE NOUVEAU. RÈGLEMENT EUROPÉEN

Une logique de défiance vis à vis du traitement des données personnelles et des atteintes à la vie privée s'est installée[165]. Dans le but de protéger les personnes physiques, les traitements de données personnelles se voit encadrés de façon de plus en plus stricte et à une échelle de plus en plus globale. La transmission massive de données entre des organismes sis sur des territoires nationaux différents est partiellement à l'origine de la volonté de l'Union Européenne d'unifier les réglementations et les régulations sur le territoire européen.

D'un point de vue plus politique, c'est aussi l'occasion pour l'Union Européenne d'imposer ses conditions aux géants du net et de restaurer une partie de sa souveraineté numérique.

La volonté de protéger les données personnelles ex-ante comme ex-post, qui visaient à contenir à un périmètre raisonnable les actions des exploitants, notamment en terme de profilage, n'a que peu séduit ces derniers, qui ont entrepris une intense activité de lobbying : quatre mille amendement ont été déposés contre le projet, dont certains rédigés clé en main par les exploitants[166].

[164] Benhamou Françoise, « L'État et l'internet. Un cousinage à géométrie variable », Esprit, 2011/7 (Juillet), p. 96-110. DOI : 10.3917/espri.1107.0096. URL : http://www.cairn.info/revue-esprit-2011-7-page-96.htm , vue le 30/08/2017

[165] Lancelot Miltgen Caroline, « Dévoilement de données personnelles et contreparties attendues en e-commerce : une approche typologique et interculturelle », Systèmes d'information & management, 2010/4 (Volume 15), p. 45-91. DOI : 10.3917/sim.104.0045. URL : http://www.cairn.info/revue-systemes-d-information-et-management-2010-4-page-45.htm , vue le 30/08/2017

[166] http://www.cil.cnrs.fr/CIL/spip.php?article1976 , vue le 30/08/2017

3.1 Harmonisation sur le territoire européen

Permettre à tous les individus résidant sur le territoire européen de bénéficier de droits et d'une protection unifiés face au traitement et à l'exploitation économique de leur données personnelles est l'un des objectifs du règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données. Ce règlement entrera en vigueur en mai 2018 dans tous les pays de l'Union Européenne. A partir de mai 2018,

> « la protection conférée par le présent règlement devrait s'appliquer aux personnes physiques, indépendamment de leur nationalité ou de leur lieu de résidence, en ce qui concerne le traitement de leurs données à caractère personnel[167] ».

Cette réglementation protège les individus indépendamment de leur nationalité, en privilégiant comme champ d'application le lieu de résidence des responsables de traitement - c'est-à-dire les organismes décidant des modalités de traitement des données personnelles - de leurs possibles sous-traitants ou des personnes ciblées par ces traitements.

« En pratique, le droit européen s'appliquera donc chaque fois qu'un résident européen sera directement visé par un traitement de données, y compris par Internet [168]». Aucun organisme ne pourra plus profiter au sein de l'Union Européenne d'une différence de législation nationale entre les états membres.

3.2 Les droits des personnes physiques

3.2.1 Renforcement des droits des utilisateurs ex-post

Le règlement statue dans son tout premier article que « [la] protection des personnes physiques à l'égard du traitement des données à caractère personnel est un droit fondamental [169]». Cela annonce clairement l'intention première du règlement : exercer le pouvoir coercitif[170] de la législation européenne afin de renforcer les droits des personnes et leurs possibilités de régulation, notamment ex-post.

---

[167] Lancelot Miltgen Caroline, « Dévoilement de données personnelles et contreparties attendues en e-commerce : une approche typologique et interculturelle », Systèmes d'information & management, 2010/4 (Volume 15), p. 45-91. DOI : 10.3917/sim.104.0045. URL : http://www.cairn.info/revue-systemes-d-information-et-management-2010-4-page-45.htm , vue le 30/08/2017

[168] La CNIL. URL : https://www.cnil.fr/fr/reglement-europeen-sur-la-protection-des-donnees-ce-qui-change-pour-les-professionnels , vue le 30/08/2017

[169] Règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, paragraphe 63 URL : http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN , vue le 30/08/2017

[170] Règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, paragraphe 63 URL : http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN , vue le 30/08/2017

3.2.1.1 Le droit d'accès

L'accès d'une personne à ses données personnelles est un droit décrit dans le paragraphe 63 du règlement :

> « Une personne concernée devrait avoir le droit d'accéder aux données à caractère personnel qui ont été collectées à son sujet et d'exercer ce droit facilement et à des intervalles raisonnables, afin de prendre connaissance du traitement et d'en vérifier la licéité. [...] En conséquence, toute personne concernée devrait avoir le droit de connaître et de se faire communiquer, en particulier, les finalités du traitement des données à caractère personnel, si possible la durée du traitement de ces données à caractère personnel, l'identité des destinataires de ces données à caractère personnel, la logique qui sous-tend leur éventuel traitement automatisé et les conséquences que ce traitement pourrait avoir, au moins en cas de profilage. Lorsque c'est possible, le responsable du traitement devrait pouvoir donner l'accès à distance à un système sécurisé permettant à la personne concernée d'accéder directement aux données à caractère personnel la concernant.[171] »

Le responsable de traitement est tenu d'informer la personne auprès de laquelle il collecte des données à caractère personnelles du contexte relatif au traitement de ses données et donc de tous les droits possédés par la personne sur ses données, y compris son droit d'accès. Le responsable de traitement ne peut ni facturer la mise à disposition de ses données à la personne concernée, ni s'opposer à sa demande. Il est au contraire tenu de faciliter l'application de ce droit, particulièrement dans la mesure où le droit d'accès 'une personne sur ses données lui permet d'exercer un autre de ses droits : le droit de rectification, qui lui permet de faire modifier par le responsable de traitement des informations fausses ou incomplètes[172].

Néanmoins, une utilisation répétitive peut autoriser à faire porter par le demandeur les frais d'envois de copies supplémentaires. Ces frais doivent rester raisonnables et si la demande est faite en ligne et sans autre exigence de la part du demandeur, le responsable de traitement doit effectuer l'envoi « sous une forme électronique d'usage courant[173] ». Il n'a pas le droit d'utiliser des formats volontairement rares ou difficiles de lecture pour limiter le droit d'accès du demandeur.

3.2.1.2 Le droit à l'oubli

Le règlement réaffirme dans l'article 17, le droit à l'oubli (ou le droit à l'effacement) des personnes : « La personne concernée a le droit d'obtenir du responsable du traitement l'effacement, dans les meilleurs délais, de données à caractère personnel la concernant et le responsable du traitement a l'obligation d'effacer ces données à caractère personnel dans les meilleurs délais » dans la mesure où le traitement de ces données ne s'avère pas absolument nécessaire.

[171] Ibid, article 16.

[172] Ibid, article 16

[173] Ibid, aricle 15.

Le traitement de données est considéré comme nécessaire dans cinq cas. D'abord, s'il participe « à l'exercice du droit d'expression et d'information » : un homme politique, par exemple, s'il est mis en cause par un journal en ligne, ne peut pas invoquer le droit à l'oubli pour faire retirer des articles gênants. Les traitements de données pour « respecter une obligation légale [...] ou pour exécuter une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement[174] » ou servant « à la constatation, à l'exercice ou à la défense de droits en justice » sont considérés comme nécessaire et sont donc non-opposables : un particulier ne pourra pas s'opposer à ce que les données personnelles de sa déclaration de revenu soit traitées pour déterminer le montant de son impôt, ou à la prise de note d'un greffier pendant un procès.

Le droit à l'oubli ne peut pas non plus s'appliquer si les données sont traitées pour « des motifs d'intérêt public dans le domaine de la santé publique » - on ne peut pas s'opposer à un diagnostic, par exemple - ou à « des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique, ou à des fins statistiques ». Néanmoins, ces traitements nécessaires doivent garantir le respect des droits de la personne. Cela passe par des « mesures techniques et organisationnelles [qui] peuvent comprendre la pseudonymisation[..]. 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[175] ».

3.2.1.3 Le droit à la limitation

L'utilisateur, conformément à l'article 18, s'il constate un traitement illicite ou une inexactitude dans les données, si les données lui sont « encore nécessaires [...] pour la constatation, l'exercice ou la défense de droits en justice », même si les responsables de traitement n'en ont plus l'utilité, et lors d'un litige « pendant la vérification portant sur le point de savoir si les motifs légitimes poursuivis par le responsable du traitement prévalent sur ceux de la personne concernée », peut demander la limitation du traitement de ses données personnelles[176]. Cela signifie que le responsable de traitement est limité à la seule conservation des données, et se voit dans l'obligation d'obtenir le consentement explicite de la personne concernée pour tout autre traitement, ainsi que d'informer l'utilisateur en cas de levée de la limitation. Comme tous les droits, il connaît une exception : « pour la constatation, l'exercice ou la défense de droits en justice, ou pour la protection des droits d'une autre personne physique ou morale, ou encore pour des motifs importants d'intérêt public de l'Union ou d'un État membre », les données d'un traitement limité peuvent être de nouveau exploitées.

[174] Ibid, article 89.

[175] Ibid, article 18.

[176] Ibid, article 18.

3.2.1.4 Le droit d'opposition

Certains traitements de données peuvent avoir des conséquences négatives sur la vie des individus, comme les spams. Afin de lutter contre ces pratiques, l'article 21 statue qu'une personne peut faire valoir son droit d'opposition « à tout moment, pour des raisons tenant à sa situation particulière, à un traitement des données à caractère personnel la concernant » :

- si ces données sont traitées dans un objectif de prospection ;
- si ce traitement est fondé sur la nécessité de « l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable du traitement

  [ou si le traitement] est nécessaire aux fins des intérêts légitimes poursuivis par le responsable du traitement ou par un tiers, à moins que ne prévalent les intérêts ou les libertés et droits fondamentaux de la personne concernée qui exigent une protection des données à caractère personnel [177]».
- si les traitements de données servent des finalités de recherche ou statistiques

Dans le premier cas, lorsqu'une personne utilise dans ce cadre son droit d'opposition, le responsable de traitement doit cesser le traitement des données à caractère personnel[178]. Dans les autres cas, la demande de l'utilisateur est étudiée, puis accédée ou contestée par le responsable de traitement.

3.2.2 Le droit à la portabilité des données

En France, les changements portent surtout sur la portabilité des données et les données personnelles des mineurs. Par exemple, un responsable de traitement de données ne pourra plus refuser de communiquer à une personne les données personnelles qu'elle a fourni, à condition que « la personne concernée a consenti au traitement de ses données à caractère personnel pour une ou plusieurs finalités spécifiques », ou que « le traitement est nécessaire à l'exécution d'un contrat auquel la personne concernée est partie ». Si ce traitement concerne des données sensibles, l'utilisateur ait donné son accord explicite.

Pour illustrer ce droit renforcé à la portabilité, mettons-le en situation. Une personne décidé de changer de messagerie. En faisant valoir son droit à la portabilité, elle peut exiger du service qu'elle quitte de se voir restituer les données qu'elle a fourni, comme son carnet d'adresses ou le contenu de ses mails. Lorsque la faisabilité en est raisonnable pour le responsable de traitement, le service que l'utilisateur quitte transfère directement les données au nouveau service de messagerie de l'utilisateur : « Lorsque la personne concernée exerce son droit à la portabilité des données [...] elle a le droit d'obtenir que les données à caractère personnel soient transmises directement d'un responsable du traitement à un autre [179]».

[177] Ibid, article 6.

[178] Ibid, article 21.

[179] Ibid, article 20.

Les responsables de traitement, à partir de 2018, ne pourront plus refuser à leurs utilisateurs de leur restituer leurs données. Actuellement, ils n'y sont pas systématiquement tenus, et rendent parfois volontairement difficile de quitter leurs services : quitter la messagerie Google, c'est renoncer à une grande partie des services de la firme et perdre l'historique de ses mails.

3.3 La responsabilisation des entreprises

Ce règlement européen visait en partie les géants du net Google, Apple, Facebook et Amazon (GAFA)[180]. De grands procès avaient déjà conduit l'action des Etats contre ces entreprises pionnières[181][182], qui ont développé leur commerce autour du traitement des données de leurs utilisateurs. Au niveau national, des actions ont déjà été menées contre ces géants, mais ce règlement, en imposant des droits renforcés et harmonisés sur tout le territoire, permet à l'Union Européenne d'imposer des régulations ex-ante comme ex-post orientée vers les exploitants destinées à encadrer l'action des exploitants de données personnelles sur le territoire européen et de sanctionner toute contravention. L'unification permet également de limiter les défenses juridiques des GAFA : elles ne peuvent désormais plus s'abriter derrière la législation de leur pays d'origine, les Etats-Unis.

3.3.1 La protection des données

3.3.1.1 Conformité

Les exploitants doivent se montrer transparents sur leurs activités relatives au traitement des données à caractère personnel. Le règlement consacre la pleine responsabilité de l'exploitant sur tout traitement des données et sur les processus et techniques qui permettent ces traitements. Il doit également « être en mesure de démontrer que le traitement est effectué conformément au présent règlement[183] ». Afin de pouvoir garantir la conformité du traitement qui est fait des données à caractère personnel, la rédaction de code de conduite est encouragée par le règlement, et une « boîte à outils » consistant en des documents et procédures simplifiés est mise à la disposition de l'exploitant[184] par les autorités, qui permettront de déterminer la bonne conformité ou non conformité d'un traitement de données.

Les responsables de traitement sont aussi tenu de mener avant le traitement des données une analyse d'impact, dont les modalités sont décrites dans le règlement, afin de déterminer les risques relatifs au traitement des données

[180] «L'Union européenne resserre l'étau sur les GAFA au nom de la vie privée », Clément Bohic http://www.itespresso.fr/union-europeenne-resserre-etau-gafa-vie-privee-146026.html , vue le 30/08/2017

[181] «La France inflige une amende record à Google », http://www.lefigaro.fr/secteur/high-tech/2014/01/09/32001-20140109ARTFIG00310-la-france-inflige-une-amende-record-a-google.php , vue le 30/08/2017

[182]« Facebook condomnée par la Cnil » , http://www.lemonde.fr/pixels/article/2017/05/16/donnees-personnelles-facebook-condamne-par-la-cnil-a-150-000-euros-d-amende_5128370_4408996.html , vue le 30/08/2017

[183] Règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, article 24. URL : http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN , vue le 30/08/2017

[184] http://www.cil.cnrs.fr/CIL/spip.php?article2862 , vue le 30/08/2017

personnelles[185], de prendre conscience de potentielles défaillances et de décider les mesures à mettre en oeuvre pour y pallier, selon la gravité du risque encouru. De plus, l'unification de la législation sur tout le territoire européen simplifie grandement les efforts de conformité des exploitants.

3.3.1.2 Le délégué à la protection des données

Le présent règlement rend obligatoire l'existence dans chaque organisme exploitant des données personnelles, d'un délégué à la protection des données, qui est chargé de l'application et de la conformité du règlement au sein de ces organismes[186]. Il dispense pour cela des conseils, au responsable de traitement mais aussi, s'ils en font la demande, à d'autres acteurs, particulièrement « en ce qui concerne l'analyse d'impact relative à la protection des données et [vérifie] l'exécution de celle-ci ».

Ce délégué est unique pour toute l'Union Européenne et en lien avec l'autorité de régulation des données personnelles dans le pays où il est principalement implanté. Il permet à l'entreprise de s'assurer de la conformité de ses décisions avec le règlement européen, et aux autorités européennes concernées[187] (la CNIL en France, par exemple) une plus grande rapidité de décision.

Les devoirs et missions d'un délégué sont consignés dans le règlement. Il doit être associé « d'une manière appropriée et en temps utile, à toutes les questions relatives à la protection des données à caractère personnel [188]» par le responsable du traitement (et le sous-traitant s'il existe) qui doit l'aider à exercer ses fonctions en lui « fournissant les ressources nécessaires [...] ainsi que l'accès aux données à caractère personnel et aux opérations de traitement, et lui permettant d'entretenir ses connaissances spécialisées ». Le règlement précise que le délégué ne peut faire l'objet de pression ou de pénalisation à cause de son activité.

3.3.2 Le devoir d'informer

Ce sont les exploitants - c'est-à-dire les responsables de traitements - qui portent l'information jusqu'à l'utilisateur pour lui permettre d'exercer une régulation ex-ante avec une bonne connaissance de la situation et de ses droits. La loi leur impose de traitement de façon loyale et transparente les données des personnes[189]. Par exemple, le droit d'opposition d'une personne doit être porté à

[185] Règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, article 35. URL : http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN , vue le 30/08/2017

[186] Ibid, article 39.

[187] Ibid.

[188] Ibid, article 38

[189] Règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, paragraphe 60 URL : http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN , vue le 30/08/2017

connaissance de celle-ci « [au] plus tard au moment de la première communication avec la personne concernée » et le droit d'opposition à la prospection et au profilage lié doit être « présenté clairement et séparément de toute autre information[190] ». La loi force les exploitants à répondre à chaque demande d'un individu concernant l'application de ses droits, dans un délai maximum d'un mois s'il ne souhaite pas y répondre positivement[191].

Lors de la collecte des données à caractère personnel, les responsables de traitement doivent fournir à la personne concernée identité et coordonnée, des renseignements sur les finalités du traitement de ces données (et réinformer la personne si ces finalités changent) et la base juridique sur laquelle le traitement repose, les intérêts légitimes de l'exploitant à effectuer ce traitement[192], si ce traitement est licite et quelles conditions le rendent licite[193]. L'exploitant doit également informer la personne concernée si ses données à caractère personnelles sont susceptibles d'être transférées et si ce transfert répond aux exigences de l'Union Européenne. S'il y a lieu, il faut également fournir à l'utilisateur les coordonnées du délégué à la protection des données[194].

3.3.3 La sécurité par défaut

Le règlement inaugure aussi la *privacy by design* comme obligation des exploitants dans l'article 25 :

> « [...]le responsable du traitement met en œuvre, tant au moment de la détermination des moyens du traitement qu'au moment du traitement lui-même, des mesures techniques et organisationnelles appropriées, telles que la pseudonymisation, qui sont destinées à mettre en œuvre les principes relatifs à la protection des données, par exemple la minimisation des données, de façon effective et à assortir le traitement des garanties nécessaires afin de répondre aux exigences du présent règlement et de protéger les droits de la personne concernée. [...] Le responsable du traitement met en œuvre les mesures techniques et organisationnelles appropriées pour garantir que, par défaut, seules les données à caractère personnel qui sont nécessaires au regard de chaque finalité spécifique du traitement sont traitées. En particulier, ces mesures garantissent que, par défaut, les données à caractère personnel ne sont pas rendues accessibles à un nombre indéterminé de personnes physiques sans l'intervention de la personne physique concernée. [195]»

Cela signifie que, quelque soit l'exploitant, et quelles que soient la finalité du traitement de données qu'il a entrepris, les moyens et processus de traitement doivent être déterminés dans une démarche de respect des droits de la personne sur

[190] Ibid, article 1.

[191] Ibid, paragraphe 59.

[192] Règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, article 13 URL : http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN , vue le 30/08/2017

[193] Ibid, article 6.

[194] Ibid, article 6.

[195] Ibid, article 25.

ces données et de mise en conformité avec le règlement. Le respect des droits de la personne et la protection des données à caractère personnel n'est pas dans des mesures qui s'ajoutent au processus de traitement, mais au cœur du traitement lui-même.

3.3.4 Les sanctions

Leurs obligations s'accroissent et avec elles les sanctions prévues en cas de violation de la protection des données sont augmentées : alors que l'amende maximale que peut infliger la CNIL est actuellement de cent-cinquante mille euros, « le non-respect d'une injonction émise par l'autorité de contrôle [...], fait l'objet, [...] d'amendes administratives pouvant s'élever jusqu'à 20 000 000 EUR ou, dans le cas d'une entreprise, jusqu'à 4 % du chiffre d'affaires annuel mondial total de l'exercice précédent, le montant le plus élevé étant retenu [196]».

Facebook a atteint en 2016 un chiffre d'affaires de vingt-sept milliards six-cent-trente-huit millions de dollars ; une amende de quatre pourcents représenterait pour la firme une perte d'un milliard cent millions de dollars, soit un dixième du bénéfice net[197].

Aux sanctions pécuniaires s'ajoutent pour les entreprises qui faillissent à protéger les données à caractère personnel l'obligation pour le responsable de traitement de communiquer ces défaillances à l'autorité de régulation immédiatement (au-delà de trois jour, le responsable est tenu de justifier son retard[198]), mais aussi à toutes les personnes concernées[199]. Cette obligation, qui force les exploitants à dénoncer publiquement leurs propres failles, porte atteinte à leur réputation et les met face au jugement des consommateurs. Cela permet de renforcer la possibilité d'une régulation ex-post orientée sur les individus.

3.4 Cas particulier : les traitements de données nécessaires à l'action de l'Etat

De nombreux droits des personnes vis à vis du traitement de leurs données à caractère personnel ne s'appliquent pas dans le cadre d'une action étatique. Si un particulier peut légitimement demander à Facebook de faire disparaître une publication mentionnant, par exemple, ses revenus, qu'il en soit ou pas l'émetteur premier, la même demande ne peut pas être faite à l'administration fiscale, qui a besoin de traiter ces données dans le cadre de sa mission. Le fait qu'une administration ou une collectivité puisse obliger une personne à fournir des

[196] Règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, article 83 URL : http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN , vue le 30/08/2017

[197] « Facebook : + 54 % de chiffre d'affaires en 2016 » , http://www.lesnumeriques.com/vie-du-net/facebook-54-pourcent-chiffre-affaires-en-2016-n60139.html , vue le 30/08/2017

[198] Règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, article 33 URL : http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN , vue le 30/08/2017

[199] Ibid, article 34.

données à caractère personnel ne signifie pas que ces données soient visibles de toutes les personnes relatives ou intégrées à cette administration.

Certaines données personnelles, voire sensibles collectées par des services étatiques ne correspondent pas toujours à une mission pleinement administratives. Les attentats que connait l'Europe - et que connaître la France - ont conduit à de nouvelles mesures législatives, notamment la loi française sur le renseignement, parue au JORF le 24 juillet 2015 et modifiée le 4 octobre de la même année[200].

La loi sur le renseignement permet, si des individus sont soupçonnés de terrorisme, aux services de renseignements français de recourir à des méthodes d'accès à l'information auparavant réservées à l'exercice de l'appareil judiciaire[201], ce qui signifie que les méthodes utilisées pour retrouver le coupable d'une infraction avérée peuvent maintenant être utilisées pour surveiller des individus suspectés d'avoir l'intention de commettre une infraction. Cela comprend des techniques comme la captation d'images dans des lieux privés ou de données informatiques ou encore l'accès aux réseaux des FAI. Cette dernière disposition vise à tracer les individus, et à avoir accès aux métadonnées de leur correspondance électronique. La liste n'est ici pas exhaustive.

C'est une loi controversée, tenant notamment à ses origines : en 2013, un rapport parlementaire établit que les services de renseignement agissent dans un cadre « extra-légal [...] extraordinairement flou[202]« et que «le retard accusé par la France dans ce domaine paraît indéfendable et nuisible ». Le même rapport préconise la création d'une inspection des services de renseignement, ce qui en dit long sur l'absence de contrôle des services de renseignent qui a précédé.

Les détracteurs de cette loi dénoncent la surveillance massive des citoyens qu'elle permet, mais aussi le fait que la loi n'instaure pas un contrôle suffisant sur ses services de renseignements[203]. Les services de renseignement doivent obtenir l'aval du premier ministre avant de mettre en application une des techniques de surveillance nouvelles légitimées par la loi. Le premier ministre donne ou refuse l'autorisation après avis de la Commission nationale de contrôle des techniques de renseignement, dont le rôle est exclusivement consultatif, puisque « Dans les cas d'urgence absolue, l'autorisation de mettre en œuvre une technique de renseignement pourra être délivrée sans avis préalable de la commission. Elle devra néanmoins en être immédiatement informée, et pourra recommander son interruption[204] ».

La loi sur le renseignement et les atteintes à la vie privée qu'elle permet ne doivent exister qu'à titre exceptionnel. Un article sur un site public énumère les finalités justifiant d'un recours à de tels moyens :

> « la sécurité nationale, les intérêts essentiels de la politique étrangère et l'exécution des engagements internationaux de la France, les intérêts économiques et scientifiques essentiels de la France, la prévention du

[200] https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000030931899&dateTexte=20170826 , vue le 30/08/2017

[201] http://www.vie-publique.fr/actualite/panorama/texte-discussion/projet-loi-relatif-au-renseignement.html, vue le 30/08/2017

[202] http://www.assemblee-nationale.fr/14/pdf/rap-info/i1022.pdf , vue le 30/08/2017

[203] https://sous-surveillance.fr/#/ , vue le 30/08/2017

[204] http://www.vie-publique.fr/actualite/panorama/texte-discussion/projet-loi-relatif-au-renseignement.html, vue le 30/08/2017

terrorisme, [...], la prévention de la criminalité et de la délinquance organisées, la prévention de la prolifération des armes de destruction massive [...], la prévention des violences collectives de nature à porter gravement atteinte à la paix publique [205]».

Le règlement européen, s'il est destiné à remplacer la loi « Informatique et Libertés », ne modifie en rien l'application de la loi sur le renseignement, puisque les activités relatives à la sécurité nationale d'un pays n'entrent pas dans le champ d'application du droit de l'Union Européenne.[206]

RÉAPPROPRIATION ET PATRIMONIALISATION DES DONNÉES PERSONNELLES

1. LE RÔLE ACTIF DE L'UTILISATEUR-SUJET

Il est vrai que jusqu'à présent, l'individu a été présenté comme peu actif, voire comme victime de la part des collecteurs de données personnelles. Or, si la loi le désigne comme « sujet » de la donnée, il n'en est pas moins un acteur dans le processus de collecte.

1.1 L'individu et la collecte de ses données personnelles

1.1.1 La fin de la privacy

Le premier moteur de la régulation des pratiques de grand acteur privé est la protection de la vie privée des individus, mise en péril par la tendance des entreprises du web à aspirer toutes les informations afin de profiler l'utilisateur et de vendre ses besoins à des annonceurs : si l'utilisateur achète une table de ping pong, il est très probable qu'il voit apparaître jusqu'à plusieurs jours après son achat, au cours de sa navigation, des publicités pour des filets ou des raquettes. Plus sujette à caution qu'une table de ping pong, une recherche sur une maladie peut permettre de supposer l'état de santé d'un utilisateur. Les objets connectés changent le monde l'assurance en permettant de tracer l'individu dans son quotidien et d'évaluer son potentiel de risques. Tous ces exemples, ainsi que les échecs répétés des autorités régulatrices à sanctionner efficacement les contrevenants, tendent à démontrer que la vie privée, également désignée par le terme anglophone « *privacy* », tend à disparaître.

Antonio Casilli concentre sa réflexion sur Facebook pour montrer que ce ne sont pas les usages qui ont changé en premier, mais les paramètres des plateformes.

[205] Ibid.

[206] Règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, paragraphe 16 URL : http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN , vue le 30/08/2017

Les dénonciations des atteintes caractérisées à la vie privée de la part de la plateforme ont même été parfois assez violentes pour faire reculer le plus grand des réseaux sociaux. Les plateformes ont donc été actrices du glissement des informations personnelles du domaine exclusivement privé à un espace public.

1.1.2 Une gestion rationnelle du capital social

Il n'est pas question de nier que les individus exposent volontairement certains aspects de leur vie privée sur Internet, mais de remettre à leur place les comportements des utilisateurs. Le dévoilement de soi sur le web est «interprété comme une forme d'« individualisme expressif [...], visant à produire et entretenir des identités numériques », et les différences observées de comportements des utilisateurs entre eux, en fonction de leur âge, de leur sexe ou leur statut socioéconomique, ne permettent pas d'affirmer l'abandon total de la *privacy*.

Au contraire, Casilli défend l'idée que loin d'avoir abandonné leur *privacy*, les utilisateurs alimentent un capital social, qui « désigne [...] l'acquisition, via des relations médiatisées [par] les TIC, de ressources matérielles, informationnelles ou émotionnelles » et s'approprient la logique marchande de gestion des données personnelles des grandes plateformes en l'appliquant à leur réputation. Si la gestion de l'image se fait dans une logique entrepreneuriale dans laquelle l'objectif à atteindre et l'amélioration de sa propre réputation, alors l'utilisateur de réseaux sociaux adopte une démarche rationnelle en dévoilant certains détails de sa vie privée, dans l'intention de faire venir à lui de nouveaux contacts. Ces nouveaux « amis », en partageant ses contenus, deviendront alors des vecteurs plus ou moins conscients de l'expansion de sa réputation.

En choisissant une approche stratégique, l'utilisateur reprend une partie de son pouvoir sur ses données. Il ne partagerait plus ce qu'il souhaite partager instinctivement, mais parce qu'il souhaite établir des connexions avec ses contacts - actuels et à venir - et remplacerait l'envie de partager un contenu par un « processus dynamique d'évaluation de la situation, d'adaptation au contexte, de catégorisation du contenu que les individus sont prêts à partager avec leurs connaissances ». Ce processus vise à retenir toute information susceptible d'avoir sur son émetteur un impact social négatif. Si l'utilisateur pense que la communauté des personnes susceptibles de lire ses contenus n'en sera pas favorablement influencée dans l'opinion qu'elle se fait de lui, alors il renonce à publier ce contenu.

Le profil d'un individu est une construction sociale, résultat d'une inter-influence entre ce que l'individu estime bénéfique de paraître et ce que la communauté de ses contacts attend et est prête à accepter de lui. Il oriente son action en réfléchissant sur le contenu qu'il rend visible et à quelle audience il la destine : « la sélection détermine à quelle personne un contenu donné est révélé, tandis que l'influence détermine quel contenu est révélé à une personne donnée ».

Cette forme d'auto-censure est d'autant plus pratiquée si le lien avec le groupe de destination du contenu est étroit car le risque de sanction sociale est plus élevé : entrer en conflit avec un membre de sa famille proche porte plus à conséquence sur la vie de l'individu que blesser la sensibilité d'une personne tellement peu rencontrée qu'elle lui est presque inconnue. L'individu nivelle la

portée des données qu'il partage, si ce n'est dans une réflexion de coût/bénéfices, au moins dans l'intention de minimiser les risques d'impact négatif sur son capital social.

1.1.3 La privacy by negociation

La vie privée adopte un nouveau modèle sur le web : la *privacy by negociation,* qui consiste en une « nouvelle vision de la privacy comme négociation incessante, dans un cadre de complexité sociale et technologique ».

C'est une transformation importante de la notion de vie privée, qui n'est plus le centre fermé de l'intimité d'un individu, mais un espace de négociation, d'interrelation. La volonté de conserver certaines informations privées secrètes, au plus près de soi, n'a pas abandonné les individus. Ils se sont simplement adaptés à un environnement qui pour fonctionner leur demande de plus en plus de données personnelles, en modulant ce qu'ils partagent et à qui ils choisissent de le partager. La logique est la suivante : mettre en lumière un certain type de contenu pour mieux dissimuler ce qu'il est souhaitable de conserver par-devers soi. Cela est issu d'une « négociation collective, conflictuelle et itérative, visant à adapter les règles et les termes d'un service aux besoins de ses utilisateurs ».

Les utilisateurs protègent leur vie privée en recourant aux régulations ex ante et ex post orientée vers les individus. Les utilisateurs adoptent une démarche de régulation ex ante, en paramétrant la confidentialité du service (*privacy by using*), et en négociant ce qu'ils souhaitent dire d'eux-mêmes avec ce qu'attend la communauté destinataire et la manière dont elle réagit ; mais aussi en «équilibrant la terreur» c'est-à-dire en faisant part aux plateformes abusives de leur mécontentement, soit une régulation ex post.

1.2 La possibilité de se passer des GAFA

La lutte contre les GAFA et les abus de collecte et de traitement sur les données personnelles ont fini par porter quelques fruits. Ceux-ci se trouvent d'abord dans les comportements des usagers, dont il a été montré qu'ils n'étaient pas majoritairement insouciants mais au contraire le résultat d'un calcul complexe, fonction de la personnalité de l'individu, du souci de sa réputation, de ses considérations sur les données privées, de ce qu'il veut faire percevoir à autrui et de l'adaptation à la réaction d'autrui.

Un autre de ces fruits – et non le moindre – réside dans le développement de marchés alternatifs de la donnée personnelle, relativement proches du marché plus connu, plus ostentatoire aussi, qu'est celui instauré par les GAFA, en qu'ils sont gratuits, eux aussi.

1.2.1 Refuser la collecte

1.2.1.1 Des moteurs de recherche non-intrusifs

Les moteurs de recherche revendiquent désormais de ne pas tracer leurs utilisateurs, et l'utilisent comme un argument de différenciation. Le méta-moteur américain Duckduckgo, par exemple, ne collecte par défaut aucune donnée personnelle sur ses utilisateurs. Les données issues du paramétrage du moteur de recherche ne sont pas collectées mais enregistrées et stockées sur l'ordinateur de l'utilisateur[207]. Les seules données collectées ne sont pas personnelles et servent à améliorer la recherche. Pour se financer, il n'a cependant pas dérogé à une autre norme instaurée par les GAFA : il recourt à la publicité.

StartPage, un méta-moteur de recherche néerlandais, suit le même modèle de confidentialité que le précédent et offre de plus dans ses règles de confidentialité une solide formation de l'utilisateur aux risques engendrés par une collecte massive de données. StartPage possède un fonctionnement particulier : c'est une interface qui anonymise complètement l'utilisateur avant de lancer sa requête via Google. Cela permet à l'utilisateur de lancer sa recherche un moteur particulièrement puissant, sans voir la liste de résultats pertinents modifiée parce les résultats « pertinents pour lui ». La firme propose, tout comme Google, un service de messagerie accessible directement de la page d'accueil.

Un moteur de recherche français dans la même veine, Qwant, se positionne aussi sur le respect de la vie privée. Il prend le contrepied de Google et des deux précédents moteurs de recherche en proposant une interface chargée, mais fonctionnelle. Il a surtout la particularité de recourir à son propre moteur de recherche, qui se veut neutre et refuse de censurer un résultat. Existe une version visant les plus jeunes et destinée à les préserver de certains résultats indésirables[208].

D'autres moteurs de recherche proposent aux internautes de consacrer l'argent créé par leurs clics à une cause qui leur tient à cœur : Ecosia plante des arbres[209], Lilo soutient financièrement des projets sociaux[210], et il ne s'agit là que d'exemples parmi d'autres.

Ces moteurs de recherche, s'ils sont toujours minoritaires, se distinguent par une forte augmentation de leur fréquentation. De 2013 à 2016, Duckduckgo est passé d'une place de 1908ème site web le plus fréquenté à 643ème[211] ; Qwant sur la même période enregistrait un passage de 3,5 millions de visites mensuelles à 27 millions et atteint depuis juillet 2017 les 40 millions d'utilisateurs[212].

---

[207] https://duckduckgo.com/privacy#s4 , vue le 30/08/2017

[208] «Qwant Junior, un moteur de recherche sécurisé pour les enfants » , https://www.solidatech.fr/utiliser/ressources/qwant-junior-un-moteur-de-recherche-securise-pour-les-enfants, vue le 30/08/2017

[209] https://info.ecosia.org/what , vue le 30/08/2017

[210] https://www.lilo.org/fr/#, vue le 30/08/2017

[211] https://fr.wikipedia.org/wiki/DuckDuckGo#Adoption , vue le 30/08/2017

[212] https://fr.wikipedia.org/wiki/Qwant , vue le 30/08/2017

1.1.2.2 D'autres initiatives

Il n'y a pas que les moteurs de recherche qui promettent un plus grand respect de la vie. Le réseau social Diaspora est un logiciel libre qui ne diffuse pas de publicité, ni ne revend les données de ses utilisateurs. Les données des utilisateurs sont chiffrées, et il est possible pour eux de ne pas stocker leurs données sur le serveur de Diaspora, mais sur un serveur privé[213].

D'autres utilisateurs estiment leur navigation polluée et ralentie par un trop plein de publicités, et craignent également pour la sécurité de leur navigation lorsqu'ils sont confrontés à un grand nombre de publicités sous toutes leurs formes : bandeau, encart, pop-up. Ils utilisent alors un bloqueur de pub. Certains, comme DontTrackme sont respectueux de la vie privée de leurs utilisateurs. En revanche, le plus populaire, l'allemand Adblock, édité par la société Eyeo[214], tombe sous le coup des politiques de confidentialité des navigateurs auxquels il est agrégé. De plus, il est aussi rémunéré par des annonceurs pour laisser passer des publicités « acceptables ».

La régulation marchande, si elle n'en est qu'à ses débuts, commence donc à émerger dans les usages, les utilisateurs informés se tournant de plus en plus vers ces moteurs de recherche alternatifs.

1.3 Pour une redistribution de la valeur produite : le *digital labor*

1.3.1 Définition

Le terme de *digital labor* désigne « les activités numériques quotidiennes des usagers de plateformes sociales, d'objet connectés ou d'applications mobiles ». Un clic, un like, un commentaire, sont autant d'usages communs sur le web qu'on en viendrait presque à oublier que ces activités simples, et tellement répétitives qu'elles en deviennent un réflexe, sont « assimilables au travail, parce que productrices de valeur, faisant l'objet d'un quelconque encadrement contractuel et soumises à des métriques de performance »[215].

1.3.2 Un travail invisible et émietté

Il ne se vit pas forcément comme un « travail », puisqu'il lie à la fois la détente et une activité productrice de valeur pour les entreprises en ligne : c'est un travail « éminemment cognitif qui se manifeste à travers une activité informelle, capturée et appropriée dans un contexte marchand en s'appuyant sur des tâches médiatisées par des dispositifs numériques »[216].

[213] https://diaspora-fr.org/, vue le 30/08/2017

[214] «Adblock : en 2016, l'utilisation des bloqueurs de pub a augmenté de 30 % », Nelly Lesage http://www.numerama.com/tech/230518-adblock-en-2016-lutilisation-des-bloqueurs-de-pub-a-augmente-de-30.html, vue le 30/08/2017

[215] Cardon Dominique et Casilli Antonio A. *Qu'est-ce que le Digital Labor ?*. Louise Merzeau. 1ère édition. Paris : Ina, 2015. 101 p. Etudes et controverses. ISBN 978-2-86938-2299.

[216] Ibid.

Évaluer une publication ou recopier le texte présent sur une image correspond à la fois à une pratique tout à fait banale et ordinaire de n'importe quel utilisateur du web ; c'est aussi du *digital labor*, une tâche nécessitant une intelligence humaine, capable de créer du sens, quand les machines ne peut agir que de façon logique. Casilli prend l'exemple de la plateforme de micro-travail MTurk, qui fait exécuter par ses travailleurs, contre une légère rémunération, des tâches simples pour les individus et trop complexes pour les algorithmes, comme la reconnaissance d'images.

1.3.3 Redistribution de la chaîne de production

Le numérique a changé la distribution des rôles de production. Dans le système de production des entreprises de web, ce sont les données produites par les consommateurs qui sont valorisées. En conséquence, les consommateurs sont aussi les travailleurs inconscients du service qu'ils utilisent. La redistribution des rôles dans la chaîne de productivité bouleverse d'autant plus le travail traditionnel que le *digital labor* n'a pas de limites temporelles ou quantitatives : tant que l'utilisateur est connecté, il produit de la donnée. Toutes ces données produites par les individus de manière similaire, incluent tous les utilisateurs au sein d'un même « processus de travail ».

S'il convient de qualifier le *digital labor* de travail, il alors faut examiner ce qu'il en est de l'exploitation et de l'aliénation des travailleurs[217]. Alors que le niveau d'exploitation, c'est-à-dire de la quantité de plus-value apportée par les travailleurs, est important, l'aliénation est plutôt faible. Plus exactement, l'aliénation au sens traditionnel, celle qui dépossède l'individu de ses capacités sociales ou intellectuelles, est faible. Pour cause, les travailleurs n'ont pas l'impression de travailler, puisqu'ils sont dans des conditions qu'ils considèrent extérieures au travail : difficile en effet, d'admettre d'un adolescent arrimé à son téléphone portable qu'il produit un « travail». Et pourtant, son activité est génératrice de plus-value pour les plateformes en ligne qu'il fréquente. En revanche, une nouvelle forme d'aliénation existe : les travailleurs perdent en partie la « maîtrise sur leurs données personnelles et leurs propres contenus » au profit de leur exploitant.

1.3.4 La répartition des richesses produites

Se pose alors la question de la répartition de la production, extrêmement inégalitaire par nature, puisque le travail des *digital laborers*, intégré à leur pratique du web, est le plus souvent gratuit. Casilli propose de pallier à cette inégalité ; non pas par la marchandisation directe de l'activité numérique, qui conduirait à une « privatisation de la privacy » ; mais par un système de redistribution par la fiscalité. Une redistribution de ce genre se heurte à la réalité des conflits de juridiction, et à la pratique d'exil fiscal commune aux GAFA.

Dans la mesure où ce n'est pas la donnée individuelle qui a de la valeur pour les plateformes en ligne, mais le réseau montrant toutes les relations, les

[217] Dominique Cardon et Antonio A. Casilli, Qu'est-ce que le Digital labor ?, Paris, Ina Éditions, 2015, 104 p. ISBN 978-2-86938-2299

résonances, les redondances et les parallèles entre toutes ces données, la redistribution par l'impôt aurait toutefois le mérite de redistribuer à la communauté, une partie des bénéfices créés par cette communauté, en tant que communauté[218].

1.4 La mort numérique

Si les données peuvent théoriquement perdurer indéfiniment, à grand renfort d'émulation et de migration, le web voit régulièrement partir certains de ses utilisateurs. Après une vie numérique d'exploitation de ses propres données, l'internaute disparaît en même temps que la personne physique dont il était une partie. Ses données sont une fois extraites indépendantes de l'internaute, indépendante de l'individu duquel elles sont originaires. Elles peuvent lui survivre et dans ce cas, il convient de laisser aux personnes et à leurs héritiers un pouvoir de décision et un moyen d'action sur leur devenir.

1.4.1 Prévoir la mort numérique

Il est établi que les individus produisent des données tout au long de leur navigation sur le web. Ces données, qu'elles aient été produites par les plateformes ou par l'individu, ne s'évanouissent d'elles-mêmes. Sur Facebook comme sur Twitter il est possible de retrouver des publications vieilles de plusieurs années. Inévitablement, certaines de ces publications sont le fruit d'une personne aujourd'hui décédée.

La volonté de disposer de ces données après la mort peut s'expliquer par une tendance à la prévoyance, qui pousse les individus à préparer leur décès de la façon la plus complète possible[219]. Ces derniers sont alors cibles du marché de la mort numérique. D'autres, sont un peu plus méfiants à l'idée de confier leur volonté testamentaire à un service en ligne, et préféreront choisir de consigner sur un document papier leurs codes, afin de faciliter les choses pour leurs proches. Un autre facteur est l'attachement de la personne à ses données. Pour un utilisateur régulier des réseaux sociaux, par exemple, l'ensemble de ses publications s'approche d'une œuvre personnelle. L'utilisateur y accorde une importance sentimentale, au même titre que ses biens matériels et il souhaite en disposer de la même façon.

1.4.2 Les différents services de la mort numérique

Se pose la question, pour les internautes, de ce que vont devenir leurs données, les traces qu'ils ont laissées d'eux, après leur disparition. Les GAFA, qui représentaient en 2010 50% du temps de navigation des internautes, sont donc

[218] Dominique Cardon et Antonio A. Casilli, Qu'est-ce que le Digital labor ?, Paris, Ina Éditions, 2015, 104 p. ISBN 978-2-86938-2299

[219] Guillemot Samuel, Gourmelen Andréa, « Quand les entreprises s'emparent de la mort numérique, qui sont les consommateurs potentiels ? », *Revue française de gestion*, 2017/1 (N° 262), p. 123-145. DOI : 10.3166/rfg.2017.00114. URL : http://www.cairn.info/revue-francaise-de-gestion-2017-1-page-123.htm

particulièrement touchées par la problématique de la mort de leurs utilisateurs, ne serait-ce que parce que conserver un compte inactif à jamais ne leur rapporte rien.

Des outils permettant aux utilisateurs de prendre des dispositions testamentaires vis à vis de leurs comptes, ont vu le jour chez Google et Facebook, mais les services relatifs à la mort numérique sont plus divers. Certains services offres des coffres-forts numériques, destinés à accueillir un document contenant des dispositions testamentaires relatives plus générales, et éventuellement des documents numériques jugés précieux, comme Edeneo[220].

Certains services ont choisi un tour plus personnel, comme Safe Beyond[221], qui permet de préparer un ou plusieurs mails à envoyer aux proches après le décès. D'autres services se sont emparés des possibilités du web en matière de multimédia et proposent à leur client de ne pas écrire leurs volontés, mais de tourner une vidéo destinée à leurs légataires, voire comme Stone Story de créer et rédiger des contenus textes, vidéos, photos, retraçant la vie de l'individu, accessibles sur le site de l'entreprise, ou grâce à un QR code placé sur la pierre tombale, le tout rendu disponible le jour de la mort du client.

1.4.3 Le cadre juridique

Ce sont des dispositions récentes qui encadrent en France la mort numérique. C'est une loi du 7 octobre 2016 qui offre enfin un cadre légal à la mort numérique des individus, et ouvre le droit aux personnes qui souhaitent préparer leur décès de choisir une personne de confiance, à laquelle sera confiée l'application des « directives générales, lorsqu'elles portent - sur l'ensemble des données concernant une personne ; ou particulières, lorsque ces directives ne concernent que certains traitements de données spécifiques »[222] auprès des responsables de traitement désignée par la personne défunte.

La loi légitime les dispositions des individus prises directement auprès des responsables de traitement de données par le défunt, de son vivant et stipule par la même qu'occasion que l'acceptation des conditions générales et de confidentialité ne peut valoir accord pour un traitement précis des données personnelles déterminé par l'entreprise exploitante (Google, ou Facebook).

Si la personne n'a de son vivant pris aucune décision, ses héritiers peuvent demander aux entreprises exploitantes de données personnelles de clôturer le compte du défunt, et éventuellement d'y avoir accès, si la succession l'exige.

[220] Site de l'entreprise Edeneo https://secure.edeneo.fr/ , vu le 30/08/2017

[221] Site de l'entreprise Safe Beyond, https://www.safebeyond.com/ , vu le 30/08/2017

[222] Site de la CNIL, https://www.cnil.fr/fr/ce-que-change-la-loi-pour-une-republique-numerique-pour-la-protection-des-donnees-personnelles, vu le 30/08/2017

2. L'INTERNAUTE COMME OBJET DE RECHERCHE

2.1 La propriété de la donnée personnelle

2.1.1 La donnée comme une extension de soi

Pierre Bellanger estime que les données «sont une extension de la personne». Il va jusqu'à associer l'association suivante :

> «à la manière du sang, [la donnée personnelle] est un soi hors de soi. Engager le transfert ou la cession de données personnelles d'autrui, indissociables ou déductibles des siennes, en échange de l'accès « gratuit » à un service s'apparente, par conséquent, au trafic d'organes.»

Les données seraient indissociables de la personne, lui appartiendraient si intimement que la connaissance d'un individu ne pourrait être monnayée, que la valeur d'une telle connaissance ne saurait être sujette à commercialisation tant elle est liée à l'individu et à son choix privé de la partager ou non. Il va jusqu'à considérer qu'un individu ne peut agréer à la cession de ses données : dans la mesure où son quotidien est perclus de capteur de données personnelles à travers les objets connectés (montre, téléphone, réveil, maison intelligente), il serait automatiquement réduit à un collecteur de données.

Pour Bellanger, « L'autorisation individuelle réfléchie à chaque captation, déjà aléatoire, n'est plus possible dans les faits » dans la mesure où un flux de données permanent est désormais assuré entre l'individu et l'exploitant des objets connectés. Il est indiscutable que les nouvelles technologies permettent de collecter de façon continue des informations personnelles et même extrêmement sensibles : une montre connectée peut calculer le rythme cardiaque, par exemple. Une fonction utile pour les coureurs, mais également une information précieuse pour les assurances, informées de l'hygiène de vie de l'utilisateur mais aussi possiblement de problèmes médicaux. Cette approche considère la personne concernée par les données personnelles comme la propriétaire de ces données.

2.1.2 La propriété juridique des données personnelles

La loi, même la plus progressive en matière de sécurité et de protection de la personne physique sur Internet, s'est bien gardée de définir un propriétaire des données. Au contraire, quand bien même ces données n'existeraient pas sans les individus dans leur contenu, ni sans les collecteurs sur leur support, aucune de ces deux parties n'a la propriété de ces données. Les collecteurs sont des «responsables de traitement» tandis que les individus sont les «sujets» de la donnée. La loi objective la donnée en ce qu'elle n'en fait pas un objet sujet à propriété, mais un objet exploité autour duquel s'articulent les intérêts de parties et des enjeux moraux.

Le Conseil National du Numérique s'est ouvertement prononcé contre l'instauration d'une propriété sur les données, arguant que les revenus dégagés par cette propriété - et donc le droit de vendre ses données directement aux plateformes, serait peu important. De plus, ce serait «[nier] le rapport de force

entre consommateurs et entreprises» que de faire peser la responsabilité entière des ces données à l'usager. Le Conseil estime que cela conduirait à un renforcement des inégalités entre utilisateurs avertis et amateurs, au détriment de ces derniers, moins bien formés à protéger efficacement leurs données.

2.2 Les traces numériques de l'individu sur les plateformes sociales

2.2.1 Une collecte massive des traces numériques

En janvier 2017, la somme des utilisateurs actifs mensuels des neufs premiers réseaux sociaux dépassent les cinq milliards. Ces réseaux sociaux, par définition, reposent sur des données personnelles des individus, qu'ils collectent et conservent dans des *datacenters*. Des données permettant d'abord d'identifier, de caractériser (identifiant, adresse mail, âge, sexe, coordonnées IRL, sexe), puis d'autres, créées par l'expression de l'individu et permettant de le connaître (déclarations sous forme de billets d'humeur, partage de photos, simple like ou commentaire).

Les principaux collecteurs de données personnelles sont de grands groupes privés. Leur pratiques de conservation de données ont souvent soulevé des débats ; en 2011, un autrichien a déposé plusieurs recours contre Facebook, qui avait conservé même les éléments de son activité qu'il avait manuellement effacé. La firme était également soupçonnée de collecter des informations sur des non-utilisateurs.

En mettant de côté les conflits juridiques, on apprend que Facebook collecte des centaines de données sur chaque individu : Max Schrems, le plaignant, a demandé à Facebook de lui faire parvenir tout ce qui le concernait et a reçu, après trois ans d'utilisation du réseau social, un document de mille deux-cents pages. Le réseau social conserve les publications des individus sa plateforme et construit leur mémoire «numérique» en liant les contenus, les métadonnées associés et les relations entre les différents acteurs. C'est ce qui permet le profilage.

La mémoire de l'individu selon les plateformes est multiple et sans doute redondante. Ce sont les algorithmes qui collectent, analysent et stockent les données personnelles des individus pour en tirer une véritable base de données, une mise en réseau de l'individu, qu'il convient bien sûr de capitaliser auprès des annonceurs.

2.2.2 Des données finalement peu exploitées

On pourrait penser que des algorithmes suffisamment puissants pour massivement collecter des données des personnes pourraient participer activement à la mise en mémoire des utilisateurs, voire à une mémoire quasi exhaustive de leur identité numérique. Pourtant, 90% des données produites en 2013 étaient des données non structurées, et donc très difficiles à exploiter, puisque difficilement soumises à la recherche par contenu. Par exemple, une photographie est difficilement exploitable par un algorithme. Composée de bit, c'est à dire de 0 et de 1, une photographie ou une vidéo ne correspondent pas aux standards de recherche d'un web sémantique. Si on en croit l'infographie de l'entreprise Visiativ

- spécialisée en édition et en intégration de solutions logicielles - 88% des données disponibles ne seraient pas analysées et sur toutes ces données disponibles, seulement un tiers pourraient générer de la valeur après avoir analysée. L'IDC rapporte qu'en 2013, seules 5% des données disponibles sur le web avaient été analysées. Il semble au vu de ces chiffres que les exploitants de données soient loin d'analyser les données de leurs usagers de façon optimale.

2.2.3 Les traces numériques ne peuvent pas faire mémoire

Pour Louise Merzeau, il n'est pas possible d'associer les données récoltées sur l'individu à une représentation de lui-même. Le fait que des données soient collectées à propos d'un individu, partout et à chaque instant de sa navigation, ne renseigne pas sur l'individu, mais au contraire en déconstruit l'identité. De plus, les traces numériques d'une personne ne dépendent pas - ne dépendent plus - uniquement de lui, et même assez peu. Merzeau différencie trois identités qui constituent les «niveaux» de profondeur dans la collecte des données personnelles sur les réseaux sociaux. L'identité déclarative est celle qui est provient de l'individu, des traces numériques qu'il laisse volontairement ; les traces d'activités collectées automatiquement forment l'identité agissante ; enfin, l'identité calculée correspond à une sorte de bilan d'activité de l'individu et «comptabilise ses scores, ses « amis », ses visites, sa production».

Les individus ne sont pas les premiers producteurs de données qui les concernent. Aux données automatiquement générées par des comportements traceurs, il faut ajouter les informations en provenance d'autres individus que le sujet des données. Il est certes toujours possible de contextualiser la mémoire d'un individu en la mettant en relation avec d'autres éléments ; pour autant, il semble erroné de penser que construire la mémoire numérique d'un individu puisse se satisfaire de données disséminées et dont la crédibilité de la source ne peut toujours être vérifiée. De plus, l'individu en tant qu'utilisateur du web ne peut qu'être influencé par les usages qui sont ceux de l'espace dans lequel il s'exprime. Les algorithmes qui en sont à l'origine structurent son expression, puisqu'ils la limitent (Facebook choisit la police et taille des caractères de publications, par exemple).

Afin de coller aux exigences de traitement des algorithmes mis en place par diverses plateformes, les données personnelles subissent une «déliaison [...] qui permet de redistribuer la personne dans les interactions». Chaque parcelle collectée à propos de l'individu ne peut participer à en construire l'identité numérique, puisqu'elle se retrouve éclatée, afin de garantir une «granularité qui permette une indexation des données». Ne reste alors que l'»ombre digitale» et non la mémoire de leurs utilisateurs.

2.3 Archiver les données personnelles des internautes sur le web

2.3.1 L'intérêt de conserver des données personnelles

L'archivage du web touche aux individus. Par exemple, presque tous les sites de presse en ligne proposent la possibilité de commenter et avec les commentaires s'affichent l'identifiant de leur auteur, parfois son âge, son pays, sa région ou son adresse e-mail. Archiver le web revient donc aussi à conserver une partie de l'activité des individus sur le web, même lorsque cette activité est annexe ou anodine. Face à la masse de sites web à archiver, il est impossible d'anonymiser les données de chaque page. Les données personnelles sont dans ce cas archivées de façon inévitable, parce qu'elles font partie d'une page donnée.

2.3.1.1 La BnF fait mémoire de l'internaute

Existent aussi un archivage volontaire de données personnelles - rappelons ici qu'une donnée personnelle est une donnée pouvant conduire, par elle-même ou croisée avec d'autres informations à l'identification d'un individu - qui peut concerner les blogs, par exemple.

Il s'agit de choses qui ont été mises le net, mais en comptant sur ce que Dominique Cardon appelle le «clair-obscur» : l'information est théoriquement accessible à tous, mais en pratique, le cercle des personnes susceptibles d'y accéder est très restreint puisque ne disposant d'un moteur de recherche ou permettant à l'utilisateur de choisir l'étendue de la diffusion qu'il souhaite donner à son contenu.

Ces données, que Marie-Anne Chabin appelle : les données de particuliers forment selon elle un «support d'un sens de la mémoire chez les individus»[223]. Conserver la mémoire individuelle, c'est conserver une mémoire directe, de l'individu qui assiste aux grands évènements de son époque et pouvoir recontextualiser ces grands évènement à travers la subjectivité d'un regard. Cette approche de la mise en mémoire de l'individu ne contredit pas nécessairement la pensée de Merzeau sur l'impossible mise en mémoire de l'individu par les systèmes. Il s'agit ici d'archiver des éléments de la vie numérique d'un individu, pas de tenter de faire mémoire d'une personne. De plus, il ne s'agit pas ici d'archiver les données involontairement produites par un internaute - son historique de navigation, ses préférences d'affichage - mais ce qu'il a choisi d'exprimer sur un sujet donné.

Le web est le médium de l'instantanéité, où tant d'informations se succèdent qu'elles sont obsolètes très rapidement. Les réactions des personnes pour un événement précis peuvent être archivées sans perdre leur spontanéité : il est devenu possible d'archiver l'émotion.

[223] « Mémoire numérique : une amnésie programmée ? Compte-rendu » Camille http://www.webarchivists.org/2013/05/memoire-numerique-une-amnesie-programmee-compte-rendu/ , vue le 30/08/2017

2.3.1.2 Le web comme espace d'autobiographie

Archiver les données des personnes, c'est aussi faire mémoire de la personne. Conway dote la mémoire autobiographique de trois connaissances.

Les périodes de vie, qui sont «de longues époques mesurées en années ou décennies et composées d'informations générales sur les lieux, les personnages, les activités». C'est ce qu'on peut rapprocher des blogs (et parfois de Facebook) dans lesquels les créateurs de contenus se livrent sur leur passé, témoignent d'expériences plus ou moins lointaines.

Les événements généraux, eux, «correspondent à des périodes plus restreintes faites de jours, semaines, mois [...]. Ils incluent aussi des événements répétés ou liés par un thème commun». Le podcast, s'il n'est pas limité à cela, permet aux podcasteurs de partager avec leurs abonnés ce qu'on peut considérer comme des événements généraux : certains expatriés racontent des épisodes marquant de leur vie dans leur pays d'accueil, par exemple.

La dernière connaissance de la mémoire selon Conway se trouve dans les «détails d'événements spécifiques» qui sont des «détails sensoriels» courte, voire de très courte durée qui »répondent à des caractéristiques émotionnelles spécifiques». Ces détails peuvent être rapprochés des tweets, dont les cent-quarante caractères sont suffisants à l'expression de détails sensoriels. Il est d'ailleurs facile d'imaginer sous la forme d'un tweet ou d'un statut Facebook l'exemple donné par Bernard Croisile pour illustrer ce qu'est un détail d'évènements spécifiques : «J'ai vu l'exposition Edward Hopper à Boston».

2.3.2 De la conservation à but commercial à l'archivage

Les données s'orientent, se réarrangent sur la nébuleuse en fonction des liens qui sont faits entre elles. L'approche sur la propriété et la régulation sur la collecte et l'utilisation commerciale diffère un peu lorsqu'il s'agit de faire mémoire. Facebook ou Google scannent l'activité de leurs utilisateurs à la recherche de mots-clés ou d'indices qui leur permettraient de gagner de l'argent en vendant vos envies à des annonceurs. La partie de leur activité qui consiste à conserver les anciennes publications des individus consiste à relier entre elles le maximum d'information possible, afin de tirer un profilage de l'individu, qu'ils pourront capitaliser dans une intention commerciale.

On peut ainsi différencier la conservation passive qui est faite de l'activité sur le net d'un individu des initiatives d'archivage actif du web, dans un but patrimonial et de recherche, qui sont celles de la BnF ou de la bibliothèque du Congrès. Ces institutions ont la responsabilité d'archiver le web, où des sites disparaissent en moyenne au bout de cinq ans et où chaque jour le flux de données à alimenter chaque jour le fonds existant ne cesse d'augmenter. Un site en est permanence étendu, modifié, commenté, supprimé, réorganisé, migré. Archiver le web, c'est fixer un fonds par nature mutant. Face à cette tâche de Sisyphe, la BnF, dépositaire en France du dépôt légal, a laissé de côté «l'idéal traditionnel d'exhaustivité [pour] la représentativité et l'échantillonnage raisonné».

2.3.3 Le web admissible au dépôt légal

Internet, et son application le web, sont utilisés en permanence et de plus en plus, impacte la vie de tout un chacun, parce qu'ils s'imposent désormais dans tous les aspects de la vie : des machines avec un accès à Internet équipent les écoles, les entreprises, les particuliers, qui s'en servent pour acheter, vendre, communiquer, s'exprimer, publier, partager, lire, réagir, s'informer, signer des pétitions, chercher un emploi. A tel point qu'en France, la loi pour une République numérique portait en 2016 le droit au maintien de la connexion Internet pour les ménages en difficulté, sept ans après que le Conseil constitutionnel ait considéré l'accès à Internet comme un droit fondamental, parce que devenu une composante importante de la liberté d'expression. L'omniprésence d'Internet dans la vie des personnes et son utilisation comme moyen d'expression en fait aussi un lieu de mémoire, d'autant qu'il est possible de faire un parallèle entre les champs lexicaux de l'édition et ceux propres à l'expression sur Internet : l'utilisateur, lorsqu'il souhaite s'exprimer, «édite» un contenu, «met en forme» son texte, et «publie» des «articles», des «billets» et des «commentaires». L'archivage du web s'inscrit alors dans les «adaptations successives du dépôt légal institué par François Ier (1537) aux différents supports de la production intellectuelle française, selon une évolution logique puisqu'une partie de cette production n'est disponible aujourd'hui qu'en ligne».

La pratique de l'archivage du web reçoit le soutien de l'Union Européenne puisque le règlement européen 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, précédemment évoqué, permet aux activités archivistiques de déroger aux droits des individus vis à vis de leurs données personnelles(droit à l'effacement, à la rectification, droit d'opposition), une mesure qui vise à simplifier le cadre juridique flottant dans lequel les archivistes du web ont évolué. En France, la première campagne d'archivage du web par la BnF se fait en 2002, à l'occasion de la campagne présidentielle, mais il faut attendre la loi DADVSI de 2006 pour que son action soit intégrée à un cadre juridique grâce à l'élargissement du dépôt légal au web.

2.3.4 Un inversement des pratiques archivistiques

Le dépôt légal traditionnel oblige les producteurs à faire parvenir à l'autorité dépositaire une ou plusieurs copies de chacune de leur œuvre. Pour le web, la logique est renversée : ce sont les archivistes qui vont chercher le contenu à collecter. Alors que le dépôt légal visait la conservation exhaustive des productions intellectuelles françaises, la masse des données et leur grande fluctuation poussent les archivistes à adopter une autre approche, consistant celle-ci à proposer un panorama le plus fidèle possible du web à un moment donné.

L'archive du web se distingue d'une archive classique en ce qu'elle diffère par sa nature de ce qu'elle permet de conserver. Le web est objet dynamique en perpétuelle mutation ; son archive en est une photographie, prise à un moment

donné. Louise Merzeau différencie les archives faites d'un document classique et celles de documents nativement numériques : »la prise en charge d'un document qui témoigne des indices d'une activité et qui est catégorisé dans un ensemble documentaire (définition classique de l'archive), et d'un matériau qui est un média en soi et qui est donc éditorialisé dans un ensemble médiatique (définition nouvelle de l'archive numérique)».

Afin de donner une «avec une chronologie interne, qui reconstitue l'environnement de lecture et de navigation original au moment du passage du robot sur le site capturé». Comme pour n'importe quel fonds, les collections ne sont pas le résultat d'une initiative unique. Les collectes larges visent à rassembler le maximum de sites ; les collectes ciblées qui permettent de compléter les premières ; les collectes projets s'intéressent en profondeur à un sujet donné, selon une logique événementielle. Celle-ci exige de l'archiviste de prioriser les évènements et les sujets, ce qui peut entrer en contradiction avec la mission première d'archiver toutes les productions intellectuelles que portent le rôle de dépositaire du dépôt légal.

La priorisation ignore donc une partie du contenu disponible à l'archivage sur le web. La trop grande masse de contenu à traiter a transformé le rôle de la BnF, qui au départ conservait toutes les oeuvres envoyées, sans discrimination d'aucune sorte. L'archivage du web, selon Marie-Anne Chabin, consiste à éliminer des documents redondants ou jugés peu dignes d'intérêt, dans le cadre d'une «destruction raisonnée».

Enfin, les archives du web, fixées dans leurs multiples formats de lecture, se retrouvent inévitablement confrontées à l'obsolescence des formats, qu'il est possible de comparer à une péremption. Au-delà d'un nombre limité de nouvelles versions d'un logiciel, les premières ne sont plus lisibles. Alors que l'archive doit pérenniser un contenu, il s'avère paradoxalement qu'archiver un contenu revienne à poser le point chronologique qui mènera le document à sa péremption. Les migrations successives retardent l'obsolescence des données, mais contiennent le risque d'une perte en authenticité, voire d'une perte des données.

3. L'ACCORD ENTRE LA BIBLIOTHÈQUE DU CONGRÈS ET TWITTER

3.1 Twitter comme support d'étude de l'opinion

Les média sociaux, en apportant à leurs utilisateurs un espace d'expression personnelle et de communication dans lequel ces derniers se sont massivement engouffrés, ont également ouvert la porte à la recherche. La multiplicité des utilisateurs, la facilité avec laquelle les chercheurs peuvent échanger avec eux en cas de besoin - proposer un questionnaire, par exemple - le grand nombre d'informations laissées publiques, et la pénétration du numérique dans la vie de tout un chacun alimentent chaque jour d'immenses bases de données dont le rôle est de garder trace de l'individu numérique, laissant entrevoir de nouvelles

possibilités de recherche. Ne manque plus qu'à transformer ces données en savoir, par leur étude, leur traitement, leur croisement.

Twitter a particulièrement intéressé les chercheurs, surtout dans la mesure du l'opinion. Il s'agit d'une plateforme permettant à ses inscrits de publier des billets, autant de fois qu'ils le souhaitent, limités uniquement par un maximum de cent-quarante caractères ; la plateforme, très populaire, enregistre aujourd'hui cinq-cent-quatre millions de tweets journaliers. Twitter a été largement étudié afin de mesurer des audiences, des tendances et d'analyser certains comportement en ligne. Il est même devenu un des champs d'affrontement des campagnes politiciennes, et un outil de mesure de l'opinion publique. Les travaux de Tumasjan tendent même à montrer une « corrélation positive entre le nombre de tweets évoquant les candidats et les partis ayant concouru aux élections législatives allemandes de 2009 et leurs résultats électoraux effectifs », et d'autres chercheurs s'intéressent à Twitter, en tant que construction à penser comme en tant que base de données extensible et auto-alimentée.

3.2 L'accord

La bibliothèque du Congrès, aux Etats-Unis, dépositaire du dépôt légal, a considéré que, dans la mesure où les média sociaux tenaient lieu d'espace de communication et d'expression originale des personnes, et qu'ils tendaient à se substituer aux supports classiques qui composent traditionnellement les collections des bibliothèques (journaux, livres papier, etc.). Logiquement - et conformément à la tradition de la bibliothèque du Congrès de collecter des récits personnels de l'Histoire - venait la nécessité d'archiver les média sociaux et a conclu en avril 2010 un accord avec le réseau de micro-blogging Twitter.

L'accord est le suivant : Twitter donne à la bibliothèque du Congrès la copie de tous les tweets publics publiés depuis la mise en ligne du site jusqu'à la date de l'accord, puis par la suite continue de fournir copie des tweets publiquement postés sur sa plateforme. La bibliothèque du Congrès peut rendre ce fonds de tweets accessible sous trois conditions : en premier lieu, la mise en accessibilité ne peut intervenir avant que les tweets aient atteint une ancienneté de six mois ; ensuite, la bibliothèque ne peut permettre sur son site web le téléchargement d'une partie substantielle du fonds ; pour finir, les chercheurs souhaitant accéder à ce fonds doivent s'engager à ne pas en faire un usage commercial où à le redistribuer.

3.3 Difficultés techniques

3.3.1 Pérennisation

La bibliothèque du Congrès se lance ici un vrai défi : les vingt-et-un milliards de tweets publics publiés les quatre premières années de l'existence de Twitter représentent un volume colossal : vingt terabytes, soit l'équivalent de trois cent mille heures de musique en streaming, c'est à dire un peu plus de trente-quatre ans. En décembre 2012, c'étaient cent-soixante-dix milliards de tweets qui avaient été versés la bibliothèque du Congrès.

Pour relever le défi de pérenniser et de rendre accessible ces tweets, désormais élevés au rang de documents, la bibliothèque a fait appel à la société Gnip, spécialisée en analyse de données et rachetée par Twitter en 2014. C'est Gnip qui reçoit les tweets publics en flux continu, puis qui les segmentent pour ranger dans un même fichier tous les tweets envoyés sur la même heure. Ces fichiers sont ensuite communiqués à la bibliothèque du Congrès, qui après avoir vérifié les fichiers, et en avoir capturé les statistiques, les copies sur deux bandes magnétiques conservées en des emplacements distincts, supprime les fichiers numériques. Ce processus, automatisé, semble permettre la bibliothèque d'accueillir et de conserver durablement et efficacement les tweets.

3.3.2 Accessibilité et visibilité

En revanche, rendre accessible un si grand volume de données relève d'un tout autre niveau de complexité technique. D'abord, depuis la signature de l'accord, le volume de tweets publiés a été multiplié par dix. Ensuite, il s'agit pour la bibliothèque de la première collection à être alimentée en flux continu et aussi massivement. Le transfert des tweets a nécessité « d'optimiser l'infrastructure technique et le workflow établis pour les autre contenus numériques ».

La bibliothèque du Congrès explique qu'archiver Twitter et archiver des sites web requiert une approche différence : en dehors de la rapidité d'expansion de la collection, la bibliothèque doit aussi distinguer les tweets originaux des retweets automatiques et des retweets manuels, afin d'éviter le bruit ; ceux liés à une conversation, ainsi que les tweets contenant un lien vers un autre contenu, des tweets accompagné d'un fichier, quel qu'il soit, qui ne peuvent pas tous être compris par eux-mêmes. A cela s'ajoute la multiplicité des demandes - faire remonter tous les tweets contenant un terme précis, ou tous ceux d'une année, d'une institution - et des besoins des chercheurs.

3.3.3 Un fonds en changement constant

L'archiviste est confronté à en mutation perpétuelle : changement de pseudo, ajout ou suppression de tweets, ajout ou suppression de compte. Il faut rappeler que la bibliothèque du Congrès n'a obtenu que le droit d'archiver les tweets publics. A chaque fois qu'un utilisateur supprime un tweet, clos ou privatise son compte, les tweets correspondants conservés par les services de la bibliothèque doivent être supprimés du fonds.

La nouvelle de l'archivage de Twitter par une institution publique avait soulevé des débats sur la propriété des tweets. Il est vrai que certains points de droit devenaient ambigus : par exemple, tous les tweets ne revêtent pas suffisamment d'originalité pour relever du droit d'auteur. La décision d'accorder ou non la propriété intellectuelle, qui revient à un juge, devient alors sujette à caution. De plus, tous les tweets archivés ne sont pas émis par des habitants des Etats-unis, aussi se posait le problème du champ de la juridiction et de l'application de la loi. Ces questions, si elles ont le mérite d'avoir été posées, n'ont pas modifié le projet. D'abord, même si les conditions générales de Twitter

certifiaient à l'utilisateur que le contenu de ses tweets restait sa propriété, elles stipulaient également que « l'utilisateur accorde à Twitter une licence mondiale non exclusive, libre de redevance avec le droit de sous-licencier, utiliser, copier, reproduire, traiter, adapter, modifier, publier, transmettre, afficher et distribuer le Contenu à tous les médias ou à toutes les méthodes de distribution (connues à présent ou développées ultérieurement) ».

Ces conditions donnent à Twitter tous les droits patrimoniaux relatifs au droit d'auteur. De plus, la bibliothèque du Congrès est dépositaire du dépôt légal et à ce titre le droit d'auteur ne peut limiter la collecte qui est faite de la production de contenus sur le territoire états-unien.

3.3.4 Twitter : une représentativité ciblée

Il a été mentionné plus haut l'intérêt du monde la recherche pour Twitter et la vertu de représentativité des personnes s'exprimant sur ce medium. Pourtant, sociologiquement, l'utilisateur type de Twitter « diffère très nettement de celui de la population dans son ensemble, avec une surreprésentation de cadres, de diplômés et d'étudiants ».

Par la suite, les corrélations entre les débats sur la twittosphère et les résultats électoraux ont été expliqués de la manière suivante : Twitter est le médium de personnes prescriptrices – « leaders d'opinion » - et ce sont ces personnes qui influencent l'opinion publique positivement ou négativement vis à vis d'un candidat.

Un travail de panélisation de la twittoshpère datant de 2014 et mené par Julien Boyadjan tend à valider l'hypothèse d'une forte concentration de leaders d'opinion sur Twitter, mais réfute les hypothèses selon lesquelles la twittospère serait politisée - beaucoup de panélisés ne se sont donnés la peine de tweeter sur un sujet politique qu'en période d'élections – ou ses opinions dotées d'un caractère - quelque soit la méthode de mesure, l'issue de la présidentielle française de 2012 n'avait été devinée – ou représentatif – les travaux de Boyadjan confirmant la forte similitude sociologique entre les utilisateurs, majoritairement masculin, jeunes et ayant suivi ou suivant un cursus universitaire, les rapprocheraient plus d'une minorité privilégiée que d'une représentation populaire.

3.4 Continuité du projet

Trois ans après que le contrat ait été conclu entre Twitter la bibliothèque du Congrès, un outil de recherche n'avait toujours été mis à disposition des chercheurs, et une recherche dans la base de données accueillant uniquement les tweets jusqu'à 2010, demandait une journée entière. Le processus de recherche est complexe et coûteux en terme de consommation de données : son accélération nécessiterait d'investir massivement dans des centaines de nouveaux serveurs, une dépense à laquelle une institution publique ne peut faire face. Il s'agissait alors pour la bibliothèque de mettre en place une solution d'accès bêta en attendant que la technologie, en attendant que la loi de Moore fasse son effet et réduise les coûts de stockage, ou qu'une solution adaptée soit trouvée. En 2016, six ans après le début de l'archivage de Twitter, les équipes de la bibliothèque du Congrès

travaillaient toujours sur l'élaboration d'une solution qui permettrait de rendre visible et accessible aux chercheurs les milliards de tweets conservés par la bibliothèque.

CONCLUSION

Les travaux scientifiques du XXème siècle ont amené les chercheurs à développer de nouvelles technologies de transmission de données, dont la commutation de paquets. L'optimisation de la communication qu'ont permis ces recherches ont donné naissance à Internet. L'utilisation d'Internet est tellement répandue que sa pratique devient un réflexe : rechercher un emploi, une information, contacter ses amis, tout cela est possible grâce au réseau des réseaux, qui s'est installé et presque standardisé en tant qu'outil de communication public ou privée, instutionnelle ou d'affaire. L'utilisation d'Internet et de son application principale, le web sont devenues des pratiques courantes, normales, à tel point que les pouvoirs publics français ont débloqué des financements pour lutter contre la "fracture numérique". Il s'agit de l'inégalité qui distingue les personnes disposant d'un accès facile à Internet de celles qui n'y ont pas, ou difficilement accès. Le terme de fracture est intéressant en ce qu'il matérialise bien la cassure sociale entre les accédants et non-accédants à Internet : le réseau de communication supporte un espace d'expression, devenu vecteur de lien social.

Pourtant, si les utilisateurs ont pu transformer Internet en lieu social, les entreprises qui s'y sont installées n'ont rien de public ou d'altruiste. Elles fonctionnent sur un principe de gratuité fiducière pour les utilisateurs, mais ne sont pas entièrement gratuite pour autant. L'internaute, lorsqu'il utilise un service en ligne, échange contre la gratuité d'accès à ce service des informations qui le concernent, sous la forme de données. Cela est particulièrement le cas avec les quatre plus grosses entreprises de service en ligne : Google, Amazon, Facebook et Apple, désignées par l'acronyme GAFA. Ces entreprises, qui collectent toutes sortes de données en grande quantité sur l'utilisateur sans systématiquement l'en informer, les capitalisent ensuite de deux façon. Les données sont directement revendues à des annonceurs, ou les annonceurs paient les GAFA pour pouvoir profiter de leur notoriété en plaçant une publicité sur leurs pages.

Les consommateurs ne sont plus clients, mais source de profit pour les entreprises en ligne. Les méthodes de collecte et de traitement de ces données ont souvent été décriées. Dans un marché peu, voire pas encadré, les recours se faisaient difficilement. C'est pourquoi l'appareil législatif s'est emparé du problème pour assurer un niveau de protection suffisant de la vie privée aux personnes physiques lors de leur utilisation d'Internet. Les GAFA, répondaient à la loi californienne, ce qui les rendait difficilement sujettes à sanction en dehors des Etats-Unis. Vecteurs de croissance, le poids financier des GAFA s’est alourdi d’années en années, et leur chiffre d’affaires a fini par équivaloir et dépasser le PIB de petits états. La régulation nationale de la collecte et du traitement de données personnelles était alors clairement à l’avantage des GAFA, sis dans l’espace juridictionnel favorable des Etats-Unis, et puissants comme des états. Après plusieurs tentatives nationales de régulation au succès contrasté, l'Europe a décidé de mettre en place une législation commune visant à protéger de façon uniforme tous les individus en les dotant de droits de recours (droits d'opposition, de rectification, d'effacement et droit à la portabilité des données), en coordonnant les autorités de régulation nationales entre elles et en augmentant significativement

les sanctions à l'égard des exploitants de données personnelles contrevenant. Une infraction avérée au règlement européen relatif à la protection des personnes physiques à l'égard des traitements de données personnelles peut coûter à une entreprise jusqu'à 4% de son chiffre d'affaire. Dans le même temps se développent chez les utilisateurs des comportements plus prudents, et de chez les entreprises des alternatives aux GAFA, plus respectueuses des données personnelles et de la vie privée.

Un mouvement intellectuel qui ramène le web à la place de marché et les services en ligne à des entreprises, avance que chaque clic, chaque partage, participant à leur chiffre d'affaire, les utilisateurs étant les producteurs non-rémunérés des données, alors ces utilisateurs sont la force de travail de ces entreprises. Les internautes constituent une force de travail inconsciente de sa nature, exploitée gratuitement ou presque (en échange d'un accès à un service). Ce travail numérique inconscient est désigné sous le terme *digital labor*.

La BnF a vu le champ de son dépôt légal s'élargir en 2006, au moment où la loi française a reconnu le web et ses contenus comme patrimoine. Depuis, des campagnes régulières archivent des pages web dans le but de conserver sur le long terme une image représentative du web à un moment donné. La loi a apporté une réponse aux questions relatives à la propriété intellectuelle et au respect de la vie privée soulevées par l'indexation du web : l'archivage est une exception au droit d'auteur comme aux droits de la personne sur ses données personnelles. Une autre question perdure : le web social est un espace d'exposition de soi, mais aussi de mise en roman de soi. Dans le monde analogique, des institutions comme les *média* (télévision, livre, radio) fonctionnent comme des sortes d'entonnoirs qui classent les informations selon le crédit qu'il est possible de leur donner et diffusent les informations les plus solides, hiérarchisées selon la ligne éditoriale du *medium* en question. Les gatekeepers n'existent pas sur le web ; les résultats d'un moteur de recherche sont fonctions de la pertinence sémantique et de la popularité de l'information, mais n'assurent en rien sa véracité. On a même vu que l'utilisation d'un réseau social relève plus de la stratégie de la sincérité. Si l'archive doit être utile au chercheur, alors l'information qu'elle porte doit pour être digne d'intérêt, être porteuse d'un peu de vérité. Comment étudier les internautes, comment trouver une logique dans leurs pratiques, ou étudier leurs contenus, si ceux-ci faussent volontairement la donne ? Et si c'est le cas, est-il possible, et de quelle façon, de déterminer le caractère probable, sinon sincère, d'un contenu ? Comment mener dans les archives du web des recherches sérieuses si ces mêmes archives sont porteuses d'informations erronées et de semi-réalités ?

SOURCES

SOURCES JURIDIQUES :

Site de la Commission Nationale Informatique et Libertés (CNIL), https://www.cnil.fr/fr/

Site du service de diffusion en ligne de la loi française et européenne, https://www.legifrance.gouv.fr/

Site du service de diffusion en ligne des règlements et directives européens, http://eur-lex.europa.eu

Règlement 2016/679 du Parlement Européen et du Conseil, du 27 avril 2016, relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, URL : http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Site de Thiébault Devergranne, docteur en droit, https://www.donneespersonnelles.fr/

Site de Stanford, répertoire des affaires contre Google, http://fairuse.stanford.edu/case/field-v-google-inc/

POLITIQUES DE CONFIDENTIALITÉ :

Politique de confidentialité d'Apple, https://www.apple.com/privacy/privacy-policy/

Politique de confidentialité de Facebook, https://www.facebook.com/privacy/explanation

Politique de confidentialité d'Amazon, https://services.amazon.com/content/Privacy_Policy.htm/ref=asus_privacy_fnav

Politique de confidentialité de Google, https://privacy.google.com/?hl=fr

Liste d'entreprises certifiées par le Privacy Shield, https://www.privacyshield.gov/participant_search

SOURCES JOURNALISTIQUES

Presse généraliste

Site de *Le Monde*, http://www.lemonde.fr/

Site de *Libération*, http://www.liberation.fr/

Site de *Le Figaro*, http://www.lefigaro.fr/

Presse spécialisée

Site du *Journal du Net*, http://www.journaldunet.com/

Site de *Macgo*, https://www.macg.co/
Site de *Ulyces*, http://www.ulyces.co/
Site de *Znet*, http://www.zdnet.fr/

LE DATA EN CHIFFRES :

Site de l'entreprise Markentive, https://www.markentive.fr/blog/
Site de l'entreprise Marketingland, http://marketingland.com
Outil de mesure d'audience de sites web, http://www.trackalytics.com/

SOURCES UNIVERSITAIRES :

Site de l'Enssib, http://www.enssib.fr
Site de Supinfo, http://www.supinfo.com/

ENTRETIENS :

Antonio Casilli. *Les données numériques : Un enjeu d'éducation et de citoyenneté*. Retranscription de l'audition devant le CESE (Section de l'Education, de la Culture et de la Comm.. 2014. URL : https://hal.archives-ouvertes.fr/hal-01068525/document

Cardon Dominique, « La démocratie Internet. Entretien avec Dominique Cardon », *Transversalités*, 2012/3 (N° 123), p. 65-73. DOI : 10.3917/trans.123.0065. URL : http://www.cairn.info/revue-transversalites-2012-3-page-65.htm

BIBLIOGRAPHIE

LES DONNÉES PERSONNELLES

Latour Bruno. La fin des moyens. In: Réseaux, volume 18, n°100, 2000. Communiquer à l'ère des réseaux. pp. 39-58. DOI : 10.3406/reso.2000.2211 URL : www.persee.fr/doc/reso_0751-7971_2000_num_18_100_2211

Ollion Étienne, Boelaert Julien, « Au delà des *big data*. Les sciences sociales et la multiplication des données numériques », *Sociologie*, 2015/3 (Vol. 6), p. 295-310. URL : http://www.cairn.info/revue-sociologie-2015-3-page-295.htm

Bellanger Pierre, « Les données personnelles : une question de souveraineté », *Le Débat*, 2015/1 (n° 183), p. 14-25. DOI : 10.3917/deba.183.0014. URL : http://www.cairn.info/revue-le-debat-2015-1-page-14.htm

Protection des données

La privacy

Rey Bénédicte, « Les intelligences numériques des informations personnelles. Vers un changement de perspective pour garantir le droit à la vie privée ? », *Les Cahiers du numérique*, 2014/1 (Vol. 10), p. 9-18. DOI : 10.3166/lcn.10.1.9-18. URL : http://www.cairn.info/revue-les-cahiers-du-numerique-2014-1-page-9.htm

Valjavec Emmanuel, « Internet, un nouvel espace de liberté sous surveillance », *Études*, 2013/3 (Tome 418), p. 317-327. URL : http://www.cairn.info/revue-etudes-2013-3-page-317.htm

Lancelot Miltgen Caroline, « Dévoilement de données personnelles et contreparties attendues en e-commerce : une approche typologique et interculturelle », *Systèmes d'information & management*, 2010/4 (Volume 15), p. 45-91. DOI : 10.3917/sim.104.0045. URL : http://www.cairn.info/revue-systemes-d-information-et-management-2010-4-page-45.htm

Régulation

Dumont Béatrice, « La régulation à l'échelle communautaire. Une analyse économique des instruments et institutions de la protection des données au sein de l'UE », *Réseaux*, 2011/3 (n° 167), p. 49-73. DOI : 10.3917/res.167.0049. URL : http://www.cairn.info/revue-reseaux-2011-3-page-49.htm

Lancelot Miltgen Caroline, « Dévoilement de données personnelles et contreparties attendues en e-commerce : une approche typologique et interculturelle », Systèmes d'information & management, 2010/4 (Volume 15), p. 45-91. DOI : 10.3917/sim.104.0045. URL : http://www.cairn.info/revue-systemes-d-information-et-management-2010-4-page-45.htm

Rallet Alain, Rochelandet Fabrice, Zolynski Célia, « De la *Privacy by Design* à la *Privacy by Using*.

L'environnement législatif et normatif

Articles

Anciaux Arnaud, Farchy Joëlle, « Données personnelles et droit de propriété : quatre chantiers et un enterrement », *Revue internationale de droit économique*, 2015/3 (t. XXIX), p. 307-331. DOI : 10.3917/ride.293.0307. URL : http://www.cairn.info/revue-internationale-de-droit-economique-2015-3-page-307.htm

Bensamoun Alexandra, Zolynski Célia, « *Cloud computing* et *big data*. Quel encadrement pour ces nouveaux usages des données personnelles ? », *Réseaux*, 2015/1 (n° 189), p. 103-121. DOI : 10.3917/res.189.0103. URL : http://www.cairn.info/revue-reseaux-2015-1-page-103.htm

Cottin Stéphane, « Le règlement européen sur la protection des données personnelles et ses implications pour les professionnels de l'I&D », *I2D – Information, données & documents*, 2017/2 (Volume 54), p. 20-22. URL : http://www.cairn.info/revue-i2d-information-donnees-et-documents-2017-2-page-20.htm

Revues

« La consécration par la CJUE d'un droit de déréférencement par les moteurs de recherche : principe, exceptions et mise en œuvre », *LEGICOM*, 2015/1 (N° 54), p. 89-105. DOI : 10.3917/legi.054.0089. URL : http://www.cairn.info/revue-legicom-2015-1-page-89.htm

« Regards croisés droit/économie », *Réseaux*, 2015/1 (n° 189), p. 15-46. DOI : 10.3917/res.189.0015. URL : http://www.cairn.info/revue-reseaux-2015-1-page-15.htm

EXPLOITATION COMMERCIALE DES DONNÉES PERSONNELLES

L'économie des GAFA

Articles

Bastard Irène, Cardon Dominique, Charbey Raphaël *et al.*, « Facebook, pour quoi faire ? Configurations d'activités et structures relationnelles », *Sociologie*, 2017/1 (Vol. 8), p. 57-82. DOI : 10.3917/socio.081.0057. URL : http://www.cairn.info/revue-sociologie-2017-1-page-57.htm

Benyayer Louis-David, Chignard Simon, « Focus - Les enjeux économiques de l'ouverture des données : pas de marché, pas de valeur », *Informations sociales*, 2015/5 (n° 191), p. 36-39. URL : http://www.cairn.info/revue-informations-sociales-2015-5-page-36.htm

Cardon Dominique, « L'ordre du Web », *Médium*, 2011/4 (N° 29), p. 191-202. DOI: 10.3917/mediu.029.0191. URL : http://www.cairn.info/revue-medium-2011-4-page-191.htm

Cardon Dominique, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », *Réseaux*, 2013/1 (n° 177), p. 63-95. DOI : 10.3917/res.177.0063. URL : http://www.cairn.info/revue-reseaux-2013-1-page-63.htm

Chevallier Marc, « Comment encadrer les géants du Net », *Alternatives économiques*, 2014/7 (N° 337), p. 86-86. URL : http://www.cairn.info/magazine-alternatives-economiques-2014-7-page-86.htm

Ceruzzi Paul E, « Aux origines américaines de l'Internet : projets militaires, intérêts commerciaux, désirs de communauté », Le Temps des médias, 2012/1 (n° 18), p. 15-28. DOI : 10.3917/tdm.018.0015. URL : http://www.cairn.info/revue-le-temps-des-medias-2012-1-page-15.htm ,

Dahan Michel, « Une guerre économique d'une violence inédite », *Le journal de l'école de Paris du management*, 2014/3 (N° 107), p. 36-42. DOI : 10.3917/jepam.107.0036. URL : http://www.cairn.info/revue-le-journal-de-l-ecole-de-paris-du-management-2014-3-page-36.htm

Granjon Fabien, « Du (dé)contrôle de l'exposition de soi sur les sites de réseaux sociaux », *Les Cahiers du numérique*, 2014/1 (Vol. 10), p. 19-44. DOI :

10.3166/lcn.10.1.19 - 44. URL : http://www.cairn.info/revue-les-cahiers-du-numerique-2014-1-page-19.htm

Revues

Cecere Grazia, Le Guel Fabrice, Rochelandet Fabrice, « Les modèles d'affaires numériques sont-ils trop indiscrets ? Une analyse empirique », *Réseaux*, 2015/1 (n° 189), p. 77-101. DOI : 10.3917/res.189.0077. URL : http://www.cairn.info/revue-reseaux-2015-1-page-77.htm

Monographie

Cardon Dominique, *La Démocratie Internet. Promesses et limites*. 1ère édition. Paris : Seuil, 2010. 102 p. La République des idées. ISBN : 9782021026917

Cardon Dominique, Casilli Antonio A., Qu'est-ce que le Digital labor ?, Paris, Ina Éditions, 2015, 104 p. ISBN 978-2-86938-2299

La mort numérique

Guillemot Samuel, Gourmelen Andréa, « Quand les entreprises s'emparent de la mort numérique, qui sont les consommateurs potentiels ? », *Revue française de gestion*, 2017/1 (N° 262), p. 123-145. DOI : 10.3166/rfg.2017.00114. URL : http://www.cairn.info/revue-francaise-de-gestion-2017-1-page-123.htm

ARCHIVAGE DU WEB

Articles

Banat-Berger Françoise, « Les archives et la révolution numérique », *Le Débat*, 2010/1 (n° 158), p. 70-82. DOI : 10.3917/deba.158.0070. URL : http://www.cairn.info/revue-le-debat-2010-1-page-70.htm

CNES, « Pour des données numériques durables », dans Qualité espace, mars 2006, n° 43, [en ligne] : http://www.cnes-multimedia.fr/cnes_fr/qualite/43-retourexperience.pdf

Gebeil Sophie, « Quand l'historien rencontre les archives du Web », *Revue de la BNF*, 2016/2 (n° 53), p. 185-191. URL : http://www.cairn.info/revue-de-la-bibliotheque-nationale-de-france-2016-2-page-185.htm

Mussou Claude, « Et le Web devint archive : enjeux et défis », *Le Temps des médias*, 2012/2 (n° 19), p. 259-266. DOI : 10.3917/tdm.019.0259. URL : http://www.cairn.info/revue-le-temps-des-medias-2012-2-page-259.htm

Paloque-Berges Camille, « Les sources nativement numériques pour les sciences humaines et sociales », *Histoire@Politique*, 2016/3 (n° 30), p. 221-244. DOI : 10.3917/hp.030.0221. URL : http://www.cairn.info/revue-histoire-politique-2016-3-page-221.htm

Paloque-Berges Camille, Schafer Valérie, « Quand la communication devient patrimoine », *Hermès, La Revue*, 2015/1 (n° 71), p. 255-261. URL : http://www.cairn.info/revue-hermes-la-revue-2015-1-page-255.htm

Monographie

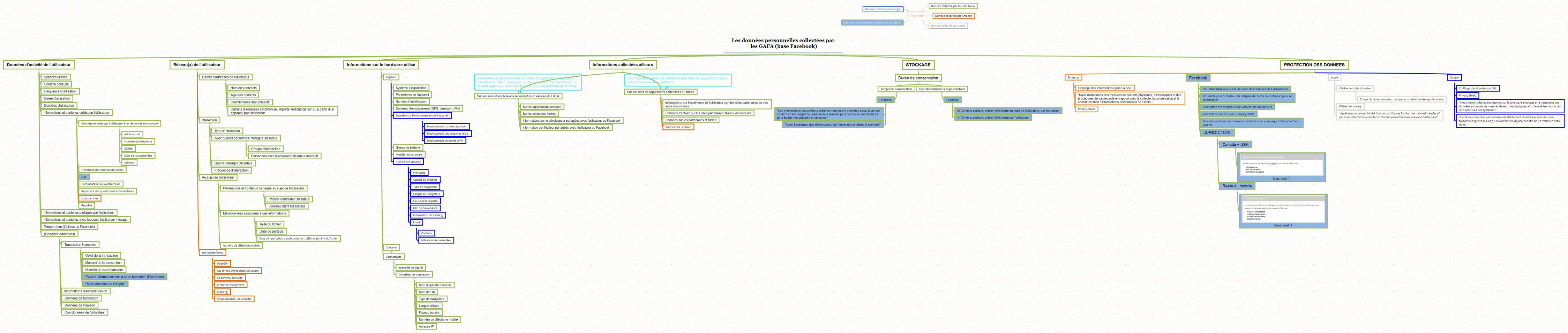
Banat-Berger, Françoise, « Les fonctions de l'archivistique à l'ère du numérique », dans Delpierre Nicolas, Hiraux Françoise, Mirguet Françoise, Les chantiers du numérique : Dématérialisation des archives et métiers de l'archiviste, 2012, Harmattan.

ANNEXES

Table des annexes

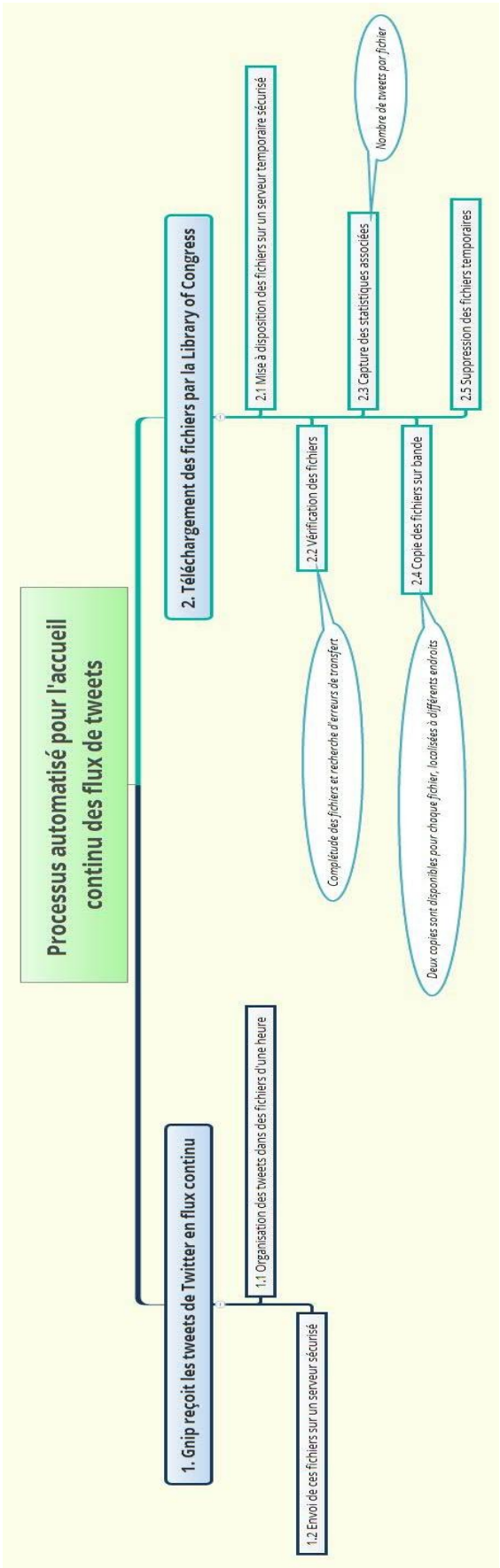
A.1. LES DONNÉES PERSONNELLES COLLECTES PAR LES GAFA

Ce travail a été fait en prenant pour base de travail les politiques de confidentialités des GAFA. La première politique de confidentialité à avoir été consultée est celle de Facebook, et le tableau a été complété par la suite. Il ne repose sur aucune autre source que les déclarations des GAFA dans leurs propres politiques de confidentialité. Pour Apple, c'est la politique de confidentialité associée au dernier Iphone qui a été choisie, pour des raisons d'actualité.


Donnée collectée par Google
Donnée exclusivement collectée par Facebook
Légende
Donnée collectée par tous les GAFA
Donnée collectée par Amazon
Donnée collectée par Apple
Les données personnelles collectées par les GAFA (base Facebook)
Données d'activité de l'utilisateur
Services utilisés
Contenu consulté
Fréquence d'utilisation
Durée d'utilisation
Données d'utilisation
Informations et contenus créés par l'utilisateur
Données remplies par l'utilisateur à la création de son compte
Adresse mail
Numéro de téléphone
Civilité
Date de naissance/âge
Adresse
Historique des commandes/achat
Like
Commentaire sur la plateforme
Réponse à des questionnaires/formulaires
Liste d'envies
Requête
Informations et contenus partagés par l'utilisateur
Informations et contenus avec lesquels l'utilisateur interagit
Température (Celsius ou Farenheit)
(Données bancaires)
Transaction financière
Objet de la transaction
Montant de la transaction
Numéro de carte bancaire
"Autres informations sur la carte bancaire" (Facebook)
"Autre données de compte"
Informations d'authentification
Données de facturation
Données de livraison
Coordonnées de l'utilisateur
Réseau(x) de l'utilisateur
Carnet d'adresses de l'utilisateur
Nom des contacts
Age des contacts
Coordonnées des contacts
Carnets d'adresses synchronisé, importé, téléchargé sur ou à partir d'un appareil, par l'utilisateur
Interaction
Type d'interaction
Avec quelles personnes interagit l'utilisateur
Groupe d'interaction
Personnes avec lesquelles l'utilisateur interagit
Quand interagit l'utilisateur
Fréquence d'interaction
Au sujet de l'utilisateur
Informations et contenus partagés au sujet de l'utilisateur
Photos identifiant l'utilisateur
Contenu citant l'utilisateur
Métadonnées associées à ces informations
Taille du fichier
Date de partage
Date d'importation, synchronisation, téléchargement du fichier
Numéro de téléphone mobile
Sur la plateforme
Requête
Les temps de réponses des pages
Le contenu consulté
Erreur de chargement
Scrolling
Cheminement URL complet
Informations sur le hardware utilisé
Appareil
Système d'exploitation
Paramètres de l'appareil
Numéro d'identification
Données d'emplacement (GPS, bluetooth, Wifi)
Données sur l'environnement de l'appareil
Emplacement d'autres appareils
Emplacement des antennes relais
Emplacement de points Wi-Fi
Niveau de batterie
Modèle du hardware
Activité de l'appareil
Plantages
Activité du système
Type de navigateur
Langue du navigateur
Heure de la requête
URL de provenance
Information de scrolling
Email
Contenu
Métadonnées associées
Contenu
Connectivité
Intensité du signal
Données de connexion
Nom d'opérateur mobile
Nom de FAI
Type de navigateur
Langue utilisée
Fuseau horaire
Numéro de téléphone mobile
Adresse IP
Informations collectées ailleurs
Recourrent à Facebook tous les sites et applications disposant d'un bouton "like" ,"partage" ou "se connecter via Facebook", ou faisant appel aux services de mesure ou de publicité de la firme
Sur les sites et applications recourant aux Services du GAFA
Sur les applications utilisées
Sur les sites web visités
Informations sur le développeur partagées avec l'utilisateur ou Facebook
Information sur l'éditeur partagées avec l'utilisateur ou Facebook
Sont des partenaires de Facebook des sites ou applications avec lesquels Facebook a collaboré
Par les sites ou applications partenaires ou filiales
Informations sur l'expérience de l'utilisateur sur des sites partenaires ou des sites annonceurs
Données d'activité sur les sites partenaires, filiales, annonceurs
Données sur les partenaires et filiales
Données de livraison
STOCKAGE
Durée de conservation
Temps de conservation
Facebook
"Les informations associées a votre compte seront stockées jusqu'à ce que ce dernier soit supprimé, sauf si nous n'avons plus besoin de vos données pour fournir nos produits et services."
"Aussi longtemps que nécessaire pour fournir nos produits et services"
Type d'informations suppressibles
Facebook
Contenu partagé, publié, téléchargé au sujet de l'utilisateur, par les autres
Contenu partagé, publié, téléchargé par l'utilisateur
PROTECTION DES DONNEES
Amazon
Cryptage des informations grâce à SSL
"Nous maintenons des mesures de sécurité physiques, électroniques et des procédures de sauvegarde en rapport avec la collecte, la conservation et la communication d'informations personnelles de clients. "
Privacy Shield
Facebook
Peu d'informations sur la sécurité des données des utilisateurs
Possibilité pour l'utilisateur de désigner des "amis de confiance" pour se reconnecter
Vérification des comptes et des activités des utilisateurs
Transfert de données par le privacy shield
Sécurité spécifique des mineures : redirection vers une page "d'éducation" aux parents
JURIDICTION
Canada + USA
Si vous habitez aux États-Unis ou au Canada...
Veuillez contacter Facebook, Inc. en ligne ou par courrier à l'adresse :
Facebook, Inc.
1601 Willow Road
Menlo Park, CA 94025
Sous-sujet 1
Reste du monde
Si vous habitez ailleurs...
Le contrôleur de données en charge de vos informations est Facebook Ireland Ltd., que vous pouvez contacter en ligne ou par courrier à l'adresse :
Facebook Ireland Ltd.
4 Grand Canal Square
Grand Canal Harbour
Dublin 2 Irlande
Sous-sujet 1
Apple
Chiffrement des données
Coupe l'accès au contenu, mais pas aux métadonnées qui l'entoure
Diferential privacy
"Apple uses approved Model Contractual Clauses for the international transfer of personal information collected in the European Economic Area and Switzerland"
Google
Chiffrage des données par SSL
Privacy Shield
"Nous menons des audits internes sur la collecte, le stockage et le traitement des données, y compris les mesures de sécurité physiques, afin d'empêcher tout accès non autorisé à nos systèmes."
"L'accès aux données personnelles est strictement réservé aux salariés, sous-traitants et agents de Google qui ont besoin d'y accéder afin de les traiter en notre nom. "

A.2. LE SCHÉMA D'ARCHIVAGE DES TWEETS PAR LA BIBLIOTHÈQUE DU CONGRÈS

Table des matières



GLOSSAIRE

Donnée : une donnée au sens informatique est une information numériquement codée, formée de 0 et de 1.

Donnée personnelle : se dit d'une donnée qui, seule ou croisée avec d'autres informations, peut conduire à identifier une personne physique

Donnée sensible : sous-catégorie de donnée personnelle, qui ajoute à la définition précédente la notion d'intime. Une donnée personnelle ne peut être collectée sans le consement de la personne physique qui en l'objet.

Privacy by design : de développement d'applications et de services qui prend en compte dés sa conception la protection des données. Elle repose sur le responsable de traitement.

Privacy by using : utilisation prudente, éclairée et raisonnée des applications et services à sa disposition par une personne physique.

Table des matières

Table des matières

TABLE DES MATIÈRES